# SECURITY ENHANCED APPLICATIONS FOR INFORMATION SYSTEMS

Edited by **Christos Kalloniatis**

**INTECHOPEN.COM**

**Security Enhanced Applications for Information Systems**
Edited by Christos Kalloniatis

**Published by InTech**
Janeza Trdine 9, 51000 Rijeka, Croatia

**Notice**
Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

**Publishing Process Manager** Romina Skomersic
**Technical Editor** Teodora Smiljanic
**Cover Designer** InTech Design Team

First published May, 2012
Printed in Croatia

A free online edition of this book is available at www.intechopen.com
Additional hard copies can be obtained from orders@intechopen.com

Security Enhanced Applications for Information Systems, Edited by Christos Kalloniatis
p. cm.
ISBN 978-953-51-0643-2

# Contents

# Preface

One of the main challenges that modern Information Systems are dealing with is the protection of security for both the external users that take advantage of the various services offered as well as the stakeholders and internal users. Security is dealt in every level of system development from the analysis stage through the implementation and testing stages. In every stage a number of methods and techniques have been proposed trying to fulfill the basic security concerns namely confidentiality, integrity and availability.

Nowadays the rapid development of new information infrastructures increases users' dependability on Information Systems and this can lead to a vulnerable information society based on insecure technologies. Indeed, more and more users access services and electronically transmit information which is usually disseminated over insecure networks and processed by websites and databases, which lack proper security protection mechanisms and tools. This may have an impact on both the users' trust as well as the reputation of the system's stakeholders. Designing and implementing security enhanced systems is of vital importance.

Therefore, this book aims to present a number of innovative security enhanced applications, it is titled "Security Enhanced Applications for Information Systems" and includes 11 chapters. This book is a quality guide for teaching purposes as well as for young researchers since it presents leading innovative contributions on security enhanced applications on various Information Systems. It involves cases based on the standalone, network and Cloud environments.

**Christos Kalloniatis**
Department of Cultural Technology and Communication,
University of the Aegean,
Greece

# 1

# Web and Database Security

Jiping Xiong, Lifeng Xuan, Jian Zhao and Tao Huang
*Zhejiang Normal University,*
*China*

## 1. Introduction

In recent years, with the frequent occurrence of security incidents, enterprises and organizations have now realized the importance of designing a safety information system. Today, information systems are heavily relied on web and database technologies, thus the risks and threats those technologies faced will also affect the security of information systems. Web and database security technologies can ensure the confidentiality, integrity and usability of data in information system, and can effectively protect the security and reliability of information system. Therefore, in order to better secure the information systems, we need to learn Web and database security-related knowledge. This chapter covers extensively practical and useful knowledge of web and database security.

This chapter can be divided into three parts: advanced security threats, the principles of safety design and safety audit; Advanced security threats section contains cross-site scripting (XSS) attacks, AJAX and SQL injection attacks and other security threats, which will be presented in detail; the principles of safe design section describe the general safety design principles to help design information systems security; last section describes the manual and automatically audit methods, and general security audit framework to help readers to understand more clearly.

## 2. Advanced security threats

### 2.1 Web security threats

### 2.1.1 AJAX security

As Web applications become increasingly complex, it is required for the performance of Web services is also increasing. AJAX (Asynchronous JavaScript and XML) (Garrett, 2005) technology is mainstream technology of Web2.0 that enables the browser to provide users with more natural browsing experience. With asynchronous communication, user can submit, wait and refresh mode freely, update partial page dynamically. So it allows users to have a smooth experience similar in desktop applications.

However, a variety of Web applications has brought us countless convenience, produced a series of security problems. When the introduction of AJAX technology, because of its inability to solve the security problems, the traditional Web security problems still exist, along with elements of the composition and structure of AJAX features, will lead to new security threats. In recent years, adding AJAX elements in sites has become a very popular trend, and most websites are typical AJAX-based applications. As most of the website builders just enjoy the conveniences of AJAX technology, little is known about its security threat, resulting in most of the AJAX application sites have different levels of security risks. Here, we summarize and analysis the AJAX security threats.

1. Security Threats of AJAX Technology

a. The Deficit of JavaScript Language

JavaScript is a widely client-side scripting language, originally designed and implemented by Netscape, and it has been widely used to reduce the burden on the server. JavaScript scripting language features determine its presence in all kinds of security risks:

- JavaScript is an interpreted language. In the interpretation process, every error is a runtime error. Run-time error can only be found during runtime. If somewhere in the code the programmer has left a Bug, but the logic of the code at run time is not running to the area, then the bug will not be found, which leaving significant risks to the application. To detect, locate the error position of interpreted language is quite difficult.
- JavaScript is a weak typing language. Weak typing languages do not need to declare variables at the time the programmer declare the variable. This flexibility often easily leads to many problems.
- JavaScript code has dynamic nature. It can be dynamically generated code, and used the eval-function dynamic execution; or you can directly modify the existing function. Once the attacker can gain control of the JavaScript code, he can overwrite the other user-defined methods and even the browser built-in method, thus cause many serious malicious behaviours.

b. Problems of Asynchronous

Asynchronous communication is the highlights and core idea of AJAX technology. But asynchronous will also introduce a series competition problems.

2. Issues of AJAX Framework

a. Explosion of Client-Side Logic

Programming client-side logic using JavaScript will bring the client-side logic to public. Users can easily through the browser's View Source feature to see the client code.

b. Incomplete Server

Most AJAX programmers validate user input at client-side, though it reduces the burden of server, it lefts room for security risks.

3. Traditional Web Security Threats of AJAX

AJAX framework gives users a good experience in desktop’s application, users no longer have such a long wait for the server to response and refresh the page. However, this feature also poses a problem: the user does not know what the current request was sent, did not even know the current request was sent. This feature allows many of the traditional Web attacks in a more intimate manner. Main traditional Web security threats are (Razvan & Maria, 2010):

- AJAX framework of SQL injection;
- AJAX framework of XPath injection;
- AJAX framework for cross-site scripting attacks (XSS);
- AJAX framework for cross-site request forgery attacks (CSRF);
- AJAX framework of denial of service attack (DOS).

4. Security Threats Introduced by AJAX

The introduction of AJAX, initially to solve the user to submit a request in the browser when the server response is required after a long wait, refresh the entire page to the next step of the problem. But AJAX technology to bring convenience, but also introduces some security threats.

a. JSON Injection and JSON Hijacking

JSON (JavaScript Object Notation) (Crockford, 2006) is widely used lightweight data transmission and exchange format in AJAX. JSON is based on a subset of JavaScript and developed from JavaScript Array and Object, and the adopted text format is completely independent with language. The data of JSON and can be transmitted cross-platform. Therefore, JSON Injection and JSON Hijacking are current two aspects of security threats.

b. Trust Crisis of AJAX Proxy

For security reasons, JavaScript code is limited to running in a sandbox, JavaScript also prohibit access to third-party domain. But sometimes you need to call in the AJAX third-party services, such as components Mashuop procedures. Solution to this is to build an AJAX proxy that the Web server to create a Web service, only forwarding calls to third-party Web service request. AJAX proxy allows the client calls the Web service as a third party may also have to provide AJAX proxy servers and third-party server, a crisis of confidence. First, the attacker can access through the AJAX proxy direct access to many previously unavailable resources. Meanwhile, the attacker via AJAX proxy attack on third-party Web server, you can also hide the source of the attack, showing up as if from the AJAX proxy attack (Anley, 2002).

c. Disclosure of User Data

AJAX technology gives users a better browsing experience, but some AJAX-based application inadvertently brought the disclosure of user data. AJXA technology is widely used in such situations: a user registers a mailbox, enter the account he wants to use, after he moves to the next input section, the browser prompts the user name input box: the account that you are applying has been applied, please re-apply. This design reflects good human-based design rule, the user does not have all the information before being prompted to fill out and submit the account has been to apply for. But in this way the user data will be leaked in the unconscious. False malicious attacker by entering any letters, numbers, combined to form the account, you can immediately know whether the mailbox has been registered. If you know the mailbox already exists, there may occur spam, or send e-mail containing the malicious XSS code may be so. Enumeration by simple repetition, the attacker can even know the mail server name of all existing accounts, which will undoubtedly bring great threats.

### 2.1.2 Cross site scripting

Cross-site Scripting (also known as XSS or CSS) occurs when dynamically generated Web pages display input that is not properly validated. In XSS, malicious attackers acted as normal visitors upload Malicious Script as JavaScript codes etc. to Web server by utilizing the bugs of utility programs or codes in the Web server. Attackers also send URL links including malicious script to objective users. When Web users visit the pages containing malicious script or open the received URL links codes in the Web sites, users' browsers will auto-load and execute the malicious script codes. This attacking procedure indicates that XSS is actually a simple attack technology. In most cases, malicious attackers attack users indirectly by utilizing Web server, and direct attack occurs merely.

XSS is a passive attack. First of all, by utilizing the XSS bugs in the Web programs, malicious attackers construct a trap page and the malicious script can be saved in the page content or URL. The URL of this page is then announced in the BBS after embedding to e-mails or disguising attractive titles. If the users visit ULR, the JavaScript will be executed by attackers' browser. The procedure of XSS attack is shown in fig. 1.

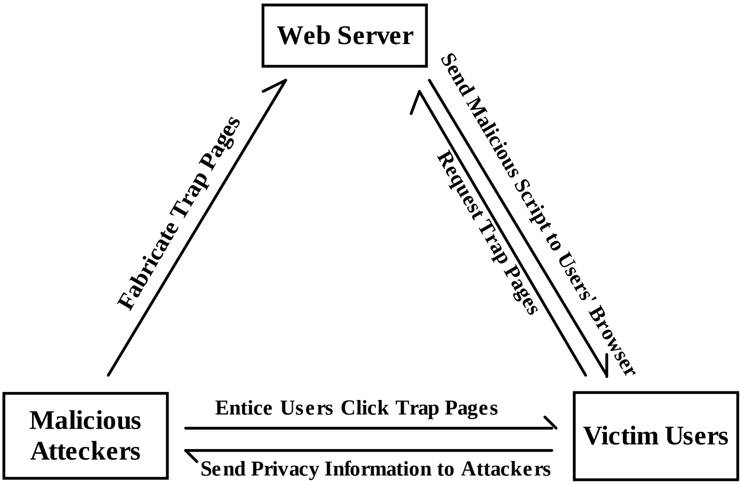

Web Server
Fabricate Trap Pages
Send Malicious Script to Users' Browser
Request Trap Pages
Malicious Atteckers
Entice Users Click Trap Pages
Send Privacy Information to Attackers
Victim Users

Fig. 1. The Process of Cross Site Scripting Attack

### 2.2 Database security threats

Database security relates to two parts: data visiting and data recovery. The first part can be realized by using a suitable authorization to make sure that the legal users can get their right data and reject all exceeding authority at the same time. The latter part means that database can recover the data securely and completely.

Recently, database is facing the problem of security hole e.g. privilege elevation, SQL inject, XSS, data leakage and improper error processing.

### 2.2.1 SQL inject

1. SQL Inject Principle

SQL Inject refers that the attackers deceive database server to execute unauthorized wilful inquire and illegal operation through adding extra SQL statement element to the end of predefined inquire statement in application programs. The essence of SQL Inject is utilizing the bugs caused by the programmers who did not detect or incomplete detect the database inquiry request, submitting malicious SQL statement and cheating server executes malicious inquiry. At last, the attackers can get the sensitive data or control the whole website.

The main reason for SQL injection attack to succeed is that when dynamically generating SQL statement commands, websites only directly using the subscribers inputted data without any verification.

2. Process and Methods of SQL Attack

In SQL attack process, the attackers firstly trial the SQL inject bugs in application programs by design inputs. The executive SQL statements are then imported to control implement programs. After obtaining the database information, the attackers acquire the administration authority of server system.

a. Discovery of SQL Inject Bugs

Discovery of SQL inject bugs brings necessary information to further attack. Before SQL attack, the attackers need to identify the aim database platform and decide what SQL attack statements or methods should utilize (Halfond et al., 2006).

The common methods for SQL inject bugs are as follows:

- Add single quotes etc. characters to the end of submitted inquiry. So that attackers judge if the inject bugs exist depending on estimating database type of prompt message return from the server.
- Push `and 1=1` and `and 1=2` to the end of submitted inquiry. Bugs exist when `and 1=1` stay regular and `and 1=2` go wrong, and then illegal inquiry and other malicious behavior occur. It means that illegal inquire statements can be added after inquiring and assaultable bugs existed.
- A regular method to find bugs is to judge database style by the build-in variable and function of database.

b. SQL Inject Bugs Utilization

After finding out the bugs, attacks including illegal inquire database, obtaining secret information and users data, controlling database and server system occur.

- Speculating table name and field name. Attackers take advantage of SQL statement, such as `and (select count(*) from TestDB.dbo.tablename)>0;` to guess table name. If the table name have existed, the webpage returns to regular.
- Obtaining field value. After getting table name and column name, field value is generated by utilizing ASCII word-by-word decoding.

After these two steps, the attackers can get data, user name, password and information in database.

- Further attack. From definition and principle of SQL inject, we know that attackers are interested in administrators authorizations rather than their account numbers. The administrator authorizations will bring further attack and acqire higher authorization for attackers to add Trojan to webpages.

## 3. Security design principle

### 3.1 Web security design principle

We should obey the following principle when designing and deploying computer network security.

1. Balance Analysis Principle of Demand, Risk and Cost

According to the existing technology, there is hardly a perfect Safety network. We should do some qualitative and quantitative analyse to threaten and possible risk network faced. The standards and measures are then made to confirm security policy of system.

2. Principle of Comprehensiveness and Integrity

We can analyse security issues of network and formulate specific measures by using views and methods of system engineering. Multi-methods make a prefer security measure. A computer network including links of human being, device, software and data take an important role in network security, and only analyse and treat in whole view can they obtain valid and executive measures.

3. Consistency Principle

Consistency Principle means that network security issues should concurrent exist with the whole network operating cycle, and security architecture should keep in line with network security. Actually, network security strategies should taken into consideration in the beginning of network construction rather than at the end of this procedure with characters of facility and low cost.

4. Principle of Security and Reliability

Guarantee of system security is very important. In procedure of design and implement, specify measures are adopted to ensure security of information secure product and technology proposal. By strict technology administration and redundancy configuration of device, quality of product and reliability of system can be guaranteed.

5. Principle of Advanced Technology

Advanced technology system and standard technology are required in security design.

6. Principle of Easy-operation

Security measures are manually completed. Complexed measures can always lead to high requirement for administrators, and low security. Otherwise, the measures should be friendly to operation of system.

7. Principle of Adaptability and Flexibility

Security measures must change with the developing of network performance. Characters of easily to adapt and modify are required.

8. Multiple-protection Principle

Perfect security protection methods merely existed, so that a multiple-protection system is constructed to protect each layer. When one layer is broken, any other layers can still protect information.

Methods as installing fire wall, setting up isolation region for protected resource, encrypting the sensitive information being stored and transmitted, providing identity authentication and building secret passage, providing digital signature for audit and tracking to software without any security guarantee are adopted to ensure Web service security.

1. Install Fire Wall

The most popular security method is providing an isolation region to LAN or website. Fire wall of LAN is a function module inside computer or network equipments between innernet and Internet. Its purpose is to provide security protection to an innernet or host and control access objects, so it can also called access control technology. There are two operation mechanisms for fire wall e.g. packet filtering and agency. Packet filtering aims at the service provided by host of special IP address. Its basic principle is to intercept and capture IP packet of IP layer in network transmission, then find out resource address and destination address, source port and destination port of IP packet. Whether to transmit IP packet is based on fixed filtering principle.

Agent is achieved in the application layer, the basic principle is to construct an independent agent program for Web services, and client program and the server can only exchange information by their own agent programs rather than allow them to interact directly with each other.

2. Encryption for Confidential Information

This method is particularly effective to protect confidential information, which can prevent wiretapping and hacking. Transmission encryption in Web services is in general achieved in the application layer. When WWW server sends confidential information, firstly, it selects keys to encrypt the information, based on the receiver's IP address or other identification; After browser receives the encrypted data, it decrypts the encrypted data according to source address or other identification of the information in IP packet to get the required data. In addition, transmission, encryption and decryption of information at the IP layer also can be achieved by encrypting and decrypting the whole message to ensure information security at the network layer.

3. Provide Identity Authentication for the Client / Server Communication and Establish A Secure Channel

Currently some network security protocols e.g. SSL and PCT have appeared, which are based on the existing network protocol. These two protocols are mainly used for not only protecting confidential information but also preventing other unauthorized users to invade their own host.

SSL protocol is a private communication and includes technology of authentication, signature, encryption for the server, which can not only provide authentication for the server but also provide authentication for the client according to the options of the server.

SSL protocol can run on any kind of reliable communication protocols, e.g. TCP, and can also run in application protocols e.g. HTTP, FTP, Telnet etc. SSL protocol uses X.509 V3 certification standards, RSA, Diffie-Hellman and the Fortezza-KEA as its public key algorithm and uses the RC4-128, RC-128, DES, 3-layer DWS or IDEA as its data encryption algorithm. The authentication scheme and encryption algorithm provided by PCT are more abundant than SSL, and it makes improvements in some details of the agreement.

IPSec protocol is used to provide end to end encryption and authentication services for public and private networks. It specifies all kinds of optional network security services, and the organizations can integrate and match these services according to their own security policy, and they can build security solution on the framework of the IPSec. The protocol provides three basic elements to protect network communications, the basic elements are "Authentication Header", "Encapsulating Security Payload" and "Internet Key Management Protocol".

HTTPS protocol (Secure Hypertext Transfer Protocol), which is built on its browser for compressing and decompressing the data, and returns the result which is back to the network.

4. Digital Signatures for the Software

Many large companies use digital signature technology for their software, and claim that they are responsible for the security of their software, especially e.g. Java applets, ActiveX controls, which will bring risks to Web services. Digital signatures are based on public key algorithms, using their private key to sign its own released software, and are authenticated by using the public key. Microsoft's Authenticode technology is used to identify a software publisher and prove that it has not been damaged. Authenticode is software for client, which monitors the ActiveX control, Cab files, Java applets, or download of executable file, and look for the digital certificate to verify in these files, and then show warning words, the certificate organization's name and other information to the user for possible security problems. Digital signature can protect the integrity of the software, and it is sensitive to illegal change of the software in the transfer process.

### 3.2 Database design principles

Users enter into the database system through the database application program when users firstly access the database, database applications deliver the username and password which is submitted by the user to the database management system for certificating, after determining their legal status, users are allowed to enter. They also must pass the authentication when operate objects, tables, views, triggers, stored procedures etc. in the database. How can users operate in application and database is depended on rights allocation and constraints of accessing control.

1. Secure Database System Model

Criteria based on security database, you can create a simple security database system model which is divided into four layers: system layer, including data access, encryption and decryption algorithm; function layer is the key to the whole system, including key distribution mechanism, fast indexing mechanism and derive control; interface layer is directly user-oriented, which includes the function of user authentication, authorization management, database maintenance and query management; At application layer, users can manage database through not only interactive ways but also command mode.

2. Management Strategy of Database

a. Access Control

Access control is the rights control of user access to all kinds of resources of the database, which is divided into two stages: one is security account identification, the other is the access permission identification. In the security account authentication phase, the user logins for the authentication, if it's successful, he can connect to the SQL Server, otherwise it will reject the connecting requirement. Access license verification refers to that after the user connect to SQL Server, the system determine whether they have license to access to the database according to the user account stored in the database and correspond to server login identification. Access control can prevent the illegal users.

b. Database License

After the legal users access to the server and database, the database access mechanism will control the legal users to operate the data objects. First of all, statements in the database license will limit the database user to carry out some SQL statements. Secondly, objects in the database license will limit the database user to carry out some tasks of the database objects.

c. Establish Data Security by Using System Stored Procedures

As the database administrator, if you want a user to have a select right rather than the delete right, at this time you can achieve the goal by establishing stored procedures, thus protecting the safety of the data.

d. Establish Data Security by Using the View

If the administrators give users the permission to access the database tables and form a too large user access area, it will cause threats brought by users to data security of the database. To avoid this situation, you can achieve data security view through the way of establishing data view.

e. Establish Data Security by Using the Database Role

This role is used for setting license at a time that number of database users can access to the database, if permission is not deployed properly, it will threat data in the database directly. As an administrator, you should be very careful when you give permission to the public role.

f. Data Backup

Data backup is principal work in the course of daily management of the database. When the server or database system breaks down, the original data is difficult to recover without a backup strategy. Therefore, the database should be installed in security zone of their intranet, and can not be connected to the Internet directly. In addition, different computers should implement backup strategies to protect data security when people deal with abnormal failure.

g. Database Encryption

Database encryption requires that database cryptography changes plaintext into cipher-text, and cipher-text data stored in the database. Cipher-text is decrypted to get clear information when queries, so data will not be leaked even if the hardware store is stolen, thus the database system security is greatly improved, of course, the cost also increases. Response to attacks from the network level, the database mainly uses many ways e.g. installing a firewall, doing intrusion detection etc. to improve its safety performance. Firewall resists the incredible connections from outside. Intrusion detection systems are generally deployed in firewall, and detect abnormalities on the network and the host through Network packet interception analysis or Analysis of log.

h. Audit Trail and Attack Detection

The audit function records all database's operation in the audit log automatically when the system works, attack detection system analyses and detects attempt of internal and external attackers according to the audit data, and reproduces events which leads to the status of the system, find vulnerabilities of the system by analyzing, and then trace the relevant responsible person.

## 4 Security audit

### 4.1 Definition of security audit

Security audit is based on certain security policy, improving system's performance and safety by recording and analyzing historical events and data. Security audit includes all actions and instruments, e.g. testing, assessing and analyzing all of the weak links in the network information system to find the best ways to let the business run normally, based on the maximum guarantee of safety. It is to ensure the safe operation of network systems and prevent confidentiality integrity and availability of the data from being damaged, prevent intentional or unintentional human error and detect criminal activity on the network. The network status and processes can be targeted to recorded, tracked and reviewed by using the audit mechanism, and find safety problems. In addition, the audit can provide the basis of making filtering rules for online information, if the harmful information is found in the website, it will be added into the list of route filtering, to reject all information of IP addresses on the filtering list through information filtering mechanism. Fig.2 gives a brief overview flowchart of security audit.

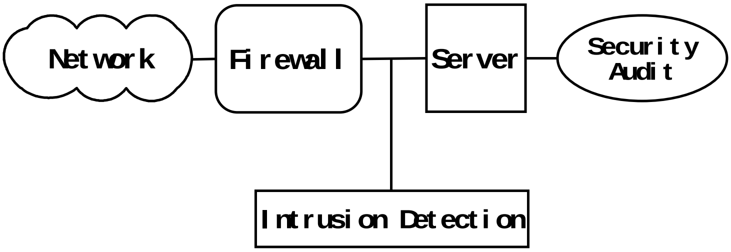


Fig. 2. Situation of Security Audit System in Network

Security audit techniques use one or several security testing tools (generally referred to as scanner), first of all, it will scan loopholes and inspects security vulnerabilities of the system, then achieve the inspection report about the weak link of system, at last it will take security protection and emergency measures according to the response strategies.

Traditional security audit has the function of "old records", pay attention to the audit afterwards and emphasize the deterrent of the audit and verification of security incidents. With the change of United States national information security policy, doing the so-called "defense in depth strategy information" in the information infrastructure is put forward by Information Assurance Technical Framework (IATF), this strategy requires security audit system to participate in the active protection and response. In modern time, network security audit is an all-round, distributed, and multiple-level strong audit concept, which breaks the previous concept of "log" and other shallow level security audit, and it's consistent with the requirements of protecting, detecting, replying and recovering (PDRR) dynamic process, which is put forward by IATF. It can protect and response to the information actively on the basis of improving the breadth and depth of audit.

1. Based on the objects of audit, security audit is divided into:

- Operating system of audit;
- Application system of audit;
- Equipment of audit;
- Network application of audit;

2. Based on the ways of audit, security audit is divided into:

- Distributed audit: audit information is stored in the server and security equipment, and system security administrator will review it. Distributed audit is applied to enterprise information system which demands less with information security protection.
- Centralized audit: audit information in the server and security equipment is collected, collated, analyzed and compiled into the audit report. Centralized audit is applied to enterprise information system which demands more with information security protection.

3. Based on control mechanism of audit, security audit is divided into:

- Host based audit. Host based control mechanism can control the specified host system, its control ability is in detail;
- Network based audit. Network based control mechanism can real-time monitor network security risks, to realize the comprehensive protection of intranet resources;
- Combination of host and network based audit. It can not only monitor host but also the network.

4. The emphasis of the information system security audit are mainly the following types:

- Network communication system: It mainly includes analysis, recognition judgment and record of the typical protocol in the flow of network, intrusion detection of Telnet, HTTP, Email, FTP, online chat, file sharing etc, as well as for traffic monitoring, recognition and alarm of anomaly traffic and network equipment operation monitoring.
- Important server host operating system: It mainly includes audit of the startup of system, running situation, the administrator login, operation situation, system configuration changes (e.g. the registry, the configuration file, the user system) as well as a worm or virus infection, the resource consumption; audit of hard disk, CPU, memory, network load, processes, operating system security log, system events, access to the important document.

- Main server host application platform software: It mainly includes the audit of the running of the important application platform processes, Web Server, Mail Server, Lotus, middleware system, health status (response time) etc.
- Main database operation audit: It mainly includes the audit of the database process operation conditions, violated access behaviour to operate the database directly by passing the application software, the database configuration changes, data backup operations and other operations of maintenance and management, to access and change important data, and data integrity.
- Main application system audit: It mainly includes the audit of office automation system, document flow and operation, webpage integrity, interrelated service systems etc. The relevant business system includes normal operation of business system, important operations of setting up or stopping the user, authorized change operation, data submission , processing, access and publishing operation, business process etc.
- Main regional network client: It mainly includes audit of virus infection situation, file sharing operation through the network, operation of copying or printing file, the situation of unauthorized connect to Internet through the Modem, installation and operation of non business abnormal software.

## 4.2 Execution of security audit

### 4.2.1 Process of security audit

Process of Security audit can be divided into two modules, including the collection of information and the security audit, the structure is shown in the following fig.3.

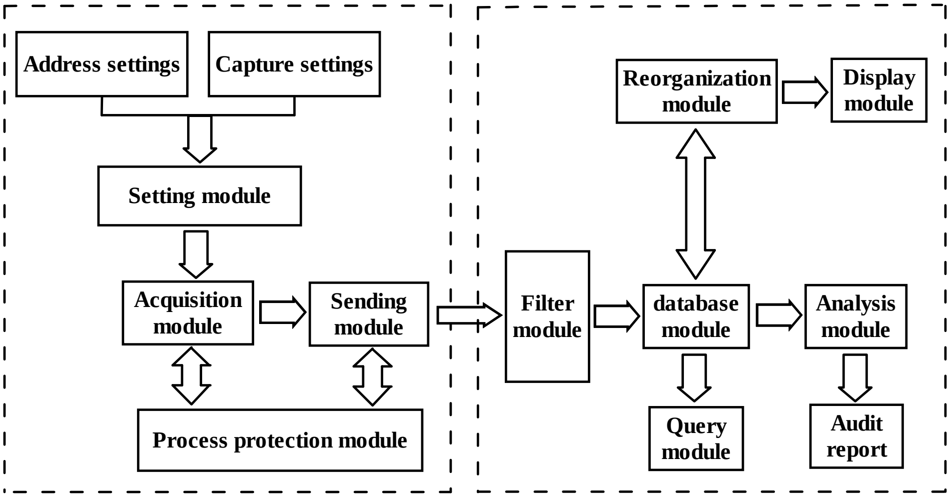


Fig. 3. Architecture of Security Audit System

Information gathering side runs on the system server. Information collecting side transmits information to the security audit terminal through a special security channel, security audit client runs on a stand-alone computer separately. To ensure the safety of audit side, the computer which operates safety audit module is not access to the server's LAN. To ensure the communication between audit module and information collection module, we do a dual-channel design, the user can set any ways to communicate, and the system can automatically switch to another way when one way can not work normally.

Information collection module collects the data which needs to be audited, including network packet, process information, port information, file access information, and modify the registry information, as well as a variety of server logs, such as WWW logs and security logs. In order to process conveniently later, all data is stored in the audit module of the database.

### 4.2.2 Key technologies of network security audit system

1. Analysis of Data Source of Network Security Audit System

For security audit system, selection of incoming data is the key problem to be solved, data source of the security audit can be divided into three categories: Based on the host, based on network and other channels. In order to select the appropriate data sources, it analyzes each class of data source respectively as follows.

a. Data Source Based on the Host

Data sources of the network security audit based on the host , including audit records of the operating system, system log, log information of application system and information based on the target.

b. Data Source Based on the Network

Network data is the most common source of information in the current network security audit system and commercial intrusion detection system. The basic principle is that when the network data stream transmitting in the network, using a special data acquisition technology to collect the data transmitted in network as the data source of security audit system.

c. Other Data Source

Data source from other safe products mainly refers to log files produced by safe products e.g. firewall, authentication system which are operated independent in the target system. These data sources also should be considered by security audit system.

Data source from network device e.g. a network management system, using information provided by SNMP (Simple Network Management Protocol) as data source. Out-of-band data source refers to data information provided by the artificial way, which is contrived and non-systematic, e.g. recording what happened in system environment manually, including hardware error information, system configuration information, system crash, other kinds of natural hazard events etc. Out-of-band data source may play an important role for later analysis.

In general, it will improve the performance of security audit system if active log of the network and its safe device are used as audit data source.

2. Functions of Network Security Audit System

a. Data Acquisition and Storage Capability

Data collection captures data-packets based on the data link layer. It filters out the packets without audit and saves selectively according to the defined policies and system analysis requirements. Data acquisition and data files are generated to provide data source the network security audit system. It is the key link of the network security audit system, and is the basis of data analysis and processing. Because the system only access the external computer and the user network audit, it is not necessary to collect or store intenal network data.

b. Log Data Management Capabilities

The log data with sustainable growth are very large, even a small network produces over 3 G network logs per day. Integrated mechanism of backups, recovery and processing is constructed for management of network security logs rather than simply delete.

c. Feature of Automatical Analysis and Statistical Reports Generation

The network will generate a lot of daily log information, and it is difficult for administrators to process these huge amount of work. A visualized analysis and statistical reports automatically generated mechanisms need to be provided to ensure that administrators can find a variety of network anomalies and security events effectively.

d. Data Analysis and Processing Functions

System access to external networks, achieve the user's computer and the contents of the audit network behavior via processing and analyzing to the data collected and preserved. The core is protocol analysis. Web content audit system includes web audit, mail audit, FTP audit and user log etc. The function data play a decisive role in audit results.

e. Function of Real-time Network Status Monitoring

Real-time monitoring function mainly includes analysis, identification, judgement and record of typical protocol in network traffic, intrusion detection for Telnet, HTTP, Email, FTP, Internet chat, file sharing etc. flow monitoring, and identification and alarming of unusual flow.

f. Network Service Control Function

Network service control function achieves control of host and service for user access to network services, to be able to support the operations of user authorization, settings of white list host and user access rules.

3. Network Security Audit System Architecture

Network security audit system mainly consists of three modules.

a. Data Collection Module

Data collection module acquires network packet of users' operation by monitoring and core filtering technology depending on imaging feature of switchers and user-defined strategies. The key to realize this module is to acquire accurate and complete packet. Data integrity of data acquisition modules is determined by the exactness and completeness of audit results.

b. Packet Processing Module

Protocol analysis is a key step in the data packet processing. Main job of packet processing module is to capture the data packets and determine the protocols e.g. TELNET, FTP and other protocols it belongs based on their header information. According to the formats, transmission mode and message content, it make the user's operation to restructure, restore, and finally it restore user data and submit to the audit module.

c. Data Audit Module

According to the rules defined format, the achieved user information e.g. TELNET and FTP commands, SQL statements, manipulate objects, operating keywords will match with the user-defined strategy in rule base. Responses are made according to the matching results, and the audited data are recorded into audit logs. The rule base is generated based on the visit strategy deployed by authorized administrator. The authorized administrators formulate or modify the strategy, and issue it to the next rule base. In process of utilizing audit system, administrators gather experience and add novel strategies constantly according to the issues in system usage making the rule base more and more abundant.

#### 4.2.3 Security audit approach

1. Methods Based on Rule Base

Rulebase based security audit method is the process below. Administrators extract feature of attack behaviors, and then push them into rulebase after represent by script language. When executing security audit, network attack behaviors are detected after the comparison and matching operations e.g. keywords, regular expression, fuzzy approximation degree between the above rule base and network data. But these rules are only fit for certain specific types of attacks or attack software, and failures of rule base are generated when new attack or upgraded software turns up.

2. Mathematical Statistics Based Method

Method of mathematical statistics is to create a statistic description for object firstly e.g. average value or variance of network traffic. And then value of characteristic quantity under normal circumstances is calculated to compare with actual network packet. If actual value is far different with regular value, the attacks is then occurred. However, the biggest problem of mathematical statistics is how to set the thresholds of statistics i.e. cut-off point between the normal and abnormal value, which often depends on the administrator's experience that inevitably prone to false and omission.

3. New Method Based on Network Security Audit System : Learning Data Mining

The biggest drawback of the above two methods is that the known intrusion patterns are hand-coded inevitably, and can not applicable to any unknown intrusion patterns. So people start to pay attention to the the data mining method owning the learning ability. Data mining is the process of analyzing mass data completely including data preparation, data preprocessing, establishment of mining model, model evaluation and interpretation. It is an iterative processing and can get a better model by continuously adjust methods and parameters. The main idea of the network security audit system is to find *normal* network communication patterns from *normal* network communication data and then achieve the purose of detecting web accack behaviors by relewance analysis with the regular attack rule base.

Firstly, the system collects data from collection points, and put the data into database after processing. Invasion evens are detected by executing engine of security audit to read in rulebase. The invasion is then recorded into invasion time database as well as the regular network visiting data are recorded in regular network database, and regular visiting patten can be abstracted by data mining. Latest rulebase is acquired from old rulebase, invasion events and regular visiting patterns. The above procedure is repeated and self-learning constantly until achieving a stable rule base. Data mining technology extracts the regular visiting mode semi-automatically from the mass of the normal data, which can reduce the human perception and experience participation which declining the possibility of misinformation.

#### 4.2.4 How do security audit

1. Establish Audit System

An audit system which leads to excellent audit should to be developed to ensure that auditors do their work on a regular basis. In the audit system, auditors should clearly know what the audit objects are. The main focus is the enterprise's information protection e.g. the server, backbone switches, routers and security devices.

2. Focus on Safety Auditors fostering

Safety audit involves massive products and wide content. The basic information of the audit come from operating system, network systems, security devices, applications system etc. Security auditors are not only necessary to understand the operating system knowledge, but also should be familiar with network protocols, database, virus infection mechanism. Moreover, auditor should understand the basic situation of application systems as well as master the work principles of servers, switches and security devices, especially the understanding a variety of security policies of information systems deeply. Thus, in the audit processing, the analysis of massive information can be developed and the observing and thinking ability can be cultivated.

However most of the enterprise information system security audit work is just begin without any own professionals. Although security audit can be conducted by professional security company or buying excellent audit software, it is still harmful in term of the safety and long-term development. The audit process involves a number of important enterprise information especially the system security weaknesses. Serious threaten will occur when criminals turn up or the workers are in low ability to analyze the weak links. From the above analysis, the security audit work should be accomplished by the professionals in the enterprises.

From the angles of security audit requirement for auditors and situation enterprise internal personnels, the enterprise should lay emphasis on the foster of information system administrators, network administrators, security guards especial the safety audit personals. Because all security policy, security system and security measures are developed by human beings, personnels with high quality and ability are required in developing management standards of enterprise information system and ensuring enterprise information security.

3. Reasonable Structure of the Security Audit System

The premise of improving the safety audit is to build a security audit system in line with business needs. In building a security audit system, the follow issues should be considered:

a. The profundity and scope of audit. The audit profundity and scope determine the complexity of the audit system, which are also the basis of audit products selection.
b. Problems of data sources. An audit system operation is based on data from the system at all levels, and how to obtain the data sources of audit system is the most critical issue.
c. Relationship with the original systems. To ensure the normal operation of the original system go smoothly in the realization of the audit, and the least modification and the minimum impact on system performance make the audit perfect.
d. Eliminate of audit function ignore. If the audit system is easily bypassed, it would lead to serious problems.
e. Effective utilization of audit data. In establishing an audit system, the lack of deep utilization of audit data will lead to weak audit system effection.

Network security is accompanied with the production of computer, especially the present popular network, security problem is emphasized by at all levels of sectors and industries especially the area of intranet security. Network security is a huge and complex dynamic system, hardware equipments provide basic security for the network, but a system which continues to improve can find a kind of dynamic equilibrium only with the help of network security audit system by doing real-time audit and effective evaluation to the system which has been established and discovering the potential safety hazard in time. These problems will become hot spots for future security research in building a solid and reliable network security audit system.

Computer network security audit is a very complex and extensive research subject, as an indispensable part in integrity security framwork, it is a complement for a firewall system and a intrusion detection system. It involves a wide range of knowledge. With the complexity of computer operating system and network communication technology increasing, the complexity of network security audit is also increasing. How to improve network security audit system performance of various technologies and how to build a strong network security audit system need to further constantly explore and research.

## 5. Conclusion

Web and database technologies are in a rapid evolution roadmap, for example web3.0 and graph database (Angles, 2008) are getting more and more attention. At the same time, related security issues will appear, but the fundamental security rules will remain the same. In this chapter, we briefly overview the advanced web and database securities, the security design principles and security audit rules and methods. Due to the limitation of chapter length and variable programming languages, most contents in each section are general guide lines and rules. When deploying practical information systems, we need to map those rules to real implementation. Information systems can be more secured if we know and apply those technologies.

## 6. Acknowledgment

This work is partially supported by Zhejiang Provincial Youth Natural Science Foundation (Q12F020022) and Zhejiang Educational Foundation.

## 7. References

Jesse James Garrett (Feb 2005). Ajax: A New Approach to Web Applications. Available from http://adaptivepath.com/ideas/ajax-new-approach-web-applications

Raducanu Razvan & Moisuc Maria (2010). The security of Web 2.0 and digital economy. *Recent Advances in MATHEMATICS and COMPUTERS in BUSINESS, ECONOMICS, BIOLOGY & CHEMISTRY*. pp.168-170, ISSN: 1790-2769

D. Crockford (July 2006). The application/json Media Type for JavaScript Object Notation (JSON). *RFC 4627*, July 2006

Anley C. Advanced SQL Injection in SQL Server Applications. *Next Generation Security Software Ltd*. 2002.

Halfond W G ; Viegas J and Orso A (2006). A classification of SQL-injection attacks and countermeasures. *Proceedings of the IEEE International Symposium on Secure Software Engineering*, Mar. 2006

Renzo Angles, Claudio Gutierrez. Survey of graph database models. *Journal ACM Computing Surveys*, Volume 40 Issue 1, February 2008

# 2

# Cyber Security

Barry Lunt, Dale Rowe and Joseph Ekstrom
*Brigham Young University*
*USA*

## 1. Introduction

Prior to HTML, browsers, and the WWW, computer interconnections were localized and limited. Since the early 1990s, web technologies have made it easy for everyone to access and post content on the Internet. Before long, there were thousands, then hundreds of thousands, and soon tens of millions of computers, all connected together via the Internet. As noted by Robert Metcalfe, and as later codified in what became known as "Metcalfe's Law", the value of a network goes up as the square of the number of users. Regardless of whether we accept his exact quantification of the value, there is no question that a few interconnected computers are more valuable than the same computers not being interconnected, and that many (or all) computers being interconnected has much more value than only a portion of them.

This is the situation today: essentially all desktop, notebook, netbook and tablet computers are interconnected via the Internet, and the same is true for the majority of cell phones. Additionally, even a significant portion of embedded computers are being connected via the Internet, as well as most industrial control and monitoring computers. Suffice it to say that, if the trend continues, and the evidence is very strong that it will, most computers, mobile devices, and even embedded systems either are or soon will be connected via the Internet.

While this has dramatic advantages for a free and open society, there has always been an element of society that would attempt to take advantage of this openness in ways that are damaging to other computers, users, the data, or to society as a whole. The need to protect our computers, users, data, and society, from this type of abuse, is the field of information assurance and security.

## 2. Guarding our information

Most businesses today would recognize the need to follow the most economical path to maximum profit. Frequently an organization's profit margins form the primary indicators as to their success. Even government agencies must admit to being somewhat cost-driven. With the recent economic downturn and increased competition to stay one foot ahead, businesses may be tempted to consider security as an afterthought, rather than an integral part of their business models and practices. In this Chapter we will look at some of the devastating implications of this error and why every genre of organization must place security at the forefront of business planning and practice.

Consider the owner of an expensive luxury vehicle who, each day outside his workplace, leaves his doors unlocked, with the keys in the ignition. The foolhardiness of the owner is apparent, and some readers may go so far as to suggest he would deserve to have his vehicle stolen. Yet in our modern information-driven organizations, corporations and agencies that depend on their information and data in their day-to-day operations often omit security entirely from consideration. At best it is an afterthought, akin to putting a 'do not steal' sign on the aforementioned vehicle and hoping this will deter all potential criminals.

In 2010, for the first time, the worldwide cost of information and electronic data theft (excluding piracy) rose 9.3% from 2009 to surpass all other theft (Kroll, 2010). In the UK alone, the cost of cyber-crime to businesses, individuals and government cost $43 billion US dollars (2010). In the 2011 series of cyber-attacks against Sony, some analysts believe the long-term costs to be in excess of $24 billion (Sebastien, 2011). Staggering as these figures are, the truth remains that most of these breaches could have been prevented had security been integrated into the victim's plans and policies.

It may be hard to understand why cyber-attack costs can reach such staggering figures. It can often come as a surprise to a victim that the true cost of an attack can far exceed the cost of hardware technology assets, or an annual IT budget. Indeed, the failure to comprehend the true risk of attacks and associated costs is in part what has led to such a prevalence of successful breaches. To be secure requires more than a retrofitted firewall installed merely as an afterthought. Organizations must understand the true cost, impact and consequences of cyber-attacks in order to identify what steps should be taken to protect their most valuable assets.

## 3. Visualizing the cyber-landscape

The first step in better understanding cyber-attacks is to become aware of how intricately connected information systems and technology have become. A system should not be thought of as a series of devices connected by wires, but rather a combination of people, technology and networks that function within defined parameters to achieve a specified objective. As organizations begin to view their systems from this perspective, it becomes obvious why few technical measures, even if expensive and state-of-the-art, may be ineffective in ensuring their protection from a cyber-attack.

Some academics have claimed that cyberspace is defined more by social interactions than technical implementation. Morningstar and Farmer argue that the computational medium in cyberspace is an augmentation of the communications channel between real people (Morningstar & Farmer, 2003). This concept of a socially interconnected system of systems was further visualized in an IBM video published on YouTube in 2010 ("Smarter Leaders vPanel: Tackling Urban Traffic with Social Computing", YouTube, 2010; http://www.youtube.com/watch?v=-thvI-IjwgY). These interconnections between social computing and cyber-security are perhaps the most overlooked aspects in providing effective security. From a defensive standpoint, we should treat cyberspace as the nexus that allows for the potential and very real connections among international organized crime, terrorists, hackers, foreign intelligence agencies, military and civilians.

The balance between usability and security is a fundamental concept that encourages security professionals to be mindful of the user needs. Even so, the visualization of social interactions using technology presents a new challenge for those responsible for cyber-security planning. Understanding the possible motivations and means behind a cyber-attack can better equip enterprises to prepare for and respond to an attack. Research has shown that on average, the cost of cyber-crime is reduced by 38% by companies which implement Governance, Risk Management and Compliance (GRC) measures across their enterprise (Ponemon Institute, 2011).

The mistake of assuming security is someone else's problem often comes with tragic consequences. It is not the responsibility of engineers, consultants, IT professionals or even management to undertake alone, but is the responsibility of every user. Granted, there are many specific roles required in security planning, but if the plan does not include each and every user as a member of the security team, it will be doomed before it has even been implemented.

The domain of cyber-security is highly subject to external pressures. These definitional forces include the following (Agresti, 2010): 1) *Rebranding exercise* – the former term "information assurance and security" is being replaced by "cyber-security", as the term "cyber" creeps further into many technologies of our era; 2) *organizational imperative* – the Internet has become essential for most modern companies; 3) *cyberspace domain* – this portion of our lives is now ubiquitous and pervasive and must be understood from that perspective; and 4) *national defense priority* – our potential vulnerability to cyber attacks is of increasing importance.

Focusing further on the last of these definitional forces – *national defense priority*, Agresti states:

"Progress in cybersecurity depends on attaining a richer, more quantitative, and more visually rendered understanding of cyberspace's size, scope, contours, composition, architecture, properties, traffic patterns, oversight, end points, and – ultimately – its vulnerabilities to malicious activities.

"The national defense sector faces the entire spectrum of security challenges: defects and malicious code in software on individual workstations, insider threats, vulnerabilities in networks, malicious intent, and attribution." (Agresti, p. 103)

## 4. The cyber-security arena

In 2010, the US government received on average, 60 million attempted cyber-attacks per day. This problem is not limited to government; Facebook recently announced that it receives over 600,000 attempted cyber-attacks per day (Enzer, 2011) although this is small compared to their one billion daily logins. Staggering as these numbers are, they show the volume of attacks that companies and even individuals now face while connected to the Internet. Managers, IT professionals, and IT security professionals must take a holistic view of security in their planning. This is crucial if a company is to survive amidst today's onslaught of cyber-attacks.

The cyber-security arena has expanded dramatically. Cyber-security now includes mobile phones, embedded computers (widely employed in our infrastructure), cloud computing, and all types of data storage. And cyber-crime has become a business, operating without borders, and has become increasingly difficult to arrest (BCS Security Forum, 2010).

## 5. The motivation for cyber attacks

An important question to ask IT professionals is how they would make a system completely secure from any and all attacks. The authors have done this in a number of settings, and have observed that many individuals reach the conclusion that disconnecting all network interfaces and powering down the system is the only way to ensure security. However, physical security – when viewed as a component of cyber-security - suggests that the only true way to be totally secure is to not exist in cyberspace at all!

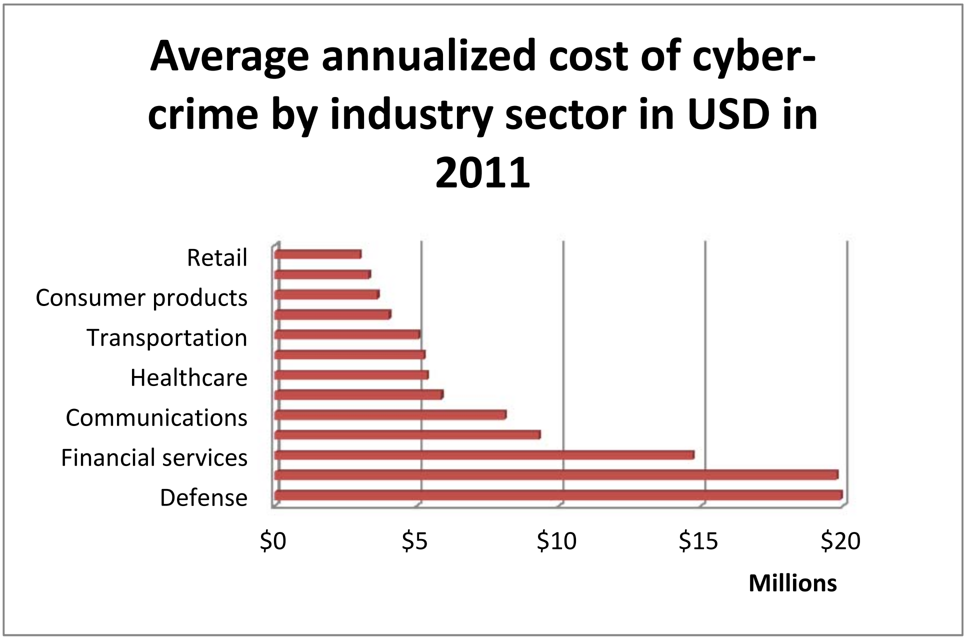


Table 1. Cost of Cyber Crime. Data sourced from the 2011 Ponemon Institute Research Report

The point of this discussion is to illustrate that every system is a target. Information is one of our most valuable assets and wherever it is stored, transmitted or processed it becomes a target for cyber-attackers. In November 2011, the US Office of the National Counterintelligence Executive drew significant media attention for their 2009-2011 congressional report which reported the headline "Foreign Spies Stealing US Economic Secrets in Cyberspace". (Security Counterintelligence, 2011). The 31-page report is rife with examples of industrial espionage by foreign state actors using a variety of cyber-attacks. Among the responses from the accused, the Chinese government pointed out the West's tendency to outsource information technology projects to Asia. This tit-for-tat accusation and response is by no means a recent development, with China and Russia frequently mentioned in counterintelligence reports over the last twelve years (ibid).

Very few press reviews of the preceding congressional report covered the alarming discussion of a growing number of other unnamed countries, as well as activist and terrorist groups, who are increasing in their cyber-attack capabilities. Today we see an evolving marriage of capability and intent from groups with far more sinister objectives than espionage. Some of these groups have sought to wreak havoc and damage critical infrastructure and have shown no aversion to the taking of human life.

Until recently, the principal threat to the private sector has been from 'traditional hackers' – skilled individuals seeking information freedom, money or fun, and 'script kiddies' – using tools created by others primarily for personal entertainment. The aforementioned events and reports indicate a very obvious shift of intentions. The cost-to-benefit tradeoff of a successful cyber-attack, and the availability of the internet as a delivery mechanism, effectively arms the masses. With the right skills, anyone, anywhere, can launch a potentially devastating cyber-attack. Several of these attacks were discussed in a recent whitepaper that analyzed the cyber-attack capabilities and vulnerabilities of Libya under the anti-Gadaffi uprising (CSFI, 2011). For example, a SCADA targeted cyber-attack against Libya's oil refineries could limit Gadaffi's funding, but risks severe economic damage to already-struggling countries such as Italy and Ireland who are dependent on Libya for most of their oil.

In the last few years, studies have highlighted the vulnerability of critical infrastructure to cyber-attack. Nuclear plants, electric smart-grids, gas pipelines, traffic management systems, prison systems, and water distribution facilities have all been identified as at risk from a cyber-attack. Fortunately at the time of this publication, actual attacks like these remain the subject of academic discussion. Many security analysts fear this situation will be short-lived.

It should be clear by now that there is no such thing as an uninteresting target for cyber-attackers. We know that certain industries and organizations may be targeted more persistently and receive more attacks than others, but should realize that every system and organization is at risk. Understanding the motivation an attacker may have to attack our systems can help us to be more prepared for the eventuality of an attack.

In summary, the motivation for cyber-attacks may include:

- Intellectual property theft
- Service disruption
- Financial gain
- Equipment damage
- Critical infrastructure control & sabotage
- Political reasons
- Personal entertainment

In the next section, we shall see how recent cyber-attacks are being targeted to realize these objectives and describe their potential impact to information systems and organizations.

The actors that typically have these motivations can be categorized as: organized groups; loosely-organized groups; and lone wolves. These categories are points in a continuum.

An example of an organized group would be the espionage organization of a nation (such as the CIA); an example of a criminal organized group would be the Russian Business Network. These groups are typically highly organized, they pursue specific objectives, and they are well funded.

More recently, there has been a surge in the category of loosely-bound groups with varying motivations. Some of the best-known of these groups include Lulzsec and Anonymous. Collectively these groups are responsible for dozens of the highest-profile attacks in recent times (Wikipedia, 2012). Indeed, many of the aforementioned attacks against Sony came from one of these groups (Security Curmudgeon, 2011). Their targets range from governments, to corporations, to religious institutions (to date having hacked the Vatican twice). Self-labeled as part of the 'Antisec' movement, they encourage other groups to join their cause and represent a politically and geographically diverse group of individuals with skills ranging from basic script kiddie, to more advanced exploitations. Recently, a new group known as The Consortium (BBC News Technology, 2012) claimed affiliation with Anonymous in a hack against a pornography website resulting in the loss of subscriber information. While some may argue that these groups have political motives, it appears that they seek organizations with a low-security profile to publically embarrass at every opportunity.

A lone wolf or solo hacker, often incorrectly stereotyped as a basement-dwelling spotty teenager, can in some instances pose an equal threat. An example of the lone wolf includes the case of the Scottish systems administrator, Gary McKinnon, and is perhaps one of the more famous of these. Driven by self-curiosity he hacked into multiple US government agencies before being apprehended (Boyd, 2008). Such hackers are greatly assisted by organizations or individuals that provide tools for creating malware.

## 6. Cyber-attack types

In a sample study of 50 organizations conducted in 2011, researchers found that on average a successful cyber-attack occurs over than 70 times per year, or on average, 1.4 times per week. This represents an increase of 44% from 2010. If this growth continues, fifteen years from now organizations will be responding to a **successful** attack every 30 minutes (Ponemon Institute, 2011).

The exact type of attack can vary in type and sophistication. Fortunately, many of these attacks are fairly simple in nature. Automated vulnerability probes along with known and recognizable self-propagating malware (worms) form the bulk of attack attempts. These are generally easy to detect and prevent using standard off-the-shelf firewalls, and intrusion protection/detection systems. The primary danger in these attacks is the noise they generate, which can make it difficult to locate the more serious threats. In excess, however, they can constitute a Denial of Service (DoS), or Distributed Denial of Service attack (DDoS), leading to a much more serious degradation of service, unpredictable behavior and even complete loss of service. Although relatively infrequent, DoS and DDoS attacks are one of the most costly types of attack.

Another type of cyber attack against infrastructure is stealing Internet access. An example of this type of security compromise is the case of Ryan Harris, the owner of TCNISO. His company produces products that enable users to steal Internet service (Poulsen, 2009).

One very successful form of attack today focuses on exploiting vulnerabilities in websites and web applications. These attacks pose the greatest danger to most organizations due to the relative simplicity with which they may be attempted and with the immense volumes of valuable information that can be stolen if successful. Many websites are connected to backend databases, which not only contain information that may be of interest to criminals, but provide an entry point into the organization's internal network. The latter form of attacks are known as pivoting attacks and enable the attacker to pivot from a principal entry point to attack other systems deeper in an organizations infrastructure. Pivoting attacks are a severe form of web-based attacks as they allow attackers to completely bypass perimeter security controls at the network edge.

Web attacks involve the attacker identifying a potential vulnerability in a web system. There are several types of vulnerabilities that allow for different forms of attacks. The most common of these are cross-site scripting (XSS) and SQL injection.

Cross-site scripting allows an attacker to plant malicious code in an organization's website and from there attack clients visiting a company's site, stealing passwords, subverting network traffic, and monitoring communications. In many instances, XSS attacks enable attackers to leverage further vulnerabilities in client web browsers to install malicious software on the visitor. Thus unknowingly a visitor of an infected site can become themselves infected, and in some instances, part of a group of infected computers known as a botnet. This form of client infection is known as a drive-by-download and is one of the principal ways attackers gain control of systems. Controlled systems can be used for a variety of purposes including sending unsolicited e-mails (SPAM), targeted cyber-attacks against organizations, and DDoS attacks. Using a victim's system to attack another victim is known as an indirect attack and can be done with relative anonymity.

The vulnerability to these type of attacks can be easily reduced by careful website programmers who include checks to validate the length of user-entered information, and remove any illegal characters. Failing to do this introduces a significant probability that the site is vulnerable to both cross-site scripting and SQL injection attacks.

An SQL injection permits the attacker to access and manipulate a backend database, revealing customer records, intellectual property and even opening routes deeper into the organization's network. Most experts agree that SQL injection attacks were used in most of the 21 independent successful attacks against Sony that occurred between 21 April and 7 July 2011 (Security Curmudgeon, 2011). Targeted attacks of this nature currently form the majority of successful cyber-attacks and are the most cost-effective for attackers.

A further category of attacks are known as Advanced Persistent Threats, or APT's. These attacks are becoming more common as attackers become more skilled, knowledgeable and resourceful in infiltrating specific networks for a specific purpose. Their title reflects the danger posed by these attacks. APT's can be technically advanced and contain advanced attack techniques, use an advanced combination of simpler attacks for a specific purpose, or both. They are persistent, indicating that the attacker has a defined objective and often will not quit until their goal is realized. This can often lead to attacks being multi-pronged, where the organization's systems and security are studied and monitored for months before an actual attack, or series of attacks, take place. APT's pose a significant threat with a high probability to succeed and be damaging to an organization. This can indicate external funding or support that provides resources for the development and deployment of the attack.

The only positive aspects of APT's are that they are targeted against a specific organization and hence are much less prevalent than other threat types. In other terms, they are akin to the sniper who studies his prey and observes its habits. The sniper waits, sometimes for days, for the perfect moment to take his shot, with a high degree of accuracy. It is very difficult to locate the sniper before the attack, and after the attack, the damage is localized but still significant, and often costly. Non-APT attacks in contrast may be thought of as 'the shotgun approach', or 'spray and pray' tactic of many video gamers. The attacker will point in a general direction, and blast away, hoping to hit something. With enough shots, a kill is guaranteed. These attackers generate a lot of noise, and can do a lot of damage if they are lucky enough to land a hit. If unsuccessful, an attacker will often move on to another target. Success at a low cost, against any target, is more important than any specific target.

Understanding the type of attack in the context of its objective and sophistication allows those responsible for information systems to gain insight to the potential damages caused. This next section looks at some of the costs a cyber-security breach can incur.

## 7. Cost of a successful cyber-attack

By our nature, humankind often finds it easier to respond or retaliate than to plan and prepare. Analyzing every potential outcome of a scenario can consume significant time and resources. Surely it is cheaper to only respond to the successful cyber-attacks than commit resources to risk management and incident response?

Recent history has shown this idea to be erroneous. It is often impossible to calculate the precise damage of a cyber-intrusion. The consequences of an attack can be far-reaching and long-term. The damage may often be irreparable; no amount of money can undo what has been done. Some of the effects of a cyber-intrusion include:

- Financial loss from service unavailability
- Loss of customer/client confidence
- Market shift to competitors
- Lawsuits and liabilities from those who have had information stolen
- Cost of recovery
- Cost of security measures to prevent a repeat attack
- Cost of staff or consultants to investigate and identify the method of attack
- Fines from regulatory bodies
- Cost of informing customers of theft
- Theft of intellectual property
- Loss of human life

The effects of cyber crime are listed above. Some previous sections have said other things about cost in specific instances. Many successful cyber attacks have been widely reported in the media, yet the frequency of successful cyber attacks continues to increase, along with associated costs.

In their second annual report, the Ponemon report (Ponemon Institute, 2011) had the following key takeaways:

- Cyber crimes can do serious harm to an organization's bottom line. We found that the median annualized cost of cyber crime for 50 organizations in our study is $5.9 million per year, with a range of $1.5 million to $36.5 million each year per company. This represents an increase in median cost of 56 percent from our first cyber cost study published last year.
- Cyber attacks have become common occurrences. The companies in our study experienced 72 successful attacks per week and more than one successful attack per company per week. This represents an increase of 44 percent from last year's successful attack experience.
- The most costly cyber crimes are those caused by malicious code, denial of service, stolen devices and web-based attacks. Mitigation of such attacks requires enabling technologies such as SIEM and enterprise governance, risk management and compliance (GRC) solutions. (Executive Summary, p. 2).

The time it takes to resolve a successful cyber attack is a key factor in the cost. The sooner the organization detects, analyzes and contains the attack, the lower their recovery and post-recovery costs will be, and the lower the overall cost will be. Therefore, it is important that all organizations constantly be on the alert against cyber attacks.

Table 2, taken from this Ponemon report, gives the average annualized cyber crime cost, weighted by the attack frequency. While the institutions studied in this report are not necessarily representative of the industry as a whole, the data are highly informative.

| Type of Attack | Average Annualized Cost |
|---|---|
| Denial of service | $187,506 |
| Web-based attacks | $141,647 |
| Malicious code | $126,787 |
| Malicious insiders | $105,352 |
| Phishing & social engineering | $30,397 |
| Stolen devices | $24,968 |
| Botnets | $1,727 |
| Malware | $1,579 |
| Viruses, worms, trojans | $1,517 |

Table 2. Types of attacks and their associated costs (Ponemon Institute, 2011).

Table 1 includes only the direct costs. There are also indirect costs, including increasing frustration on the part of computer users, increased time spent on working with necessary security measures, lost business opportunities, and a tarnished reputation.

Worldwide, cyber crimes have cost in the neighborhood of $388 billion in 2010, according to the 2011 Norton Cybercrime Report (Norton, Inc., 2011). This figure includes both direct and indirect costs, and is a staggering amount. And unfortunately, this figure has only been increasing for the past several years, with no sign of major improvement.

## 8. Interested parties

The list of interested parties has grown in direct proportion to the number of connected entities. It would probably be much easier to list those who need not be concerned with cyber crime, because IT has become an integral part of companies, government agencies, the military, our infrastructure (including water, electricity, roads, bridges, natural gas, etc.), health care, research, and our personal lives, and because a very large portion of all IT is now connected via the Internet. Anyone who has any device connected via the Internet – and this includes cell phones, MP3 players, computer gaming equipment, and ALL forms of computers – anyone with any of these devices should be concerned about cyber crime. All of these devices have fallen victim to cyber crime, and it is unlikely that this will change.

Perhaps the parties most difficult to convince of this are the general public. Only a small percentage of the population is deeply aware of how easily the security of their connected devices can be compromised. And because only a small percentage are affected by cyber crime each year, the general public remains relatively uncommitted to deploying the latest and best security software. And as long as this remains true, it is inevitable that cyber crime will continue to increase.

## 9. Outstanding issues

The most difficult issue in any defensive endeavor is knowing what to defend against. One of the most famous large investments that failed to protect against the real risk was the Maginot line. It was a set of fortifications and tank obstacles designed to give the French time to mobilize. The line proved useless for defending France because the Germans simply conquered Belgium and went around the defenses. This led to the adage that "generals always fight the last war, especially if they have won it" (Kemp, 1988). Unfortunately there are numerous examples of companies that sit behind their line of firewalls believing that they are safe while the enemy simply goes around their defenses.

The canonical example of this class of problem in cyber-security is called a “zero-day exploit”. An exploit is actual code that acts on a vulnerability or combination of vulnerabilities. A zero-day exploit is an exploit that takes advantage of vulnerabilities that are unknown or considered to pose insufficient risk to worry about; then a malefactor figures out an automated exploit and hundreds or thousands of computers are compromised in a few hours. The computers join a botnet. We can see that an opponent that exploits a vulnerability the first time has the advantage of surprise. No matter how rapid the response by software developers and security vendors to a zero-day exploit, the black hats have a significant window of opportunity to attack vulnerable systems until a remediation and/or a signature for the malware is deployed to the defenses on the platform. Cyber-security will always be a race between malefactors who want to compromise systems and the vendors, developers, and legitimate users of computing systems who want to secure their systems.

A major hurdle is that decision makers often think like the French government before WWII, they think their large investment in firewalls will protect them while the reality is that new software and hardware are continuously being deployed to add functionality and remediate vulnerabilities and no static defense can provide protection in a dynamic environment. Experience teaches that the fixes often create new vulnerabilities. At the same time malefactors are continuously searching for vulnerabilities and creating exploits for the vulnerabilities that they isolate. Thus the problem becomes one of continuously defending a relatively slowly changing target from an unknown, rapidly moving and evolving attacker.

In the current world of IT, attackers have a huge advantage. The majority of machines deployed in businesses and homes run the same platform software. Microsoft platforms got the reputation for having poor security because their platform provided a large set of targets that made the value of an exploit much greater. Finding vulnerabilities and developing exploits is a technically demanding and uncertain process. A large monoculture to attack provides the incentive to invest in exploits. There is now an active underground market in zero-day exploits that are sold to the highest bidder. An active market provides incentives for skilled individuals to invest time and expertise to create "products" that are in demand.

## 10. Implementing security

A likely question at this stage is what can be done? How can we realistically and affordably protect our information under this continuous barrage of attacks? Often in these circumstances, managers may find themselves facing the responsibility to choose between large numbers of different technology-based solutions. This can quickly overwhelm, and actually create more problems than it solves. In order to implement effectual security controls, we must first understand the risks posed by different threats to our business model.

There is no shortage of security frameworks for analyzing risk and implementing security controls, and plenty of excellent books for a variety of audiences on this topic. For the purposes of this chapter, we shall present security implementation from a greatly simplified model that should enable an organization to effectively prepare and respond to security threats.

The Cyber Security space can be broken down into three areas, or domains. These are:

- Prepare
- Defend

- Act

These domains should not be seen as sequential steps in which each is terminated prior to the commencement of the next, but rather three continual processes that form the foundation of organizational security.

### 10.1 Prepare

Preparation includes planning, risk assessment, policy, business continuity planning, countermeasure deployment, training, education and accreditation. These are all essential in optimizing our readiness for cyber attacks.

Accreditation is a particularly interesting term in this context. Security accreditation is management acceptance of the risks associated with a system. This is no small responsibility in the event of an attack. To increase assurance and reduce associated risk, a thorough penetration test should be carried out as standard part of an accreditation process. Conducting a penetration test is effectively paying someone to hack your organization's systems. A skilled penetration tester will be able to locate vulnerabilities and advise on cost effective ways to reduce their risk. Organizations should be careful of individuals marketing themselves as penetration testers without the appropriate skills. A tester should carry recognizable certifications (GIAC, CEH, etc.) and be a member of an accredited or approved organization (such as (ISC)2) that requires a member code of ethics.

After the test, a report should be provided which will indicate the specific vulnerabilities found with suitable fixes, and recommend process improvements that will reduce the risk of future vulnerabilities going unchecked.

### 10.2 Defend

In the context of defending against cyber attacks, defensive processes include ongoing risk mitigation, service and device hardening, and incident detection. A recent study (Schwartz, 2011) showed that up to 96% of organizations are unaware they have been hacked. Believing themselves either untargeted or immune to cyber attacks, they remain blissfully ignorant of information theft, espionage and other malicious attacks taking place right under their noses. Organizations must ensure, at a bare minimum, that they are able to detect security incidents when they occur. To fail in this opens the door to potentially expensive lawsuits and even criminal proceedings depending on the type of information that has been lost.

### 10.3 Act

Finally, we should establish procedures and protocols to ensure that in the event of an incident we act appropriately. We avoid the use of the term 'react', as it tends to carry a negative connotation of a knee-jerk 'reaction' that is ill conceived and inflammatory. Actions in response to a cyber-attack should be carefully planned to facilitate the effective response that minimizes expense and collateral damage. The word act is hence deliberate and suggests that organizations should be proactive rather than reactive.

The continual application of these three domains cannot be emphasized enough. External consultants who are experienced, certified security professionals can be invaluable resources in maintaining an effective cyber-security posture and ensuring our businesses remain unhindered by an attack they were unprepared to handle.

## 11. Current research

Historically the attackers have also had the advantage that the majority of home PC owners and many businesses have been lax in applying fixes and upgrading their platform software. Thus attackers can have years to find and exploit vulnerable machines. Buffer overflow and other code injection attacks often depend on the static layout of the code and data in memory for their effectiveness. Historically network risks were mitigated by building a fortress around systems. This approach led to network architectures with components with names like DMZ (Demilitarized Zone), a boundary location that has both public and private addresses so that "bastion hosts" could be hardened to live in the DMZ while normal systems would be deployed behind the "firewall". This provides a static environment that allows an attacker almost unlimited time to search for a vulnerability in the attack surface. The advent of APT attackers that patiently probe for years against a target of particular interest make these fortress designs vulnerable. Just as WEP-based wireless networking was vulnerable to attack because it used static encryption keys, static networks that can be mapped over time are more vulnerable than more dynamic designs.

In order to defeat these threats in a slowly evolving infrastructure, some new products and research results demonstrate that significant gains in security can be achieved by adding random dynamic behavior to systems. Starting with Windows Vista and improved in Windows 7 and Server 2008 SP1, the operating system loads the parts of the operating system into different random locations every time it boots (Microsoft, 2011). Microsoft does not claim that this eliminates the threat of attacks - it just makes it significantly more difficult.

Vendors have begun to sell network appliances that randomize the footprint of the network by using Network Address Translation (NAT) technology and randomizing outbound connections over a set of IP addresses, as well as other dynamic behavior (Masking Networks, 2011).

The military is looking at many similar approaches to improve the security of its networks, especially combat control systems (Baker et al, 2011; Jones, 2011; Okhravi, et al, 2011; Wright, 2011). In November 2011, the Defense Advanced Projects Research Agency (DARPA) announced plans to increase cyber-security research by 50% (Hoover, 2011).

The next generation of networks may be significantly more robust, as could hardware and software systems. This will probably be accomplished by introducing more and more random behavior into the operational characteristics of systems which will overcome many of the disadvantages of our current environment of the majority of systems being identical platform software deployed on identical hardware connected in static networks running on a single vendor's equipment (Jajodia, et al, 2011). Much will depend on decision makers recognizing the threats and being willing to invest both intellectual and financial capital in understanding the risks and applying appropriate defensive technologies.

## 12. Conclusion

At this point in time, the outlook for cyber security is not as rosy as the authors would prefer. Attackers continue to find new ways to exploit weaknesses, while developers continue to fix the known problems and attempt to develop new operating systems and applications with fewer vulnerabilities. Because there is very ample motivation (the chance for success is high), and because the possibility of being caught is relatively low (the risk is not terribly high), the area of cyber security continues to attract many black hats with many motivations.

In many senses, it is just like physical (aka "kinetic") warfare. As soon as one side develops a new weapon, the other side begins to develop a counter-weapon or a work-around. Additionally, the more aggressive side continues to probe the target for all possible points of weakness, and exploits these weaknesses when found. History has shown that this cycle of probing, exploiting, developing, and counter-developing can continue *ad infinitum*. Our belief is that the white hats will continue to make life in the cyber-world tolerable, minimizing risk and continuing to make improvements that provide major advantages to offset the associated problems. It will take a great deal of effort by many parties to keep cyberspace widely useful.

## 13. References

Agresti, William W., "The Four Forces Shaping Cybersecurity", *Computer*, Feb 2010, pp 101-104.

Baker, Jill, et al., "Winning in Cyberspace: Air Force Space Command's Approach to Defending the Air Force Network", *High Frontier*, v 7 #3, May 2011.

BBC News Technology, "Porn site breached in hack attack", Mar 12, 2012, http://www.bbc.co.uk/news/technology-17339508.

BCS Security Forum, *ISNOW*, Winter 2010, pp 6-13.

Boyd, Clark, "Profile: Gary McKinnon", BBC News, July 30, 2008, http://news.bbc.co.uk/2/hi/technology/4715612.stm.

CSFI (Cyber Security Forum Inititive), "Project Cyber Dawn – Libya", Apr 17, 2011, www.unveillance.com/wp-content/.../Project_Cyber_Dawn_Public.pdf.

Enzer, Georgina, "Facebook unveils extent of cyber-attacks on site", *ITP.net*, Oct 30, 2011, http://www.itp.net/586887-facebook-unveils-extent-of-cyber-attacks-on-site.

Greene, Tim, "Cybercrime costs rival those of illegal drug trafficking, *Network World*, Sept 7, 2011, http://www.networkworld.com/news/2011/090711-cybercrime-250580.html.

Hoover, J. Nicholas, “DARPA Boosts Cybersecurity Research Spending 50%”, Information Week, Nov 7, 2011, www.informationweek.com/news/government/security/231902495.

Jajodia, S.; Ghosh, A.; Swarup, V.; Wang, C.; Wang, X.S. (Eds.), Moving Target Defense: Creating Asymmetric Uncertainty for Cyber Threats Series: Advances in Information Security, 1st Ed, 2011.

Jones, Andrew T., “Preparing the Air Force for Computer Network Operations”, *High Frontier*, v 7 #3, May 2011.

Kemp, Anthony, *The Maginot Line: myth and reality*. Military Heritage Press. p. 14, 1988

Kroll (2010). Global Fraud Report. Global Fraud Report - Annual Edition. USA, Kroll Consulting.

Masking Networks, “Network Address Vulnerabilities”, white paper 2011, http://www.maskingnetworks.com/network-masking-technology/network-address-vulnerabilities

Microsoft, “Microsoft ASLR: Loading DLLs at a different location every boot”, Feb 9, 2011, http://blogs.technet.com/b/virtualization/archive/2011/02/09/windows-7-and-windows-server-2008-r2-sp1-add-new-virtualization-innovations.aspx

Morningstar, Chip and F. Randall Farmer. The Lessons of Lucasfilm’s Habitat. The New Media Reader. Ed. Wardrip-Fruin and Nick Montfort. The MIT Press, 2003. 664-667.

Norton, Inc., “Norton Cybercrime Report”, 2011, http://www.symantec.com/content/en/us/home_homeoffice/html/ncr/.

Okhravi, Hamed, et al., “Achieving Cyber Survivability in a Contested Environment Using a Cyber Moving Target”, *High Frontier*, v 7 #3, May 2011.

Ponemon Institute, “Second Annual Cost of Cyber Crime Study: Benchmark Study of U.S. Companies”, Aug 2011.

Poulsen, Kevin, “Feds Charge Cable Modem Modder With ‘Aiding Computer Intrusion’”, Wired, Nov 2, 2009.

Schwartz, Matthew J., “Most Businesses Don’t Spot Hack Attacks”, *Information Week*, Oct 5, 2011; www.informationweek.com/news/security/attacks/231900054.

Sebastien, M. (2011, 25 May, 2011). "Infrographic:  Cost of Sony's data hack could reach $24 billion." PR Daily. Retrieved 9/13/11, from http://www.prdaily.com/Main/Articles/Infographic_Cost_of_Sonys_data_hack_could_reach_24_8359.aspx.

Security Counterintelligence, Office of the National Counterintelligence Executive, “Foreign Spies Stealing US Economic Secrets in Cyberspace: Report to Congress on Foreign Economic Collection and Industrial Espionage, 2009-2011”, Oct 2011.

Security Curmudgeon, “Absolute Sownage: A concise history of recent Sony hacks”, June 4, 2011, http://attrition.org/security/rants/sony_aka_sownage.html.

U.C. Office, London, “The Cost of Cyber Crime”, 2010.

Wikipedia, “Timeline of events involving Anonymous”, http://en.wikipedia.org/wiki/Timeline_of_events_involving_Anonymous.

Wright, John D., “Air Force Cyberspace Strategic Planning Factors”, *High Frontier*, v 7 #3, May 2011.

YouTube, http://www.youtube.com/watch?v=h2br2_twHfw, 2010

3

# Development of an e-Learning Recommender System Using Discrete Choice Models and Bayesian Theory: A Pilot Case in the Shipping Industry

Amalia Polydoropoulou and Maria A. Lambrou
*University of the Aegean, Business School,*
*Department of Shipping, Trade and Transport, Korai, Chios,*
*Greece*

## 1. Introduction

The field of e-learning and self-learning has rapidly evolved during the past decade mainly because of major advances in telecommunications and information technologies, in particular the widespread use of web and mobile applications. Furthermore, the work environment conditions in most industries have become extremely demanding and competitive; therefore various forms of life-long learning appear to play an important role for employees' career development, as well as for companies' productivity improvement and human resources' efficiency. The flexibility, and cost effectiveness that e-learning offers is very significant, in most cases.

In an environment with abundant educational and training institutions the development of efficient new procedures to better guide students or trainees for selecting suitable learning materials is a challenging and open issue. In particular, the development of efficient e-learning recommender systems, such as an electronic training advisor who will help individuals in choosing the appropriate e-learning courses matching their particular characteristics, preferences and needs and based on their expected professional and personal development, is an open and promising research and development area (Adomavicius & Tuzhilin, 2005; Brusilovsky, 2002; Kim et al, 2009; Ricci & Werthner, 2006; Zanker & Jessenitschnig, 2006;).

An e-learning recommender system can help trainees overcome basic constraints that e-learning users increasingly face:

- the uncertainty and insecurity of the individuals that a particular educational offer will have the best possible outcome related to their career objectives.
- the difficulty of finding and choosing courses among a large number of courses offered by different institutions, universities and organizations, based on their individual training needs;
- the perception regarding the lack of interesting content or relative courses per case;

This chapter presents an innovative methodology for the development of a training advisor for e-learning environments. We consider e-learning personalization issues and present an e-learning recommender framework based on discrete choice models and Bayesian theory (Chaptini, 2005; Greene, 1993).

## 2. The context

The development of the presented intelligent training advisor, took place within the scope of the SLIM-VRT E.U. research project and was applied in the shipping industry. A field study was conducted for requirements data collection, with 5000 questionnaires and 24% response rate. The techniques used offer the advantage of a comparatively detailed, user-focused e-learning attributes modelling framework (advanced choice theory) and a competent system learning capability (Bayesian theory), that improves over time the performance of the recommender system itself .

In the following, we explain development of the advisor system in the context of involved and affected stakeholders, as shown in Fig.1, including:

1. the training society;
2. the firms and organizations;
3. the learners/trainees or end-users;

These groups interact in a common environment and the actions of one group directly affect the other. Obviously, these major groups of stakeholders in a particular market, for instance the shipping industry, operate in a distinct, dynamic Economic, Technological and Legislative environment.

The methodology we propose is implemented in the shipping industry, where seafarers have specific learning needs in order to adapt and perform successfully in a continuously changing organizational, business, and employment environment (Theotokas & Progoulaki 2007; Progoulaki & Theotokas, 2008; Progoulaki et al, 2005).

Common career paths in the shipping industry may require alternate career changes and job rotation such as acquisition of different positions onboard, work on different types of ships, and employment, ashore, at an office position related with the shipping and port operations management. The training process should take into account particularities of the seafaring profession such as having long intervals between two employment periods, being far away from the family when employed onboard, experiencing alienation form the broader society, etc. Thus, the context of the maritime e-learning is highlighted by those specific characteristics underpinning the shipping market and educational environment. Today, the following features are pertinent to the shipping workforce training aspects:

- There is a decline of interest in the seafaring profession and shortage of ex-ship officers for shore-based positions
- The impact of new technologies in terms of reorganization of crew duties is important
- Maritime personnel needs to be polyvalent
- The safety of life at sea and the protection of marine environment have been a basic concern of the maritime community

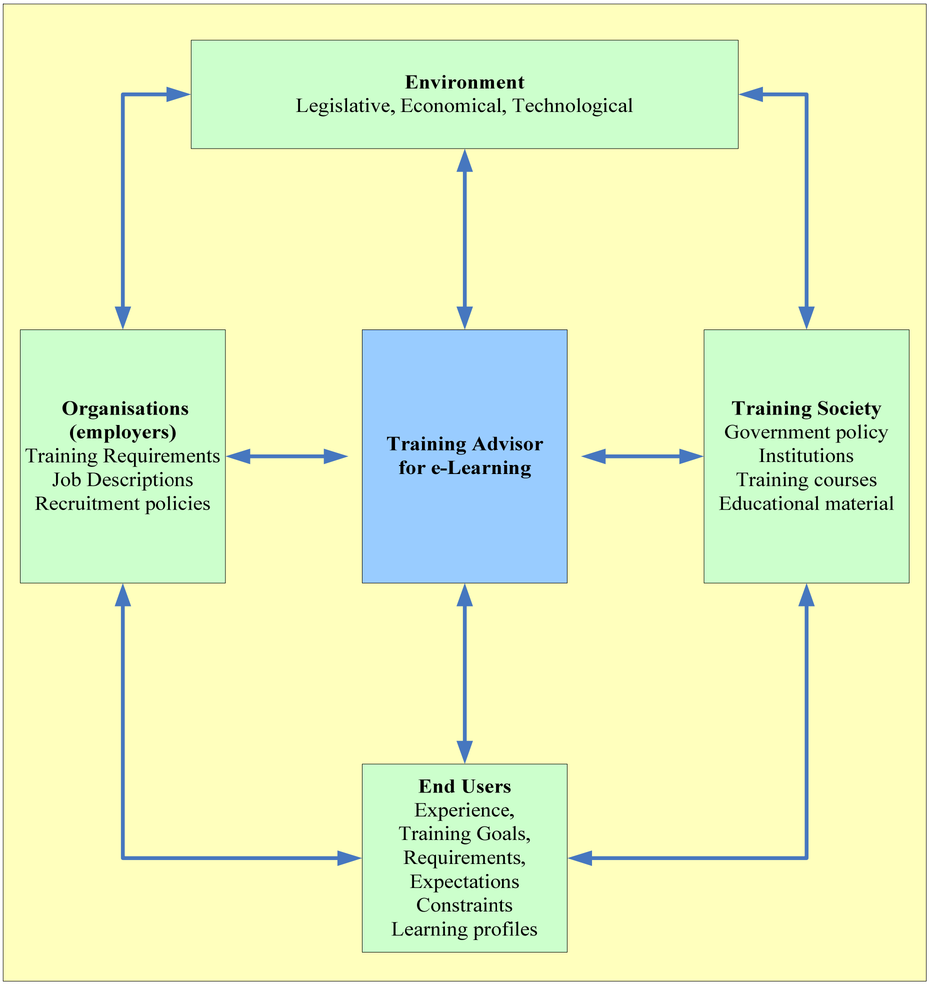


Fig. 1. Stakeholders in an e-learning Environment

- There is need to offer to the seafarers additional knowledge-skills as well the motivation to improve their soft skills, in particular (team work, leadership, communication, negotiation skills)
- There is need for applications that help the diffusion of tacit knowledge that crewmembers may have

To address these needs in a systematic and valid manner we firstly developed a tool for accumulating the basic knowledge regarding the seafaring profession and training environment, registered the information of the interviewed sample population during the data collection phase, and used statistical analysis to support the choices of each individual separately.

Four major groups of potential e-learning users are identified: a) Junior and Senior Captains; b) Junior and Senior Engineers; c) Marine students and d) Office personnel. Each group has special characteristics, interests and needs with regard to their career perspectives (continuous employment onboard, get promoted or switch career from the ship to the office), and respective training needs, which are specifically taken into account in the design of the recommender system.

Seafarers' individual preferences for course attributes and learning methodologies vary. The overall learning methodology they favour should comprise the following characteristics to meet the user needs: (a) Adaptive-blended learning (combined traditional lectures, e-learning in the office, e-learning on board with instructor); (b) Cooperative learning (practice on board, emergency drills with peers); (c) Contextual and participatory learning (via simulation, or Virtual Reality (VR) case studies); and (d) Provision of a variety of training material: textbooks, notes, videos, and VR cases.

In order to demonstrate the particular framework proposed hereafter, we developed an e-learning platform with training advisor capabilities for the shipping sector entitled SLIM-VRT. SLIM-VRT enables the existing and potential maritime students, employees, employers and authorities in the shipping industry to receive training in a way that is user-friendly, flexible, and learning effective. The overall SLIM-VRT system presents three major innovative dimensions: a) dynamic education program and content generation according to the user individual needs, especially in terms of the training advisor recommender capabilities, as explained hereafter; b) new pedagogical methodologies emphasizing collaborative and contextual learning, on the basis of case studies; and c) use of innovative and user-friendly information technologies emphasizing on VR tools.

## 3. E-learning recommender systems

Today, e-business applications and e-services are commonly taking advantage of advanced information and communication technology and methodologies to personalize their interactions with users. Personalization aims to tailor services to individual needs, and its immediate objectives are to understand and to deliver highly focused, relevant content, services and products matched to users' needs and contexts (Adomavicius & Tuzhilin, 2005; Brusilovsky, 2002; Ho, 2006; Kim et al, 2009; Kim et al, 2002; Ricci & Werthner, 2006;). In e-services personalization and web adaptation have been employed in many different ways: (i) the personalization service can be designed and used as an advice-giving system to provide recommendations to each individual and to generate up-sell and cross-sell opportunities (ii) personalization services are used to (dynamically) structure the index of information, product pages based on click-stream analysis to minimize the users' search efforts, where personalized content based on the user's profile is generated. The users can personalize not only the content but also the interface of the application used (Brusilovsky & Maybury, 2002; Burke, 2000; Rashid, 2002 ).

Applications of personalization technology are found to be useful in different domains. These include information dissemination, entertainment recommendations, search engines, medicine, tourism, financial services, consumer goods and e-learning (Adomavicius & Tuzhilin, 2005; Garcia-Crespo et al, 2011; Kim et al, 2009; Papanikolaou, & Grigoriadou, 2002).

According to (Papanikolaou, & Grigoriadou, 2002) Adaptive Educational Hypermedia Systems aim to increase the functionality of hypermedia by making it personalised to individual learners. The adaptive dimension of these systems mainly refers to the adaptation of the content or the appearance of hypermedia to the knowledge level, goals and other characteristics of each learner. Learners' knowledge level and individual traits are used as valuable information to represent learners' current state and personalise the educational system accordingly, in order to facilitate learners to achieve their personal learning goals and objectives. Nowadays, most e-learning recommender systems consider learner/user preferences, interests, and browsing behaviours when analyzing personalization considering different levels of learner/user knowledge A distinctive feature of an adaptive e-learning system is a comprehensive user model that represents user knowledge, learning goals, interests, and pertinent contextual features that enable the system to distinguish among different user groups and feasible learning service solutions (Balabanovic, 1998; Lee, 2001; Sarwar, 2000).

Over the last 10 years, researchers in adaptive hypermedia and Web systems have explored many user modelling and adaptation methods, whereas a number of them already have been applied to the e-learning domain. The pre-Web generation of adaptive hypermedia systems explored mainly adaptive presentation and adaptive navigation support and concentrated on modelling user knowledge and goals. Empirical studies have shown adaptive navigation support can increase the speed of navigation and learning, whereas adaptive presentation can improve content understanding. The Web generation emphasized exploring adaptive content selection and adaptive recommendation based on modelling user interests. The mobile generation is now extending the basis of the adaptation by adding models of context and situation-awareness (location, time, computing platform, quality of service).

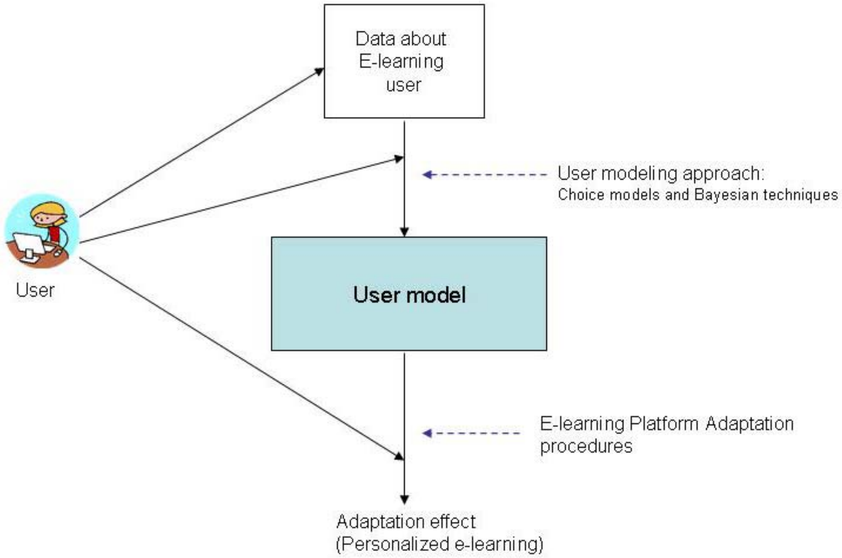

Data about
E-learning
user
User modeling approach:
Choice models and Bayesian techniques
User
User model
E-learning Platform Adaptation
procedures
Adaptation effect
(Personalized e-learning)

Fig. 2. Adaptive e-Learning System

An adaptive system collects data for the user model from various sources that can include implicitly observing user interaction and explicitly requesting direct input from the user (Brusilovsky & Maybury, 2002). The user model is used to provide an adaptation effect, that is, tailor interaction to different users in the same context. Adaptive systems often use intelligent technologies for user modelling and adaptation.

In specific, user modelling could be considered as the key process that enables recommendation systems to generate offers to users according to their needs, using content-based, filtering based and hybrid techniques (Adomavicius & Tuzhilin, 2005). More analytically, in the case of recommender system based on an intelligent system, the recommendation process is commonly implemented by a rule-based system that maintains a collection of knowledge facts, also denoted as the knowledge base (registering attributes such as student's preferences, interests and knowledge levels). The knowledge of an intelligent system can be expressed in several ways. One common method is in the form of "if-then-else" type rules. A simple comparison of the input to a set of rules will implement the desired actions, in order to deduct the recommendation. An intelligent system platform can be utilized for an e-learning recommendation system, such as the Java Expert System Shell (JESS), which represents its knowledge not only in the form of rules but also as objects. This allows rules to use pattern matching on the fact objects as well as input data.

In the following, we present a particular e-learning recommender framework based on advanced choice models and Bayesian techniques (Chaptini, 2005; Greene, 1993), which is considered as an intelligent recommender system. The presented e-learning recommender model is applied and tested in the shipping e-learning environment, it is though proposed for adaptation and use in different e-learning settings, also e-service environments favouring intense personalization and recommendation value-adding features. The techniques used in our system offer the strong competitive advantage of a comparatively detailed, user-focused e-learning attributes modelling framework (advanced choice theory) and a competent system learning capability (Bayesian theory), that improves over time the performance of the recommender system itself .

## 4. Training advisor development: A pilot case in the shipping industry

As mentioned above, the particular training advisor framework proposed hereafter, was developed as integrated with the SLIM-VRT e-learning platform for the shipping sector.

### 4.1 Methodological framework

The training advisor we developed is primarily characterized by a detailed, user-focused modelling approach, as denoted in Fig. 3.The user model that is the employees' preferences for e-learning is a function of (a) the Trainees Profile, which includes their socioeconomic characteristics, educational profile, working experience, and learning profile; and (b) The Training Package characteristics. Each training package is composed of one or more courses with specific attributes such as training method, duration of training, location of training, assessment method, training material, training institution, and training cost.

Individuals' preferences with regards to the training package are formed based on their training objectives and career expectations. That is, individuals may want to be promoted, change career, or stay in the current position, but enhance certain skills or improve their

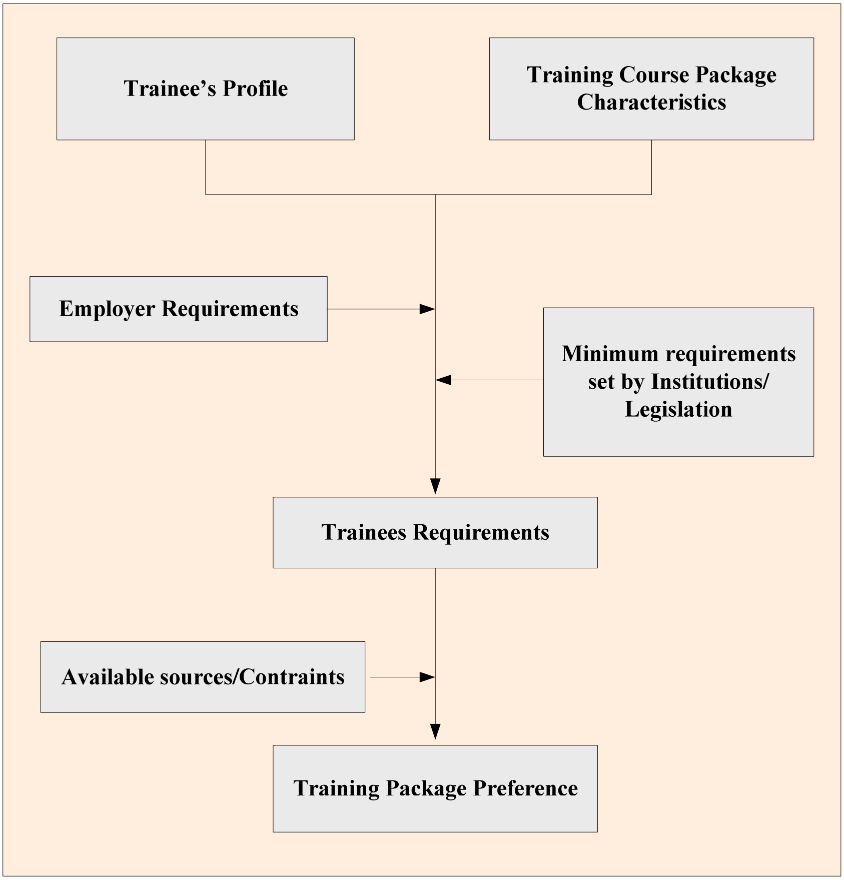

Trainee's Profile
Training Course Package Characteristics
Employer Requirements
Minimum requirements set by Institutions/ Legislation
Trainees Requirements
Available sources/Contraints
Training Package Preference

Fig. 3. E-learning Recommender Methodological Framework

knowledge on specific topics. The choice to adopt a "self-training for work" package is a function of these objectives, as well as of the resource availability and of situational constraints.

Resource availability includes training institution and course offerings, as well as job position openings by shipping companies and institutions. The situational constraints include the Employers' Requirements and the Institutional-Legal Minimum Requirements for each job position.

The underlying assumption in this recommender model framework is that employees try to choose the training package with the maximum utility. Thus, assuming a motivation to follow a "self learning for work course", they choose, among all the courses offered, the one that maximizes their utility, given their characteristics, the course attributes and the motivational and situational constraints. Given the source of motivation, employees select for consideration, from the menu of all –at present- available and non available courses, the ones that best address their needs. For example, based on the utility maximization assumption, an increase in flexibility is expected to increase the preference for the training package, while an increase in costs should decrease its preference.

Different groups of individuals are expected to have a different level of sensitivity to each of these attributes. In particular, it is expected that middle aged professional seafarers (captains and engineers) are to demonstrate higher willingness to pay for increased flexibility for courses that "lead" to shore positions. However, the extent to which an employee is willing to incur "training costs" depends on how much these costs are affected by the available income or expectations for promotion or reorientation of her/his career.

The characteristics of the job, particularly the available free time onboard may also constrain the consideration set of courses. Age, gender, marital status, profession and previous working experience are considered to have a strong effect on the perceived impact of "self training for work". For example, older learners are likely to be well-established in their career and have minimal preferences for new courses. On the other hand, young individuals (students in marine academies and personnel of shipping companies) are expected to appreciate more the flexibility and the "internet facilities" allowed by a "self learning for work" package. The training advisor developed provides as recommended training package, the one that maximises the utility of the user.

### 4.2 Training advisor logic

This section describes the e-learning recommender workflow implemented based on advanced choice models and Bayesian techniques for its core computational part. The proposed system architecture can be thought of as divided into two main parts according to system operation procedures, which is the front-end and back-end parts. The front-end part manages the interaction and communication with learners and records learner behavior, whereas the back-end part performs the analysis of learner preferences, skills and selects appropriate course materials for learners based on estimated learner ability. A main component of the latter subsystem is the *Training Advisor*, the logic and capabilities of which is detailed in the following section.

More specifically, our training advisor's structure and workflow can be seen as composed of 9 main steps, depicted in Figure 4.

First, at the entry of the trainee in the system a registration process is made (step 1). After the successful completion of registration, the user follows a series of steps (step 2-5) in order to receive detailed advices from the system regarding the educational package which suits in his/her needs (step 6).

In the end of the process, a trainee can review certain elements of the proposed courses (objectives, level, cost, etc.) and he/she can decide if he/she will attend one or more courses of the proposed package (step 7). Provided that the trainee has completed a full educational life cycle (step 8) this education process can be evaluated as well as the advisor usefulness (step 9).

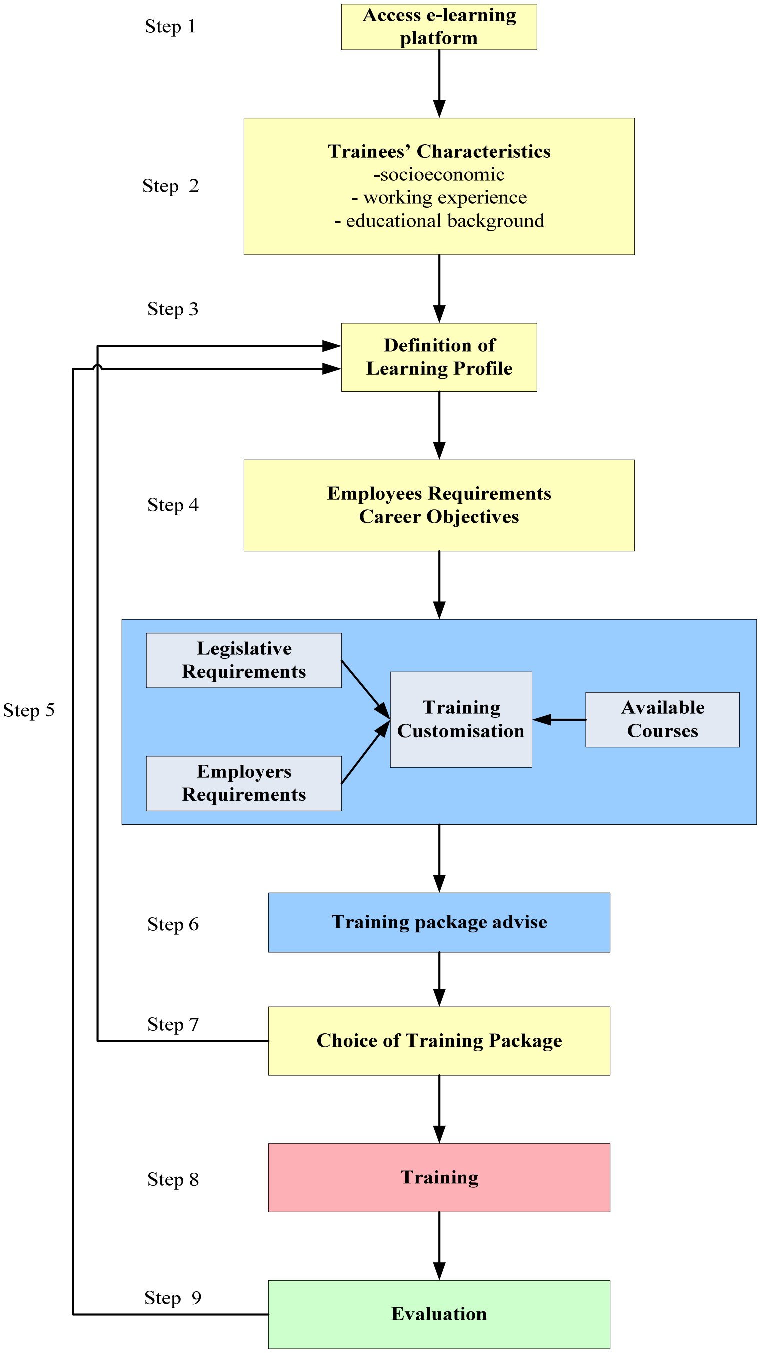


Fig. 4. Workflow of the Training Advisor

## 5. The model

In order to develop the *Training Advisor*, we used theories of individual choice behaviour analysis (Ben-Akiva & Lerman, 1992; ChoiceStream, 2004; Chaptini, 2005). Discrete choice analysis is the modelling of individuals' choices from a set of mutually exclusive and collectively exhaustive alternatives. A decision maker is modelled as selecting the alternative with the highest utility among those available at the time the choice is made. An operational model consists of parameterised utility functions in terms of observable independent variables and unknown parameters the values of which are estimated from a sample of observed choices made by decision makers when confronted with a choice situation.

The framework for choice theories can be viewed as an outcome of a sequential decision-making process that includes the following steps: 1) Definition of the choice problem; 2) Generation of alternatives; 3) Evaluation of attributes of the alternatives; and 4) Choice Model.

### 5.1 The choice problem

In our case, the decision makers are seafarers and employees of the shipping industry. These decision makers face different choice situations and have widely different tastes.

The recommender system acts as an automated training advisor. It facilitates a bundle of courses choice selection task by recommending courses that would satisfy employees' personal preferences and suit their abilities and interests.

The underlying hypothesis is that trainees perceive courses as a bundle of attributes. The utility of a course to a particular individual is a function of its attributes. Once those attributes are defined, discrete choice models can be used to calculate the utility of a set of courses, and the bundle of courses with the highest utilities would be selected.

### 5.2 Generation of alternatives

All the courses offered at various training institutions define the universal choice set of alternatives. This includes courses that are feasible during the decision process. The feasibility of an alternative is defined by a variety of constraints such as course offering and scheduling requirements and prerequisites.

The additional complexity of the problem stems from the fact that the training advisor needs to recommend a combination of courses that may be offered by different institutions.

### 5.3 Identifying attributes of the alternatives

The main hypothesis is that a course can be represented by a set of attributes that would define its attractiveness to a particular trainee (see Table 1). Courses are considered to be heterogeneous alternatives where decision makers may have different choice sets, evaluate different attributes, and assign diverse values for the same attribute of the same alternative.

| ATTRIBUTES | LEVEL 1 | LEVEL 2 | LEVEL 3 |
|---|---|---|---|
| COURSE | Compulsory (STCW/95, IMO Model Course, National Education Authority) | Non Compulsory | |
| INSTITUTE | Governmental (Marine, Academy, University ) | Non Governmental (Helmepa, Private Training, Manufacturer) | |
| ASSESSMENT PROCEDURES | Exams | Without Exams (Self Evaluation, Practice) | |
| COST | Fees (Paid By Trainee) | Without Fees (Cost Paid By Government Or/And Employee) | |
| LOCATION OF TRAINING | Ashore (Institute's Quarters, In-house Training) | Onboard | Ashore And Onboard |
| DURATION OF TRAINING | Academic Semester | Short Courses / Seminars | Flexible |
| TRAINING METHODOLOGY | Classic Lectures (In Class, In Office, Onboard With Instructor) | Self Learning (By Distance Or E-Learning With The Support Of Training Multimedia, Virtual Case Studies Etc) | On The Job Training (Practice Onboard, Emergency Drills, Simulation) |
| TRAINING MATERIAL | Printed (Textbooks, Notes, Manuals) | Digital (C.D, Diskettes, Internet, Virtual Equipment) | Audiovisual (Video, Audiocassettes) |

Table 1. Attributes of the Alternative Choice

### 5.4 Choice model

The choice model involves the estimation of a preference function, based on the attributes presented above. The estimation is based on stated preferences data, which are expressed responses to hypothetical scenarios presented to the employees.

With regards to the development of the training advisor it is important to estimate models in a relatively short time frame in order to deliver online recommendations.

Therefore, two models are developed:

*Training Needs Module* - An offline model system: that relies on advanced choice models to estimate the base parameters for different trainee's profiles; and

*Training Advisor Module* - A real time model system: that is based on the above models and in real time, estimates and customizes the parameters to each individual using Bayesian Techniques, and give fast recommendations. Bayes's theorem calculates the probability of a new event on the basis of earlier probability estimates which have been derived from empirical data. A key feature of Bayesian methods is the notion of using an empirically derived probability distribution for a population parameter. The Bayesian approach permits the use of objective data or subjective opinion in specifying a prior distribution. With the Bayesian approach, different individuals might specify different prior distributions. Bayesian methods have been used extensively in statistical decision theory. In this context, Bayes's theorem provides a mechanism for combining a prior probability distribution for the states of nature with new sample information, the combined data giving a revised probability distribution about the states of nature, which can then be used as a prior probability with a future new sample, and so on. The intent is that the earlier probabilities are then used to make ever better decisions. Thus, this is an iterative or learning process, and is a common basis for establishing computer algorithms that learn from experience (Greene, 1993).

### 5.5 Training needs module

This section presents the development of the training needs module and estimation results of the training advisor offline models that are used in our approach.

### 5.6 Data collection

In order to identify the seafarers' needs and develop the SLIM-VRT training advisor a field study was conducted. The target sample included members of seafarer's unions, shipping office's personnel and people working at shore based activities related to shipping sector. A total of 5000 questionnaires were sent to crew and shore based personnel, as well as to students of marine academies. 1195 completed questionnaires were received, corresponding to a 24% response rate. From these completed "employee's questionnaires", 59% (710 seafarers and employees) were Greek, 10% (115 seafarers and employees) were from the U.K, 7% (85 seafarers and employees) were from Spain and the rest 24% (285 seafarers and employees) were from other countries (Norway, Ukraine, Egypt, the Philippines, India, etc.). The multinational and multicultural character of the sample represents the decision making behaviour of major nationality groups in the shipping industry. The questionnaire included one to two Stated Preferences Experiments for Self-Learning for work. In each scenario individuals were presented with a course, described by several attributes and were asked to state their preference for following such a course.

A total of 1664 observations were used for estimating the preference models. Table 2 presents the distribution of the observations of the course attributes included in model estimations as independent variables.

We can see a very good distribution of the observations among the levels of attributes. This suggests a successful distribution of the stated preferences experiments between subjects.

Table 3 presents the distribution of observations of the dependent variable, or preference rating, taking the value of 1 if the individual is most unlikely to take the course and 7 if the individual is most likely to take the course.

| Attributes | Levels | Distribution of Observations |
|---|---|---|
| Training methodology | Classic lecture | 552 |
|  | Self learning | 682 |
|  | On-the-Job | 430 |
| Training material | Printed | 811 |
|  | Digital | 548 |
|  | Audiovisual | 305 |
| Duration of Training | Academic semester | 432 |
|  | Flexible | 795 |
|  | Short seminars | 437 |
| Institute | Governmental | 1355 |
| Cost | No fee | 705 |
|  | 100-500 Euros | 528 |
|  | >=500 Euros | 431 |
| Location | Ashore | 739 |
|  | On-board | 476 |
|  | Both ashore and on-board | 449 |
| Certification | Certificate | 1123 |
| Assessment procedure | Exams | 783 |

Table 2. Number of Observations Administered at Each Level

| Preference Rating | Level | Frequency | Percent |
|---|---|---|---|
| 1 | Most Unlikely | 159 | 9,6 |
| 2 | More Unlikely | 112 | 6,7 |
| 3 | Unlikely | 137 | 8,2 |
| 4 | In The Middle | 413 | 24,8 |
| 5 | Likely | 298 | 17,9 |
| 6 | More Likely | 269 | 16,2 |
| 7 | Most Likely | 276 | 16,6 |
|  | **Total** | 1664 | 100,0 |

Table 3. Choice of Course

### 5.7 Model estimation results

Regression models were run with as dependent variable the choice of course and as independent variables the attributes of the course.

These result in regression-type models of the following form:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ..... + \beta_k X_k$$

or

$$y = \beta_0 + \beta_1 SelfLearning + \beta_2 OnTheJob + \beta_3 VR + \beta_4 Govermental + \beta_5 Flexible$$
$$+ \beta_6 NoFees + \beta_7 GT500Euros + \beta_8 Certificate + \beta_9 Exams + \beta_{10} Ashore \,\&\, onBoard$$

Table 4 presents a generic model estimated with all the available observations.

| Coefficient Number | Variable | Estimated Coefficient | t-stat |
|---|---|---|---|
| 0 | Constant | 3.54 | 10.9 |
| 1 | Self-Learning | 0.36 | 1.8 |
| 2 | On-the-job Training | 0.42 | 2.0 |
| 3 | Digital | 0.25 | 1.5 |
| 4 | Governmental | 0.31 | 1.6 |
| 5 | Flexible | 0.23 | 1.7 |
| 6 | Without fees | 0.10 | 1.0 |
| 7 | Greater than 500 Euros | -0.84 | -2.8 |
| 8 | Certificate | 0.78 | 5.3 |
| 9 | Exams | -0.26 | -1.8 |
| 10 | On Board and Ashore | 0.35 | 1.6 |
| **Summary Statistics** | | | |
| Number of observations: 1664 | | | |
| Rho-bar squared = 0.2 | | | |

Table 4. Model Estimation Results

The estimated results demonstrated the following:

- Individuals prefer self-learning and on-the-job training over classical lectures
- Digital material are favoured over printed ones
- Trainees prefer studying at governmental institutions, such as universities
- There is a preference over flexible courses adjusted to the user needs rather than courses offered for a full academic semester
- Individuals prefer not to pay for receiving the courses and are especially negative towards following a course with cost more than 500 Euros
- Getting a certificate is very important for the trainees
- There is negative attitude towards courses that have exams as the assessment procedure
- Individuals would prefer courses that are offered both on-shore and off-shore.

The above-mentioned results have been tested by a panel of experts and were found consistent with the current situation and emergent trends in the maritime education and employment environment, and our a priori hypothesis regarding the behaviour of seafarers.

The equation implemented with regards to the preference rating of each course (y) is therefore the following:

$$y = 3.54 + 0.36 SelfLearning + 0.42 OnTheJob + 0.25 VR + 0.31 Govermental + 0.23 Flexible \\ + 0.10 NoFees - 0.84 GT500 Euros + 0.78 Certificate - 0.26 Exams + 0.35 Ashore \,\&\, onBoard$$

Similar equations are estimated for different user groups. These groups were defined based on the opinion of the experts' panel used for this purpose. The categorization is based on the following characteristics: (1) Age; (2) Education; (3) Years of working experience; (4) Current Job (Engine, Deck, other); (5) Learning styles; and (6) Soft skills.

### 5.8 Training advisor module

The basic models developed for each group of trainee, as described in the previous section, need to be customized for each respondent. Bayesian theory is used to provide the suggestions of the courses that match the preferences of the individuals.

For each Group the following information, the following outputs of survey regressions are saved to be used in the Training Advisor:

$$s = \begin{cases} \overline{\beta}_g = \text{estimated coefficients for each group g} \\ s_g = \text{standard deviation} \end{cases}$$

$\Sigma_{\overline{\beta}}$ = variance-covariance matrix of the prior $\overline{\beta}_g$

The steps followed are:

**Step 1.** Subject profile

A number of questions are asked at the beginning of the session regarding the characteristics of the trainees. A number of these characteristics (the $X$'s) define the Group in which each individual belongs.

Assume $g = 1,...,G$ number of groups.

**Step 2.** Elicitation

The individual is presented with a sample of courses (using always the same attributes) and is asked of his preferences.

Assume:

$N$ = number of experiments presented to the individual, and
$y_n$ (Nx1 vector) = the ratings of course n.

**Step 3.** Creation of Individualized Data

The new data is ( $y_n$ , $X_n$ ) where:

$y_n$ (Nx1 vector) = the ratings of the courses.
$X_n$ (1xK matrix) = attributes of course n
K= number of attributes

In the new table the respondent has $N$ ratings. To each rating the K attributes of the course are appended.

**Step 4.** Develop the personalized preference equation for each individual

Bayesian updating is used to calculate the personalized coefficients of the preference equation as follows:

$$\tilde{\beta} = \left[ s_g^2 \Sigma_{\overline{\beta}_{(n)}}^{-1} + X'X \right]^{-1} \left[ s_g^2 \Sigma_{\overline{\beta}_g}^{-1} \overline{\beta}_g + X'y \right]$$

where:

$$s = \overline{\beta}_g, s_g, \Sigma_{\overline{\beta}_g} = \text{outputs of the survey} - \text{regressions of group g, and}$$

$(y \;\;, X \;\;)$ = the new data

$y$ is Nx1

$X$ is NxK

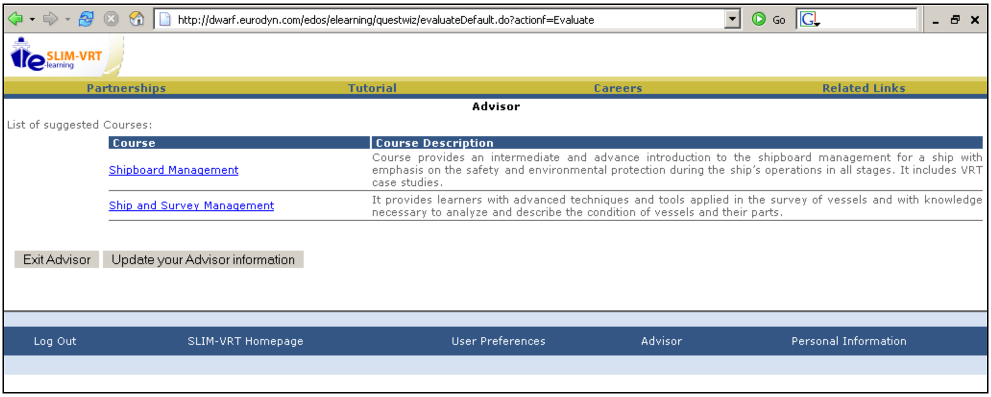
- Calculate new coefficients
- Update preference equation with new coefficients
- Calculate Course Ratings
- Calculate the rating of each course by applying the personalized equation of Step 3.

**Step 5.** Preference Score of Bundles of Courses

A number of potential bundles of courses exist based on expert judgment. Apply ratings to each course of the bundle. Sum-up the ratings.

**Step 6.** Recommend

Present to the trainee the bundle of courses with the highest rating (Fig. 5).



http://dwarf.eurodyn.com/edos/elearning/questwiz/evaluateDefault.do?actionf=Evaluate

Advisor

List of suggested Courses:

| Course | Course Description |
|---|---|
| Shipboard Management | Course provides an intermediate and advance introduction to the shipboard management for a ship with emphasis on the safety and environmental protection during the ship's operations in all stages. It includes VRT case studies. |
| Ship and Survey Management | It provides learners with advanced techniques and tools applied in the survey of vessels and with knowledge necessary to analyze and describe the condition of vessels and their parts. |

Fig. 5. E-learning system recommendation

## 6. Conclusions

This chapter presents a modelling methodology for developing an e-learning recommender system. The proposed methodology includes the definition of a mathematical user model, as formulated in the context of the shipping industry and its employment and training environment. The implementation of the central component of this recommender system, namely the Training Advisor, is explained as based on discrete choice models and Bayesian theory. In particular, the development of an e-learning recommender system, such as the electronic training advisor proposed will help trainees in choosing the appropriate e-learning courses matching their particular characteristics, preferences and needs and based on their expected professional development. The training process as assisted by the proposed training advisor; it takes into account the peculiarities of the seafaring profession, applicable career paths and respective seafarers' training needs. To formally model these requirements we developed a knowledgebase for accumulating the basic knowledge regarding the shipping work and training environment and registered the information of a 5000 users- sample population, furthermore we used statistical analysis to support the choices of each individual separately. In specific, our e-learning recommender framework is based on advanced choice models and Bayesian techniques and is considered as an intelligent system that can be tested and reused in different e-learning settings, favouring intense personalization and recommendation value-adding features. The foundational techniques used in our system offer the strong competitive advantage of a comparatively detailed, user-focused e-learning attributes modelling framework (advanced choice theory) and a competent system learning capability (Bayesian theory), that improves over time the performance of the recommender system itself .

## 7. References

Adomavicius, G., Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering,* Vol. 17, No 6, pp. 734–749.

Balabanovic, M. (1998). Exploring versus Exploiting when Learning: User Models for Text Representation. *User Modeling and User-Adapted Interaction*, Vol. 8, No 1, pp. 71-102.

Ben-Akiva, M., S. Lerman, S. Discrete Choice Analysis: Theory and Application to Travel Demand. The MIT Press, Cambridge, MA, 1985.

Brusilovsky P., Maybury, M. T. (2002). From Adaptive Multimedia to the Adaptive Web, *Communications of the ACM* , Vol. 45, No 5, pp. 31-33.

Brusilovsky, P. Adaptive and Intelligent Technologies for Web-Based Education, in: Rollinger, C. & Peylo, C.Eds., *Special issue on intelligent systems and tele-teaching,* Kunstliche Intelligenz, 2002, pp. 19–25.

Burke, R. (2000). Knowledge-based Recommender Systems. In: A. Kent (ed.): *Encyclopedia of Library and Information Systems*. Vol. 69, Supplement 32.

Chaptini, B. Use of Discrete Choice Models with Recommender Systems, PhD thesis, MIT, 2005.

ChoiceStream (2004). Review of Personalization Technologies: Collaborative Filtering vs. ChoiceStream's Attributized Bayesian Choice Modeling.
http://www.choicestream.com/pdf/ChoiceStream_TechBrief.pdf

García-Crespo, Á., López-Cuadrado, J.L., Colomo-Palacios, R., González-Carrasco I., Ruiz-Mezcua, B. (2011). Sem-Fit: a semantic based expert system to provide recommendations in the tourism domain. *Expert Systems with Applications*, Vol. 38, No 10 (2011), pp. 13310–13319.

Greene, W.H. Econometric Analysis, Second Edition, Macmillan, 1993.

Ho, S.Y. (2006).The Attraction of Internet Personalization to Web Users. *Electronic Markets*, Vol.16, No 1, pp.41–50.

Kim, H.K., Kim, J.K., Ryu, Y.U. (2009). Personalized Recommendation over a Customer Network for Ubiquitous Shopping. *IEEE Transactions on Services Computing,* Vol. 2, No. 2, pp. 140-151.

Kim, H.K., Cho, Y.H., Kim, W.J., Kim J.R., Suh, J.H. (2002). A personalized recommendation procedure for Internet shopping. *Electronic Commerce Research and Applications,* Vol.1, pp.301–313.

Lee, M.G. (2001). Profiling Students' Adaptation Styles in Web-Based Learning, *Computers & Education*, Vol. 36 , pp. 121–132.

Papanikolaou, K.A., Grigoriadou, M. (2002). Towards New Forms of Knowledge Communication: The Adaptive Dimensions of a Web-Based Learning Environment. *Computers & Education*, Vol. 39, pp. 333–360.

Progoulaki M., Theotokas I. (2008). Human resource management and competitiveness in the shipping industry: a resource based view, *Proceedings of International Association of Maritime Economist (IAME) Annual Conference,* Dalian, China.

Progoulaki M., Polydoropoulou A. and Theotokas I. (2005). Multicultural Training and e-Learning in Shipping Industry, *Proceedings of ICODL 2005, 3rd International Conference on Open and Distance Learning 'Applications of Pedagogy and Technology'*, Patras, Greece

Rashid, A. M., Albert, I., Cosley, D., Lam, S.K., McNee, S., Konstan, J.A. (2002). Getting to know you: Learning new user preferences in recommender systems, Proceedings of the International Conference on Intelligent User Interfaces, pp. 127–134.

Ricci, F., Werthner, H. (2006). Recommender systems. *International Journal of Electronic Commerce*, Vol. 11, No 2, pp. 5–9.

Sarwar, B., Karypis, G., Konstan, J.A.,Riedl, J.T. (2000). Analysis of Recommendation Algorithms for e-Commerce, *Proceedings of the Second ACM Conference of Electronic Commerce*, pp. 158-167.

Theotokas I., Progoulaki M. (2007). Cultural diversity, manning strategies and management practices in Greek shipping. *Maritime Policy and Management*, Vol. 34, No. 4, pp. 383-403.

Zanker, M., Jessenitschnig, M. (2006). Case-studies on exploiting explicit customer requirements in recommender systems. *User Modeling and User-Adapted Interaction*, Vol. 19, pp. 133–166.

# Intrusion Detection and Prevention in High Speed Network

Kuo Zhao and Liang Hu
*Jilin University,*
*China*

## 1. Introduction

With the rapid development and comprehensive application of network technology, network security problems gradually appear serious· Traditional firewall technologies can't provide sufficient security protection against various attacks and intrusions (Anderson, 1980), while intrusion detection systems (IDS) are faced with compromise between false alarms and false positives (Denning, 1987). In this chapter, we investigate intrusion detection and intrusion prevention in high speed network, introduce related technology and our research results.

### 1.1 The information system and system security

Information system is an integrated set of components for collecting, storing, processing, and communicating information. Information systems are more than just computer programs. Though information and communications technologies are playing an increasing role in meeting organisations' information needs, an information system is a much more general concept. It refers to the wider systems of people, data and activities, both computer-based and manual, that effectively gather, process, store and disseminate organisations' information. Of course, system security is essential for information system. In another words, security is the most reliable foundation for information system.

### 1.2 The actual condition of information system security

With the development of Internet, the world economy has been deeply communed together. The nation is just like a huge network computer, and computer network has been the foundation and life vein of a nation's economy. As the entire society increasingly relies on network infrastructures, network security also changes for the worse seriously. It is very difficult for traditional security policies or mechanisms (such as authentication, cryptography and firewall) to prevent network attacks. The whole society needs new technology to solve those problems.

The openness of the system network, the security hole of the network protocol, the defects of the software…Those drawbacks make the network security worse than worse. According to the recently research and report, people found the details and data of network attack easily. The high occurrence probability makes the problem urgent (Allen et al., 2000).

Cases are known, the network security is the most reliable foundation for network applications. Every country, for commercial or military purposes, spared a lot to study network security. Research on this issue

Although there are various measures to protect safety, they are not the keys to all kinds of attack. For instance:

1. A perfect design of software safety is impossible.
2. Encryption technology itself has some problems, and those shortcomings may lead to keylogger activities. Moreover, people may misunderstand the arithmetic.
3. The security hole of the network protocol.
4. The contradiction between the availability and the safety is always one of those contradictions running through the long developing process of computer technology.
5. The complex security system is usually difficult to configure. The blander of wrong configuration will leave some hidden danger.
6. The system log and the audit have massive data. They need automatic mode to work with those information.
7. Staff members may abuse the safety system.

This chapter presents the corresponding research work on the intrusion detection and intrusion prevention in large-scale high-speed network environment and is organized as follows: firstly, a distributed extensible intrusion prevention system is provided, then various packet selection models for intrusion detection systems based-on sampling are illustrated, again the design and implementation of a high-speed traffic collection platform based-on sampling on FPGAs and the research of trusted communication protocol for intrusion prevention system are presented, last we draw concclusions.

## 2. DXIPS: A distributed extensible intrusion prevention system

### 2.1 Related technology

### 2.1.1 Snort_inline

The Snort_inline IPS is a modified version of the famous Snort IDS. It receives packets sent from the Netfilter firewall with the help of the lipipq library[1], compares them with Snort signature rules and tags them as drop if they match a rule, then finally sends them back to Netfilter where the Snort_Inline tagged packets are dropped.

There are 5 available default actions in Snort, alert, log, pass, activate , and dynamic:

1. alert-generate an alert using the selected alert method, and then log the packet
2. log-log the packet
3. pass - ignore the packet
4. activate - alert and then turn on another dynamic rule
5. dynamic - remain idle until activated by an activate rule , then act as a log rule

There are three rule options more than Snort's:

1. Drop – The drop rule tells iptables to drop the packet and log it via usual snort means
2. Sdrop – The sdrop rule tells iptables to drop the packet. Nothing is logged.

[1] Libipq library is a development library for iptables userspace packet queuing

3. Reject – The reject rule type tells iptables to drop the packet; log it via usual snort means; and send a TCP reset if the protocol is TCP or an ICMP port unreachable if the protocol is UDP.

### 2.1.2 Netfilter

Along with the development of the technology used in Linux firewall, Netfilter comes into being. It is well known that Netfilter is a framework that provides hook handling within the Linux kernel for intercepting and manipulating network packets.

Netfilter is made up of five hook functions towards IPV4, IPV6, and IPX. The identifiers of all hooks for each supported protocol are defined in the protocol-specific header file. The following five hooks are defined for IP Version 4 in <linux/netfilter_ipv4.h>:

NF_IP_PRE_ROUTING (default value is 0): incoming packets pass this hook in the ip_rcv() (linux/net/ipv4/ip_input.c) function before they are processed by the routing code

NF_IP_LOCAL_IN (default value is 1): all incoming packets addressed to the local computer pass this hook in the function ip_local_deliver()

NF_IP_FORWARD (default value is 2): all incoming packets not addressed to the local computer pass this hook in the function ip_forward()

NF_IP_LOCAL_OUT (default value is 3): all outgoing packets created in the local computer pass this hook in the function ip_build_and_send_pkt()

NF_IP_POST_ROUTING (default value is 4): this hook in the ip_finish_output() function represents the last chance to access all outgoing (forwarded or locally created) packets before they leave the computer over a network device

There are five return values in the hook functions:

NF_DROP (default value is 0): The active rules list processing is stopped, and the packet is dropped

NF_ACCEPT (default value is 1): The packet is passed to the next packet filter function in the rules list. Once the end of the list has been reached, the packet is released by okfn() for further processing

NF_STOLEN (default value is 2): The packet filter function withholds the packet for further processing, so that the active rules list processing is stopped. In contrast to NF_DROP, however, the packet does not have to be explicitly dropped

NF_QUEUE (default value is 3): The function nf_queue() (net/core/netfilter.c) puts the packet in a queue from which it can be removed and processed (e.g., by a user space program). Subsequently, nf_reinject() has to be invoked to return the packet to the Linux kernel for further processing by netfilter

NF_REPEAT (default value is 4): In contrast to NF_ACCEPT, rather than a continuation of processing at the next packet-filter function, the current filter function is invoked again

### 2.1.3 Iptables

Iptables is used to set up, maintain, and inspect the tables of IP packet filter rules in the Linux kernel. Several different tables may be defined. Each table contains a number of built-in chains and may also contain user-defined chains.

A firewall rule specifies criteria for a packet, and a target. If the packet does not match, the next rule in the chain is the examined; if it does match, then the next rule is specified by the value of the target, which can be the name of a user-defined chain or one of the special values ACCEPT, DROP, QUEUE, or RETURN.

ACCEPT means to let the packet through. DROP means to drop the packet on the floor. QUEUE means to pass the packet to userspace. (How the packet can be received by a userspace process differs by the particular queue handler. 2.4.x and 2.6.x kernels up to 2.6.13 include the ip_queue queue handler. Kernels 2.6.14 and later additionally include the nfnetlink_queue queue handler. Packets with a target of QUEUE will be sent to queue number '0' in this case. Please also see the NFQUEUE target as described later in this man page.) RETURN means stop traversing this chain and resume at the next rule in the previous (calling) chain. If the end of a built-in chain is reached or a rule in a built-in chain with target RETURN is matched, the target specified by the chain policy determines the fate of the packet.

### 2.2 Architecture of DXIPS

DXIPS is a distributed network IPS based on the combination of Snort_inline and Netfilter firewall configured by IPtables, and it provides the ability to detect malicious network traffic, drop or reject attack packets, and perform intrusion detection and prevention on 4-7 layers of network protocol.

Hierarchical structure is applied to the architecture of DXIPS, which consists of three layers:

Intrusion prevention layer: monitor the network traffic passing by and perform intrusion detection and prevention.
Server layer: collect log data and save to readable format.
Control layer: analysis console and perform data display.

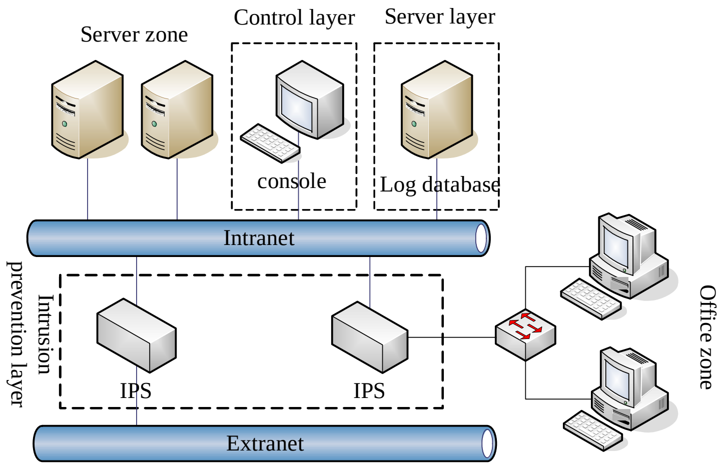


Fig. 1. Architecture of DXIPS

### 2.3 Construction and implementation of DXIPS

DXIPS is composed of four modules: intrusion prevention module, log record module, central control module and communication module. These four modules coordinate with each other and perform intrusion prevention in distributed network environment.

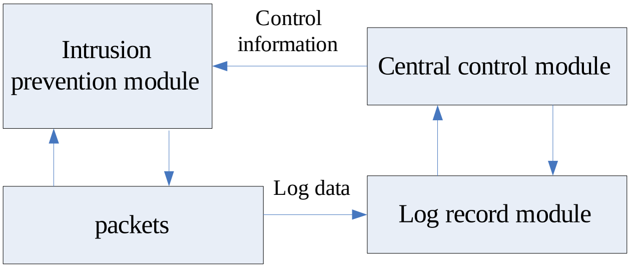


Fig. 2. Construction of DXIPS

Intrusion prevention module runs on intrusion prevention layer, and it is responsible for capturing packets, intrusion detection and prevention. This module is to be deployed at the crucial position of network, such as the network link between intranet and extranet, thus all the packets are to be monitored. This module is based on the combination of Snort_inline and Netfilter firewall configured by IPtables, and consists of three submodules such as packets capture, packets detection and response.

Log record module runs on server layer and aims for log collection and formatting. The collected logs include intrusion detection log generated by Snort_inline and firewall log by the configuration of IPtables.

Central control module is the core of the whole system and runs on control layer. It is in charge of coordinating other modules and performing management operations such as the configuration of IPS, management of log server, data analysis and load balancing.

Communication module has the ability to provide secure and reliable communication channels between modules.

#### 2.3.1 Intrusion prevention module

Intrusion prevention module is to be deployed at the crucial position of network and needs the ability of routing, that is, packets with correct route information should be delivered from source address to destination address. Thus normal communication between intranet and extranet is to be done. The part of function of routing in DXIPS makes use of the default route module in Linux system, and the command is as follows:

“echo 1>/proc/sys/net/ipv4/ip_ forward”

When packets are captured by intrusion prevention module, the function of IP forward is active, and kernel will deliver packets based on address information of packets and routing table information.

1. Packets capture

Packets capture procedure is to pass packets from kernel space to user space, and includes event generator, Netfilter hook program, IPtables, ip_queue kernel module and netlink interface.

During the procedure of packets capture, intrusion prevention module looks on the filtered packets by Netfilter firewall configured by IPtables as data source. It may reduce the amount of packets to be detected by performing intrusion detection on the specific traffic with regard to security policy. IPtables, which is a packet filter firewall running on data link layer and network layer, may firstly filter packets according to security policy, then pass the filtered packets to Snort_inline and perform intrusion detection.

Specific types of packets are to be accepted by Snort_inline according to the configuration of IPtables. For example, SMTP traffic is to be monitored by Snort_inline according to the configuration of IPtables. The following parts present the commands:

```
“iptables -A FORWARD -m state --state ESTABLISHED,RELATED -p tcp --dport 25 -j QUEUE”

“iptables -A FORWARD -m state --state ESTABLISHED,RELATED -j ACCEPT”

“iptables -A FORWARD -p tcp --dport 25 -m state --state NEW -j QUEUE”
```

2. Packets detection

Packets detection procedure is to perform intrusion detection corresponding with rules set, and includes event analyzer, libipq library, Snort_inline and rules set.

Intrusion prevention module captures packets according to security policy and parses the structure of packet based on protocol specifications, then executes data formatting and performs string matching on packets with rules set.

Fig.3 illustrates the packets detection procedure. Snort_inline executes initialization based on command line parameters, and working mode is specified by parameters. Then two dimensional linked lists are constructed based on rules by parsing rules library. Again, Snort_inline performs intrusion detection on packets by calling the function “ProcessPacket( )” after packets are accessed cyclically from structure “ip_queue”. Function “ProcessPacket( )” firstly executes protocol parsing and set structure “Packet”, then calls detection function “Detect( )” to perform detection according to rule linked lists in certain turn, lastly returns detection results.

Corresponding response actions based on packet detection results may be executed. There are 5 available default actions in Snort, alert, log, pass, activate, and dynamic. In addition, there are additional options which include drop, sdrop and reject in Snort_inline. Action “drop” is to make IPtables drop the packet and log the packet; action “sdrop” is to make IPtables drop the packet but does not log it; action “reject” is to make IPtables drop the packet, log it, and then send a TCP reset if the protocol is TCP or an ICMP port unreachable message if the protocol is UDP.

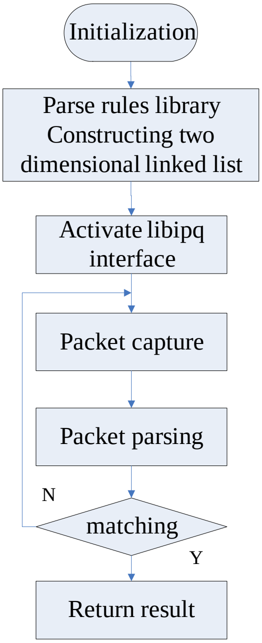

Initialization
Parse rules library Constructing two dimensional linked list
Activate libipq interface
Packet capture
Packet parsing
N
matching
Y
Return result

Fig. 3. Flow diagram of packets detection

### 2.3.2 Log record module

Log record module is applied to record and audit the time when event happened and corresponding information of subjects and objects, and it aims for providing sufficient information for detailed analysis of security events later.

Log record module collects log of intrusion prevention module, saves it to log server, and provides analysis data for security policy specified by central control module. The collected logs include intrusion detection log generated by Snort_inline and firewall log by the configuration of IPtables.

Snort_inline will generate corresponding log based on specific parameter of rule action such as alert, log, drop and reject. The default directory of log files is "/var/log/snort".

The Netfilter firewall log may be accessed with LOG target of IPtables. To make use of LOG target, ipt_LOG module is needed to be loaded as follows:

"/sbin/modprobe ipt_LOG"

For instance, all the connection information passing by may be recorded as follows:

"#iptables -A FORWARD -j LOG"

LOG target is dedicated for recording detailed packet information such as IP header and other useful information. The default directory of log of IPtables is "/var/log/message", which is done by the "syslogd"daemon.

There are three manners for log server to collect intrusion detection log of Snort_inline and firewall log of Netfilter configured by IPtables:

**All the nodes of intrusion prevention module directly interact with log server**

This can be done by outputting log records directly to log database with Snort_inline output plugin and LOG target of IPtables. The defect of this manner is to improve the system overhead in that intrusion prevention module needs to execute log operations as well as perform intrusion detection and prevention. Additionally, this manner compromises the extensibility of the whole system for the duplicated implementation of log recording.

**Log server accesses logs saved in every intrusion prevention nodes at regular time**

Though this manner decreases the overhead of intrusion prevention module, it compromises real time performance of the whole system. Log server can't get log information in time, thus central control module is unable to monitor the protected network in real time.

**Apply specialized log collection daemon**

These daemons are responsible for receiving and dispatching logs of various intrusion prevention nodes and saving to log server in distributed network environment. Compared to the preceding manners, this manner shortens handling time, reduces system overhead, and provides better extensibility.

### 2.3.3 Central control module

Central control module is the core of the whole system, in charge of coordinating various modules, and conducts centralized management. It locates at central position of the whole system, and connects with intrusion prevention module and log record module. The main functions of central control module include:

Customized policy: central control module is able to establish corresponding policies based on intrusion detection and prevention logs. Log management: central control module can analyze intrusion prevention logs and firewall logs saved in log server in real time and provide corresponding statistical information.

System management: central control module has the ability to manage various intrusion prevention nodes by console, such as dynamical management of intrusion detection rules and firewall rules, management of log server, network traffic statistics and load balancing.

### 2.3.4 Communication module

It is an important component for secure and reliable communication among various modules of DXIPS. This module mainly refers to the communication among central control module, intrusion prevention module and log record module.

To achieve the interoperation among modules, there is a need of general communication protocol, which contributes to simplifying the management of DXIPS. For example, if central control module is to disable the intrusion prevention function of certain node, a control message is sent to this node by console, and then this node returns an acknowledgement message after it quits normally.

The general communication protocol is derived from standard TCP/IP protocol with farther encapsulation, and it is an application layer protocol.

| | |
|---|---|
| Header | Payload |

Fig. 4. Customized protocol unit

Customized protocol header is composed of version, type and total length. Type field specifies the type of message such as control message, log message or acknowledgement message according to specific functions. Meanwhile, type field specifies the format of data.

| | | |
|---|---|---|
| Version | Type | Total length |

Fig. 5. Format of customized protocol header

The format of payload field varies according to type field. The format of a sample log message is as follows:

| | | | | |
|---|---|---|---|---|
| Node ID | Log level | Log time | Message length | Message content |

Fig. 6. Format of payload field of a sample log message

Customized communication protocol provides the better extensibility for later new function to be updated or new communication needs. Also it supports encryption manners for secure and reliable communication.

### 2.4 Deployment of DXIPS

DXIPS provides flexible deployment strategies and better extensibility. It can be expediently deployed in distributed network environment corresponding with various securities prevention needs, and presents a comprehensive protection.

#### 2.4.1 Border prevention deployment

With the rapid development of Internet, great deals of business applications rely on Internet. However, there are various security threats, such as worms, computer viruses, spywares and DDoS attacks, towards the entrance to the Internet of an intranet due to the openability of Internet.

To minimize external network security risks, intrusion prevention module may sit on the entrance to Internet, examine traffic, and block malicious or suspect code in real time. As shown in figure 7.

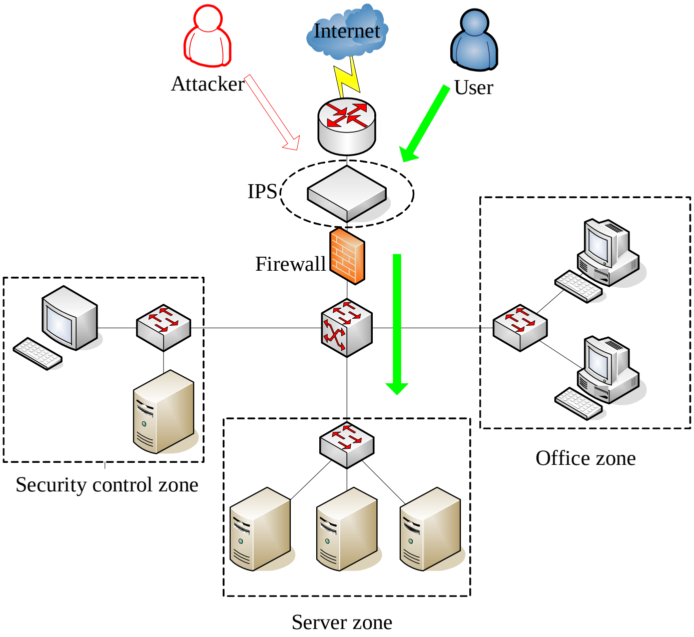


Fig. 7. Border prevention deployment

### 2.4.2 Key zones prevention deployment

External security threats, such as worms, computer viruses, spywares, may be introduced into intranet by inconsciently users. To minimize internal network security risks, intrusion prevention module may sit on the entrance to office zone and key server zone, block malicious traffic in real time, filter worms, computer viruses and spywares from office zone, and protect the key network server. As shown in Fig. 8.

### 2.4.3 Hybrid prevention deployment

To minimize external and internal network security risks simultaneously, intrusion prevention module may sit on the entrance to key network link, block malicious traffic in real time, and protect the key network resources.

Meanwhile, intrusion prevention module may sit on the entrance to key network link in bypass mode, which can be treated as intrusion detection system, and performs analysis and detections on intranet. As shown in figure 9.

### 2.5 Related work

There are various research and commercial work on the design and implementation of IPS. Tan and Weinsberg attempt to improve string-matching algorithms for intrusion detection and prevention on large-scale high-speed network traffic; Drinic (Drinic & Kirovski, 2004) and Weaver (Weaver et al., 2007) present the hardware implementation of IPS based on field programmable gate arrays; Green uses a generic and reliable model to anticipate future

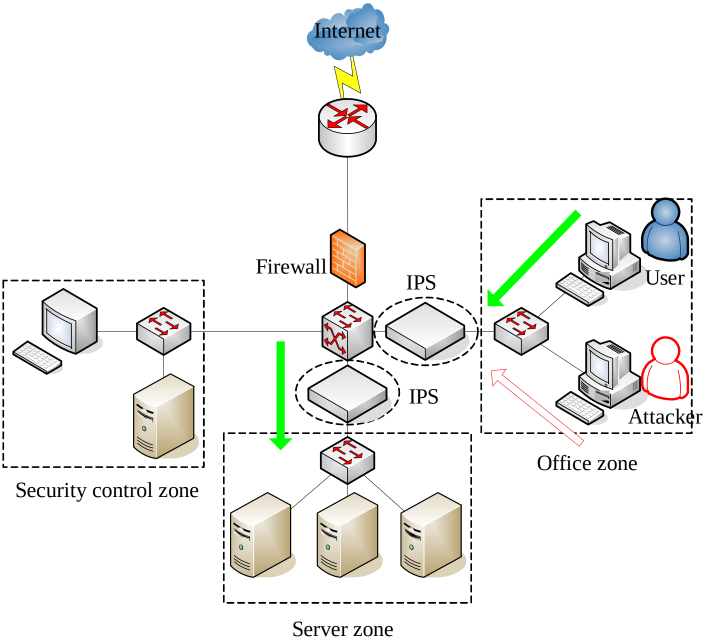


Fig. 8. Key zone prevention deployment

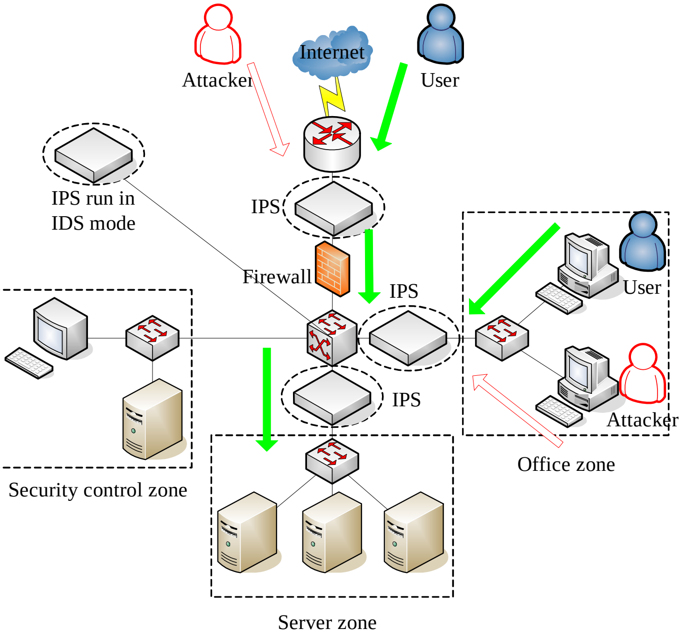


Fig. 9. Hybrid prevention deployment

attack scenarios; Uppuluri provides a practical approach to detect and prevent race condition attacks (Uppuluri et al., 2005). What's more, there are many commercial IPS products available such as TippingPoint IPS, ISS IPS, Cisco IPS and NetKeeper IPS, and these representative products are online, network-based solution, designed to accurately identify, classify, and stop malicious traffic, including worms, spyware/adware, network viruses, and application abuse.

### 2.6 Summary

This subsection presents the design and implementation of DXIPS, a distributed extensible intrusion prevention system, which is composed of intrusion prevention module, central control module, log record module and communication module. And DXIPS provides an extensible architecture of intrusion detection and prevention in distributed network environment.

## 3. Packet selection model for intrusion detection system based-on sampling

### 3.1 Packet selection model

There are many sampling model of statistics, while there are few applied to computer network especially high speed network. For real time demands of network packets processing, packet selection model have to conform to high precision and simple applications. In this paper, systematic sampling, Poisson sampling and stratified sampling methods are applied to network packets selection.

### 3.1.1 Systematic sampling

Systematic sampling describes the process of selecting the starting points and the duration of the selection intervals according to a deterministic function. This can be for instance the periodic selection of every n-th element of a trace but also the selection of all packets that arrive at pre-defined points in time. Even if the selection process does not follow a periodic function (e.g. if the time between the sampling intervals varies over time) we consider this as systematic sampling as long as the selection is deterministic.

The use of systematic sampling always involves the risk of biasing the results. If the systematics in the sampling process resembles systematics in the observed stochastic process (occurrence of the characteristic of interest in the network), there is a high probability that the estimation will be biased. Systematics (e.g. periodic repetition of an event) in the observed process might not be known in advance.

### 3.1.2 Poisson sampling

Poisson sampling is an example of random additive sampling. In this sampling, samples are separated by independent, randomly generated intervals that have a common statistical distribution. In general, Poisson sampling avoids synchronization effects and yields an unbiased estimate of the property being sampled.

If sampling function obeys an exponential distribution with rate $\lambda$, that is $F(t) = 1 - e^{-\lambda t}$, then the arrival of new samples cannot be predicted (and, again, the sampling is unbiased).

Furthermore, the sampling is asymptotically unbiased even if the act of sampling affects the network's state. Such sampling is referred to as Poisson sampling. The sample size during interval *T* obeys Poisson distribution with rate *λ* at which the singleton measurements will on average be made, that is $P_k(T) = (\lambda T) k^k e^{-\lambda T} / k!$ .

To generate Poisson sampling intervals, one first determines the rate *λ* at which the singleton measurements will on average be made (e.g., for an average sampling interval of 30 seconds, we have *λ*= 1/30, if the units of time are seconds). One then generates a series of exponentially-distributed (pseudo) random numbers $E_1$, $E_2$,…, $E_n$ The first measurement is made at time $E_1$, the next at time $E_1$+$E_2$, and so on.

### 3.1.3 Stratified sampline

Before sampling, the whole population is first divided into mutually exclusive subgroups, called stratum. Let *N* be the number of population unit, *L* be the number of strata and $N_1$,$N_2$,…,$N_L$ represent the size of each stratum, then $N = \sum_{h=1}^{L} N_h$ . If the sample is taken randomly from each stratum, the procedure is known as stratified random sampling.

For stratified sampling, there are some factors supposed to determine such as stratified characteristics, stratum number, stratum border, sample size allocation and variance within strata. The concrete discussions of these factors are as follows:

Stratified characteristics are the base of stratified sampling. Network packets may be stratified by protocol type, TTL or total length field of IP header. The selection of stratified characteristics relates to the type of intrusions. As a example of ICMP sweep attack, the characteristics of this attack is the generation of lots of ping packets with light payload suddenly, then the proportion of ICMP packets ascends and the average length of packets decreases. If the packet length is selected for stratification, sound detection results are to be achieved.

From [12], the gain of stratified sampling increases with the more stratum number; when the stratum number increases to 3 from 2, the gain won’t improve too much; the gain will appear to decrease with the increase of stratum number; when the stratum number is beyond 4, the gain of stratified sampling tend to be stable.

The determination of stratum border is based on the stratified strategy. Sample with similar characteristics may be classified into certain stratum in stratified sampling, which leads to the less variance in certain stratum. The simplest way to determine stratum border is based on the type of packets such as TCP, UDP and ICMP, and stratum border is naturally determined by protocol type of IP header of packet. If packet is classified into certain stratum according to the total length field, the stratum border is determined by the variable interval of the actual value $L_i$ ( $L_{\min} \le L_i \le L_{\max}$ ). If every stratum is based on $L_i$, it is obvious that this is impractical. In this paper, there are hundreds kinds of packets related to total length field. If every stratum is based on total length, the efficiency of this implementation is too bad. So we make use of the optimum stratified method based on the cumulate square root of stratified variable distribution [13], shown as Table 1.

| Total length of packet | Number | $\sqrt{f}$ | Cumulate $\sqrt{f}$ |
|---|---|---|---|
| $L_1$-$L_2$ | $M_i$ | $\sqrt{M_1}$ | $\sqrt{M_1}$ |
| $L_2$-$L_3$ | $M_2$ | $\sqrt{M_2}$ | $\sqrt{M_1}+\sqrt{M_2}$ |
| … | … | … | … |
| … | … | … | … |
| $L_{n-2}$-$L_{n-1}$ | $M_{n-1}$ | $\sqrt{M_{n-1}}$ | $\sqrt{M_1}+\sqrt{M_2}+\ldots+\sqrt{M_{i-1}}$ |
| $L_{n-1}$-$L_n$ | $M_n$ | $\sqrt{M_n}$ | $\sum_{i=1}^{n}\sqrt{M_i}$ |

Table 1. Cumulate square root table

Let $R=\sum_{i=1}^{n}\sqrt{M_i}$ , if the stratum number is 5, then a new stratum is generated at intervals of *R/5*, and so the four stratum border point is to make the value of cumulate $\sqrt{f}$ closed to the following value: *R/5*, *2R/5*, *3R/5*, *4R/5*, and these four values are stratum border points of stratified sampling.

For stratified sampling, it is supposed to solve the problem of sample size allocation in strata given fixed population size. Because the precision of stratified random sampling relates to sample size allocation and variance within strata, the stratification or sample size allocation would affect the efficiency of stratified sampling directly. Generally, proportional allocation is the better if the means between strata vary greatly, while optimum allocation is the better if the standard deviations between strata vary greatly. In practical sample procedure, we tend to select the proportional allocation method because optimum allocation is only toward to single target variable. In fact, there are usually more target variables. The optimum allocation of single variable may be not proper to other variables. Proportional allocation refers to allocate the size within strata based on each stratum weight.

What's more, sampling strategy in a stratum is also to be determined. In practice, both systematic sampling and random sampling are to be applied.

### 3.2 Experiments and discussion

To test the performance of different sampling strategies, we conduct experiments in our campus network. The network traffic is captured by a host linked to the router of a C class subnet. Experimental results are shown in Figure 10 and Table 2 as follows.

| | Population | Systematic sampling | Poisson sampling | Stratified sampling |
|---|---|---|---|---|
| UDP | 82.47% | 82.14% | 82.85% | 82.5% |
| TCP | 11.87% | 11.92% | 11.94% | 11.73% |
| ICMP | 5.66% | 5.95% | 5.21% | 5.77% |

Table 2. Packet type proportion

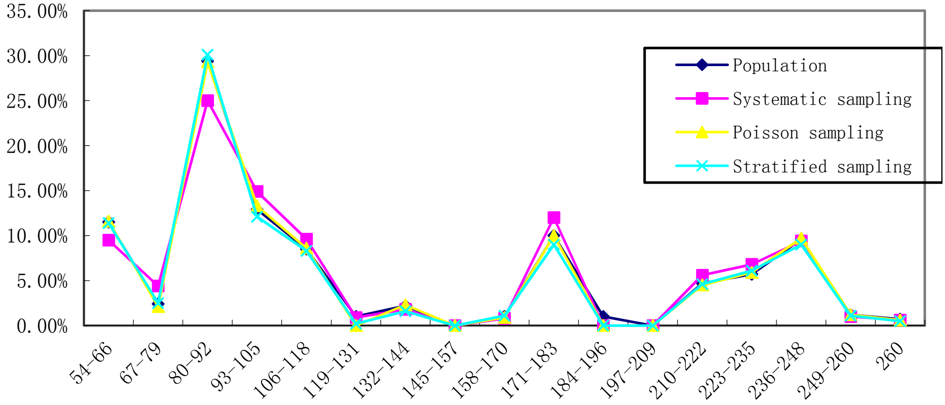


Fig. 10. Packet length distribution curve

Figure 10 is the packet length distribution curve, and table 2 shows the proportion of different type of packets in population and separate sampling strategies. From figure 1, it can be seen that the distribution of packet length in Poisson sampling and stratified sampling conforms to the distribution of packet length in population. Though there is a little diversity between the distribution of packet length in systematic sampling and the distribution of population, this diversity is also in the acceptive scope of precision. Table 2 shows the proportion of different kind of packet with diverse sampling strategies accords with population and these sampling strategies are applicable.

In practice, sampling may also be applied to intrusion detection by estimating the value of certain measure in population. The following parts briefly present the relative efforts with the average length of packets as the measure.

Assume the mean of population $X$ is $\mu$, variance is $\sigma^2$. Let $X_1$, $X_2$, $X_3$, …, $Xn$ are the sample of population $X$, then $E\left(\overline{X}\right)=\mu, D\left(\overline{X}\right)=\sigma^2/n$, and $X \sim N\left(\mu,\sigma^2\right)$, so $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ also obeys normal distribution, that is $\overline{X} \sim N\left(\mu,\sigma^2/n\right)$ [14].

For given confidence $1-\alpha$, let $X_1$, $X_2$, $X_3$, …, $Xn$ are the sample of population $N\left(\mu,\sigma^2\right)$, $\overline{X}$ and $S^2$ are the corresponding mean and variance of sample, then

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1) \tag{1}$$

and right side distribution $t(n\text{-}1)$ doesn't rely on any other parameters, then

$$P\{-t_{\alpha/2}(n-1)<\frac{\overline{X}-\mu}{S/\sqrt{n}}<t_{\alpha/2}(n-1)\}=1-\alpha \tag{2}$$

that is

$$p\{\overline{X} - \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1) < \mu < \overline{X} + \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1)\} = 1 - \alpha$$

so the confidence interval of $\mu$ with given confidence $1-\alpha$ is

$$(\overline{X} \pm \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1)) \tag{3}$$

The measured value of average length of packet with systematic sampling is 134.0631, standard variance is 73.607 and total number of packet *n* is 92647. From the distribution table of *t*, the value of *t(n-1)* appears to be a constant when *n* run to infinite.

We estimate the average packet length of population by systematic sampling and ensure the confidence of this length is 95%. From (3), we get the confidence interval of average packet length of population with confidence 95% is $(134.0634 \pm 0.47)$. If the error is no more than 0.94, the confidence of this error is 95% considering any value in this interval as the estimate of packet average length in population. From preceding experimental results, the value of average length of packet in population is 134.4601, which is in the interval of $(134.0634 \pm 0.47)$.

It can be seen from experiments that the average length of packet in normal traffic tends to be stable. When ICMP sweep attack appears, there are lots of packets with short length, and average length of packet is obviously various [15]. So the average length may be considered as the measure to detect intrusions.

With the change of user behavior and network topology, the characteristic of network may also vary. So the data related to the behavior of certain network during some time may be chosen to characterize the normal characteristics rather than all the passed data. Assume the granularity of time is *T*, $L_i$ is the average length of packet in the *ith* time interval. Anomaly actions are to be detected by comparing current sampled value and the value of preceding *i-1th* time interval. That is, if current average length of packet is in the scope of normal value calculated by preceding sampled data, there is no intrusion, and then the preceding data can be updated. Otherwise, there appears intrusion, and current data can’t be updated.

### 3.3 Summary

With the rapid development of network technology, there are more severe challenges to information security, and IDS has been an indispensable part of computer security. However, there appears packet drop for IDS especially in a high speed network environment. In this chaper, we apply packet selection model based on sampling methods of statistics to the procedure of data collection of IDS. Experiment results show that selected sample (packets) can be applied to detection and analysis for IDS in the scope of certain precision. In short, our method has the following advantages: firstly, this method exceedingly strengthens the processing performance of IDS by the means of replacing dropping packets passively with sampling packets actively especially in the large-scale high-speed network; secondly, this method has better expansibility, and various sampling strategies may be applied corresponding to different implementation.

## 4. STAMP -A high-speed traffic collection platform based on sampling on FPGAs

### 4.1 Design and implementation of STAMP

In this paper, we describe the design and implementation of a uniform high-speed traffic collection platform for intrusion detection/prevention based on sampling on FPGAs. To achieve this goal, HSTCP's architecture integrates elephant flow identification and adaptive elephant flow sampling into a FPGA prototyping board, which is a gigabit Ethernet network interface card with open hardware and software specifications.

A flow is a sequence of packets that share certain common properties (called flow specification) and have some temporal locality as observed at a given measurement point. Depending on the application and measurement objectives, flows may be defined in various manners such as source/destination IP addresses, port numbers, protocols, or combinations thereof. They can be further grouped and aggregated into various granularity levels such as network prefixes or autonomous systems. In this paper, we present flow statistics and experimental results using flows of 5 tuple (source/destination IP addresses, port numbers, and the protocol number) with a 60-s timeout value as our basic flow definition.

As many measurement-based studies have revealed, flow statistics exhibit strong heavy-tail behaviors in various networks (including the Internet). This characteristic is often referred to as the elephant and mice phenomenon (aka the vital few and trivial many rule), i.e., most flows (mice flows) only have a small number of packets, while a very few flows (elephant flows) have a large number of packets. A noticeable attribute of elephant flows is that they contribute a large portion of the total traffic volume despite being relatively few in the number of flows. In this paper, we define an elephant flow as a flow that contributes more than 0.1% of all unsampled packets.

The elephant flow identification module maintains an array of counters for every flow. Counters at certain index would contain the total number of packets belonging to all of the flows colliding into this index.

At intervals of certain time (60 s), the adaptive elephant flow sampling module would adjust the sampled rate according to the traffic load changes in the identified elephant flow. The sampled rate is based on the packet count. An AR model is used for predicting the number of packets of a certain elephant flow in the next time interval.

HSTCP is built on the Avnet Virtex-II Pro Development Board shown in Figure 11. This FPGA prototyping board includes all of the components necessary for a gigabit Ethernet network interface with embedded processors and on-board memory.

Fig. 11. HSTCP PCI card

### 4.1.1 Elephant flow identification

Identifying elephant flows is very important in developing effective and efficient traffic engineering schemes. In addition, obtaining the statistics of these flows is also very useful for network operation and management. On the other hand, with the rapid growth of the link speed in recent years, packet sampling has become a very attractive and scalable means to measure flow statistics.

To identify elephant flows, traditionally we have to collect all packets in the concerned network, and then extract their flow statistics. As many previous studies have indicated, however, such an approach lacks scalability. For very high speed links (say, OC-192+), directly measuring all flows is beyond the capability of measurement equipments (i.e., the requirements for CPU power, memory/storage capacity, and access speed are overwhelming).

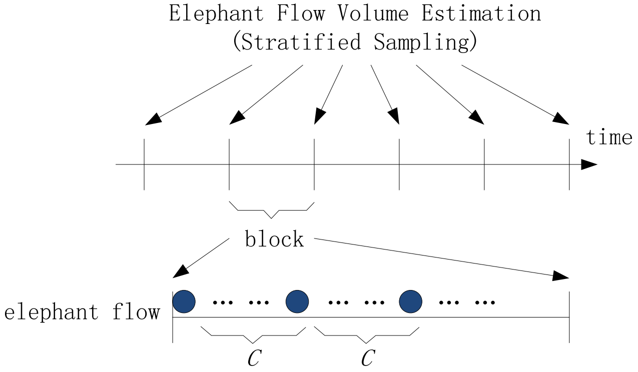


Fig. 12. Elephant flow volume estimation.

Because flows are dynamic in their arrival time and active duration, it is very hard to define a sampling interval that is valid for all elephant flows, while allowing us to adjust the sampling rate in accordance with the changing traffic condition to ensure estimation accuracy. We tackle this problem by using stratified sampling. Sequenced predetermined, nonoverlapping time blocks are called strata. In each block, systematic count - based sampling is applied, that is every Cth packet of the parent process is deterministically selected for sampling, starting from some starting sampling point, and other packets will be directly dropped. At the end of each block, flow statistics are estimated. Then, naturally, a flow's volume is summarized into a single estimation record at the end of the last time block enclosing the flow. Notice that from each flow's point of view, its duration is divided or stratified in a fixed time. The predetermined time blocks enable us to estimate the flow volume without knowing dynamic flow arrival times and their durations while adjusting the sampling rate according to dynamical traffic changes.

The elephant flow identification module maintains an array of counters. Upon the arrival of a packet, its flow specification is hashed to generate an index into this array, and the counter at this index is incremented by 1. Collisions due to hashing might cause two or more flow labels to be hashed to same indices. Counters at such an index would contain the total number of packets belonging to all of the flows colliding into this index. We do not have any explicit mechanisms to handle collisions as any such mechanism would impose additional processing and storage overheads that are unsustainable at high speeds. This makes the encoding process very simple and fast. Efficient implementations of hash functions allow the online streaming module to operate at speeds as high as OC-768 (40 Gbps) without missing any packets.

Internet traffic is known to have the property that a few flows can be very large, while most other flows are small. Thus, the counters in our array need to be large enough to accommodate the largest flow size . On the other hand, the counter size needs to be made as small as possible to save precious SRAM. Recent work on efficient implementation of statistical counters provides an ideal mechanism to balance these two conflicting requirements, which we will leverage in our scheme. For each counter in the array, say 32 bits wide, this mechanism uses 32 bits of slow memory (DRAM) to store a large counter and maintains a smaller counter, say 7 bits wide, in fast memory (SRAM). As the counters in SRAM exceed a certain threshold value (say 64) due to increments, it increments the value of the corresponding counter in DRAM by 64 and resets the counter in SRAM to 0. There is a 2-bit per counter overhead that covers the cost of keeping track of counters above the threshold, bringing the total number of bits per counter in SRAM to 9. For suitable choices of parameters, this scheme allows an efficient implementation of wide counters using a small amount of SRAM. This technique can be applied seamlessly to implementing the array of counters required in our data streaming module. In our algorithm 6, the size of each counter in SRAM is 9 bits and in DRAM is 32. Also, since the scheme in [23] incurs very little extra computational and memory access overhead, our streaming algorithm running on top of it can still achieve high speeds such as OC-768.

### 4.1.2 Adaptive elephant flow sampling

Traffic measurement and monitoring serves as the basis for a wide range of IP network operations and engineering tasks such as troubleshooting, accounting and usage profiling, routing weight configuration, load balancing, capacity planning, etc. Traditionally, traffic measurement and monitoring is done by capturing every packet traversing a router interface or a link. With today’s high-speed (e.g., gigabit or terabit) links, such an approach is no longer feasible due to the excessive overheads it incurs on line-cards or routers. As a result, packet sampling has been suggested as a scalable alternative to address this problem. In this paper, we have investigated two sampling techniques, namely, simple random packet sampling and adaptive weighted packet sampling.

Given the dynamic nature of network traffic, static sampling does not always ensure the accuracy of estimation, and tends to oversample at peak periods when efficiency and timeliness are most critical. More generally, static random sampling techniques do not take traffic dynamics into account; thus they cannot guarantee that the sampling error in each block falls within a prescribed error tolerance level.

In other words, under some traffic loads, static count-based sampling may be poorly suited to the monitoring task. During periods of idle activity or low network loads, a big sampling count provides sufficient accuracy at a minimal overhead. However, bursts of high activity require a small sampling count to accurately measure the network status at the expense of increased sampling overhead. To address this issue, adaptive sampling techniques can be employed to dynamically adjust the sampling count and optimize accuracy and overhead.

In this paper, we investigated adaptive sampling techniques to intelligently sample the incoming elephant flows. A key element in adaptive sampling is the prediction of future behavior based on the observed samples. The AR model described in this section predicts the packet count of the next sampling interval based on the past samples. Inaccurate predictions indicate a change in the elephant flow load and require an increased/decreased sampling count to determine the new value.

In any case, we cannot accurately choose the sampling count when the population size (total packet count of the observation time block) is unknown. We can compute the sampling probability at the beginning of a block by predicting the total packet count of a certain elephant flow. We employ an AR model for predicting the total packet count $m_h^f$ of the *h*th block of elephant flow *f*, as compared to other time series models, since it is easier to understand and computationally more efficient. In particular, using the AR model, the model parameters can be obtained by solving a set of simple linear equations, making it suitable for online implementation.

We will now briefly describe how the total packet count $m_h^f$ of the *h*th block of elephant flow *f* can be estimated, based on the past packet counts using the AR(u) model, where u is the lag length. Using the AR(u) model, $m_h^f$ can be expressed as

$$m_h^f = \sum_{i=1}^{u} a_i m_{h-i}^f + e_h$$

where $a_i$, *i*=1, …, *u*, are the model parameters, and $e_h$ is the uncorrelated error (which we refer to as the prediction error).

The model parameters $a_i$, *i*=1, …, *u*, can be determined by solving a set of linear equations in terms of *v* past values of $m_i^f$ *'s*, where *v*≥1 is a configurable parameter independent of *u*, and is typically referred to as the memory size.

Let $\hat{m}_h^f$ denote the predicted packet count of the *h*th block of elephant flow *f*. Using the AR(*u*) prediction model, we have

$$\hat{m}_h^f = \sum_{i=1}^{u} a_i m_{h-i}^f \ .$$

Using the AR prediction model, at the end of each block, the model parameters ($a_i$) are computed. The complexity of the AR prediction model parameter computation is only *O*(*v*) where *v* is the memory size.

The predicted $\hat{m}_h^f$ is then compared with the actual value of the sample $m_h^f$. A set of rules is applied to adjust the current sampling count, $\Delta C_{\text{curr.}}=c(h)-c(h-1)$, to a new value, $\Delta C_{\text{new}}$, which is used to adjust the sampling count to compare the rate of change in the predicted sample value, $\hat{m}_h^f - m_{h-1}^f$, to the actual rate of change, $m_h^f - m_{h-1}^f$. The ratio between the two rates is defined as *R*, where

$$R = \left| \frac{\hat{m}_h^f - m_{h-1}^f}{m_h^f - m_{h-1}^f} \right| .$$

The value of $R$ will be equal to 1 when the predicted behavior is same as the observed behavior. We define a range of values $R_{MIN} \leq 1 \leq R_{MAX}$, such that

if $R < R_{MIN}$, that is $\Delta C_{New} < \Delta C_{Curr.}$, then

$$\Delta C_{New} = \lfloor (R) \times \Delta C_{Curr.} \rfloor .$$

If $R_{MIN} < R < R_{MAX}$, then

$$\Delta C_{New} = 2 \times \Delta C_{Curr.}$$

If $R > R_{MAX}$, that is $\Delta C_{New} < \Delta C_{Curr.}$, then

$$\Delta C_{New} = \lceil (1 + R) \times \Delta C_{Curr.} \rceil .$$

Otherwise,

$$R_{undefined} \Rightarrow \Delta C_{New} = 2 \times \Delta C_{Curr.}$$

### 4.1.3 FPGA implementation

HSTCP is built on the Avnet Virtex-II Pro Development Board, and its architecture satisfies the constraints of the Avnet board while efficiently integrating the components of a gigabit Ethernet network interface. The MAC unit, DMA unit, inter-FPGA bridge, hardware event management unit, and PCI interface were custom designed for HSTCP. The remaining components were provided by Xilinx and used with little or no modification. Both the MAC unit and the PCI interface are built around low-level interfaces provided by Xilinx; however, those units are still mostly custom logic to integrate them into the rest of the system and to provide flexible software control over the hardware functionality.

The Xilinx Virtex-II Pro FPGA on the Avnet development board contains most of the NIC logic, including the PowerPC processors, on-chip memories, MAC controller, DMA unit front-end, and DDR memory controller. The smaller Spartan-IIE FPGA contains the PCI controller, the back-end DMA controller, and a SRAM memory controller. The SDRAM, although connected to a shared data bus between the Spartan and Virtex FPGAs, was not used because the entire bus bandwidth was needed to efficiently use the PCI interface.

To save development time, prebuilt Xilinx cores were used for several of the hardware modules, including the PCI interface, DDR controller, and a low-level MAC. However, these cores cannot be connected directly to form a working NIC. For example, although Xilinx provides a PCI core, it must be wrapped within a custom DMA unit to allow the PowerPC to initiate and manage high-performance burst transfers to/from the host system across the FPGA bridge. Similarly, although Xilinx provides a low-level MAC core, it must be outfitted with an advanced descriptor-based control system, data buffers, and a DMA unit to transfer packet data between NIC memory and the PHY. Finally, the DDR controller required modifications to function in this specific development board with its unique wiring.

The processors, memories, and hardware units are interconnected on the Virtex FPGA by a processor local bus (PLB). The PLB is a high-performance memory-mapped 100-MHz 64-bit-wide split-transaction bus that can provide a maximum theoretical bandwidth of 12.5 Gbits/s in a full duplex mode. The PLB allows for burst transmissions of up to 128 bytes in a single operation, which is used by the DMA and MAC units to improve memory access efficiency.

Also attached to the PLB is a small memory-mapped control module used to route descriptors to or from the MAC and DMA hardware assist units. This module also provides a central location for hardware counters, event thresholds, and other low-level control functions. By attaching this central control unit to the PLB, either PowerPC processor can manipulate these important NIC functions. The control unit takes 18 PowerPC cycles to read and 12 cycles to write a 32-bit word, primarily due to bus arbitration delays.

## 4.2 Experiment

We evaluated HSTCP using a synthetic dataset that was generated by combining the data from the 1999 DARPA intrusion detection project and 2000 DARPA "Scenario Datasets" that have been crafted to provide examples of multiple component attack scenarios instead of the atomic attacks as found in past evaluations [24] and Münchner Wissenschaftsnetz (MWN), Germany. The MWN provides Internet connectivity to two major universities and a number of research institutes. Overall, the network contains about 50,000 individual hosts and 65,000 registered users. The trace "mwn-cs-full" that we analyzed is a 2-h trace including the full payload of all packets to/from one of the CS departments in MWN, with some high-volume servers excluded.

The 1999 DARPA dataset that we used consisted of 5 weeks of TCPdump data. Weeks 1 and 3 have normal attack-free network traffic. Week 2 consists of network traffic with labeled attacks, while weeks 4 and 5 contain 201 instances of 58 different attacks, 177 of which are visible in the TCPdump data. The 2000 DARPA dataset includes two recently created scenario datasets that address the needs of mid-level correlation systems. Each includes several hours of background traffic and a complete attack scenario. Attacks and background traffic were run on the same testbed used in the 1999 evaluation, but with the addition of a commercial off-the-shelf firewall and demilitarized zone (DMZ) network separating the Internal and Internet networks and a Solaris 2.7 victim host.

The experimental evaluation of HSTCP has been divided into three steps. In section 4.1, we evaluate the performance of the sampling algorithm and compare its performance with the stratified random sampling algorithm. Then in section 4.2 we evaluate the performance of HSTCP for intrusion detection/prevention.

### 4.2.1 Evaluation of the sampling algorithm

Experiments were conducted to compare and evaluate the performance of the proposed adaptive sampling algorithm with the stratified random sampling algorithm.

In 1994, Leland *et al.* showed that the Ethernet traffic consisted of slowly decaying packet count bursts across all time scales. Time series that consist of such a pattern are said to exhibit the property of long-range dependence and are termed as "self-similar." Similar self-similar behavior has also been observed in wide-area Internet traffic by other researchers . One important characteristic of a self-similar process is that its degree of self-similarity can be expressed with a single parameter, namely the Hurst parameter which can be derived from the rescaled adjusted range (R/S) statistic. It is defined as follows.

For a given set of observations X1, X2, …, Xn with a sample mean $\bar{X}(n)$ and sample variance $S^2(n)$, the rescaled adjusted range or the R/S statistic is given by

$R(n)/S(n) = 1/S(n)\left[\max(0, W_1, W_2, ..., W_n) - \min(0, W_1, W_2, ..., W_n)\right]$, with

$W_k = (X_1 + X_2 + \cdots X_k) - k\bar{X}(n)$, k=1, 2, …, n. Hurst (1955) found that many naturally occurring time series appear to be well represented by the relation $E\left[R(n)/S(n)\right] \square\ \alpha_4 n^H$, as $n \to \infty$, with Hurst Parameter H "typically" about 0.73. On the other hand, if the observations Xk come from a short-range dependent model, then Mandelbrot and Van Ness (1968) showed that $E\left[R(n)/S(n)\right] \square\ \alpha_5 n^{0.5}$, as $n \to \infty$. This discrepancy is generally referred to as the Hurst effect or Hurst phenomenon.

It is very important whether the traffic data sampled by the proposed sampling scheme retains the self-similar property for various anomaly detection techniques, which may directly affect the accuracy and efficiency of detection. So we verify this based on two different parameters: the mean of the packet count and the Hurst parameter. The peak-to-mean ratio (PMR) can be used as an indicator of traffic burstiness. The PMR is calculated by comparing the peak value of the measure entity with the average value from the population. However, this statistic is heavily dependent on the size of the intervals, and therefore may or may not represent the actual traffic characteristic. A more accurate indicator of traffic burstiness is given by the Hurst parameter. The Hurst parameter (*H*) is a measure of the degree of self-similarity. In this paper we use the R/S statistical test to obtain an estimate for the Hurst parameter. We run the test on both the original and the sampled data.

In our sampling scheme, simple random sampling is conducted in every time block (strata) and this refers to stratified random sampling.

In Figures 3 and 4, we show the average sampling error for the Hurst parameter and the sample mean, respectively. As one can see in Figure 13, the stratified random sampling algorithm resulted in a higher average percent error for the Hurst parameter when compared to adaptive sampling. This could be the result of missing data spread out over a number of sampling intervals. In Figure 14, the average percentage error for the mean statistic was marginally lower for our sampling algorithm when compared with the stratified random sampling algorithm, albeit the difference was insignificant. One possible reason for this marginal difference is the inherent randomness nature of stratified random sampling algorithm—i.e., the weighted mean packets are nicely sampled randomly, which results in good estimation of mean.

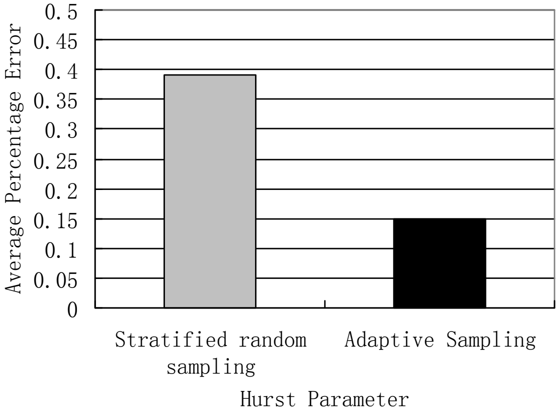


Fig. 13. Average percentage error for the Hurst parameter.

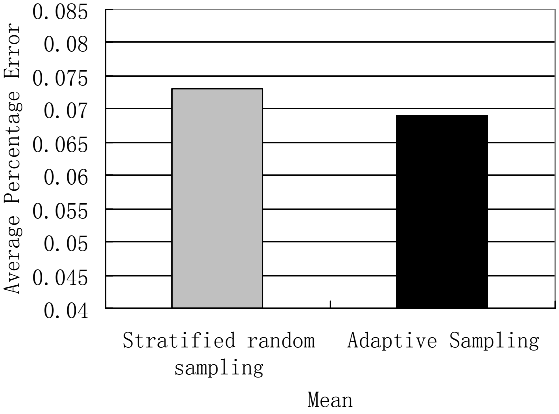


Fig. 14. Average percentage error for the mean statistic.

### 4.2.2 Evaluation of the platform for intrusion detection/prevention

The IXIA1600T (a special network test facility of IXIA Crop) is used to transmit the synthetic dataset in 1000M Ethernet. The ISS RealSecure gigabit network is selected to detect attacks (intrusions), and HSTCP is used for collecting network traffic for it. As the first commercial IDS, RealSecure has been playing an important role in this field, and it provides network intrusion detection and response capabilities that monitor the gigabit network.

The ROC (receiver operating characteristic) technique is used to evaluate the detection effect of ISS RealSecure. The ROC approach analyzes the tradeoff between false alarm and detection rates for detection systems. It was originally developed in the field of signal detection. More recently, it has become the standard approach to evaluate IDSs. In this paper, we mainly take into account the detection rates of RealSecure when its false alarm rate is under 0.2. Too many false alarms will make administrators consume unnecessary time and energy analyzing these alarms, which compromises the usability and validity.

In Figure 15, we can see that HSTCP was able to cope with the 1-Gbps network traffic, and elephant flow sampling was not initiated. The detection rates of RealSecure remarkably increased during initial phases, and then tended to be stable.

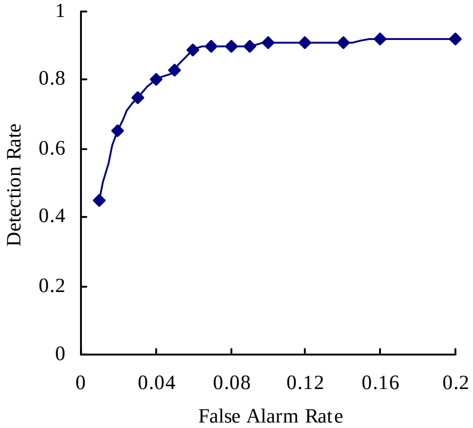


Fig. 15. The ROC curves at 1-Gbps speed.

To evaluate the performance of HSTCP for intrusion detection in the high-speed network, we used IXIA1600T to transmit the synthetic dataset at a 10-Gbps speed. Under this circumstance, HSTCP could not capture the whole traffic and tended to drop packets. In Figure16, we can see that the detection rates sharply decline without sampling. While HSTCP initiated elephant flow sampling methods to cope with the network traffic, RealSecure could still make response to the high-speed traffic. With the increase in false alarm rates, detection rates remained high and stable.

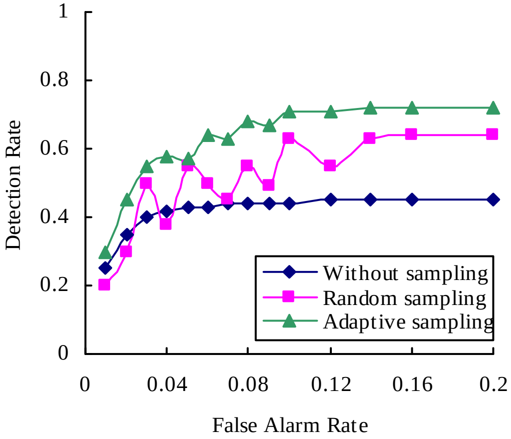


Fig. 16. The ROC curves at a 10-Gbps speed.

We compared the performance of the adaptive sampling technique and the random sampling technique. As expected, the adaptive sampling technique showed superior performance.

### 4.3 Summary

In this subsection, we have presented a uniform high-speed traffic collection platform for intrusion detection/prevention based on sampling on FPGAs—called HSTCP—that has the ability to cope with very high speed network traffic (even Tbps). By employing complete mice flows' capture and adaptive elephant flow sampling, HSTCP effectively reduces the volume of network traffic for intrusion detection/prevention without losing its intrinsic characteristics. In addition, HSTCP provides a flexible and scalable platform for network IDSs/IPSs faced to the challenge of future high-speed networks.

## 5. The research of trusted communication protocol for intrustion prevention system

### 5.1 IPS trusted communication mechanism

A trusted communication mechanism proposed in this chapter applied to the correlation between firewall and IDS in a distributed intrusion prevention system. It is mainly based on middleware technology and security Protocol standard techniques. And the middleware technology is CORBA. If CORBA underlies network layer, it may encapsulate the underlying unit; which may make the application transparent to the up-layer. In this paper, TLS is applied to the trusted data transmission between firewall and IDS.

### 5.2 The design of IPS trusted communication protocol

The design of the mechanism of trusted data communication is illustrated by figure 17:

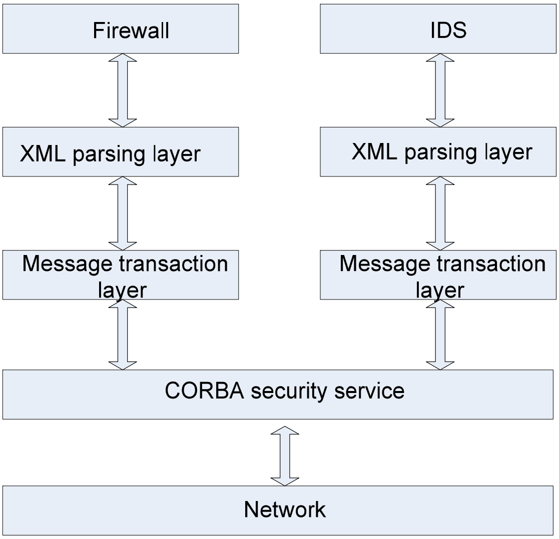


Fig. 17. The mechanism of trusted communication for IPS

**Application Layer**

The application layer mainly refers to Client and Server; in this paper it represents firewall and IDS. During the transmission, both of the firewall and IDS can be client or server. Firewall and IDS system are all in the network layer. Firewall provides the data source and a place to process the final data while IDS is responsible for receiving data request from firewall and analyzing them, and then return the processed results to the firewall.

**XML parsing layer**

This layer primarily encapsulates and analyzes the communicated data.

**Message transaction layer**

In this paper, a security protocol called TLS (Transport Layer Security) applied to supporting the security and reliability for data communicate among each other. It also may protect the privacy of the applications and users for network communication. When server and client are communicated, TLS could make sure important messages won't be sniffered or stolen by the third party. It is a successor protocol followed by SSL.

**CORBA security service**

CORBA security service (CORBASec) is an important public object service in CORBA. It constructs secure language environment between client objects and service objects, and also provides better security service [10].

### 5.3 Design of data exchange format based on XML

In a distributed intrusion prevention system, data filter module is composed of a firewall and other components. Network data processing module generally refers to IDS. In this paper, data transmitted between firewall and IDS are classified into four categories: event data, rule data, analysis result data and actions response data, and it is referred to the function units in CIDF framework [11].

The relationship among above data: Data generated by the firewall with network packets filtered called event data. Whether the event described is an intrusion event, it depends on the match or analysis of IDS based rule data. If it was a real intrusion event, then generated the analysis result data. The firewall will make a response to the analysis result data based on the corresponding strategies, and generates action response data.

**Design of event data**

Data filtered original network packets by firewall according to security strategy is event data. Therefore, this kind of data must contain the complete description of network original data that IDS can detect or analyze by those matching information in rules. In addition, it needs to contain the firewall name and the time of event happened. The reason of including firewall name is that more firewalls may be deployed in network. When detecting the network data, more than one firewall will find the same event, so we need to distinguish and analyze. In some cases, IDS need to detect what happened during some phases and then ensure whether any intrusions had happened, so the data type also contains the time of event happened. At last, to support data expansion, we need additional data applied to describing some additional description in event data, and it can be used as a reserved interface. The illustration of event data is as follows:

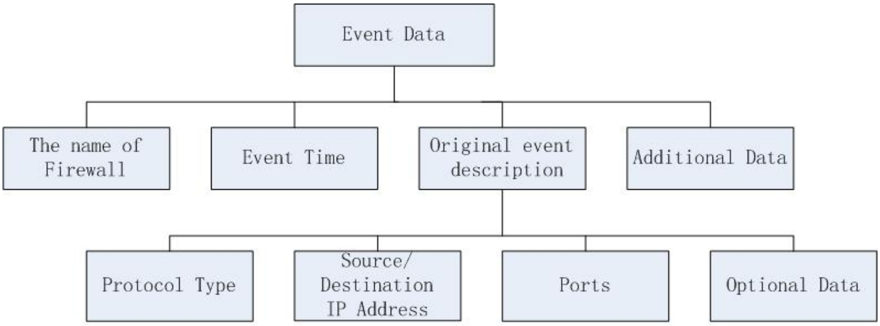


Fig. 18. The design of event data

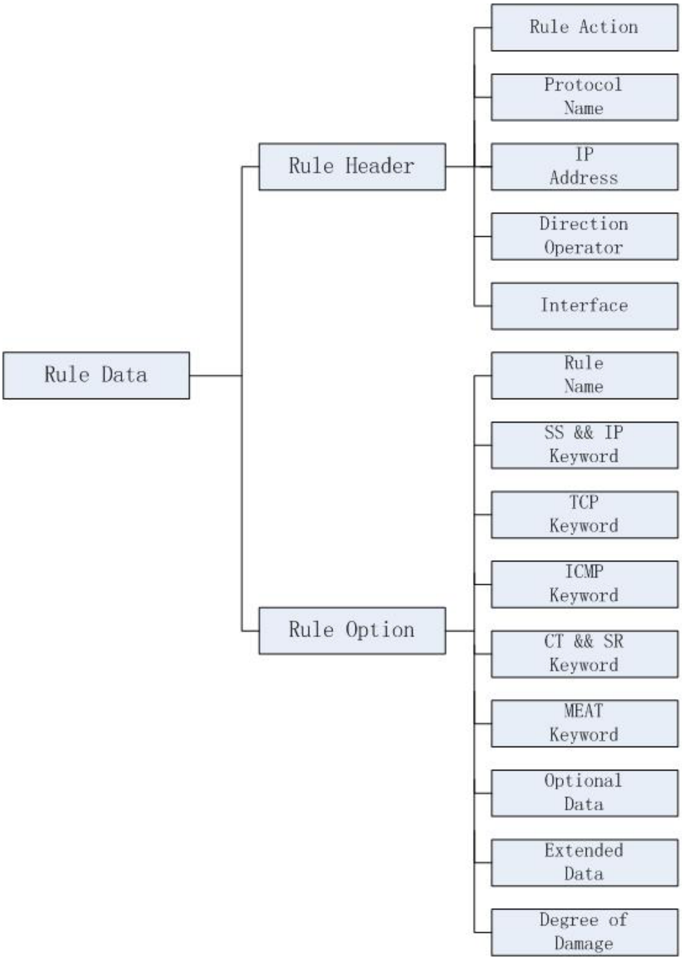


Fig. 19. The design of rule data

The XML DTD of event data is provided in details as follows, but other types are ignored.

**Design of rule data**

According to snort rules, the rule data divided into two logic parts: Rule header and rule options. The rule header contains rule actions, protocols, source and destination IP Address and mask, source or destination ports and some direction operators. The rule option contains some alert messages and the main parts of checked packets; it includes the characteristics and the priority. And the part of characteristics should be described by one or more keywords. Same as the event data, rule data also need contain the reserved interface applied to the expansion of rule data, and then put them into rule options. The design of rule data is shown in Figure.3.

**Design of analysis results data**

To analyze the data delivered to IDS and the rule data of IDS then get the analysis results data. After the detection analysis by IDS, the event can be classified into three categories: a malicious attacked event, a security event being suspected but not confirmed and normal network traffic. To the malicious attacks event, we use the obstructive response methods and record this event; it needs some alert response methods to those being suspected but not confirmed security events. Just writing down the key information to the normal network traffic is enough. However, none of these will be directly solved in IDS. IDS will put the analysis results data through the central control module and match it, then feedback to firewall if it was allowed, and firewall will tackle them according to the response actions. So the analysis results data should describe the event type, event processing action, time of the event, which IDS dealt with the event and so on. If it was the malicious or suspected event, we also need to describe the name of event, the source of event, the target of event, when attack happened and how it did affect the system. The same as the first two data structure, we also need to give the analysis results data a reserved interface which is called expansion data. The design of analysis results data is shown in Figure.20.

**Design of action response data**

According to response actions, firewall deal with the results data detected by IDS and gets the action response data. During this, the normal network traffic don’t need to be recorded, whereas put the malicious or suspected events and corresponding processing actions into the record, thus the action response data only defined to IDS. So the action response data should present which firewall has been responded to the analysis data? What is attack event called? What is the source from? What is the target? When has it happened? And how does it compromise the system? All of these relate to the corresponding attack events in the analysis result events. These records used to keep the respond data integrity. The action response data should also record the final processing results analyzed by firewall to the results data, which is called response action. In addition, the receiver should be provided either. At last, including the extended data and the implementation of data extended. The design of action response data is shown in Figure.21.

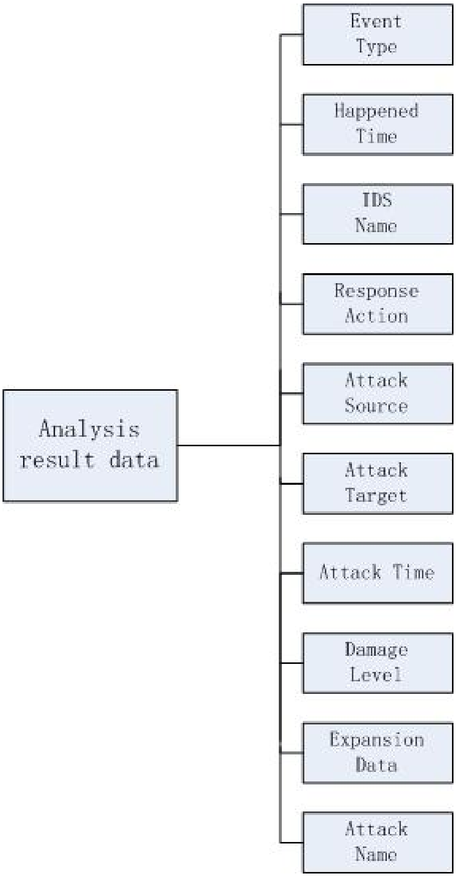


Fig. 20. The design of analysis results data

### 5.4 Related works

IDXP (Intrusion Detection Exchange Protocol) is an application layer protocol, which is applied to the data exchange among intrusion detection entities (Feinstein & Matthews, 2007), and it may support the transformation of the IDMEF (Intrusion Detection Message Exchange Format) message, unstructured text and binary dataset (Debar et al., 2007). Again, it has the security characteristics of bidirectional authentication, integrity and confidentiality based on connection-oriented protocols. However, IDXP and IDMEF are simply data exchange protocols, and they aren't adaptive to other correlation schemes.

IAP (Intrusion Alert Protocol)[9] is a transport protocol applied to the alert data of intrusion detection, which is based on TCP protocol and TLS protocol used for the secure data transmission (Gupta et al., 2001). Alert data is described in XML, and it conforms to the format specification of IDMEF.

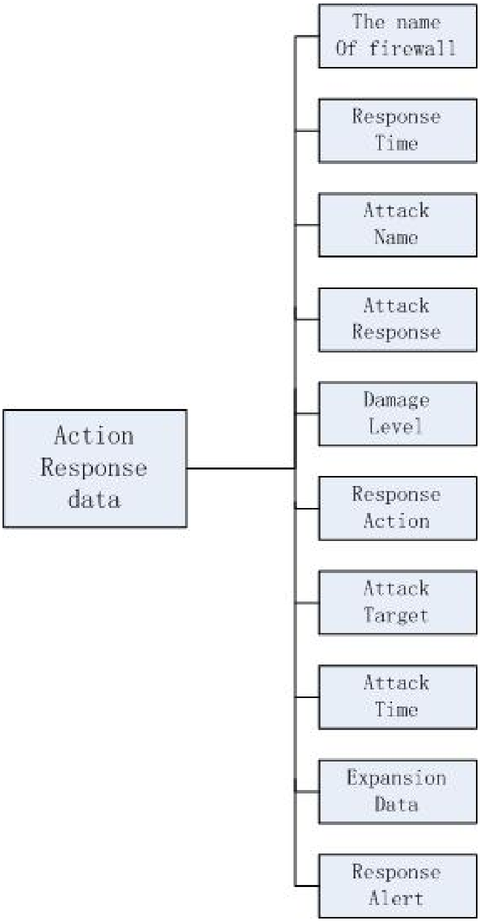


Fig. 21. The design drawing of action response data

IPIEP (Intrusion Protection Interaction Exchange Protocol) is an application layer protocol applied to the definition of exchange rules for correlation messages, while IPIMEF (Intrusion Protection Interaction Message Exchange Format) provides the definition of data format for message, and they don't have corresponding dependent relationship.

Above all, all of these related works don't provide the support to the trusted communication transmission between data filter module and network data process module in a distributed intrusion prevention system.

### 5.5 Summary

The main contribution of this subsection is to provide a trusted communication protocol between data filter module and network data processing module in a distributed IPS. The transmit data can be classified into four categories with XML technology defined respectively, and XML DTD description language is applied to the definition of these four data formats, and then a data trusted communication mechanism is provided, which divided into three layers: application layer, XML parsing layer and message transaction layer. CORBA technology is applied to solving the heterogeneous platform and load-balancing problems. TLS security protocol standard is to ensure the integrity and security during data transmission. The trusted communicated protocol also can be adaptive to network security products and network management equipments, and it contributes to security data fusion and detecting sophisticated distributed network attacks.

## 6. Conclusion

With the rapid development and comprehensive application of network technology, Internet significantly contributes to the development of the human society. Meanwhile, network security problems gradually appear serious. Traditional firewall technologies can’t provide sufficient security protection against various attacks and intrusions, while intrusion detection systems (IDS) are faced with compromise between false alarms and false positives, so intrusion prevention system (IPS) come into being. IPS may block malicious attack traffic before corresponding intrusions cause more severe damage other than simply generate intrusion alarms.

In this chaper, we investigate intrusion detection and intrusion prevention in high speed network and the main research work is as follows:

1. Based on the investigation on the recent trends of network security techniques, such as firewall and IDS, we propose a intrusion prevention scheme based on the correlation between IDS and firewall. This scheme complements the fundamental flaws of IDS and firewall, and it may provide real-time, active prevention and attempts to stop attacks, which contributes to normal transmission of legal network traffic. In this paper, we present the design and implementation of a prototype system of network IPS —— DXIPS, based on the correlation between Snort_inline and Netfilter configured by IPtables. The hierarchical architecture of this system includes intrusion prevention layer, server layer and control layer, in which intrusion prevention layer monitors the traversing traffic and conducts intrusion detection and prevention; server layer collects log data and translate them into readable formats; control layer is administrational console and perform data display. The system is design with modularization, which includes intrusion prevention module, log recording module, central control module and communication module, and the concrete implementations of these modules are presented. The deployment policies are discussed according to various applications environment. Netfilter is a built-in firewall in the kernel of Linux, which belongs to the latest fifth generation firewall. It has the capability to directly filter malicious packets in the TCP/IP stack in kernel, which improves the response performance. What’s more, DXIPS provides better scalability according to various applications environment.
2. Data collection mechanism is a key factor that affects the performance of IDS/IPS. The most current products execute per-packet detection. However, with the development and widespread of high speed networking technique, the application of IDS/IPS has been faced with serious challenges. In this paper, the sampling technique in statistics is introduced into the procedure of data collection for IDS/IPS, and the new data collection module based on sampling is proposed. Three typical sampling strategies, such as systematic sampling, Poisson sampling and stratified sampling, are applied to

network traffic collection. The packet length and type serve as the measure of anomaly detection, and simulation results show that the sample traffic is still characterized as the whole network traffic, and it may provide efficient data source for anomaly detection with the lower overhead. In a short, this method exceedingly strengthens the processing performance of IDS/IPS by the means of replacing dropping packets passively with sampling packets actively with the minor degradation of detection rates, and may improve resistant to Denial of Service attacks.

3. With the ever increasing deployment and usage of gigabit networks, traditional networks Intrusion Detection/Prevention Systems (IDS/IPS) have not scaled accordingly. More recently, researchers have been looking at hardware based solutions that use FPGA's to assist network IDSs/IPSs, and some proposed systems have been developed that can be scaled to achieve a high speed over 10Gbps. However, these solutions available have inherent limitations and unable to be applied to future high speed network (Tbps). In this paper, we present a scalable traffic sampling platform for intrusion detection/prevention on FPGA, called STAMP. The methodology is when the proposed platform is unable to capture the whole network traffic; it will initiate elephant flow sampling other than merely randomly dropping packets. Meanwhile, sampling rate is adaptive to the traffic load of elephant flow. All the captured packets are forward from STAMP to IDS via PCI bus. The noteworthy features of STAMP include: it takes the self similarity of network traffic into account with the attempts to collect malicious traffic, and improve the efficiency of network traffic sampling for IDS/IPS; it employs adaptive elephant flow sampling (AEFS) to retain inherent characteristics of network traffic, which contributes to anomaly detection; it provides a flexible and scalable platform for network IDSs/IPSs that will be faced the challenge of future high-speed network.
4. To achieve the secure and reliable transmission for the interactive data between IDS and firewall, the concept of trusted communication is introduced in this paper. We give the design and implementation of a trusted communication protocol based on XML. The design and implementation of trusted communication mechanism between firewall and IDS is presented considering each functional unit of common intrusion detection framework. The CORBA middleware is applied to data transmission, and TLS secure protocol is applied to trusted transmission between IDS and firewall. The hierarchical architecture of this protocol includes application layer, XML resolution layer and message transaction layer, in which application layer consists of client and server used to capture and analyze packets; XML resolution layer translates the data into uniform XML format and provide the base for data exchange; message transaction layer employs TLS security protocol to achieve secure and trusted communication. The data type between IDS and firewall of the proposed prototype system is composed of event data, rule data, analysis result data and response action data, and the concrete descriptions of these data based on XML DTD are also provided. The proposed trusted communication protocol has the scalability to support various network security products (such as firewall, IDS, IPS, etc.) and management facilities, and may contribute to the data fusion of these facilities and detect sophisticated distributed network attacks.

## 7. Acknowledgment

The authors would like to thank the editors and anonymous reviewers for their valuable comments. This work was supported in part by the National Grand Fundamental Research

973 Program of China under Grant No. 2009CB320706, the National High Technology Research and Development Program of China under Grant No. 2011AA010101, the National Natural Science Foundation of China under Grant No. 61103197 and 61073009, Program of New Century Excellent Talents in University of Ministry of Education of China under Grant No. NCET-06-0300, the Youth Foundation of Jilin Province of China under Grant No. 201101035, and the Fundamental Research Funds for the Central Universities of China under Grant NO.200903179.

## 8. References

Anderson, J.P. (1980). Computer Security Threat Monitoring and Surveillance, Technical report, James P Anderson Corporation: Fort Washington, Pennsylvania

Allen, J.; Christie, W.; Fithen, et al. (2000). State of the practice of intrusion detection technologies, S Technical report: CMU/ SEI2992TR2028, Software Engineering Institute, Carnegie Mellon University

Denning, D. E. (1987). An Intrusion Detection Model. *IEEE Transactions on Software Engineering*, Vol.13, No.2, (1987), pp. 222-232, ISSN: 0098-5589

Drinic, M.; & Kirovski, D. (2004). A hardware-software platform for intrusion prevention, *Proceedings of 37th annual IEEE/ACM international symposium on microarchitecture*, pp: 233-242, ISBN 0-7695-2126-6, Oregon, Portland, USA, 2004,

Weaver, N.; Paxson V.; & Gonzalez, J.M. (2007). The shunt: an FPGA-based accelerator for network iintrusion prevention, *Proceedings of 2007 ACM/SIGDA 15th international symposium on field programmable gate arrays*, pp:199-206, ISBN 978-1-59593-600-4, Monterey, California, USA, 2007

Uppuluri, P.; Joshi U.; & Ray, A. (2005). Preventing race condition attacks on file-systems, *Proceedings of 2005 ACM symposium on applied computing*, pp:346-353, ISBN 978-1-58113-964-8, Santa Fe, New Mexico, USA, 2005

Feinstein, B. & Matthews, G. (2007). The Intrusion Detection Exchange Protocol (IDXP), IETF RFC 4767, 24.10.2011, Available from http://www.ietf.org/rfc/rfc4767.txt, 2007

Debar, H.; Curry, D. & Feinstein, B. (2007). The Intrusion Detection Message Exchange Format (IDMEF), IETF RFC 4765, 24.10.2011, Available from http://www.ietf.org/rfc/rfc4765.txt, 2007

Gupta, D.; Buchheim, T.; Matthews, G. et al. (2001). IAP: Intrusion Alert Protocol IETF IDWG Draft, 24.10.2011, Available from http://tools.ietf.org/html/draft-ietf-idwg-iap-05, 2001.

5

# Challenges in Building Trusted Information Systems[1]

Serena Chan and Gregory N. Larsen
*Institute for Defense Analyses,*
*USA*

## 1. Introduction

Globalization is a phenomenon that is bringing the world closer together through the exchange of raw goods, products, services, information, knowledge, and culture. Unprecedented advancements in technology, communications, science, transport, and industry have quickened the pace of global integration. The globalization process is creating and accelerating the emergence of transnational markets. Due to the presence of a worldwide market, there is a wider range of options to choose from among the products and services for building information systems.

The global supply chain and system complexity obscure "what's in the system." Systems are vulnerable to counterfeits, malicious inserts, or negligent design flaws. In today's global environment, one cannot afford to manage risks by simply seeking to avoid risks. The traditional discourse is that of risk avoidance. However, risk avoidance is untenable in an economic environment that operates globally with great variation in performance and with rapidly changing processes and technologies of consumption of production. Risks must be actively managed. Risk reduction comes at an expense with cost, schedule, and performance impacts to building trusted information systems. It may cost less to build robustness below some threshold of concern than to eliminate the risks, but it costs more than ignoring the risks. To find the right balance between the benefits, costs, and risks associated with globalization, one needs to understand how globalization works, the issues and challenges, and the subsequent system design and policy choices.

This book chapter discusses several research areas that address the effects of globalization coupled with the increasing complexity of building trusted information systems. The growing trend of globalization demands a more inclusive and persistent approach for actively managing risks in building trusted information systems. For example, the multifaceted, transitory, and global nature of the commercial information and communications technology (ICT) marketplace is limiting visibility into the supply and suppliers. One of the main challenges is verification of trustworthy components and services in the design, development, test, production, deployment, operation, and maintenance of trusted information systems.

[1] The publication of this book chapter does not indicate endorsement by the Department of Defense (DoD) or the Institute for Defense Analyses (IDA), nor should the contents be construed as reflecting the official position of those organizations.

## 2. Globalization

Globalization is linking people and things at a faster pace than ever before. With global markets, supply chains have become more intricate, uncertain, and unpredictable. Therefore, globalization presents challenging problems to assuring the integrity of components used to build trustworthy information systems and networks. Critical information systems should be composed of parts that are trusted to do only that which is expected or specified and to do so reliably and dependably. Global supply chains are vulnerable to questions of unknown product or service provenance, which subsequently leads to questionable trustworthiness of the supplied items and the suppliers in the supply chain.

Both globalization and outsourcing are creating longer supply chains. Outsourcing creates a greater dependency on outsiders – procuring ever-more-complex and more critical products from external strategic suppliers instead of developing products in-house (Bolgar, 2010). Outsourcing projects can provide a number of benefits, including cost savings, increased productivity, improved schedule performance, and higher quality of work (Kliem, 2004). However, extended supply chains greatly increase the complexity of the supply network and decrease the visibility of risks. Nevertheless, globalization provides an opportunity to increase the security of mission critical information systems. The global marketplace can be leveraged to propagate better information assurance techniques and security practices in designing and building trusted information systems.

## 3. Information systems

An information system is specifically designed to operate on information, i.e., information is the flow variable in the system. In general, systems are designed for a purpose and have the following operational properties:

- Consume (ingest)
- Process (convert)
- Produce (output)
- Control signalling (regulate operations)
- Store (hold)

A system can be defined as a combination of hardware, software, infrastructure, and trained personnel operating to achieve specified mission objectives. This definition of system includes both the communications technology and information that is employed in addition to the way in which people interact with the technology.

Modern information systems increasingly rely on globally sourced ICT components and services. The variety and abundance in the marketplace is driven by the rapid decline in cost and the rapid increase in performance advancements. As supply is able to meet the demand for low cost and more functions, today’s information systems are increasingly complex in nature.

### 3.1 Trusted information systems

One foundation for building trusted information systems is systems assurance. Systems assurance is defined as the justified confidence that the system functions as intended and is free of exploitable vulnerabilities, either intentionally or unintentionally designed or inserted as part of the system at any time during the life cycle (NDIA, 2008). The ideal scenario where no exploitable vulnerabilities exist is unrealistic. Therefore, active risk management must be performed to reduce the probability and impact of vulnerabilities to tolerable levels of risks.

Confidence establishes trustworthiness and tolerable residual risk. Trust in any information system is really the result of the methods employed to assure confidence in the system, both in its functions and protection of the information it holds and the results it produces.

## 4. Trust & risk

Trust and risk are closely related. Trust can be described as the willingness to take risk (Mayer et al., 1995, as cited in Laeequddin et al., 2008). Trust can be defined in terms of willingness to assume risk, intention in terms of willingness to assume risk, intention to make oneself vulnerable, acceptance of risk, and readiness to assume risk (Chopra & Wallace, 2003, as cited in Zuo & Hu, 2009). Meanwhile, risk is about choice; the action that is undertaken (Bernstein, 1996, as cited in Laeequddin et al., 2008).

Table 1 sorts risks into several basic categories and lists the areas they affect (Kleim, 2004). These risks are not necessarily mutually exclusive.

| **Risk Type** | **Affected Area** |
|---|---|
| **Financial** | Budget and cost |
| **Technical** | Tools, techniques, and standards |
| **Managerial** | Decision making and reporting |
| **Behavior** | Managing and leading people |
| **Legal** | Governmental laws and regulatory considerations |

Table 1. Types of Risks and Affected Areas

(Haimes, 2006) defines the following terms that have been broadly applied to risk analysis:

- *Vulnerability* – the manifestation of the inherent states of the system (e.g., physical, technical, organizational, cultural) that can be exploited to adversely affect (cause harm or damage to) that system.
- *Intent* – the desire or motivation to attack a target and cause adverse effects.
- *Capability* – the ability and capacity to attack a target and cause adverse effects.
- *Threat* – the intent and capability to adversely affect (cause harm or damage to) the system by adversely changing its states.
- *Risk* – the result of a threat with adverse effects to a vulnerable system.

The term 'susceptibility' is missing from the above list of definitions. The authors posit that one cannot manage risk unless there is an understanding of susceptibility. Understanding threat and vulnerability is necessary but not sufficient. Susceptibility is the intersection of threat (access) and vulnerability (opportunity). A viable threat requires access and a vulnerability provides an exploitable opportunity. A risk is realized when the susceptibility occurs at a certain instance or point in time. If threat and vulnerability intersect and there are no defenses, then the consequences of the realized risk must be tolerated (Chan & Larsen, 2010).

The nature of the risk being addressed in this chapter is fundamentally different from the current view of risk. The current view of risk focuses on vulnerabilities and motives to exploit the vulnerabilities. These risks are distinct from risks for which a threat actor makes deliberate investment to create an opportunity and trigger the realization of the risk for malicious purposes. Globalization creates conditions of supply that enable malicious threat actors to enrich their opportunities to craft susceptibility which may later be triggered to produce adverse consequences.

The classes of risks that have gone unaddressed are those of supply chain exploitation, particularly exploits motivated by malice. These risks are the most difficult to detect. A malicious actor may not exploit vulnerabilities immediately but insert an opportunity to exploit at some point in the lifecycle development of the item of supply. The malicious actor's opportunity can occur at any time from cradle to grave. This type of malice is hard to detect because a threat actor has opportunity separate from the invocation of the risk to be realized. The malicious actor invests in providing "extra sauce" to the item of supply being consumed. Therefore, there is a need to counter invest to provide the ways and means that manage the risk of malicious exploitation of the supply chain.

Figure 1 illustrates susceptibility analyses as the center of gravity for risk management with respect to supply chain exploits. In a general context, any risk management effort needs a

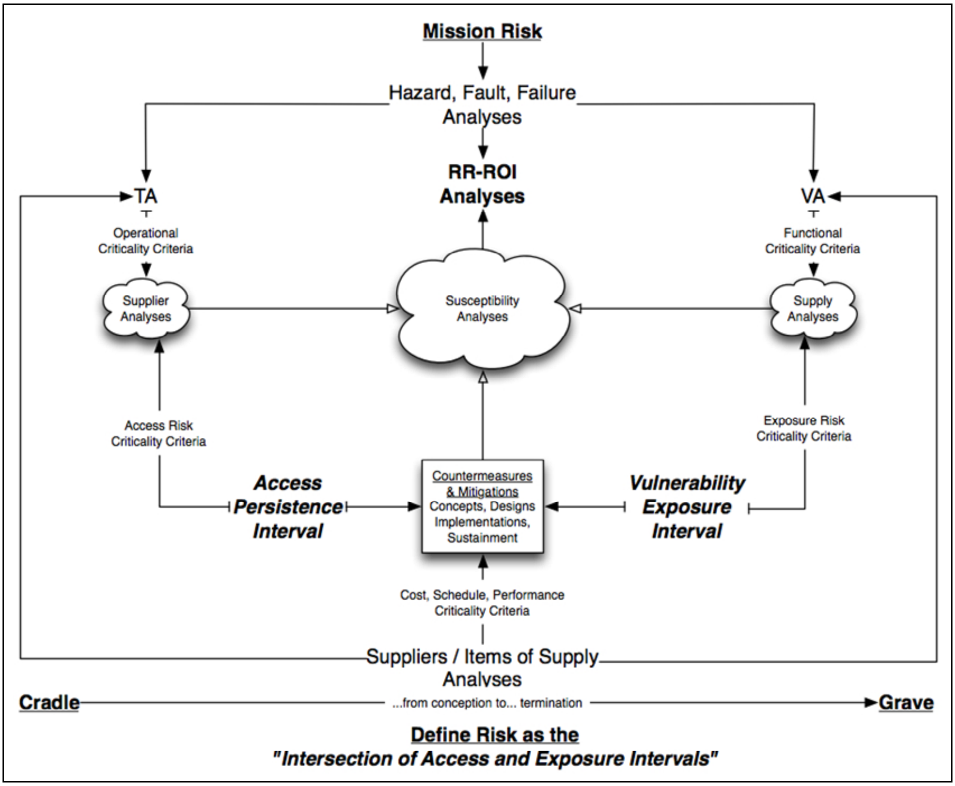

Mission Risk
Hazard, Fault, Failure Analyses
RR-ROI Analyses
TA
VA
Operational Criticality Criteria
Functional Criticality Criteria
Supplier Analyses
Susceptibility Analyses
Supply Analyses
Access Risk Criticality Criteria
Exposure Risk Criticality Criteria
Access Persistence Interval
Countermeasures & Mitigations Concepts, Designs Implementations, Sustainment
Vulnerability Exposure Interval
Cost, Schedule, Performance Criticality Criteria
Suppliers / Items of Supply Analyses
Cradle
...from conception to... termination
Grave
Define Risk as the
"Intersection of Access and Exposure Intervals"

Fig. 1. Susceptibility Analyses as Center of Gravity

calculus that brings threats and vulnerabilities into coincidence to identify a risk of concern. Combining the identified risks of concern with an evaluation of how to reduce the adverse consequences of risk realizations enables the generation of a ranked list of susceptibilities. These susceptibilities provide input to risk managers to determine some combination of investments to reduce the impact of risks to mission tolerable levels of outcome that can be selected and implemented. This general approach is applicable to any specific class of all risks and contributes to the overall trade-space of risks to successful execution of mission performance.

### 4.1 Risk management

Risk management is the process of responding to an event that offers negative or positive consequences. The goal is to maximize the gain from positive risks (opportunities) and minimize the loss from negative risks (Kliem, 2004). Risk management includes risk identification, risk analysis, and risk mitigation. Generally, the necessary steps to effective risk management are to (1) identify any potential risks, (2) assess the levels of threats, and (3) develop countermeasures and mitigations to reduce risks. Countermeasures are defined as actions or devices designed to negate or offset another whereas mitigations are defined as actions that can be taken to help reduce the impact of realized risks.

Qualitative analysis of the risk level helps to better prioritize risk (Khanmohammadi & Houmb, 2010) for obligating resources to reduce risk. After gaining a shared understanding of the impact and relative importance of risks, appropriate risk controls (Kleim, 2004) may be selected to be implemented:

- *Preventative controls* – techniques that mitigate the impact of a risk or stop it before having an impact (countermeasure).
- *Corrective controls* – techniques that involve determining the impact of a risk and require establishing measures to preclude future impacts (mitigation).
- *Detective controls* – techniques that reveal the existence of a risk and preclude future impact under similar conditions (vulnerability detection).

A traditional view in risk management is to "avoid risk." In reality, many people ignore risk because they do not understand it. However, risks cannot be ignored nor can all risks be eliminated. Attempting to eliminate all risks by applying countermeasures and mitigations is economically untenable. Residual risks will always remain. Due to these factors, active risk management is required to establish tolerable levels of trust and risk. Furthermore, many existing risk management models are not applicable to supply chain networks. There is a lack of a satisfactory framework to analyze information risks unique to a supply chain network and to provide a structure to organize the deliberations and tools for managing supplier and supply chain risks for ICT components and services bound for trusted information systems and networks.

## 5. Research areas

Globalization, more than many other factors, brings the uncertainties of an information operating environment into consideration of risks and the ways and means of countering and mitigating them. Furthermore, when the operating environment is a third party system of vulnerabilities, threats, intentions, capabilities, and risks for which every other system is dependent, the dynamics of the whole are considerably more difficult to assess, analyze, and apply actions to control or influence. The current research and development (R&D) agenda for single systems or systems-of-systems is more deliberate and explicit about particular characteristics of a system or collection of systems. While this work is improving the understanding of a "system," it is deficient in the totality of characteristics, knowledge, and techniques need to actually manage risks.

The authors suggest a paradigm of active risk management that requires a continuous feedback loop. Figure 2 illustrates a risk event timeline in three phases: pre-risk event, transitional risk event, and post-risk event. In each phase, there exist indicators and warnings to a potential risk occurrence. Vulnerability detection must occur at all times to seek out opportunities to prevent risks or mitigate risks to reduce impacts further downstream. Figure 3 illustrates the continuous feedback loop for active risk management. The activities for each of the three phases of active risk management seek to answer questions such as the ones listed below:

1. *Pre-Risk Event*: Identify and select risks to invest in doing something
   a. What can go wrong?
   b. What is the likelihood?
   c. What are the consequences?
   d. And at what time domain?
   e. What can be done and what options are available?
   f. What are the associated trade-offs in terms of all relevant costs, benefits, and risks?
2. *Transitional Risk Event*: Deploy countermeasures and mitigations
   a. What can be done and what options are available?
   b. What are the associated trade-offs in terms of all relevant costs, benefits, and risks?
   c. What are the impacts of current management decisions on future options?
3. *Post-Risk Event*: Evaluate the implemented countermeasures and mitigations, and readjust strategy if necessary
   a. What are the associated trade-offs in terms of all relevant costs, benefits, and risks?
   b. What are the impacts of current management decisions on future options?
   c. What can be done and what options are available?

Note that active risk management is a continuous cycle, with some of the same questions being asked and answered throughout. Further note that a risk must be experienced at least once, otherwise it is just theory and not practice.

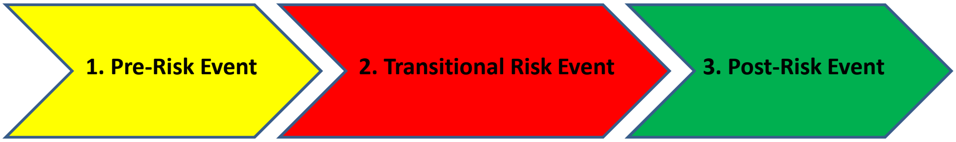


Fig. 2. Risk Event Timeline

The effects of globalization set up a rich set of challenges, issues, and opportunities for research. Globalization further begs for a broad and interdisciplinary agenda of research to relate the pace, tempo, and interaction of the environment with information systems. Current research is not driven by a totality of systems view nor does it deal with active

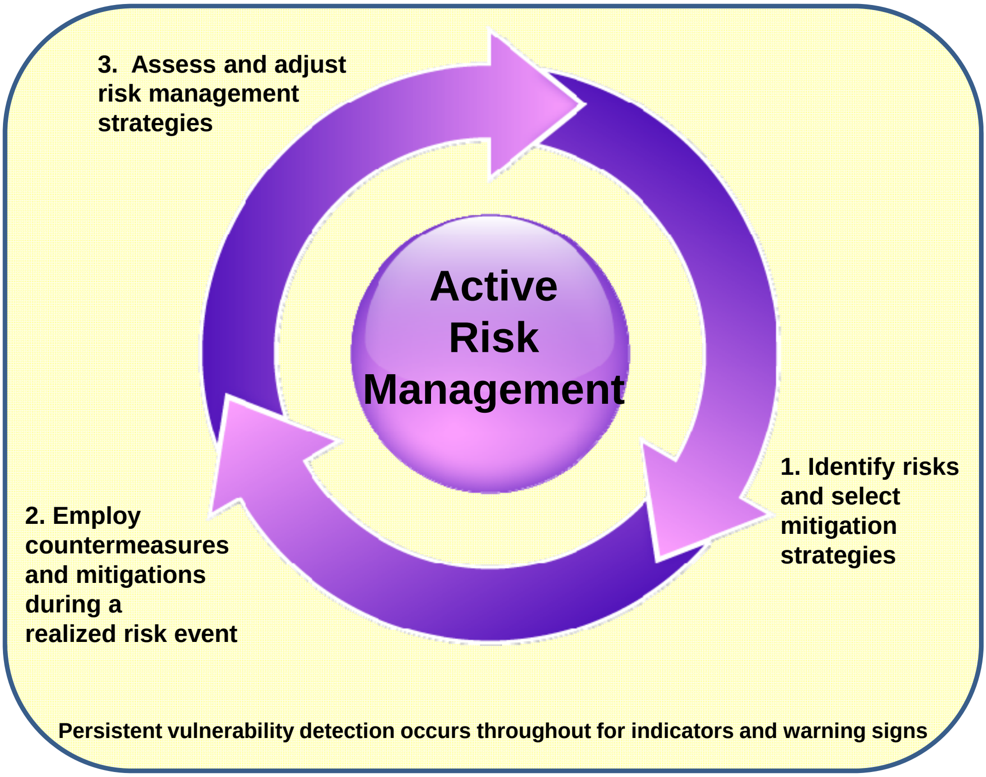

3. Assess and adjust risk management strategies
Active Risk Management
1. Identify risks and select mitigation strategies
2. Employ countermeasures and mitigations during a realized risk event
Persistent vulnerability detection occurs throughout for indicators and warning signs

Fig. 3. Active Risk Management Activities

management of risk. Modern and future dependence on information systems require a systematic development of research areas to deal with the problem space holistically. Otherwise, only partial knowledge and solutions are obtained because the focus is only on particular issues and solutions. A holistic perspective to identify risk and quality of implementation is therefore required.

### 5.1 Cyber attacks

Cyberspace is a domain constructed by man and constantly under construction (Welch, 2011). Modern information systems are connected to one another via networks. Functions essential to the computer control of the networks, information flowing or stored in the networks, and the decision support systems supported by the networks are subject to both physical damage and attacks that affect the logical realm. Table 2 maps six potential cyber attacks and their effects on information (Musman et al., 2011) and the impacted information assurance (IA) categories of confidentiality, integrity, and availability. Each 'X' represents an affirmative answer to the following question: Does the attack type, as defined, affect the IA category? Confidentiality refers to the prevention of unauthorized disclosure of data (both stored and communicated). Integrity refers to the prevention of unauthorized modification of data (both stored and communicated); detection and notification of unauthorized modification of data; and recording of all changes to data. Availability refers to the timely, reliable access to data and information services for authorized users. Availability attacks include destruction of assets and denial-of-service.

Table 2 highlights the importance of protecting integrity, yet this area is the least mature. Currently, there is a lot of research work that address confidentiality and availability. Most information systems assurance work deals with various malicious attacks that range from computer viruses, network penetration, and system breaches (Zuo & Hu, 2009). Availability can be preserved through asset diversity means (e.g., network path diversity). Confidentiality preservation mechanisms include authentication and authorization so that sensitive information is protected from unauthorized users. Encryption is a technique usually assumed to answer confidentiality and integrity issues. However, there is not much conducted research with regards to trusting the encryptor and protecting the integrity of information and data within an information system.

Figure 4 illustrates 36 potential situational-based mitigation categories that address integrity, confidentiality, and availability. Mitigations are time-dependent: pre-risk event, transitional risk event, and post-risk event. Potential risk management strategies can be classified as: prevention, remediation, mitigation, recovery, and reconstitution. Preventative strategies are usually derived from key and leading practices. Much research work is required to develop and organize the other types of risk management strategies. Figure 4 provides an illustrative framework to begin categorizing ways and means to manage risks represented by an action (applied pre-, trans-, or post-event) with an intended impact of the potential or actual loss of integrity, confidentiality, and availability.

| | | **Information Assurance Categories** | | |
|---|---|---|---|---|
| **Attack Category** | **Effect on Information** | **Confidentiality** | **Integrity** | **Availability** |
| **Degradation** | Rate of information delivery is decreased; Quality or precision of information produced by an activity is decreased | | X | X |
| **Interruption** | Information is unavailable for some time period | | | X |
| **Modification** | Information has been altered, meaning that the processes that use it may fail, or produce incorrect results | | X | |
| **Fabrication** | False information has been entered into the system | | X | |
| **Interception** | Information has been captured by the attacker | X | X | X |
| **Unauthorized Use** | Raises the potential for future effects on information | X | X | |

Table 2. Information Assurance Impacts Due to Various Cyber Attacks

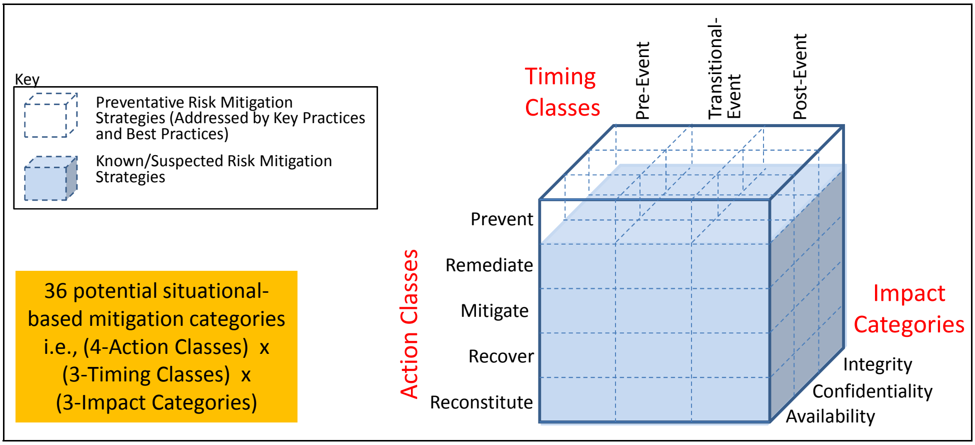

Key
Preventative Risk Mitigation Strategies (Addressed by Key Practices and Best Practices)
Known/Suspected Risk Mitigation Strategies
36 potential situational-based mitigation categories i.e., (4-Action Classes) x (3-Timing Classes) x (3-Impact Categories)
Timing Classes
Pre-Event
Transitional-Event
Post-Event
Action Classes
Prevent
Remediate
Mitigate
Recover
Reconstitute
Impact Categories
Integrity
Confidentiality
Availability

Fig. 4. Framework for Potential Risk Mitigation Strategies for Information Systems

### 5.2 Insider threat

Most information security losses are due to the theft of proprietary information, a feat usually executed by insiders. An "insider" is stereotypically an employee, contractor, business partner, or anybody who has any level of legitimate access, driven by a wide range of reasons, both rational (e.g., money, status, power) and irrational (e.g., revenge, frustration, emotion pain, other personal problems) (Chuvakin, 2003). Insiders can be categorized by their intent into non-malicious and malicious insiders.

Non-malicious insiders compromise security due to their mistakes. Non-malicious users include people who want to "explore" the network or "improve" how things work without regard to security regulations. Non-malicious users present a hazard to the enterprise because they can incorrectly destroy information, degrade the availability and integrity of computing resources, and create opportunities for outsider attackers. Non-malicious insiders may also be unwitting participants, under the control of a malicious insider who uses social engineering techniques such as direct requests, persuasion, and other forms of deception. Hackers are known to evaluate the target information system, get initial information about the protective measures, and then launch social engineering attacks to enlist insiders to do their bidding (Chuvakin, 2003).

Malicious insiders are generally motivated by greed, a need for acknowledgment, sabotage, revenge, or a desire to be irreplaceable by creating problems only they can fix. Malicious insiders act to eavesdrop on a private communication, steal or damage data, use information in violation of company policy, or deny access to other authorized users (Chuvakin, 2003).

One proposed paradigm shift is to think of the supply chain problem as an insider threat problem. Because globalization expands insider access and knowledge of critical information systems to new populations, the information systems being built today are exposed to greater insider threat risk. The insider threat problem requires research in the areas of threat identification and appropriate countermeasures and mitigation. Moreover, the insider may not only be human. The machine (e.g., any logic-bearing or programmable ICT component) is a potential insider threat in an information system and network.

Most modern ICT are programmable and expected to execute operations with a degree of predictable variability. Unfortunately, this property of programmability enables malicious intent to also be implemented. This may take the form of design or incorrect implementation of design with residual vulnerability able to be exploited at one extreme. On the other extreme, it may be the deliberate insertion of programming intended to be exploited or triggered with the intent of malicious effect. The very property of programmability that gives great flexibility and range of utility is also an intrinsic vulnerability to be exploited from the outside (the most common form of threat exploit) or from the inside (the deliberate inclusion of programming that allows the device to behave as an insider – normal behavior in all respects until triggered to behave with intent to create malicious effects).

### 5.3 Vulnerability detection

Detecting the human insider threat problem has been explored extensively. This concept depends on an understanding of correct behavior and the ability to observe the correctness of expected behavior. Humans are "programmed" through cultural norms, training, and education to behave correctly. However, human-machine systems can be exploited by human insiders to exhibit anomalous behaviors.

Modern information systems have ICT-enabled advantages such as programmability. The programmability property provides variability that leads to flexibility which simultaneously gives rise to vulnerabilities. As the machinery of information systems has become increasingly programmable, complex, interconnected, and pervasive, the machine is becoming the means of malicious insider exploitation. Table 3 presents malicious vulnerability examples that can be inflicted by human or machine to another human or machine. Traditional methods of detection (e.g., behavioral approaches) are used to detect man-made vulnerabilities (e.g., conspiracies and hacking). Machine vulnerabilities may be inherent in the hardware equipment and require human testing to detect. Globalization of the suppliers and the programmable nature of supplied items have vastly increased the opportunities for "insider behavior" implemented not by humans but by the machinery. The authors assert that an R&D agenda is required in counter-investment to understand, implement, and apply countermeasures and mitigations designed to meet what is functionally an insider threat realized in and executed by machine.

| | **Human** | **Machine** |
|---|---|---|
| **Human** | Conspiracy | Hacking |
| **Machine** | Vulnerability | Operations Research |

Table 3. Malicious Vulnerability Examples

As integrated circuit (IC) fabrication work is increasingly outsourced due to much lower costs, hardware manufacturers face significant security risks for ICs used in critical information systems and networks. Local, high-end, trusted facilities are economically unviable given the global economy. Further research work is required to address the uncertainty in provenance and hardware integrity. Example areas include digital IC fingerprinting and IC authentication tools and techniques.

### 5.4 Provenance & supply chain visibility

Organizations are addressing new threats and opportunities presented by the question: "where does this stuff come from?" Due to the magnitude of the global sourcing issue and the multi-layered nature of the global supply chain, there are more variance and unpredictable factors in the environment to control. Therefore, a high level of supply chain visibility can be incorporated into the risk management processes to reduce product and performance related errors, and enhance the quality and responsiveness to risk incident occurrence (Tse & Tan, 2011). Due to the longer supply chains, it is critical to enhance the supply chain risk visibility by examining sub-tier suppliers and adjusting the supplier assessment process with the insights gained in a cyclic manner.

Issues of provenance can be applied to both physical artifacts and to information. Provenance can be identified in in two distinct ways: the source (or derivation) of an object and the record of the derivation (Moreau et al., 2008). Much of the provenance work has been applied to artifacts, especially in archives, art, and archaeology. Provenance has recently become essential for digital documents in financial, commercial, medical, scientific, and legal contexts. Such information often originates in a remote location, gets processed by multiple parties, and resides in potentially untrustworthy storage (Hasan et al., 2009). In order to trust the information in a document, its provenance must be known because it is increasingly important to know where the information comes from and how it has been processed and handled.

More provenance research work is needed in the area of information and knowledge management, specifically electronic data. Electronic data does not usually contain historical information that would help end users, reviewers, or regulators. Process documentation is to electronic data as record of ownership is to a work of art (Moreau et al., 2008). A user's confidence in an application's electronic data can be increased by including the provenance that describes the process that led to the data's production. Digital data provenance tracking is useful for rights protection, regulatory compliance, management of intelligence and medical data, and authentication of information as it flows through information systems and networks. While significant research is being conducted in this area, the associated security and privacy issues have not been explored, leaving provenance information vulnerable to illicit alteration as it passes through untrusted environments (Hasan et al., 2009). Therefore, provenance of electronic data does not completely address or assure integrity.

### 5.5 Security & privacy

Security and privacy have different requirements but share a point of intersection; security can be achieved without privacy, and privacy cannot be preserved without security. Security is provided by a "system" that handles information. Privacy begins with an accountable action taken by a user of information machinery. If the "system" consists of a user (human) and the machinery, then the information system can be designed to holistically intersect to provide security and privacy by employing machine handling of information to achieve both security and privacy protections. However, privacy begins with the human who enters information into the machine and authorizes its use and transmission by the machine component of the "system" or "network." No amount of machine security can guarantee privacy as privacy begins with the original provider of information into the machinery of the "system."

Security is about protection, whereas privacy is about permission and use of personally identifiable information (PII). Information technology systems can be built to the highest security standards without any regard to privacy. However, once PII is collected, security measures are necessary to preserve privacy (Federal Enterprise Architecture Program Management Office, 2006). A security policy may address information classification, protection, and periodic review to ensure compliance. However, privacy policies are needed to determine how security is implemented for the purposes of protecting PII within information systems. Elements of a privacy policy include information regarding the processes of information collection, analysis, maintenance, access, dissemination, and deletion.

Information security is defined as protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction (44 U.S.C. § 3542(b)(1)). As security is crucial for ensuring privacy, an initial look at how IA measures can deal with privacy concerns is provided. Table 4 provides a mapping of the IA categories that can address the 17 privacy control families defined in *The Federal Enterprise Architecture Security and Privacy Profile* (June 2006). Each 'X' represents an affirmative answer to the following question: Can the standards and technologies of the IA category address and/or support the policies and procedures of the privacy control family? As Table 4 is a high-level view, each entry can be further examined in regards to specific standards, technologies, policies, and procedures.

| | **Information Assurance Categories** | | |
|---|---|---|---|
| **Privacy Control Family** | **Confidentiality** | **Integrity** | **Availability** |
| Policies and Procedures | X | X | |
| Privacy as Part of the Development Life Cycle | X | X | X |
| Assigned Roles, Responsibilities, and Accountability | X | X | |
| Monitoring and Measuring | | | |
| Education: Awareness and Role-based Training Programs | X | X | |
| Public Disclosure | | | |
| Notice | | | |
| Consent | | | |
| Minimum Necessary | | | |
| Acceptable Use | X | | |
| Accuracy of Data | | X | |
| Individual Rights | | | |
| Authorization | | | |
| Chain of Trust | X | | |
| Risk Management | | | |
| Reporting and Response | | | |
| Security Measures | X | X | X |

Table 4. Mapping of IA Categories That Can Address Privacy Control Families

The results of Table 4 draw out the following main findings:

1. The IA category of availability has a minimal relationship with privacy. Availability deals with system design to ensure high accessibility and redundancy of resources and capabilities.
2. While IA can address a majority of the privacy control families, IA does not address public disclosure, notice, consent, minimum necessary, individual rights, and authorization. Authorization in the privacy context refers to an individual's ability to authorize all new and secondary uses of PII not previously identified on the original collection notice. These privacy control families must therefore be addressed through another mechanism, such as business rules and processes.
3. The remaining privacy control families not addressed by the traditional IA categories are addressed by the expanded IA categories, which include assured information sharing, assured mission management, and system/network defense.

In summary, IA does not address every threat to protecting privacy and personal data. While IA relates to securing and protecting an information system, information privacy relates to an individual's right to determine how, when, and to what extent personal information will be released to another person or organization.

### 5.5.1 Insider threat

With respect to privacy, there are two types of insider threats to consider: (1) insiders who have the correct permissions and authorizations for data access may deliberately misuse information and/or provide information to unauthorized parties and (2) insiders who may inadvertently misuse information due to ignorance or carelessness that may result in improper disclosure (Waterman, 2006). The second kind of insider threat as described poses the greatest danger to the appropriate protection of privacy data. While IA can provide measures to ensure proper authorization and access control, IA cannot tackle the use of information for purposes other than the one or ones for which it was originally collected. One way to consider addressing the insider threat issue is through policies and procedures that provide education and training on the appropriate use of information and through enforcement of these policies and procedures.

### 5.5.2 Data mining tools

Some data mining tools make automatic associations in such a way that even naïve users could deduce private information from the unclassified or public pieces of data by basically exploiting the associations made available by these tools. Thus, there is a need for the development of privacy-preserving techniques for PII data management (Ferrari & Thuraisingham, 2006). Data anonymization, masking, and filtering are methods being used to protect the rights of individuals and minimial disclosure. However, even these techniques can be subverted.

### 5.5.3 Appropriate privacy policies

Privacy policies are being used by organizations to tackle the issue of PII. However, few privacy policies actually assert that your PII will remain secret or private and under your control (Poore, 1999). In reality, a privacy policy is simply an information policy that tells you what information is collected and how it is used. It does not necessarily mean that your privacy is protected and may actually specify that privacy is not provided. The inability for individuals to agree in terms of what they believe is an appropriate privacy policy or practice is a major challenge to achieving consistent protection for groups of individuals. Privacy is not absolute. There are many trade-offs in the benefits versus the risks.

### 5.5.4 Transparency

The ability to balance the privacy concerns of individuals with effective monitoring of potential insiders is a very challenging task. One significant problem is that individuals often are not aware of the information that is collected and how it is used or what has been done with it. In addition to the privacy conditions of notice, choice, use, and security, it is equally important to offer the privacy conditions of correction and enforcement. These six conditions associated with privacy are derived from the European Union's Privacy Directive and are briefly described in Table 5. Transparency protects not only the individual but the organization or company and promotes public trust.

| Condition | Description |
|---|---|
| Notice | The individual has the right to know that the collection of PII will exist. |
| Choice | The individual has the right to choose not to have the data collected. |
| Use | The individual has the right to know how data will be used and to restrict its use. |
| Security | The individual has the right to know the extent to which the data will be protected. |
| Correction | The individual has the right to challenge the accuracy of the data and to provide corrected information. |
| Enforcement | The individual has the right to seek legal relief through appropriate channels to protect privacy rights. |

Table 5. Privacy Control Conditions

### 5.5.5 Privacy implementation

Privacy breaches occur mainly due to the failure to develop the business processes around privacy. Privacy management should not be reactive. Protecting privacy and personal data should be an integral component in the development of information systems that involve PII. Privacy should be one of the most important priorities in the development of trusted information systems, not an afterthought.

Most importantly, privacy depends upon the human component of the system to handle and exchange information with the machine component in a manner that establishes accountabilities for it at the outset and at every opportunity during the use and exchange of tagged and marked information with privacy attributes. Most losses of privacy are the result of human mishandling of information or the failure of machinery to maintain security of the information as it is manipulated and handled according to privacy attributes that originate with humans. Frequently, human handling or the inability of machine components to preserve privacy attributes at exchange points are the primary reasons for privacy failures. Nonetheless, privacy failures are usually misattributed to security failures.

Privacy must be considered an integral part of the development and use of an information system. Privacy policies and procedures must be developed as part of the business process so that appropriate IA measures can be implemented to support them. Recall that security can be developed without privacy but privacy cannot be provided without security. However, security measures cannot address all privacy issues; hence privacy must be considered from the beginning. Privacy management should not be an afterthought, reactive, or piecemeal.

The following recommendations are provided to help move forward in providing both security and privacy for information systems:

1. Promote a more coordinated approach to security and privacy consistent with business objectives and the goals of efficiency and interoperability.
2. Conduct a Privacy Impact Assessment (PIA) to determine the effects of information services and sharing initiatives on individual privacy. Elements of PIAs should include the following:
    - The information that is being collected,
    - Why the information is being collected,
    - Intended use of the information,
    - With whom the information will be shared,
    - What opportunities individuals will have to provide information or to consent to particular uses of the information,
    - How information will be secured, and
    - Whether a system of records is being created under the privacy policy.
3. Develop a plan for evaluation and continued monitoring of the implementation of the privacy policy.

A significant gap in R&D is the interface between human and machine components of a system. Security is about the machine component. Privacy is about the human component and its interaction with the machine component.

In summary, an information system should have a privacy policy that publicly articulates that it will adhere to legal requirements and processes that enable gathering and sharing of information to occur in a manner that protects personal privacy interests. A well-developed and implemented privacy policy should also be transparent in order to protect the enterprise, the individual, and the public; and promotes trust. A well-developed privacy policy also ensures that appropriate IA measures can be taken to meet both security and privacy needs.

### 5.6 Supplier-Supply Chain Risk Management (S-SCRM)

Traditional research work in supply chain risk management involves activities and processes for planning, coordination, operation, control, and optimization of the supply chain. These efforts do not examine supply chain risks associated with the compromise or loss of product/service confidentiality or integrity. Supply chain exploits are the opportunities where adversaries can gain access, obtain knowledge, insert malicious code, or corrupt devices bound for information systems.

In recent years, supply chain risk management research work emphasizes the key role of managing the operational risks in multi-layered supply chains. The authors have presented an enterprise framework for characterizing supplier-supply chain risk management that captures the underlying complexity and scope of concerns to manage globalization of ICT risks (Chan & Larsen, 2010). The holistic view provides a way to manage risk by bringing intelligence mitigations, technical mitigations, and business mitigations into a tradespace to reach a collective view and to review or adjudicate decisions to obtain a tolerable risk-reduction – return on investment in making informed acquisition and procurement decisions of ICT components and services bound for trusted information systems and networks. Additionally, there have been a number of risk management decision models that may apply to managing supply chain quality risk, including supplier qualification screening, multi-sourcing, flexibility, and penalties levied for supplier non-performance (Tse & Tan, 2011).

Figure 5 illustrates a framework to assess strategy options for managing supply chain risks. Risks must meet a threshold of feasible realization to be considered for the application of countermeasures or mitigations. This figure proposes an approach to determine the feasibility of risk mitigation strategies by making a determination of susceptibility (previously defined as an intersection at a point or interval of time of a threat able to gain access with the exposure of a vulnerability). Note that this requires access and exposure to occur at the same point or during some interval of time in order to become susceptible to an adverse consequence. Thus, simply being accessible or being vulnerable is insufficient to gauge the degree of consequence. The coincidence of a threat and vulnerability may have quite different consequences depending on the time of coincidence and the particular mission operation underway. Consequences must be judged with respect to impact on mission and relative to the state of mission execution (a point or duration in time). An assessment can then be made on whether the system retains the risk and the associated impacts on mission performance or if the risk needs to be specifically countered or mitigated. Subsequently, an appropriate risk reduction strategy must be developed by both considering the existing counters and mitigations and the availability or potential for applying additional countermeasures and mitigations to diminish the impact. In some cases, the strategic outcome may require a change of operational concept, a redesign of the system or component in order to eliminate a susceptibility, alter the consequence impact, or develop more effective and affordable counters and mitigations, or in extreme cases abandon the system altogether.

In all cases, this collection of knowledge is applicable in an iterative manner to arrive at an approach that applies supplier and supply item countermeasures and mitigations at various points of risk realization – either in a general manner or very specific manner. This could be at the level of threat actor generally (avoid suppliers) or more specifically at the point of access to a vulnerability (know the supplier but assume the item of supply can become an insider threat). Similarly, one can deal with vulnerabilities. In the ideal case, exquisite knowledge and opportunity exist to deal with susceptibilities that give rise to intolerable risks. If not, then counters and mitigations can be developed and applied for an assumed or poorly understood susceptibility and the consequences implied to system performance if realized. In the end, no judgment can be satisfactory without some sense of the economic and mission impact of uncountered and unmitigated risks. At a minimum, the identification

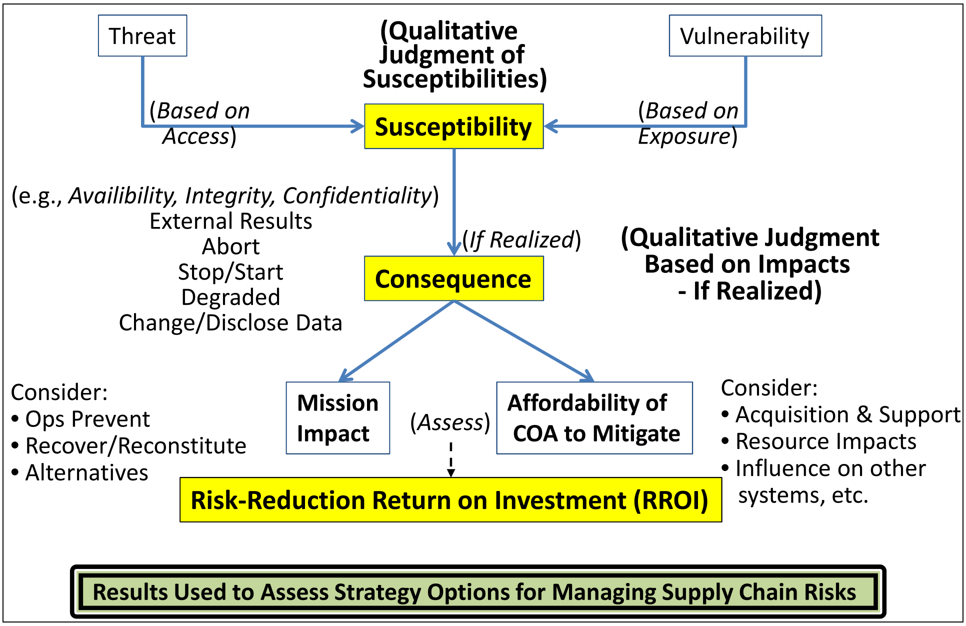

Threat
(Qualitative Judgment of Susceptibilities)
Vulnerability
(Based on Access)
Susceptibility
(Based on Exposure)
(e.g., Availibility, Integrity, Confidentiality)
External Results
Abort
Stop/Start
Degraded
Change/Disclose Data
(If Realized)
Consequence
(Qualitative Judgment Based on Impacts - If Realized)
Consider:
• Ops Prevent
• Recover/Reconstitute
• Alternatives
Mission Impact
(Assess)
Affordability of COA to Mitigate
Consider:
• Acquisition & Support
• Resource Impacts
• Influence on other systems, etc.
Risk-Reduction Return on Investment (RROI)
Results Used to Assess Strategy Options for Managing Supply Chain Risks

Fig. 5. Framework for Assessing Supply Chain Risk Management Strategies

of susceptibilities and their potential consequences gives visibility to risk managers and the opportunity to understand the worth of engineering for securing of systems and the residual risks for which security is currently not possible or not affordable. These can then be used to rationalize and justify R&D to discover, invent, and innovate security measures, which if applied achieve dependable and economic implementations able to withstand this emerging insider threat of compromised machinery and its potential to behave as intended most of the time and to change behavior and become malicious at inopportune times.

The prospect that variable human-motivated and incentivized behaviors are the only insider behaviors to withstand by applying security measures is no longer wise given both the pervasive dependencies on ICT and the complexities and difficulties of having, maintaining, and sustaining provenance and visibility of supplied items and the production and delivery processes of their suppliers. The most basic need is to advance an agenda of R&D that recognizes today's security measures and countermeasures are not scalable to the new world order. This does not make the current research efforts unnecessary, but they are clearly not sufficient. The authors use the "label" of active risk management to capture the need for scalable means of identifying, implementing, and managing the application of countermeasures and mitigations for this growing set of risks that can never be fully prevented, eliminated, or avoided; but rather must be actively managed. Furthermore, the authors contend that the engineering for active risk management will stress the ability of traditional engineering disciplines, because it is likely to require risk management to be "built-in," not applied and forgotten. In the same manner that much rhetoric surrounds the "building-in" of security versus "strapping it on" like appliqué armor, active risk management will demand advances in design techniques, new and innovative ways and means of implementing countermeasures and mitigations, and vastly improved instrumentation to know the state of risk, the state of counters for such risks, and the invocation of countermeasures and mitigations on-demand.

## 5.7 Systems security engineering

Systems security engineering is defined in MIL-HDBK-1785 as "an element of system engineering that applies scientific and engineering principles to identify security vulnerabilities and minimize or contain risks associated with these vulnerabilities. It uses mathematical, physical, and related scientific disciplines, and the principles and methods of engineering design and analysis to specify, predict, and evaluate the vulnerability of the system to security threats."

Building more secure systems (i.e., security assurance in information technology systems) calls for the following requirements:

- Well-defined system-level security requirements,
- Well-designed component products,
- Sound systems security engineering practices,
- Competent systems security engineers,
- Appropriate metrics for product/system testing, evaluation, and assessment, and
- Comprehensive system security planning and lifecycle management.

Improving the security of information systems often relies on red teaming activities (Clem et al., 2007) because they enhance knowledge in the following ways:

- Understanding adversaries and operational environments (assessing threats);
- Anticipating program risk, identifying security assumptions, and supporting security decisions;
- Exploring and developing security options, policy, process, procedures, and impacts;
- Identifying and describing consequential program security;
- Identifying and describing consequential security design alternatives;
- Measuring security progress and establishing security baselines;
- Exploring security of future concepts of operation; and
- Identifying and describing surprise, unintended consequences.

Example red team activities (Clem et al., 2007) and their objectives include the following:

- *Design assurance red teaming* – Helps ensure that a system will achieve its mission in degraded modes of operation.
- *Operational red teaming* – Helps to train staff, conduct testing and evaluation, validate concepts of operations, and identify vulnerabilities.
- *Penetration testing* – Helps to determine what access or control an insider, an outsider, or an outsider working with an insider may obtain.

The systems security engineering community needs to move toward designing for dependability, in addition to ensuring the other system "ilities" (e.g., affordability, availability, extensibility, flexibility, and scalability). Users of systems want the ability to trust that the systems (and services) will be consistently good in quality and performance.

Trustworthy and reliable is the definition of dependable. Thus, securely engineering systems need to move in the direction of being dependable.

The authors posit that dependability is one of the missing specifications that enable the systems security engineering community to be more effective through the lifecycle of a system's development and to maintain effectiveness during operations and maintenance of a system. Without a specification of necessary and sufficient trustworthiness in the context of a system's use, it is extremely difficult to provide the arguments and demonstrations of the worth of security measures. They are often then subordinated in the engineering trade-space to other systems engineering factors (the other "ilities").

Dependability may serve as the property that combines security engineering with other engineering disciplines that lead to security being "built-in" versus "strapped-on." Security cannot be an afterthought. It must be built into cost, schedule, and performance. Not all risks are equal and not all users/consumers have equal tolerance for risks. Dependability establishes the value of security measures in a way that they legitimately become part of the engineering trade-space. It places a value on security measures that enable a value of worth during operations and use and serves as a metric for how well risks have been managed. Dependability may be the basis for integrating traditional "engineering of" systems (the partitioning with a defined system boundary and application of engineering disciplines) with the larger context of "engineering for" systems (the inclusion of the system environment that leads to the specification and definition of a system boundary).

### 5.8 Engineering systems

"Engineering of" systems requires a holistic perspective that treats the operating environment of the engineering of a system concept, design, development, implementation, and support as more than an assumed and invariant actor that must merely be characterized and exploited by the system to be engineered. The operating environment "as a system" can be conceived, designed, developed, implemented, and supported to attain an advantage or benefit or present a risk. This is what distinguishes the "engineering of" versus the "engineering for" a system.

Systems engineering is defined as an interdisciplinary approach to enable the realization of successful systems. It focuses on defining user needs and required functionality early in the development cycle, documenting requirements, and then continuing with the design synthesis and system validation while bearing in mind the whole problem: operations, cost and schedule, performance, training and support, test, manufacturing, and disposal. Systems engineering includes both the business and technical needs of all users with the objective of creating a quality product that meets user needs. Systems engineering can fit within the overall engineering systems field. For example, systems engineering views the enterprise as a consideration or major influence on the system whereas engineering systems includes the enterprise as an essential part of the system.

Engineering systems is an emerging interdisciplinary academic and research field focused on addressing large-scale, complex engineering challenges with their social-political context. It takes an integrative holistic view of large-scale, complex, technologically-enabled systems with significant enterprise level interactions and socio-technical interfaces (Rhoades & Hastings, 2004). It may include components from several engineering disciplines, as well as economics, public policy, and other sciences. It is suggested that the four underlying disciplines for engineering systems are:

- Systems architecture / systems engineering and product development;
- Operations research and systems analysis;
- Engineering management; and
- Technology and policy

Engineering trusted information systems requires active risk management. The definitions presented in Section 4 often lead to assessment and analysis results that fall short of what is required to fully manage risks; in particular, risks that originate in the operating environment (OE) of the system under analysis. Recent research has addressed a portion of this shortfall under the rubric of "systems-of-systems" considerations. However, much of this research simply alters the system boundary and applies systems knowledge and technique to a collection of interacting systems. Generally, this approach is distinguished by a deliberate and explicit focus on the interconnects among systems and how to influence or modulate each individual system to gain a better understanding of how the collective whole behaves as a "single system." Although valuable and important to the understanding of complex systems, emergent properties and behaviors of an interacting whole, and the relevance and significance of the linkages among systems as a system of its own, this approach does little to advance the study of managing risks.

Uncertainty is a fundamental source of risk. Managing uncertainty is the difficulty that hinders the successful management of risks. Uncertainty arises in the environment and propagates into a system or system-of-systems and back into the environment. These interconnections are shown in Figure 6. The carrier of uncertainty is information in information systems. This suggests that the information environment acts as an autonomous system, and is a third party actor for consideration in the management of risk. This third party originates uncertainties and has casual impact on both the structural interconnection among systems and the flow variables (information elements) that interact among the interconnected systems, and within the individual systems. The operating environment becomes a critical system when risk management is the objective. It may not be assumed away, avoided or ignored. This model of interacting systems of individual vulnerabilities, threats, intents, capabilities, and risks must be adequately characterized, modelled, analyzed, and evaluated as a whole. A risk event may originate anywhere, be propagated anywhere, be realized and have impact distant from its provenance, and be countered or mitigated anywhere. This is the dynamic that defines supply chain networks and the information risks they present.

Modern dependency on ICT and the information operating environments creates challenges for today's system engineer in building trusted information systems. System complexity, coupled with the global sourcing of components and services, presents uncertainty in both the supplied items and the ways and means of producing the supplied items. The opportunities for "insider behavior" implemented not by humans but by the machinery of the human-machine system should alter the focus of R&D, the types and nature of countermeasures and mitigations implemented, and most certainly the tools and techniques of design, engineering, and test and evaluation, and operational monitoring.

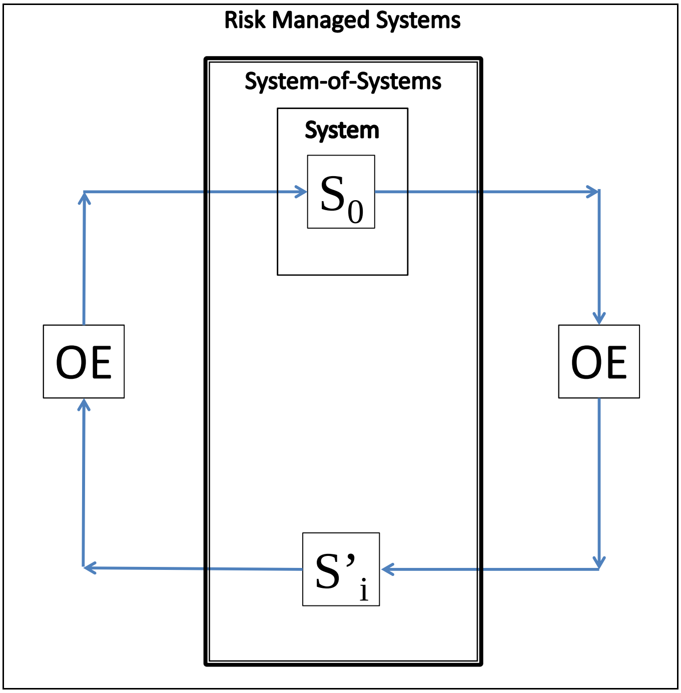


Fig. 6. Risk Managed Systems Framework

## 6. Conclusion

A high level of confidence is needed in trusted information systems. There is a current void in active risk management research. Active risk management for building trusted information systems requires the following activities:

- Understanding mission tolerance to failures of integrity, confidentiality, and availability of information throughout the life-cycle of a product and the processes producing and maintaining it;
- Understanding system criticality and priority tolerance to risks to focus resources on appropriate and adequate countermeasures and mitigations;
- Understanding dependence on critical subcomponents and designing and instrumenting for robustness of risk management during mission operations and sustainment;
- Understanding supply chain for critical components and procuring within mission risk tolerance;
- Partnering with industry to drive security (manufacturing, engineering, test and evaluation, etc.) into the processes and sub-suppliers at every production and support tier; and
- Understanding that privacy is a trust threshold only enabled in part by mechanisms of security, the association of that trust threshold with a security capability, and the ability

to maintain that security throughout the interactions among machinery components of the human-machine information system.

A research agenda in active risk management should include the following areas:

- Evolving the insider threat paradigm where the machine is the insider (both products and processes);
- Assuring information integrity and confidentiality throughout the life-cycle from idea to retirement from inventory;
- Improving supplier and supply item provenance and supply chain visibility;
- Enhancing testing and evaluation techniques for vulnerability detection of supply chain exploitation opportunities in the products and in the processes that produce and support them;
- Furthering systems security engineering to provide an understanding of the tools and techniques that discover and lead to effective and affordable countermeasures and mitigations; and
- Developing the knowledge, inventing the technologies, and producing the innovations that recognize the differences between privacy and security while enabling the individual and organization to manage the risks of highly variable thresholds of trust and integrity of information employed and applied in human-machine information systems.

These areas of research are motivated by the effects of globalization and the tempo and pace of ICT advancement and application to complex information systems. The pervasive dependence and increasing strength of dependence requires correct systems behavior. An engineering systems approach to the problem and issues provides the holistic view to bring all the pieces together to understand the interactions and any emergent behavior of the information system as a whole.

## 7. References

Bolgar, C. (2010). Boosting Protection : Strategies for Reducing Risk and Staying Ahead of Your Competitor, In : *The Wall Street Journal Supply Chain Risk Insights*, 17.10.2011, Available from http://www.supplychainriskinsights.com/archive/scri-protection

Chan, S. & Larsen, G.N. (2010). A Framework for Supplier-Supply Chain Risk Management : Tradespace Factors to Achieve Risk Reduction – Return on Investment, *Proceedings of 2010 IEEE International Conference on Technologies for Homeland Security*, ISBN 978-1-4244-6047-2, Waltham, MA, November 2010.

Chuvakin, A. (2003). Methods to Thwart Insider Attacks : Products, Techniques, and Policies, *Data Security Management*, Vol. 26, No. 1, (Feb/Mar 2003), pp. 1-11, ISSN 10967907.

Clem, J. F. ; Robbins, K. D. ; Parks, R. C. ; Mateski, M. E. & Page, K. J. (2007). *Red Teaming Quick Reference Sheet*. (25 April 2007).

Federal Enterprise Architecture Program Management Office (2006). The Federal Enterprise Architecture Security and Privacy Profile, Version 2.0, 1 June 2006.

Ferrari, E. & Thuraisingham, B. (2006). Guest Editorial: Special Issue on Privacy Preserving Data Management. *THE VLDB Journal*, Vol. 15, No, 4, (2006), pp. 291-292.

Haimes, Y.Y. (2006). On the Définition of Vulnerabilities in Measuring Risks to Infrastructures. *Risk Analysis*, Vol. 26, No. 2, (April 2006), pp. 293-296, ISSN 02724332.

Hasan, R. ; Sion, R. & Winslett, M. (2009). Preventing History Forgery with Secure Provenance. *ACM Transactions on Storage*, Vol. 5, No. 4, (December 2009), pp. 12-55, ISSN 1553-3077.

Khanmohammadi, K. & Houmb, S.H. (2010). Business Process-based Information Security Risk Assessment, *Proceedings of 2010 Fourth International Conference on Network and System Security*, pp. 199-206, ISBN 978-0-7695-4159-4, Melbourne, VIC, Australia.

Kliem, R. (2004). Managing the Risks of Offshoring IT Development Projects, *Information Systems Management*, Vol. 21, No. 3, (Summer 2004), pp. 22-27, ISSN 10580530.

Laeequddin, M. ; Sahay, B.S. & Sahay, V. (2008). Capturing the Concept of Trust Right in Supply Chain Partner's Relationship – A Conceptual Framework, In : *Journal of Knowledge Management Practice*, Vol. 11, Special Issue 1, (January 2010), Available from http://www.tlainc.com/articlsi11.htm.

Moreau, L. ; Groth, P. ; Miles, S. ; Vazquez-Salceda, J. ; Ibbotson, J. ; Jiang, S. ; Munroe, S. ; Rana, O. ; Schreiber, A. ; Tan, V. & Varga, L. (2008). The Provenance of Electronic Data. *Communications of the ACM*, Vol. 51, No. 4, (April 2008), pp. 52-58, ISSN 00010782.

Musman, S. ; Tanner, M. ; Temin, A. ; Elsaesser, E. & Loren, L. (2011). Computing the Impact of Cyber Attacks on Complex Missions, *Proceedings of 2011 IEEE International Systems Conference*, ISBN 978-1-4244-9493-4, Montreal, QC, Canada, April 2011.

National Defense Industrial Association (NDIA) System Assurance Committee. (2008). *Engineering for System Assurance*, (Version 1.0), October 2008, Available from http://www.acq.osd.mil/se/docs/SA-Guidebook-v1-Oct2008.pdf

Poore, R. S. (1999). Anonymity, Privacy and Trust. *Information Systems Security*, Vol. 8, No. 3, (Fall 1999), pp. 16-20, ISSN 1065898X.

Rhoades, D. & Hastings, D. (2004). The Case for Evolving Systems Engineering as a Field within Engineering Systems. *MIT Engineering Systems Symposium*, Cambridge, MA March 2004. Available from
http://esd.mit.edu/symposium/pdfs/papers/rhodes.pdf

Tse, Y.K. & Tan, K.H. (2011). Managing Product Quality Risk in a Multi-Tier Global Supply Chain, *International Journal of Production Research*, Vol. 49, No. 1, (January 2011), pp. 139-158, ISSN 0020-7543.

Waterman, S. (2006). Analysis: DNI debates privacy rule changes, *UPI Security & Terrorism*, 21 August 2006, Available from http://www.upi.com/Business_News/Security-Industry/2006/08/21/Analysis-DNI-debates-privacy-rule-changes/UPI-47741156196154/

Welch, L. D. (2011). Cyberspace – The Fifth Operational Domain. *IDA Research Notes*. (Summer 2011), pp. 2-7.

Zuo, Y. & Hu, W. (2009). Trust-Based Information Risk Management in a Supply Chain Network, *International Journal of Information Systems and Supply Chain Management*, Vol. 2, No. 3, (July – September 2009), pp. 19-34, ISSN 1935-5726.

# 6

# Construction of Effective Database System for Information Risk Mitigation

Kiyoshi Nagata
*Faculty of Business Administration,
Daito Bunka University,
Takashimadaira Itabashi-ku, Tokyo,
Japan*

## 1. Introduction

In the Information Technology Communication Society, the information system in any organization is always exposed to various kinds of risks, and they should prepare countermeasures against possible risks to protect their assets and secure their activities' continuity. For that purpose, several types of information risk evaluation and management systems, such as ISO/IEC 27002, MEHARIT, MAGERIT, SP800-30, OCTAVESM, etc., are proposed by institutions all over the world. Although each system has its own policy and characteristic, on the final stage after the risk evaluation was done and some serious risks were clarified, the system usually goes on the process of choosing effective and available mitigation controls against each of risks.

In our prior works, we proposed a method to choose a set of effective elements from a given database of properly valued mitigation controls and we also proposed a method of clustering these controls related to the threat path of OCTAVE's risk profile worksheet.

However we have not yet constructed any feasible database system for practical use, now the effort is in progress. For that sake, it is necessary to investigate several existent systems of mitigation controls, and to compare and analyse them.

The content of the chapter is as follows:

1. Overview and investigation of existent information risk management systems and their mitigation controls
2. Brief explanation of useful tools for the proposed total system of risk management, such as fuzzy outranking, fuzzy inference mechanism, modified structural modelling method, and c-mean clustering.
3. Review of our proposed method for choosing effective set of mitigation controls from a well-defined database of controls
4. Details of the process constructing database systems
5. Discussion and conclusion

## 2. Overview and investigation of existent information risk management systems and their mitigation controls

Throughout this chapter, we define a risk mitigation control to be a measure which could reduce the current or potential risk degree. However the risk degree is evaluated in various aspects and from different point of views, and each mitigation control has its own property, characteristic, and merit, the total process of risk mitigation can be summarized in several similar steps. In this section, we will see some risk evaluation and management methodologies.

### 2.1 Hand book of information security

According to D. Kaye, risk mitigation is a process aimed at limiting the likelihood of risks and the potential losses those risks can cause (Kaye, 2002, p.100).

The following step summarization is from the Hand Book of Information Security (Bidgoli, 2006, p.750).

- Avoid the causes

Risks are caused by many types of instances. If the risk is technological, we can avoid the risk by updating or replacing the related system by more robust and reliable one.

- Reduce the frequency

Risk is usually assessed by the frequency it occurs and the impact it may cause. By adopting a control which mainly reduces the occurrence frequency of the risk, the risk can be mitigated.

- Minimize the impact

Since the frequency of the risk can not be reduced to zero, we should consider the impact of the risk to the organization's activities as the other important factor of risk. The impact related to a risk has various aspects depending on the organization under mind, and try to minimize the impact not only from each aspects but also from the total point of view.

- Reduce the duration

The duration of the exposure to a risk may cause more serious risks. The recovery time of data or system, for instance, is important matter.

The risks are usually evaluated as the pair of two factors such as the frequency and the impact, then the second and the third steps are usual steps for risk evaluation. The cause avoidance and the duration reduction are sometimes treated as concrete measures of mitigation controls.

In the book, the risk transfer, such as insurance or outsourcing, is dealt as the different step from the risk mitigation.

### 2.2 OCTAVE-S

SEI (Software Engineering Institute) of Carnegie Melon University developed OCTAVE$^{SM}$ (Operationally Critical Threat, Asset, and Vulnerability Evaluation System) (Alberts &

Dorofee, 2003) as a security evaluation system based on organizational assets. OCTAVE-S is a variation of the approach tailored to relatively small organizations (less than 100 people) which have the limited means and unique constraints.

In the implementation guide (Alberts et al., 2005), the key differences between OCTAVE and other traditional information risk evaluation and management approaches are described as in the table 1.

Ordinary risk assessment has three important aspects such as operational risk, security risk, and technology risk. OCTAVE developers say that other evaluation systems are tend to evaluate the organizational systems and to focus on the technology. In OCTAVE, the technology is examined as the part of security practice, and other two aspects mainly drive OCTAVE approach.

| **OCTAVE** | **Other Evaluation systems** |
|---|---|
| Organization evaluation | System evaluation |
| Focus on security practices | Focus on technology |
| Strategic issues | Tactical issues |
| Self direction | Expert led |

Table 1. The key Differences

OCTAVE aims to evaluate the organization itself in aspect of information assets, threats and vulnerabilities, and focus on their practices to obtain the information security, which eventually lead the organization to strategic protection issues rather than tactical ones. The expert led system is managed by a team of experts in risk analysis, or in information technologies from outside or inside. OCTAVE is self-directed system lead by a small interdisciplinary team, called the analysis team, consist of members in the organization.

OCTAVE(-S) has three phases in each of which the analysis team outputs the corresponding matters as follows.

Phase1. Build Asset-Based Threat Profiles

Outputs: Critical assets, security requirements for critical assets, threats to critical assets, and current security practices

Phase2. Identify Infrastructure Vulnerabilities

Outputs: Key components and current technology vulnerabilities

Phase3. Develop Security Strategy and Plans

Outputs: Risks to critical asset, risk measures, protection strategy, and risk mitigation plans

Each phase has some process consist of several steps, which we show in the table2 from the guide (Alberts et al., 2005).

In the series of our research project, we first proposed a method to identify the set of critical assets from huge number of possible information related assets in correspondence of the step S2.1 in the table (Nagata et al., 2007). In the method we used FSM (Fuzzy Structural Modelling) based the modified structural modelling method described in the following section. Next we proposed a risk evaluation system for a chosen critical asset with fuzzy inference mechanism corresponding to the process S4 (Nagata, et al., 2008B).

One of important roles of any risk management system is to develop a mitigation plan in which effective and proper mitigation controls are set up. For this purpose, a method to select effective risk mitigation controls is proposed using fuzzy outranking in correspondence of the process S5 (Nagata, et al., 2009). This method works under the assumption that there is a database of mitigation controls with some kind of vector whose entries are numerical values assigned to the attributes in OCTAVE's threat path. We also proposed a method for constructing that kind of database (Nagata, 2011).

| Phase | Process | Group of Steps |
|---|---|---|
| Phase1 | S1:<br>Identify Organizational Information | S1.1:Establish impact evaluation criteria<br>S1.2: Identify organizational assets<br>S1.3: Evaluate organizational security practices |
| | S2:<br>Create Threat Profiles | S2.1: Select Critical Assets<br>S2.2: Identify security requirements for critical assets<br>S2.3: Identify threats to critical assets |
| Phase2 | S3:<br>Examine the Computing infrastructure in Relation to Critical Assets | S3.1: Examine access path<br>S3.2: Analyze technology-related process |
| Phase3 | S4:<br>Identify and Analyse Risks | S4.1: Evaluate impact of threats<br>S4.2: Establish probability evaluation criteria<br>S4.3: Evaluate probabilities of threats |
| | S5:<br>Develop Protection Strategy and Mitigation Plans | S5.1: Describe current protection strategy<br>S5.2: Select mitigation approaches<br>S5.3: Develop risk mitigation plans<br>S5.4: Identify changes to protection strategy<br>S5.5: Identify next steps |

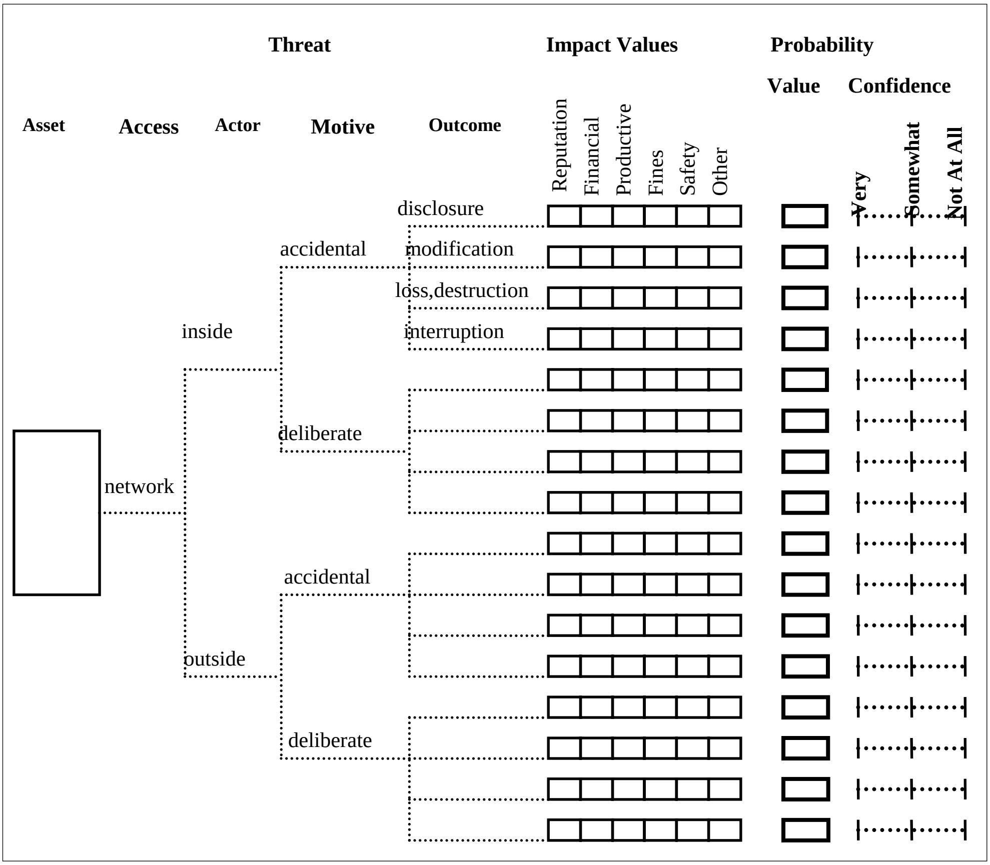
Table 2. Phase, Process, and Group of Steps in OCTAVE-S

When proceeding in risk evaluation steps, the risk profile worksheet plays a big role in order to recognize the information related threat, and to evaluate the impact and the frequency the threat may cause.

In the worksheet shown in Fig. 1, threats are classified into three types such as "Human actors", "System problems", and "Other problems" in the first place. For the human actors causing threats, the access path (network or physical), actors (inside or outside), motive (accidental or deliberate), and outcome (disclosure or modification or loss and destruction or interruption) are examined in this order. For the System problems causing threats, actors (software defects or system crashes or hardware defects or malicious code), and outcome are examined. For the "Other problems", various actors (e.g. problems related to power supply, telecommunication, third-party, natural disasters, physical configuration etc.) are examined. Each impact area of Reputation, Financial, Productivity, Fines/legal penalties, Safety and

Other (facilities) are considered for the non-negligible threats as the result of examination. According to the volume 3 of the OCTAVE-S Implementation Guide (Alberts, et al., 2005), the three impact measure (High, Medium, or Low) are adopted, and probability values are also measured as one of them (H, M, or L) by considering a frequencies such as daily, weekly, monthly, 4 times per year, 2 times per year, once per year, once very 2 years, and so on. Fig1 is an example of the risk profile worksheet for the Human Actors Using Network Access.

At first, put one of critical assets in the left-hand side box, and trace the dotted line considering the possibility of access, actor, motive, and outcome. Then, for each threat on the possible path, the impact values related to given subjects and the probability value are determined with confidence level.



(source: the Volume 5 of OCTAVE-S Implementation Guide, Version1)

Fig. 1. Risk profile worksheet for human actors with network access

We use the worksheet, but we adopt much more numerical evaluation method without loss of human related, consensus based, and organizational strategic concept. Our proposed total system for evaluation of threat is based on Modified Structural Modeling Method (MSMM), fuzzy integral, and fuzzy inference mechanism. In our system, the input values for impact values and for probability which should be marked in the box or on the scale bar as linguistic values in the OCTAVE are all numerical crisp values between 0 and 1, and the human related, consensus based, and organizational strategic concept are mounted and integrated with them in the process of fuzzification.

In the final process, selection of mitigation plans comes up, and listed up in the OCTAVE's catalogue of practices (Alberts & Dorofee, 2003, pp. 443—454).

The followings are classified groups of them.

- Strategic Practices (SP)
  - Security awareness and training (SP1)
  - Security strategy (SP2)
  - Security management (SP3)
  - Security policies and regulations (SP4)
  - Collaborative security management (SP5)
  - Contingency planning/disaster recovery (SP6)
- Operational Practices (OP)
  - Physical security (OP1): "Physical security plans and procedures (OP1.1)", "Physical access control (OP1.2)", "Monitoring and auditing physical security (OP1.3)"
  - Information technology security (OP2): "System and network management (OP2.1)", "System administration (OP2.2)", "Monitoring and auditing IT security (OP2.3)", "Authentication and authorization (OP2.4)", "Vulnerability management (OP2.5)", "Encryption (OP2.6)", "Security architecture and design (OP2.7)"
  - Staff security (OP3): "Incident management (OP3.1)", "General staff practice (OP3.2)"

In each subcategories listed above, there are several controls. For instance, SP1.1 of SP1 is "Staff members understand their security roles and responsibilities. This is documented and verified". OP2.1 contains 10 controls, e.g. OP2.1.3 is "Sensitive information is protected by secure storage such as…", OP2.1.4 is "The integrity of installed software is regularly refined", and OP2.1.6 is "There is a documented data backup plan that …".

### 2.3 ENISA

European Network and Information Security Agency, ENISA, provides risk management related documents in one of which risk mitigation is took up as a risk treatment. They define the risk treatment as a process of selecting and implementing measures to modify risk, and the process is composed of five steps such as, "Identification of Options", "Development of the Action Plan", "Approval of the Action Plan", "Implementation of the Action Plan" and "Identification of Residual Risks".

ENISA also provides a document named "Information Package for SMEs", where "SMEs" denotes "Small or Medium sized Enterprises". In the document, the risk management process is composed of four phases.

Phase1: Select Risk Profiles

The risk profiling is done using the risk evaluation matrix in which risk areas are specified as "Legal and Regulatory", "Productivity", "Financial Stability", and "Reputation and Loss of Customer Confidence". The possible risk levels are "High", "Medium", and "Low", and each level is clearly defined according to the risk area. For instance, if the organization's yearly revenue is of excess of 25 million Euros or/and financial transactions with third parties or customers are taking place as part of the business as usual process, then the risk area of financial stability is "High". If the yearly revenue exceeds 5 million Euros and not exceeds 25 million Euros, then the risk level is "Medium". Otherwise it is "Low". After identifying the risk levels for all the risk areas, the risk profile of the organization is defined as the highest level in the risk evaluation matrix overall the risk areas.

Phase2: Identify Critical Assets

In SME, the number of critical assets is fixed as five, and the analysis team choose them considering a large adverse impact on the organization caused by "disclosure" or "modification" or "loss and destruction" or "interruption" of the asset. These scenarios are same as the outcomes in OCTAVE's risk profile worksheet shown in Figure 1. The assets are categorised into "systems", "network", "people", and "applications", then the rationale and security requirement for selecting each critical asset are described. Here the security requirements are three ordinary information security aspects, i.e. Confidentiality, Integrity, and Availability.

Phase3: Select Control Cards

SME adopts OCTAVE's mitigation controls as their control cards. This phase proceeds in three steps such as "Step1: select organization control cards", "Step2: select asset base control cards", and "Step3: document list of selected controls and rationale". Here the organization control cards correspond to the mitigation controls of strategic practice (SP), and the asset base control cards correspond to those of operational practice (OP). The step1 is performed according to the risk profile in phase 1, and some control cards are selected beforehand. For instance, if the risk area "legal and regulatory" is low, then the control SP1.1 is adopted. The step2 is performed according to the critical asset category, and control card consist of security requirements and type of controls is prepared for each asset category and risk level. The table below is the list of control cards:

| Critical Assets | High Risk Cards | Medium Risk Cards | Low Risk Cards |
|---|---|---|---|
| Application | CC-1A | CC-2A | CC-3A |
| System | CC-1S | CC-2S | CC-3S |
| Network | CC-1N | CC-2N | CC-3N |
| People | CC-1P | CC-2P | CC-3P |

Table 3. Asset based control selection

For instance, CC-1A contains OP2.1.3, OP2.1.4, and OP2.1.6 for security requirement of confidentiality, integrity, and availability respectively as system and network management related controls.

Phase4: Implementation and Management

In this phase, the gap between the selected control cards and current security practice is analysed at first. Then create risk management plan, and the implementation is done.

The selection of mitigation controls is discussed both in the Phase3 and in the Phase4, and they classify controls into organizational controls shown in annex C, and asset based controls shown in annex D.

### 2.4 MEHARI

MEHARI, Method Harmonise d'Analyse de Risque, is developed by CLUSIF, Club de la Securite de L'Information Francais, aimed at providing a set of tools specifically designed for security management.

MEHARI uses a word of risk treatment measures or security services for mitigation controls, and classifies them into four categories, "Retention", "Reduction", "Transfer", and "Avoidance".

The standard scales of measures for likelihood reduction or for reduction of frequency factors are

- Efficiency of dissuasion measures
- Efficiency of prevention measures
- Efficiency of protective or confinement measures
- Efficiency of palliative measures

Each factor has four levels from level1, low or nul, to level4, very high (strong). The list of security services has more than 300 of sub-services classified into several service categories as follows.

1. Organization of security: "Roles and structures of security (01A)", "Security reference guide (01B)", "Human resource management (01C)", "Insurance (01D)", "Business continuity (01E)"
2. Sites security: "Physical access control to the site and the building (02A)", "Protection against miscellaneous environmental risks (02B)", "Control of access to office areas (02C)", "Protection of written information (02D)"
3. Security of Premises: "General maintenance (03A)", "Control of access to sensitive locations (except office zones) (03B)", "Security against water damage (03C)", "Fine security (03D)"
4. Extended Network (intersite): "Security of the extended network architectures and service continuity (04A)", "Control of connections on the extended network (04B)", "Security during data exchange and communication (04C)", "Control, detection and handling of incidents on the extended network (04D)"
5. Local Area Network (LAN): "Security of the architecture of the LAN (05A)", "Access control of the LAN (05B)", "Security of data exchange and communication on the LAN (05C)", "Control, detection and resolution of incidents on the LAN (05D)"
6. Network operations: "Security of operations procedures (06A)", "Parameters setting and control of hardware and software configurations (06B), "Control of administration rights (06C)", "Audit and network control procedures (06D)"

7. Security and architecture of systems: “Control of access to systems (07A)”, “Containment of environment (07B)”, “Management and saving of logs (07C)”, “Security of the architecture (07D)
8. IT Protection environment: “Security of operational procedures (08A)”, “Control of hardware and software configurations (08B)”, “Management of storage media for data and problems (08C)”, “Service continuity (08D)”, “Management and handling of incidents (08E)”, “Control of administrative right (08F)”, “Audits and control procedures relative to information systems (08G)”, “Management of IT related archives (08H)”
9. Application security: “Application access control (09A)”, “Control of data integrity (09B)”, “Control of data confidentiality (09C)”, “Data availability (09D)”, “Service continuity (09E)”, “Control of origin and receipt of data (09F)”, “Detection and management of application incident and anomalies (09G)”, “Security of the e-commerce sites (09H)”
10. Security of application projects and developments: “Security of application projects and developments (10A)”, “Ensuring security in the development and maintenance processes (10B)”
11. Protection of users' work equipment: “Security of the operational procedures for the whole set of users' equipment (11A)”, “Protection of workstations (11B)”, “Protection of data on the workstation (11C)”, “Service continuity of the work environment (11D)”, “Control of administrative rights (11E)”
12. Telecommunications operations: “Security of operational procedures (12A)”, “Control of hardware and software configurations (12B)”, “Service continuity (12C)”, “Use of end-user telecommunication equipment (12D)”, “Control of administrative rights (12E)”
13. Management processes: “Protection of personal information (PPI; 13A)”, “Communication of financial data (13B)”, “Respect of regulations concerning the verification of computerized accounting (VCA; 13C)”, “Protection of Intellectual property rights (IPR; 13D)”, “Protection of computerized systems (13E)”, “Human safety and protection of the environment (13F)”, “Rules related to the use of encryption (13G)”
14. Information security management: “Establish the management system (14A)”, “Implement the management system (14B)”, “Monitor the management system (14C)”, “Improve the management system (14D)”, “Documentation (14E)”

We can see that the same or similar expressions appeared in different categories such as “security of operational procedure” is in 06A, 08A, 11A, and 12A, and “service continuity” is in 08D, 09E, 11D, and 12C. This suggests the possibility of different perspective for the classification of controls.

MEHARI describes threat by similar items in OCTAVE's risk profile worksheet as shown in Fig. 1.

- Events: ”Accidents”, “Errors”, “Voluntary acts, whether malicious or not”, etc. For each of the events, following aspects are described,
  - Whether the cause is internal to the entity,
  - Whether the event is material or immaterial,

  - Any other factor that may influence the probability of the event occurring.
- Actors: rights and privileges,
- Circumstances in which the risk occurs,
  - Process or process steps: modification of files during maintenance operations,
  - Location: theft of media from one location or another, inside or outside the organization,
  - Time: actions occurring during or outside working hours.

A risk scenario is created with the different element, and risk treatment measures effective to the scenario are selected.

### 2.5 ISO/IEC

BS7799 part1 based ISO/IEC 27002 defines a security control to be a control which should ensure risks are reduced to an acceptable level. The selection of appropriate controls is dependent on organizational decisions based on the criteria for risk acceptance and the general risk management approach. Thus the acceptance level for the organization should be discussed and determined previously.

The categorization of controls in the document is shown below with corresponding number of controls in MEHARIT.

- Security Policy: "Information security policy (14)"
- Organization of Information Security(01): "Internal organization", "External organization"
- Asset Management: "Responsibility for assets (11E)", "Information classification"
- Human Resources Security (01C): "Prior to employment", "During employment", "Termination or changes of employment"
- Physical and Environmental Security (02): "Secure areas", "Equipment security (03C)"
- Communications and Operations Management: "Operational procedures and responsibilities (08A)", "Third party services delivery management", "System planning and acceptance", "Protection against malicious and mobiles code", "Bach-up", "Network security management", "Media handling, Exchange of Information", "Electronic commerce services (09H)", "Monitoring"
- Access Control (05B): "Business requirement for access control", "User access management", "User responsibilities", "Network system access control (04B)", "Operating system access control", "Application and information access control", "Mobile computing and tele-working"
- Information Systems Acquisition, Development and Maintenance: "Security requirement", "Correct processing in application", "Cryptographic controls (13G)", "Security of system files", "Security in development and support processes", "Technical vulnerability management"
- Information Security Incident Management: "Reporting information security events and weakness", "Management of information security incidents and improvement"
- Business Continuity Management (01E, 01D): "Information security aspects of business continuity management"

- Compliance: “Compliance with legal requirements (03D, 13A, 13D)”, “Compliance with security policies and standards, and technical compliance”, “Information systems audits considerations”

These controls are selected by considering the possible options including:

- applying appropriate controls to reduce the risk
- knowingly and objectively accepting risks, providing they clearly satisfy the organization's policy and criteria for risk acceptance
- avoiding risks by not allowing actions that would cause the risks to occur
- transferring the associated risks to other parties, e.g. insurers or suppliers

### 2.6 NIST

We refer to NIST SP800--30, where the total process of risk mitigation is described in four phases such as “risk mitigation options”, “risk mitigation strategy”, “an approach for control implementation, control categories, the cost--benefit analysis", and "residual risk".

The followings are risk mitigation options.

- Risk Assumption: To accept the potential risk and continue operating the IT system or to implement controls to lower the risk to an acceptable level
- Risk Avoidance: To avoid the risk by eliminating the risk cause and/or consequence
- Risk Limitation: To limit the risk by implementing controls that minimize the adverse impact of a threat's exercising a vulnerability
- Risk Planning: To manage risk by developing a risk mitigation plan that prioritizes, implements, and maintains controls.
- Research and Acknowledgement: To lower the risk of loss by acknowledging the vulnerability or flaw and researching controls to correct the vulnerability
- Risk Transference: To transfer the risk by using other options to compensate for the loss, such as purchasing insurance.

NIST also provides SP800--53, which includes a list of more than 170 recommended security controls for Federal Information Systems.

The classes of controls and their families are shown as follows.

- Management Class: “Certification, Accreditation, and Security Assessments (CA)”, “Planning (PL)”, “Risk Assessment (RA)”, “System and Services Acquisition (SA)”
- Operational Class: “Awareness and Training (AT)”, “Configuration Management (CM)”, “Contingency Planning (CP)”, “Incident Response (IR)”, “Maintenance (MA)”, “Media Protection (MP)”, “Physical and Environmental Protection (PE)”, “Personnel Security (PS)”, “System and Information Integrity (SI)”
- Technical Class: “Access Control (AC)”, “Audit and Accountability (AU)”, “Identification and Authentication (IA)”, “System and Communications Protection (SC)”

## 3. Brief explanation of useful tools

In this section, some tools based on fuzzy theory such as fuzzy outranking method, fuzzy inference mechanism, modified structural modelling method based on FSM, and fuzzy c-mean (clustering) are briefly described.

### 3.1 Fuzzy outranking method

The method to roughly compare two alternatives $a$ and $a'$ through the adoption of loose relation is called outranking. When $a$ is judged not to be inferior to $a'$ at least, it is said that $a$ outranks $a'$. When $a'$ is more preferable than $a$ or they are incomparable to each other, it is said that $a$ doesn't outrank $a'$. While these relations are valued as 0 or 1 in the conventional outranking method, such as $\mu(a,a')=1$ if $a$ outranks $a'$ and $\mu(a,a')=0$ if $a$ does not outranks $a'$, the fuzzy outranking method access the outranking degree as a value between 0 and 1. More precisely, that degree is determined using a fuzzy membership function with lower threshold value $q_i$ and upper one $p_i$, where "$i$" represents one of view points for evaluating these alternatives. Thus the corresponding value is denoted by $c_i(a,a')$ ($i$=1,…,$n$), and they are aggregated by taking the weighted average $\omega_1c_1(a,a')+\ldots+\omega_nc_n(a, a')$ with a set of certain weight $\{\omega_1,\ldots,\omega_n\}$. This index is called the "concordance index" denoted by $C(a,a')$. Another index is "discordance index" denoted by $d_j(a,a')$, which is also calculated using a fuzzy set with lower threshold value $p_j$ and upper one $v_j$. This index represents the degree of objection against the preferability to choose $a$ then $a'$. Thus $d_j(a,a')=1$ implies that the condition "$a$ outranks $a'$" is exclusively vetoed from the number $j$ point of view.

If there are discordant points of view $j_1,\ldots, j_k$, whose index are greater than $C(a,a')$, then the total outranking index $\mu(a,a')$ is calculated by the following formula:

$$\mu(a,a') = C(a,a') \times \frac{1-d_{j_1}(a,a')}{1-C(a,a')} \times \cdots \times \frac{1-d_{j_k}(a,a')}{1-C(a,a')} \tag{1}$$

### 3.2 Fuzzy inference mechanism

Fuzzy inference (Kaufman, et al., 1975; Klir & Yuan, 1995) is originally the process of formulating the mapping from a given input to an output using fuzzy logic. Then the mapping provides a basis for which decisions can be made, or patterns distinguished.

The rule of fuzzy inference is generally expressed as follows:

"IF $x$ is $A_1$ and $y$ is $B_1$ THEN $z$ is $C_1$, else IF $x$ is $A_2$ and $y$ is $B_2$ THEN $z$ is $C_2$, else IF $x$ is $A_n$ and $y$ is $B_n$ THEN $z$ is $C_n$ , else $x$ is A′ and $y$ is B′ THEN $z$ is C′ ", where $A_1,\ldots, A_n$, A′ are subsets of universe of discourse $U$, and $B_1$ ,…, $B_n$, B′ are fuzzy subsets of universe of discourse ($V$; $C_1$ , …, $C_n$, C′ are fuzzy subsets of universe of discourse $W$).

We have several types of fuzzy number such as triangular, trapezoidal, and Gaussian fuzzy numbers in mind (Inoue & Amagasa, 1998, pp. 57-66).

### 3.3 Modified structural modelling

The modified structural modelling method is developed by Cui, D. and Amagasa, M. for constructing a structural model with consensus of multi-participants (Amagasa, 2004, pp. 121-132, Nagata et al., 2008A). Here, assume that a decision group consists of several members (decision makers) with either equal or different knowledge background for a given problem.

Let GM$_k$ ($k$=1,...,$m$) denote group members, and A$_k$ ($k$=1,...,$m$) be fuzzy subordination matrices of data given by GM$_k$.

Then, mental model of GM$_k$ is embedded into a fuzzy subordination matrix on the context on basis of the relaxation of transitivity, reflexivity and symmetry by each group member (Zadeh 1965; Klir & Yuan 1995; Tazaki & Amagasa 1979). Herein, NGT and automatic generation method of subordination matrix are applied to embed entries of the matrices efficiently and effectively. In order to formulate the individual fuzzy subordination matrix with the same establishment level, the entries of the matrix embedded by group individual are normalized statistically. Then, a representative subordination matrix is formulated by integrating the fuzzy subordination matrices of group members as follows:

Let $S=\{s_1,\dots,s_i,\dots,s_n\}$ denote a system with $n$ elements, and let $A_k = [a_{ij}^k]_{n\times n}$ ($k$=1,2,…,$m$) denote the fuzzy subordination matrices in S, where $a_{ij}^k = f^k(s_i,s_j)$ (0≤ $a_{ij}^k$ ≤1, $i,j$=1,2,…,$n$, $k$=1,2,…,$m$). $a_{ij}^k$ is the grade of which $s_i$ is subordinate to $s_j$ and $m$ is the number of group members. Let $NA_k = N^k(\overline{a}_k, (\sigma_a^k)^2) = [h_{ij}^k]_{n\times n}$ ($k$=1,2,…,$m$) denote the normalized fuzzy subordination matrices calculated by the given data $A_k = [a_{ij}^k]_{n\times n}$ from group members with

$$h_{ij}^k = \frac{1}{100}\left(\frac{a_{ij}^k-\overline{a}_k}{\sigma_a^k}\times 10+50\right)\ (i,j=1,\dots,n,\ \ k=1,\dots,m),\ \text{ where }\ \overline{a}_k=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^k\ \ (k=1,\dots,m)\ \text{ and}$$

$$\sigma_a^k = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^{k^2} - \overline{a}_k^2}\ \ (k=1,\dots,m).$$

Now, the normalized subordination matrices are used to compute the representative subordination matrix which holds the data factor from group members. Let $NAR=[d_{ij}]$ ($i,j$=1,…,$n$) be a representative subordination matrix, which is computed by

$$d_{ij} = \frac{1}{m}\sum_{k=1}^{m}h_{ij}^k\ \ (i,j=1,2,\dots,n). \quad (2)$$

Next, the fuzzy reachability matrix is computed on the basis of $NAR$, and multi-level digraph is drawn as an interpretive structural model. In order to compare the structural model with mental model, a feedback for learning will be performed to group members. If an agreement among group members is obtained, the process goes ahead to documentation step. Otherwise, a threshold and fuzzy structure parameter will be modified and the process is iterated until a consenting model is derived. Here, let $p$ be the threshold, specified by $\alpha$-cut, which is defined by the modified z-value in standard normal distribution. The value of $p$ is used for controlling the percentage of subordination relations among elements which exist in the structural model to be evaluated.

Fig. 2 illustrates a flowchart of the modified structural modeling method which begins with mental model of individual group member which is determined depending on their intuition to the given problem.

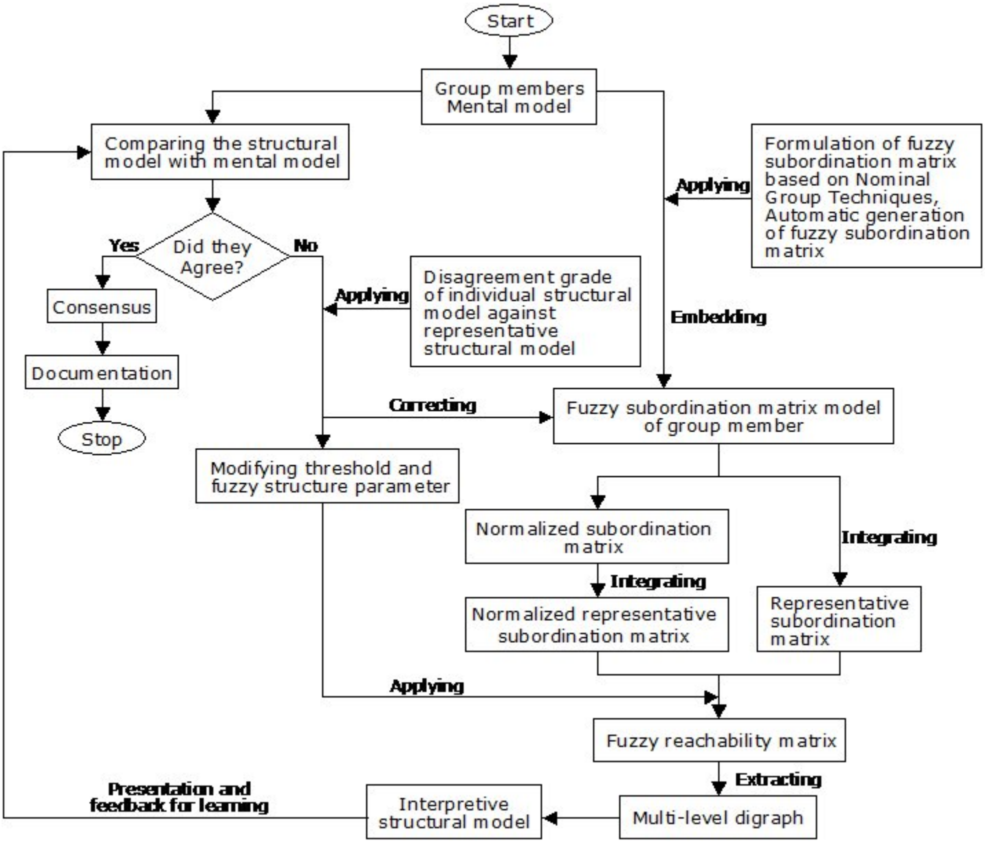

Start
Group members Mental model
Comparing the structural model with mental model
Formulation of fuzzy subordination matrix based on Nominal Group Techniques, Automatic generation of fuzzy subordination matrix
Applying
Did they Agree?
Yes
No
Consensus
Documentation
Stop
Applying
Disagreement grade of individual structural model against representative structural model
Embedding
Correcting
Fuzzy subordination matrix model of group member
Modifying threshold and fuzzy structure parameter
Normalized subordination matrix
Integrating
Integrating
Normalized representative subordination matrix
Representative subordination matrix
Applying
Fuzzy reachability matrix
Extracting
Presentation and feedback for learning
Interpretive structural model
Multi-level digraph

Fig. 2. The flowchart of modified structural modeling method

### 3.4 Fuzzy c-mean clustering

Data base of multi-attribute elements can be classified into several groups according to a fixed metric. This kind of process is call a clustering, and ordinary clustering is simply that defining a function,

$$\mu = \mu_d : D \times C \to \{0,1\}$$

satisfying the condition that for any $x \in D$ there is only one $c \in C$ such that $\mu(c,x) = 1$ .

Here $D$ is the set of all data, $C$ is the set of clusters, and $d$ represents a distance with some kind of metric e.g., Euclidean metric, maximum metric, etc. With the function above, each element has only one cluster and no overlapping of clusters.

Fuzzy c-mean clustering is represented by a function,

$$\mu = \mu_d : D \times C \to [0,1]$$

Here the value $\mu(c,x)$ between 0 and 1 indicates the degree of membership of a data $x \in D$ in a cluster $c \in C$ , and clusters can be overlapped.

Now let *n* be the number of data with $n_v$ attributes, and *s* be a number of clusters, and express each data $x_i=(x_{i1},...,x_{in})$ with values of *j*-th attribute $x_{ij}$ where $D = \{x_1,...,x_n\}$ .

When we put the set of all the cluster centers $V = \{v_1,...,v_s\}$ , the objective function which should be minimized is defined as following.

$$J(D;U,V) = \sum_{j=1}^{s}\sum_{i=1}^{n} \mu^{m}{}_{ij}\, d^2(x_i,v_j)\ , \tag{2}$$

where $U = \{\mu_{ij} = \mu_d(c_j,x_i)\}$ satisfying trivial constraints in inequalities $0 \le \mu_{ij} \le 1$ ($i$=1,...,$n$, $j$=1,...,$s$) and only one non-trivial equation $\sum_{j=1}^{s} \mu_{ij} = 1$ .

The exponential number *m* of *μ* reflects the fuzzyness of the clustering, such as setting *m*=1 implies the ordinary, not fuzzy, clustering, increasing the value of *m* means the widely overlapping of the resulted clusters.

By introducing the Lagrange multiplier λ, objective function is

$$W(D;U,V) = J(D;U,V) - \lambda(\sum_{j=1}^{s} \mu_{ij} - 1) \tag{3}$$

and optimal solutions are given at the saddle points, that is{ $\mu_{ij}$} and $v_j$ ($j$=1,...,$s$) satisfy

$$\begin{cases} \dfrac{\partial W}{\partial \mu_{ij}} = m\mu_{ij}^{m-1} d^2(x_i,v_j) - \lambda = 0 \\ \dfrac{\partial W}{\partial v_{jk}} = \displaystyle\sum_{i=1}^{n} 2(x_{ik} - v_{jk})\mu_{ji}^{m} = 0 \end{cases} \tag{4}$$

where $v_{jk}$ represents the *k* coordinate of point $v_j$, and the distance function is the Euclidean distance $d(x_i,v_j) = \sum_{k=1}^{n_v} (x_{ik} - v_{jk})^2$ .

Solving the equations above, we have

$$\begin{cases} \mu_{ij} = \left( \displaystyle\sum_{k=1}^{s} \left( \frac{d(x_i,v_j)}{d(x_i,v_k)} \right)^{\frac{2}{m-1}} \right)^{-1} \\ v_{jk} = \dfrac{\displaystyle\sum_{i=1}^{n} \mu_{ij} x_{ik}}{\displaystyle\sum_{i=1}^{n} \mu_{ij}^{m}} \end{cases} \tag{5}$$

Thus the algorithm proceeds in the following steps;

1. Load a database $D$. Determine the number of clusters $s$, fuzzification value $m$, and the error evaluation threshold $\varepsilon$
2. Set $t$=1, and give certain initial values for $\{\mu_{ij}\}$ denoted by $\{\mu_{ij}{}^{(t-1)}\}$
3. Calculate $v_{jk}^{(t)} = \sum_{i=1}^{n} \mu_{ij}^{(t-1)} x_{ik} \,/\, \sum_{i=1}^{n} \mu_{ij}^{(t-1)m}$ , and put

$$\mu_{ij}^{(t)} = \left( \sum_{k=1}^{s} \left( \frac{d(x_i, v_j^{(t)})}{d(x_i, v_k^{(t)})} \right)^{\frac{2}{m-1}} \right) \tag{6}$$

4. With the corresponding values for $\lambda = m\mu_{ij}^{m-1} d^2(x_i, v_j)$ , evaluate the difference of two values $W(D;U^{(t)},V^{(t)})$ and $W(D;U^{(t-1)},V^{(t-1)})$ by $\varepsilon$
5. If the difference value is less than $\varepsilon$, then stop and output $\{\mu_{ij}^{(t)}\}$ and $\{v_j^{(t)}\}$ as the results for $U$ and $V$. If not, increase $t$ by 1 and go back to the step 3)

In the algorithm above, we need to be careful that the fuzzification exponent $m$=1 reduces the denominator of the exponent of each terms in Σ for $\mu_{ij}^{(t)}$ to 0.

Moreover, $m$ is usually set a values between 1.4 and 2.6 (Celikyilmaz, & Turksen, 2009, p.57).

## 4. Method for choosing effective set of mitigation controls

For our proposed method for selecting set of mitigation controls from a database of controls, we assume the existence of an external database, $D$, of mitigation controls with mitigation degree, $\delta_m(T)\in[0,1]$ and $m\in D$, evaluated depending only on the type of threat path $T$. This mitigation degree should signify that adopting the control roughly mitigate the risk level from 1 to that degree.

We use the risk profile work sheet of OCATVE-S, and we suppose that determination of the set of critical assets are done, and all the possible threat path were distinguished with the risk value calculated from $(v_R,v_F,v_P,v_{Fi},v_S,v_O,p)$, the vector of impacts and probability. This is the preliminary stage of our method.

Then the process is performed according to the following steps.

**Step 1.** Determine a threat path T.

**Step 2.** Select several controls as members of the candidate set, $M\subset D$, by evaluating their initial mitigation degree dependent on $T$. One simple way to determine $M$ is setting $M=\{m\in D{:}\delta_m(T)<\delta\}$ for a definite value $\delta$.

**Step 3.** Define the desirable, but dummy, mitigation control, $a_0$, as an acceptable impacts and probability vector $(v_{R0},v_{F0},v_{P0},v_{Fi0},v_{S0},v_{O0},p_0)$.

**Step 4.** For each element $m_j\in M$, figure out its mitigation degree $d_*{}^j$ with respect to each of impacts and probability. For instance, $d_R{}^j$ represents the reduction degree with respect to the impact of reputation when $m_j$ is performed. These degrees are calculated by considering the type of assets, threat path, and impact or probability in some criteria.

**Step 5.** Calculate $a_j$= $(v_R{}^j,v_F{}^j,v_P{}^j,v_{Fi}{}^j,v_S{}^j,v_O{}^j,p^j)$ as the alternative vectors corresponds to $m_j$ by $d_*{}^j\times v_*$.

**Step 6.** Apply the fuzzy outranking method with certain threshold values of concordance and discordance indices to each of $(a_j, a_0)$ for $j=1,\ldots n$, where $n$ is the cardinality of $M$.

**Step 7.** Determine the set of effective mitigation controls $E_T$ by referring the outranking relation values $\mu_j = \mu(a_j, a_0)$. We have two versions for this. One is to determine $E_T = \{m_j; \mu_j > a\}$ as the optimal set with fixed lower boundary value α. The other is to choose the definite number of $m_i$s′ from the permutated mitigation controls in descending order.

## 5. Method for construction of effective database system

Now we propose a method composed of three phases to construct a database system with an effective clusters.

Phase I: Collecting Mitigation Controls

It seems to be patient and time-consuming works that we gather and examine all controls possible to mitigate information related risks, together with giving each of them a kind of classification index simultaneously. The classification is used to give each control a value vector of OCATVE's threat path attributes related entries in Phase II. Fortunately, we have some of existing database of controls referred in section 2 such as in ISO/IEC 27002, MEHARI, NIST SP-800, and in OCTAVE. They are already classified in view of various aspects.

Phase II: Evaluation of Controls

This phase is composed of two processes.

Process 1: Vector indication in a fixed set

Fix a set of mitigation controls with some classification. Indicate a vector whose entries are values between 0 and 1 corresponding to each of attributes in OCTAVE's threat paths to all the controls in the set. Concretely speaking, we have six possible attributes “access“ (“network“, “physical“), “actor“ (“inside“, “outside“), “motive“ (“accident“, “deliberate“) on the human actors worksheet, and four possible attributes “actor“ (“software defects“, “malicious code“, “system crashes“, “hardware defects“) on the system problems worksheet. We propose a method to indicate the values for each of attribute by applying the MSMM in the following steps,

**Step 1.** give a weight each of first level or second level classes

**Step 2.** give a weight all the controls in each class

**Step 3.** aggregate two weight values in step 1 and step2

Process 2: Evaluation and modification

In the previous process, we have controls with value vector according to each classified set. The same or similar control can be appear in some classified sets, and it could be possible that one control has more than one value vector. We need to identify those controls and examine the indicated vectors of each of them before going on the next phase. If the vectors corresponding to a control have only acceptable difference, then take a vector whose entries are the average of each entries as the final value vector of the control. If not, go back to the value vector indication steps.

Phase III: Clustering Controls

Clustering all controls using fuzzy c-mean clustering method by means of attribute vectors. Make the correspondence between each of clusters and each of threat paths by looking at the center vectors of clusters. Selecting a small set of mitigation controls is performed using this correspondence and *U* defined in subsection 3.4.

## 6. Conclusion and discussion

As the final goal of the series of information security evaluation and management system, a system to propose a set of mitigation controls effective and efficient to reduce the organizational risk level is very important. For this purpose, the construction of a feasible database of mitigation controls is necessary. In this chapter, we look over several types of controls, and proposed a method for construct the database. The resulted consists of controls with a value vector whose entries are corresponding to some of attributes on the threat path in OCTAVE's risk profile worksheet. Our idea to apply the fuzzy c-mean clustering might be helpful to choose a small set of control candidates from a huge number of controls.

For the practical use, we need to construct a feasible and real database by applying our system and to verify the effectiveness of the total system.

In our future work, we intend to apply our system to some of classified set of mitigation controls, such as in OCTAVE, ENISA, NIST SP800 and in MEHARI, to obtain an example of effective database. We also intend to define a function from a set of threat path attributes to a set of clusters resulted from fuzzy c-mean clustering.

## 7. References

Alberts, C. & Dorofee, A. (2003). *Management Information Security Risks*, Addison-Wesley.

Amagasa, M. (2004). *Management Systems Engineering*, Institute of Business Research, Daito Bunka University

Bidgoli, H. (Editor-in-Chief) (2006). *Hand Book of Information Security*,Vol. III, John Wiley & Sons.

Inoue, H. & Amagasa, M., (1998), *Fundamentals of Fuzzy Theory*, (in Japanese), Asakura Shoten.

Celikyilmaz, A. & Turksen, I. B. (2009) *Modeling Uncertainty with Fuzzy Logic*, Springer.

Kaufman, A. et al. (1975). *Introduction to the Theory of Fuzzy Subsets*, NewYork: Academic Press.

Kaye, D. (2002) *Strategy for Web Hosting and Managed Services*, John Wiley & Sons.

Klir, G. J. & Yuan B. (1995) *Fuzzy Sets and Fuzzy Logic-Theory and Application*, Prentice Hall International Inc.

Nagata, K.; Kigawa, Y.; Cui, D. & Amagasa, M. (2007). Integrating Modified Structural Modeling Method with an Information Security Evaluation System, *Proceedings of* *the 8th Asia Pacific Industrial Engineering and Management Systems Conference 2007*, T1-R02, ID68.

Nagata, K.; Umezawa, M.; Cui, D. & Amagasa, M. (2008A). Modified Structural Modeling Method and Its Application -Behavior Analysis of Passengers for East Japan Railway Company-, *Journal of Industrial Engineering and Management Systems*, Vol. 7, NO. 3, pp. 245-256.

Nagata, K.; Kigawa, Y.; Cui, D. & Amagasa, M. (2008B). Risk Evaluation for Critical Assets with Fuzzy Inference Mechanism in an Information Security Evaluation System, *Proceedings of the 9th Asia Pacific Industrial Engineering and Management Systems Conference 2008*, pp. 2630-2640.

Nagata, K.; Kigawa, Y.; Cui, D. & Amagasa, M. (2009). Method to Select Effective Risk Mitigation Controls Using Fuzzy Outranking, *Proceedings of the 9th International Conference on Intelligent Systems Design and Applications*, pp. 479-484.

Nagata, K. (2011). On Clustering of Risk Mitigation Controls, *Proceedings of 2011 International Conference on Network-Based Information Systems*, pp. 148-155.

Tazaki, E. & Amagasa, M. (1979). Structural Modeling in a Class of Systems Using Fuzzy Sets Theory, *International Journal of Fuzzy Sets and Systems*, Vol.2, No.1, pp. 87-103.

Yu, Q. H. ; Liang, G. Y. & Nagata, K. (2010). Risk Scoring Method on Business Information Management System, *Proceedings of the 11th Asia Pacific Industrial Engineering and Management Systems Conference 2010*, DVD-ROM, ID117.

Zadeh, L. A. (1965). Fuzzy Set, *Information and Control*, Vol.8, pp. 338-353.

Alberts, C.; Dorofee, A.; Stevens, J. & Woody, C. (2005). OCTAVE-S Implementation Guide, Version 1.0, CMU/SEI-2003-HB-003. 28.02.2011,Available from
http://www.cert.org/octave/octaves.html

Information technology--Security techniques--Code of practice for information security management, ISO/IEC 27002 Central, 28.02.2011, Available from
http://www.17799central.com/

MEHARI 2010: Fundamental concepts and functional specifications, 28.02.2011, Available from
http://www.clusif.asso.fr/fr/production/ouvrages/pdf/MEHARI--2010--Principles—Specifications.pdf

Recommended Security Controls for Federal Information Systems: 28.02.2011, Available from
http://csrc.nist.gov/publications/nistpubs/800--53--Rev3/sp800--53--rev3--final/_updated--errata/_05--01--2010.pdf

Risk Management:Implementation principles and Inventories for Risk Management/Risk Assessment methods and tools, 28.02.2011, Available from
http://www.enisa.europa.eu/act/rm/cr/risk--management--inventory/downloads.

Risk Management: Information Package for SMEs, 28.02.2011,Available from http://www.enisa.europa.eu/act/rm/cr/risk--management--inventory/downloads

Risk Management Guide for Information Technology Systems, 28.02.2011,Available from http://csrc.nist.gov/publications/nistpubs/800--30/sp800 --30.pdf

# Quality Model – Master Plan and DNA of an Information System

Finne Auvo
*University of Jyväskylä,
Finland*

## 1. Introduction

The goal of the chapter is to give a refined definition for the quality of information system as a technical artifact and based on that statement present a complete conceptual framework for quality modeling. Further, the chapter shows how a quality model as a master plan for information systems can drive and control the entire development process.

### 1.1 Information system and its context - Models and objects of modeling

Every theory has its surroundings and postulates. So has a theory about quality models, and it is better to make the main lines of these ideas explicit before presenting the theory itself. A human made information system (IS) as a technical artifact exists and operates always in the context of societies, organizations, personal lives etc. It is a tool used for gathering, storing, processing, presenting and exchanging (communication) information. These activities can be termed "information behavior" (Allen et al., 2011). Accordingly the context of an information system has a two-tiered structure (Figure 1). The inner tier, information behavior, is subordinate to the outer tier, human society. Information in general is used to support human activities, and technical information tools, in turn, are used to enhance the use of information.

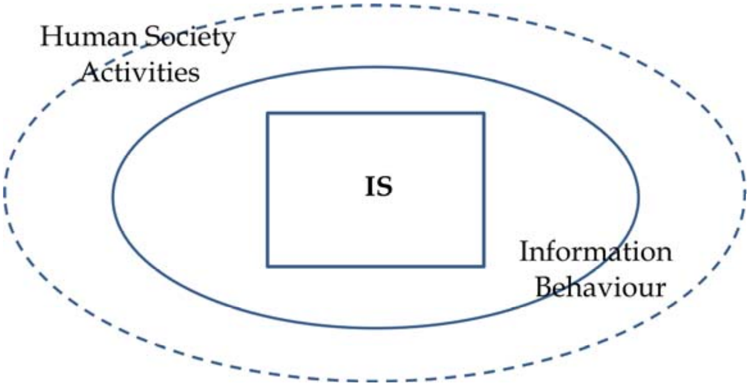


Fig. 1. Two-tiered context of information system

In this context both information and information system are supposed to have meaningful functions, and because of these purposes human actors have various wants, needs and expectations about the information itself and the supporting system's behavior. This view is implicit for example in the writings where sense-making theory is applied to information system design (e.g. Muhren et al., 2008). Sense-making provides means to understand how information in general is used by humans (see e.g. Savolainen, 2006), which in turn should be reflected in information tools design. Another theory frequently used (e.g. Silva, 2007 and Macome, 2008) to explain the interplay between information system and its context is Actor Network Theory (ATN). ATN views technology as part of a network of human actors and nonhuman artifacts. And still one interesting framework has been developed by Alter (1999, 2008). He views information systems in connection with work systems that the former serve, and in the end information systems as special cases of work systems. Human society is full of work systems "in which human participants and/or machines perform work (processes and activities) using information, technology, and other resources to produce specific products and/or services for specific internal or external customers" (Alter 2008, 451). The definition of work system comprises projects, supply chains, web sites, etc.

In everyday life and scientific literature actors' expectations for information and information systems are commonly called requirements, and a part of them more precisely quality requirements. Each IS has a life course that contains various cyclical or repeating processes, like system development, moving from one platform to another, etc. ISO/IEC 12207 (2008), for example, gives a good picture of software lifecycle processes. In the course of these processes there are several points where requirements are captured and goals set, something – usually a system or component - to fulfill the requirements modeled, implemented and put into operation, and finally the results measured and evaluated. The words "modeling" and "designing" are in this chapter used interchangeably. And a "model" means the product of modeling or designing process, an abstraction of or a blueprint for something to be realized. The difference between requirements and a model based on them is that requirements express needs and desires and are often less structured and consequently written in the form "x is needed" or "x must be y", whereas a model is structured and statements are in indicative form like "x is y".

Traditionally modeling in software engineering has centered on the system itself as a product of development, the development process or the entire system life course (Figure 2). With respect to these three alternatives this chapter focuses on information system as a technical artifact seen in its context of development and use, primarily human activities. It does not go into the area of system development process or IS life cycle models. Consequently, the quality models that are discussed can be categorized as product quality models.

Modeling can take place on different levels of abstraction. And the actual object (X) that is modeled can be any real thing, not only an entity or process, but even a state of affairs as this chapter will show. After modeling and implementing several Xs one can create a general model of Xs or of certain type of Xs that consequently constitute theories of X. Finally, instance and general level models can be used to create a meta-model, a model for X-models that can be regarded as an even higher level of theory. The other way around, higher level model can be used as template for creating lower level models. By iterating this kind of model generation, from bottom up and from top down, models on all levels grow better and better by the time and experience. Figure 3 depicts this relationship between levels of modeling, and at the same time between practice and theory.

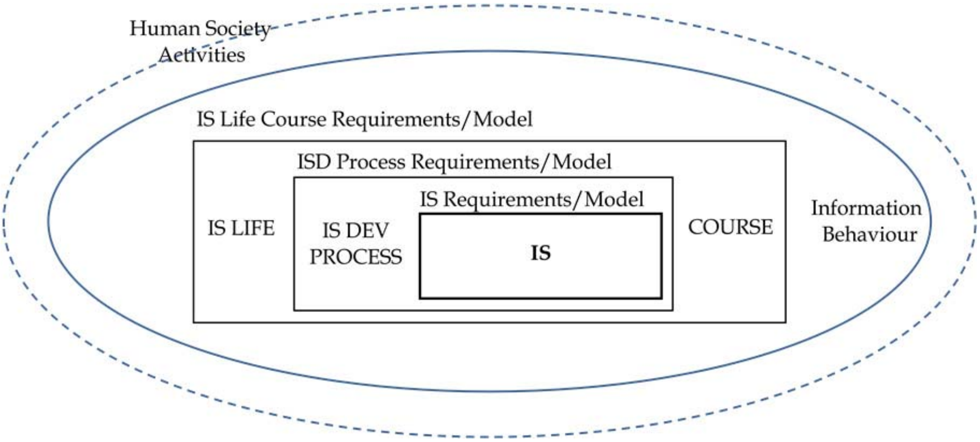


Fig. 2. Traditional objects of modeling in software engineering

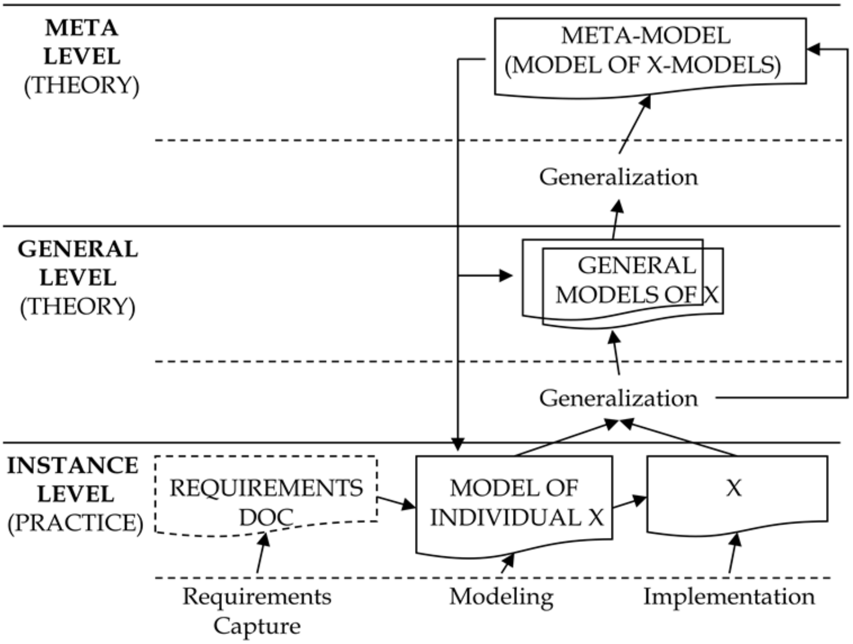


Fig. 3. Levels of modeling, practice and theory

This chapter is about quality requirements, quality models and their DNA like role in system development. In order to be complete and avoid misunderstandings, an account of quality requirements, quality models and their DNA like role in system development has to start from the meta-level.

## 2. The essence of quality - Requirements and quality requirements

Reeves and Bednar (1994) have discussed different definitions of quality in business context. They state that the concept of quality has had multiple and often unclear definitions and look more closely at four of them: quality as 'excellence', 'value', 'conformance to specifications', and 'meeting expectations'. Excellence means meeting the highest criteria in some area like intelligence, strength etc. The value aspect introduced price, or value, as an additional determinant of consumer's decision. Different compatibility requirements in production of component based machines lead to equaling quality with conformance to specifications and to making quality measurable. Finally the most pervasive definition 'meeting customer's expectations', according to Reeves and Bednar (1994), grew out of services marketing. It is also the most complex definition and most difficult to measure. Reeves and Bednar (1994) conclude that a global definition of quality does not exist and different definitions are appropriate in different contexts. The IEEE Standard Glossary of Software Terminology offers a two part definition of system quality: "the degree to which a system, component or process meets specified requirements" and "the degree to which a system, component or process meets customer or user needs or expectations. It coincides with the fourth definition discussed by Reeves and Bednar (1994).

Defining quality can be started from the viewpoint that in very general terms information system quality can be seen as determined by the existence and intensity of something pertaining to the system, identifiable and desired by actors and stakeholders. This point of view is in a way similar to the definition of quality as 'meeting expectations' above. What is desired is referenced to by using adjectives and abstract nouns and commonly interpreted as characteristics or features that reside inside and constitute an integral part of the entity (system) being described. Quality definitions like in ISO 9126 reflect this viewpoint. It gives in the annex a general definition for quality taken from ISO 8402:1994 (replaced now by ISO 9000:2000): "the totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs" (ISO 9126 (2001, 20)). Analysis of the meaning of adjectives and abstract nouns used for qualities discloses, however, that they refer actually to certain states of affairs or relationships between the system and things in its context. By taking, for example, ISO 9126 quality model's six main characteristics functionality, reliability, usability, efficiency, maintainability and portability as examples, one can see that they all refer to a relationship between information system and its context. Functionality is the capability of software to provide functions which meet stated and implied needs (ISO 9126, 7). It is clearly a relationship between the product, its users and business processes they have to carry out. Reliability is defined as the capability of software to maintain a specified level of performance under specified conditions (ISO 9126, 8). Again it is about a relationship, this time a relationship between the product, a specific instance of using it and certain conditions. Usability is the capability of software to be understood, learned, used and attractive to the user when used under specified conditions (ISO 9126, 9). This characteristic relates the product to the user and certain conditions. Efficiency is the capability of software to provide appropriate performance, relative to resources used and under stated conditions

(ISO 9126, 10). It relates the product to resources and use under certain conditions. Maintainability is the capability of product to be modified and adapt to changes in environment, requirements and functional specifications. Again it is clearly about the product in relation to its context. Finally portability is defined as the capability of software to be transferred from one environment to another (ISO 9126, 11).

The search for the essence of quality can as well be started from inside the system. Taking an internal view, any information system can be described exactly and without remnants by indicating its architectural type, programming language used, design patterns, layers and packages, by listing procedures and methods, records in the database, series of instructions, and so on. This kind of description does not, however, explain WHY these constituents are required. In some cases an element is needed to create another element. But for what is the latter needed? Following the chains of *WHY*s one comes in the end out of the system internals into some desired relationship between system and its context, even if it is just a relationship between system and individual actor. Accordingly, for to explain *WHY* a design pattern, method, etc. is required or needed, these relationships must be described and understood. After finding the raison d'être of internal constituents in system-context relationships, the constituents themselves can be viewed as contributing factors in the former.

The fulfilment of quality requirements is not only dependent of internal system characteristics. Studies on information system quality show that external things can also have an impact on the desired relationships. Narasimhaiah and Lin (2010), for example, have studied external contributors. They focus on organizational and individual human factors associated with quality attributes like reliability, ease-of-use, maintainability, usefulness and relevance. They found as external determinants of software quality things like attitude of management, responsiveness and capability of IS department and capability of users themselves. Each of the determinants was further decomposed into smaller items. Capability of users, for example, consisted of users' knowledge in the system, training received, involvement in or resistance to the system, and technical competency. When it comes to super-attributes like sustainability the importance of external contributors is evident.

Figure 4 depicts the above viewpoints. It lists along the upper part of the ellipse a number of terms used in literature for referring to those subcategories of requirements that are commonly regarded as quality requirements. Functional requirements are added to the set on basis of the discussion above. Each individual requirement (NR1, G1 and FR1) or "desire" points to a state of affairs or relationship (R1, R2 and R3) between information system and its context. At the same time these requirements point to certain concrete system features (e.g. architecture and method) and things (T1, T2 and T3) outside the system. The former, desired relationships, explain the need for the latter, concrete system features and external conditions. Accordingly, if a developer starts asking in respect of any internal system feature that is under design or implementation, why it is required, he or she finally always traces back to some desired relationship between the system and its environment. This subordinate status of system internals and externals compared to the expected system-context relationships explains the meaning of the expressions "in the first place" and "in the second place" in the following definition paragraph. The two-tiered nature of system context, discussed in section1, is represented in Figure 4 by the two circles around the system rectangle.

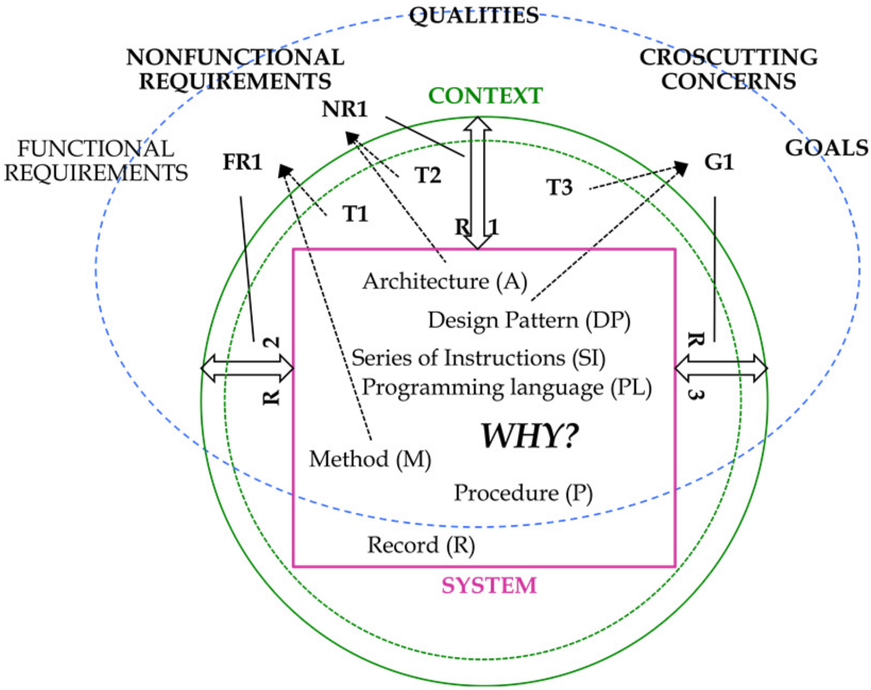


Fig. 4. Quality requirements and contributors

Quality of information system as a technical artefact in its context of development and use is in the first place determined by the existence, lack, intensity or number of desired relationships between information system or system constituent and its context. It can also be characterized as a desired state of affairs. In the second place quality of information system is determined by the existence, lack, number of or intensity of elements, inside or outside the system, contributing to the realization of the desired relationships. These elements get in a way there “justification” or “raison d’être” through the desired system-context relationships. Consequently individual qualities as well as overall quality of information system must be defined and measured in terms of these relationships.

‘Requirement’ as such is a broader notion meaning anything required, be it a certain quality, or something needed to realize it, or something else needed in the system for some reason. The main difference between quality attributes and other attributes, or quality requirements and other requirements, is that the former refer in first place to desired relationships between information system and its context and have more importance (priority) to actors than the latter, and the latter can often be derived from the former. Consequently quality requirements constitute the core of requirements for an information system. Priority, definition and measured level of implemented quality requirements are often relative to use case, actor, or some feature of the context. Therefore actors usually agree on goals regarding to what degree quality requirements need to be met.

## 3. A comprehensive quality model

Given the definition of information system quality put forward in previous section it is obvious that the design for a quality technical artifact, being created and functioning in its context of development and use, has to cover quite a number of different aspects. Describing a state of affairs or relationship takes much more than describing an entity and its attributes. Case study findings (Finne, 2011) suggest that a comprehensive model, that is needed to fully account for the quality of an information system as an end product of development process, appears to be a hybrid model with six sub-models. It can alternatively be described as an intersection of models. Figure 5 represents a meta-level view of the hybrid model.

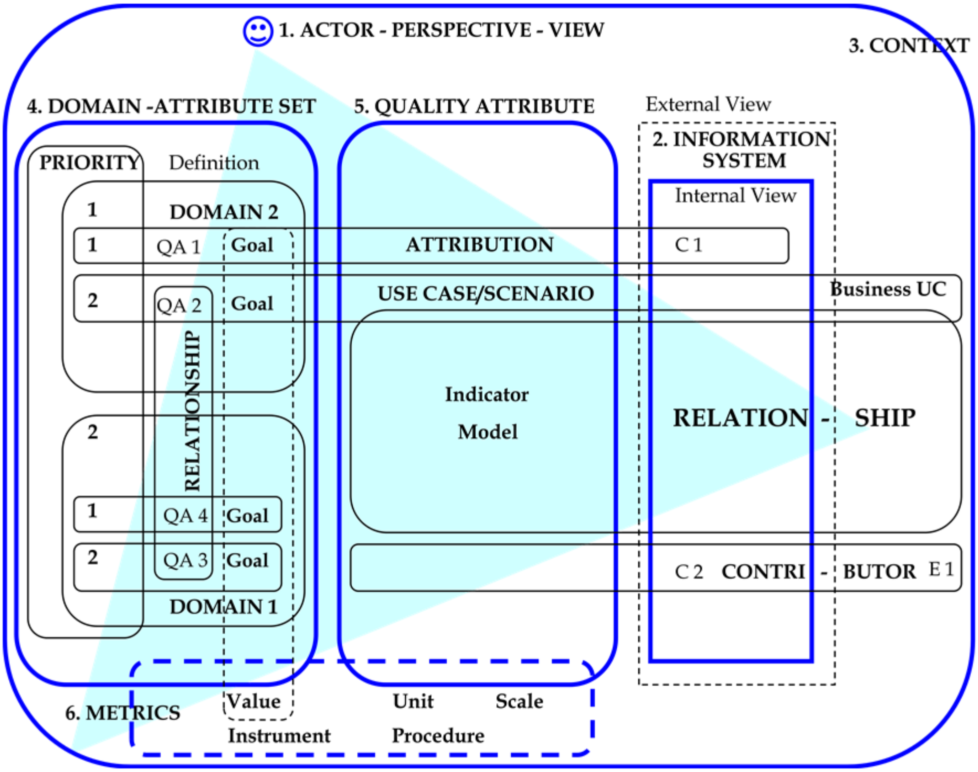


Fig. 5. A meta-level view of information system quality model

The first sub-model, 1) human actors with their perspectives and views (symbolized by the filled triangle without borders) is actually part of an activity or process (quality modeling or system development) model. Actors are typically seen as elements of activity models. Next, to gain an understanding of the target information system on necessary levels of abstraction 2) an information system model is needed. To deal correctly with 3) context one needs to model it to some extent. Next, the 4) domain and quality attribute set reflects the core requirements and areas of concern for the stakeholders, and can be viewed as a model of overall system quality. Individual 5) quality attribute models are in essence models of desired system-context relationships or states of affairs. Finally, the 6) metrics part is a model by itself. It describes procedures and means of determining if the intermediate or end product of development process complies with the design in respect of quality. Following paragraphs give a detailed definition to meta-model elements. D1 and D2 stand for two sample domains. QA1, QA2, QA3 and QA4 are individual quality attributes. C1 and C2 symbolize system constituents, and E1 symbolizes an entity in the context. UC in "Business UC" stands for use case. Goal means a goal value set for a quality attribute. The overlapping of rectangles symbolizes:

- PRIORITY: Both domains and attributes are prioritized.
- ATTRIBUTION: A quality attribute (QA1 in the figure) in the attribute set can be attributed to the information system as a whole or to some of its constituents (C1 in the figure). The attribution of qualities starts when the domain-attribute set is created and is reviewed during modeling of individual attributes.
- USE CASE/SCENARIO: Use cases comprise system and business use cases, the latter being part of the context. An attribute (QA2 in the figure) can be initially connected to use cases already during creation of domain-attribute set. Attributes can have different order of priority in connection with different use cases.
- RELATIONSHIP (on the left): Quality attributes are interrelated in many ways.
- RELATIONSHIP (on the right): The majority of quality attributes refer in first place to desired relationships between information system and its context. These relationships become visible and are defined in connection with use cases and scenarios.
- CONTRIBUTOR: Internal (C2 in the figure) and external (E1 in the figure) contributors determine for their part to what extent the desired relationships are met. Contributors are system constituents or things in the system context.
- VALUE/GOALS: Goals are target attribute values.
- METRICS: Metrics can be designed for individual quality attributes or overall quality (taking into account the whole attribute set).

People or groups of people who have some meaningful relationship with the information system under scrutiny are called *actors*. They are affected by the qualities of the information system or its products. A general naming for actors that can also be used is "stakeholders". Some actors may never use the system, but are anyway somehow interested in it or affected by it and its products. The term "*informants*", in turn, means a sub-set of actors who actually participate in quality modeling or quality measuring and give some relevant information. Actors are part of the information system context.

Human actors have certain perspectives on and views of information system and its context that are reflected in resulting quality models. Each quality model element can in fact be traced back to a particular actor or actor group. A *perspective* is characterized by the actor's background, organizational roles and activities, beliefs, values, etc., and actor's relationship to the information system on basis of them. The former are at the same time elements of the information system context. A *view* of information system and its environment, in turn, is characterized by what it excludes and what it includes, i.e. by the constituents or elements visible in the view and their relationships. The elements in a view can be activities and processes as well as other things. A view of system can be concrete (through using or creating the system) or based on system descriptions. It can be general or detailed, partial or complete and so on. A view can be affected by elements constituting the perspective or other psychological and cognitive factors that determine what an actor wants to see or how an actor interprets what is seen. In terms of quality modeling a particular view usually covers only some parts of the information system and context.

For each actor, the knowledge about context, the knowledge about information systems in general and about the specific system under modeling together with the perspectives given by the roles and work the actor participates, constitute a kind of *frame of reference* that determines the appearance of the system to the actor. Perspectives and views which are presented and shared in the course of quality modeling form the intersection of individual frames. Figure 6 depicts the main elements of actor's frame of reference. The system is depicted on a very high level of abstraction as a structure built from different constituents, having contents and acting in a particular way. T1, etc. stand for "things" and S1, etc. for "systems" in the context of system and actor.

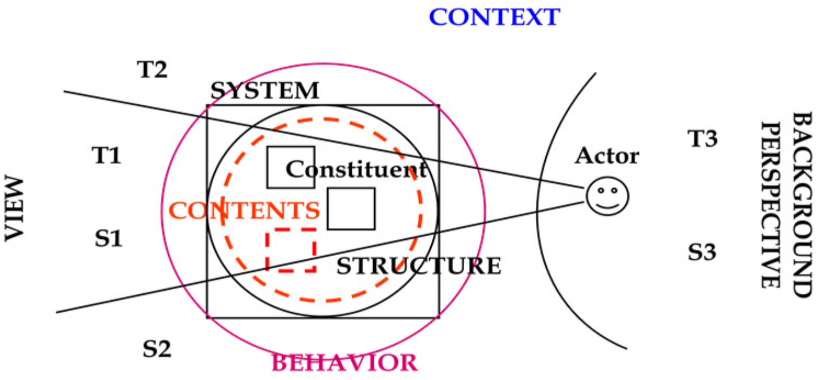


Fig. 6. Actor's frame of reference

An information system as a technical artifact consists of electronic and non-electronic components (*constituents*), their functioning (*behavior*) and relationships (*structure*). It includes the human and machine interfaces for data input and output. It is used to store, process, produce and present information (*contents*) in order to support human activities, including entertainment. Information products created by the system or serving as input into system are regarded as elements in a wider integrated system or the human information behavior as a whole. From the perspective of product quality modeling the systems elements have no advance justification. The reason or explanation for their existence comes through their ability to contribute for fulfilling quality requirements. Starting from a mere black-box view, gradually, according to the needs of attributing some qualities to lower level components and finding contributors to the qualities from inside information system, more detailed architectural descriptions are used. Good architectural descriptions are worth of gold but unfortunately a rarity in system development projects.

From the perspective of quality modeling *context* comprises all the elements in system environment that are members in the desired relationships between the system and context. These relevant entities can be human or non-human, independent of the spatial or temporal distance. Business model forms an important part of context model.

A *quality attribute* refers in first place to a desired relationship between information system and its context or to a number of connected relationships. The existence and intensity of this relationship determines the level of quality in question. In a domain-attribute set each quality attribute can be given a general *definition* and a *goal value*, and it can be *attributed* or allocated to the system as a whole or some of its constituents. Having goals is essential in defining and assessing quality. Attribution to a constituent means that the particular quality of the constituent determines in fact the same quality of the whole system. Qualities are of high importance to actors and form a prioritized set.

*Domain* is a field of thought or area of concern to the actors in connection with which a quality attribute or group of attributes is relevant. It groups together related attributes that can be viewed as individual concerns. Each domain in an attribute set or collection is given a general *definition*.

A *domain-attribute set* is a *prioritized* list of all quality attributes ascribed to the information system as a whole or to a particular constituent. The attributes in the set are given a general definition and goal value, grouped according to prioritized domains and related to each other. Attributes themselves are prioritized on the level of the whole set, inside domains or both. Priority is characteristic of quality determining attributes. Different factors, among them actors' perspectives and views, have an impact on the final selection and prioritization of attributes. Quality *domain-attribute collection*, in turn, is a general set or supply of domains and quality attributes that can be used as a source when assembling the system specific domain-attribute set. Specific attribute collections can be created for different types of systems.

Figure 7 shows as an example a domain-attribute set that was used in an EMIS (Education Management Information System) case study (Finne, 2006). Attributes sustainability and suitability are positioned in the middle of the "palette" to underline their composite nature. This kind of presentation helps to identify biases and gaps in system's quality design. In Figure 7, for example, neglected domains are 'architecture and design', 'change' and 'performance'. Predefined, general or system type specific attribute collections, in turn, can act as starting points and checklists ensuring that the experiences of similar information systems and similar environments, or information systems in general, are taken into account. At the same time it must be kept in mind that no listing can cover all possible quality characteristics relevant to all possible information systems. Similarly a fixed and non-controversial categorization of quality attributes is probably impossible.

*Business use cases*, as part of the business model, represent the processes of an organization with or without explicit reference to supporting information systems. A *system use case*, for its part, represents the use of system per se, without, in the first place, drawing attention to its connection to business processes. The term *scenario*, in turn, refers to particular circumstances or to a flow of events, other than use cases, where the role of human actors as users of an information system sometimes can be non-existent, or not focused on.

Most of the characteristics that are traditionally called 'quality attributes' refer to a desired *relationship* or set of desired relationships *between the information system*, or its constituent, *and* one or more entities in the *context*. This can be regarded as characteristic of quality

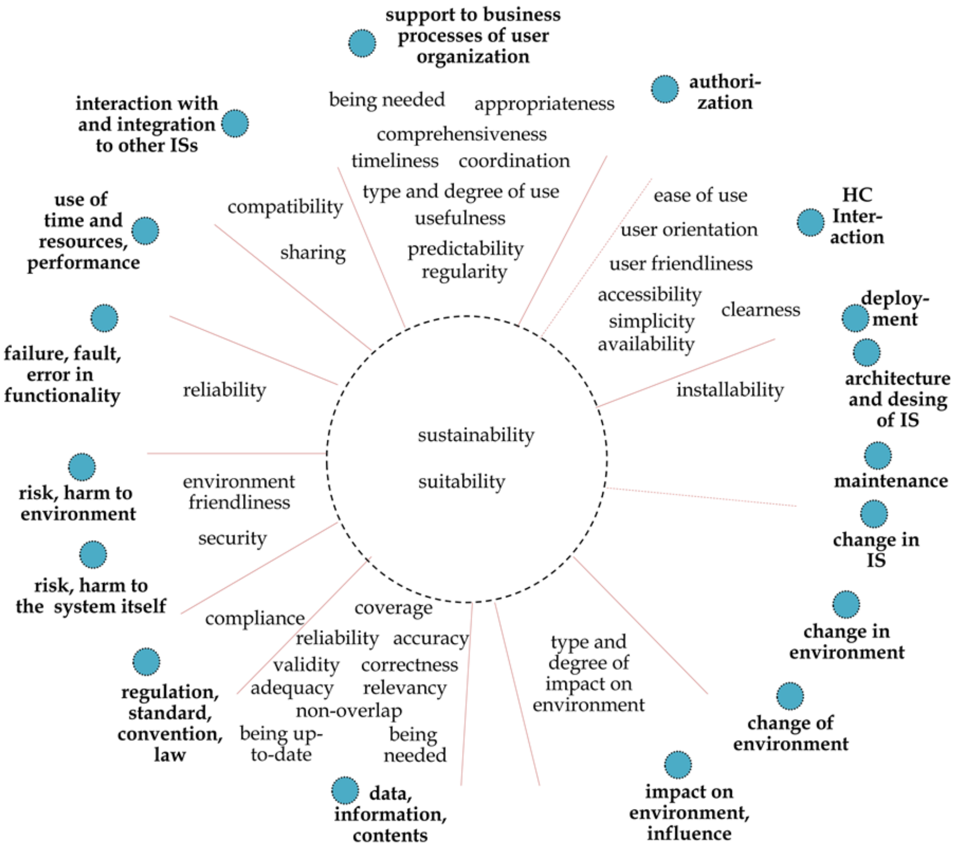


Fig. 7. An example of domain-attribute set

attributes. It can also be named a 'desired state of affairs'. In this way the system is linked with those entities in actors' minds or physically. Certain facts, called *indicators*, indicate the existence and intensity of the relationship in connection with real use cases and scenarios. A quality model can in addition list negative facts that show the lack of the desired relationship. A general formal presentation of the desired relationship is called *model*. It can be given, for example, in the form of an entity relationship (ER) model. Figure 8 gives a simplified example. It is taken from a quality modeling case study conducted in Zanzibar. It is part of the security-attribute-model and shows three entities: a land registration system (LRS), one human actor and one external connected information system. One security related attribute (security mechanism) is attached to LRS and another (ICT skills) to the human actor. When the human actor or external system tries to connect to LRS the latter either denies or allows the connection and shows and hides information according to implemented security rules. This really is a simplified example. In practice one can identify a number of additional relevant attributes. And the description of relationship is more advanced.

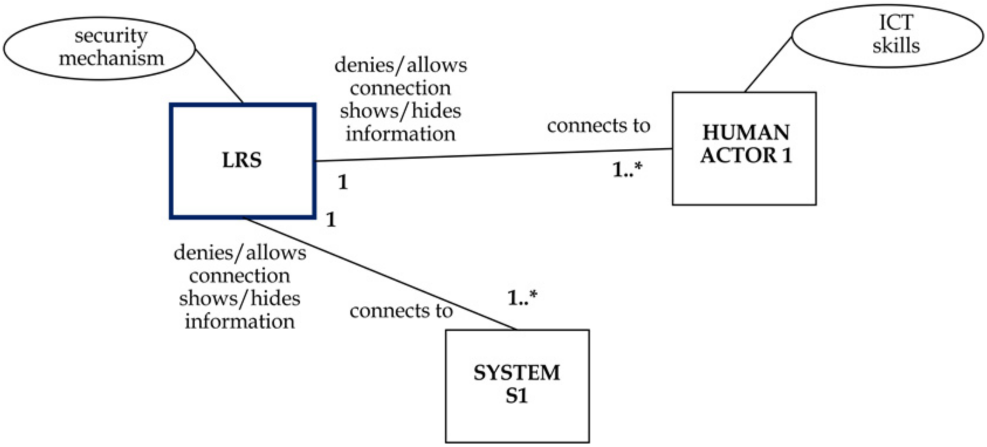
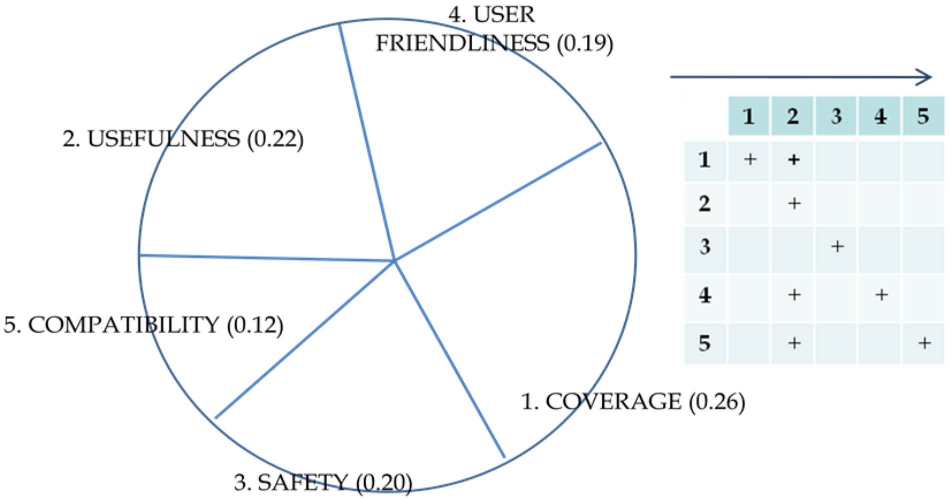

security mechanism
ICT skills
denies/allows connection shows/hides information
connects to
LRS
1
1..*
HUMAN ACTOR 1
1
denies/allows connection shows/hides information
connects to
1..*
SYSTEM S1

Fig. 8. A simplified ER-model depicting the security system-context relationship

*Contributor* is a thing *inside or outside* the information system that affects in a positive way a desired relationship between system and its context. It can be alternatively called “factor”. The description and understanding of system environment, discussed above, is important in identifying the external contributors. The internal things, in turn, are usually system constituents, behavior (functioning) or structures and consequently part of the system model. Desired relationships are the “raison d’être” of contributors and quality attributes can be said to refer in the second place to the latter. What are the contributors in reality with respect to each quality is in the end a subject of empirical study. Thereafter, based on achieved theory system developers can instantiate the contributor elements in system design.

Attribute relationships are a feature of domain-attribute set. These relationships can be identified by comparing indicators, models, contributors and measured values. Indicator sets, for example, can be separate, overlapping, or one included in another. Contributors, in turn, can be indifferent to one another, cooperating (supporting), or conflicting. In theory there can be as many kinds of attribute relationships as there are relationship types between indicators or contributors. Figure 9 shows a way to depict diagrammatically attribute relationships. It is taken from the same EMIS case study (Finne, 2006) as Figure 6 above. First, the circle at left represents five top ranked quality attributes. The angle of the slices represents the relative priority value (as a number in brackets) of attribute. In the matrix “+” sign stands for a positive contribution. As one can see from the matrix, the attributes are quite independent. Only usefulness is clearly influenced by most of the other attributes. Different signs in the intersection of rows and columns can symbolize different relationships. A similar matrix is used by Khaddaj and Horgan (2005).

A complete arrangement for measuring the existence and degree of a quality, i.e. of a desired relationship between system and its context, includes selection of *instrument, unit, scale, actors, and measurement procedure*. The *object of measurement* is in first place the existence and degree of certain desired system-context relationship represented by the indicators, and in the second place the existence and properties of internal and external contributors. As



|   | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | + | + |   |   |   |
| 2 |   | + |   |   |   |
| 3 |   |   | + |   |   |
| 4 |   | + |   | + |   |
| 5 |   | + |   |   | + |

Fig. 9. Example of relationships between top five attributes

was noted above in connection with actor's perspective and view, the objects of measurement are in each case defined, observed and measured in a certain *frame of reference*. The things that cause relativity, i.e. difference of results compared to another frame of reference, constitute together the set of axes of the frame. They include, for example, business and system use cases, business processes, observer's organization and role in it. Relativity is characteristic of information system quality.

There exist many alternative ways of representing graphically the model of a specific quality attribute, or even the entire quality model of a particular system. In general, however, it is often impossible to fit all information into one diagram. This problem has to be solved by representing only core features graphically, by attaching textual descriptions or distributing the information on several diagrams. Figure 10 exemplifies how little information actually fits in a diagram that can be viewed without scrolling. It is again taken from the EMIS study and it summarizes some core features of 'coverage' quality. The attribute refers to the concern of covering with the system all the information that potential users need. It is ranked as number one concern. The general definition has been written by the researcher. Only one indicator is shown in the upper right corner. The model reflects the perspectives and views of researcher, 11 selected informants and the EMIS development team. Attribute belongs to the data-domain and is positively related to usefulness. Of the context elements only an unnamed business process (P1) is visible in the diagram. It represents the large number of business processes where actors use information provided by the system. Coverage has been defined with respect to process data use case. It refers to use of EMIS for processing statistical information for example in order to create different statistical presentations. 100 per cent coverage is set as goal. Measuring instrument, unit and procedure are indicated in the lower left corner of the diagram.

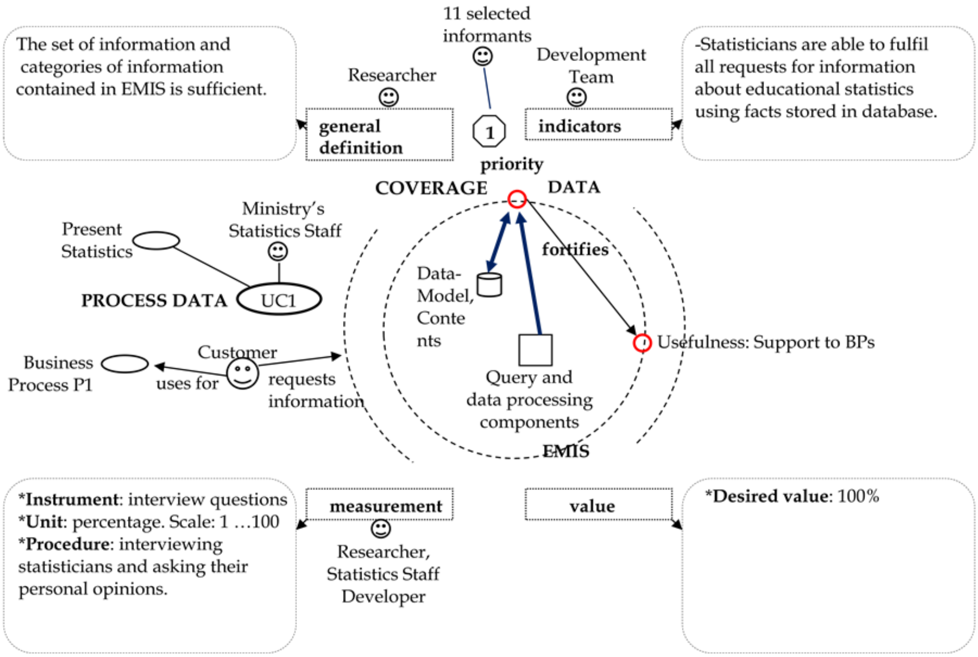

The set of information and categories of information contained in EMIS is sufficient.
11 selected informants
Researcher
Development Team
-Statisticians are able to fulfil all requests for information about educational statistics using facts stored in database.
general definition
1
indicators
priority
COVERAGE
DATA
Present Statistics
Ministry's Statistics Staff
PROCESS DATA
UC1
fortifies
Data-Model, Contents
Usefulness: Support to BPs
Business Process P1
Customer
uses for
requests information
Query and data processing components
EMIS
*Instrument: interview questions
*Unit: percentage. Scale: 1 …100
*Procedure: interviewing statisticians and asking their personal opinions.
measurement
Researcher, Statistics Staff Developer
value
*Desired value: 100%

Fig. 10. Graphical representation of coverage attribute model

## 4. Quality driven development - Quality model as master plan of information systems

Given the view on quality requirements constituting the core of requirements for an information system, an endeavor to fulfill them must also be the core process in system development. During the last ten years the software developer community has seen a rise of different approaches characterized by the word "driven". The experiences in case studies (e.g. Finne, 2011) using the comprehensive model suggest introducing still another approach that could be called simply "quality driven development". It does not only mean that quality design and implementation must be a truly integral part of software development that cannot be given up. It means raising the quality modeling from the role of being just a separate component to the status of an umbrella like driving force. If stakeholders in the end want a quality system, quality is also the issue to start with. All other goals and decisions should be subordinate to it and all other design elements in line with quality design. In this sense quality drives the whole development process, not only for example, architecting. And if quality is understood as realization of a set of desired and most essential system-context relationships, then a quality model can be seen having a DNA-like role and carrying the "genetic information" of system.

In an era that pursues agility it is often feared that the introduction of extra models cause too much overhead to development work in terms of calendar time, money and other resources that are badly needed to address more fundamental challenges. But what can be more fundamental than capturing, prioritizing and realizing core system requirements. Figure 11 shows how the creation and use of quality model elements are positioned in relation to object-oriented software development process (as presented by Jacobson et al., 1999).

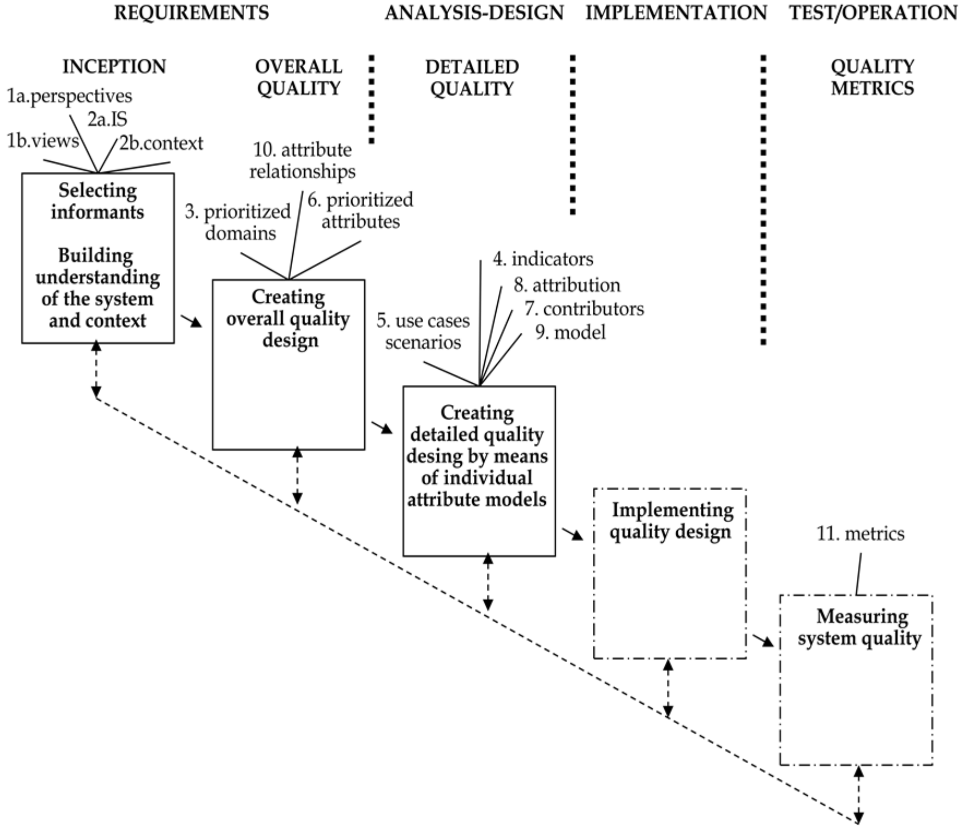


Fig. 11. Quality modeling in relation to object-oriented development process

Awareness of context as well as perspectives and views of actors who participate in system development is mandatory in any kind of development approach. IS, the technical artifact, must always be designed. Further, no system exists separate from requirements that are sometimes conflicting and need to be prioritized. Use cases and scenarios are important stuff even in agile methodologies. Finally every system must be tested, assessed and compared to requirements. In these respects quality modeling does not bring any extra burden to the picture. The point in quality driven development is that quality of an information system is understood to be in first place determined by the existence and intensity of desired system-context relationships and consequently as well defined and measured in terms of these relationships. The quality meta-model, presented in previous section, provides a template for designing the relationships and relating the resulting model to other models needed in system development process. The point is further that a system specific quality model is the arch-model in system development and never left out of sight, and that everything what is done is done in the name of realizing the quality goals. The following guidelines for applying the quality meta-model are based on experiences of three case studies in East-Africa.

From Figure 11 one can one can see how overall quality design coincides with requirements capture in unified software development model. Jacobson et al. (1999) divide requirements into two categories: functional and non-functional requirements. Functional specifications tell what the system is supposed to do for the users. Non-functional requirements, in turn, correspond to what are traditionally called quality attributes, like performance, availability, etc. The core quality 'usefulness' in fact covers the functional requirements in Jacobson's model. Consequently a well designed and prioritized system specific attribute set with goal values and attribute relationships, can cover the whole range of requirements and act as a guide and driving force for the whole development process. The importance of integration between functional and other requirements is seen for example by Kotonya and Sommerville (1996).

The first task during a phase that can be called "*inception*" is always formation of an actor group that carries out quality modeling. This group is so important that it is positioned as the first element in quality meta-model followed by the initial understanding of the system and its context. The latter things have an impact on the selection of first group members. Even the first understanding of the system is inevitably an interpretation made by some human actor(s). If the group is formed when the project is initiated and most of the members participate also into other development activities of the same system, it guarantees that quality modeling will be integral and dominant part of the whole software development process. The actor-informant group is properly formed if all essential actor-stakeholder categories are represented. Case studies showed, however, that it is often difficult to engage all relevant stakeholders. Therefore, it is practical to start the work with most immediate system users and expand the informant group later according to needs and possibilities. If relevant, cultural aspects have to be taken into account when actor group is formed. Some studies in developing countries (e.g. Thanasankit and Corbitt, 2000, using a Thai case) show that social structures and hierarchies can be tall. All kinds of decisions must be approved by managers or committees. In these kinds of environments actor group must cover not only people with knowledge but also people with power. Before starting the actual quality modeling the informant group must be aware of the perspectives (step 1a) its members possess and the views (step 1b) they can have of the target system and its context. After that comes the task of gaining initial understanding of system and context.

The information system can first be described (step 2a) to actors more or less as a "black-box" or by simple structural diagrams. The purpose of this view is to turn attention from the very beginning to the desired relationships between system and its context, i.e. quality requirements. It resembles the Taylorian view where, for example, messages stored in information system have no inherent value, and the value of entire system emerges only within a context (see Scholl et al. 2011, 790). The description of information system will gradually become more detailed and transparent during "attribution" and "contributors" steps. After initial system description follows the initial description of the system context (step 2b). It is a more important task in the initial phase than description of the system itself. Entities in the environment, including different human actors, determine what is required of the relationships between system and context. *A* suitable "meta-model" or theory of context would guide in focusing on the most relevant features of environment with respect to quality modeling.

After inception phase the overall quality design can start. It consists of creating a prioritized system specific set of domains (areas of concern) and quality attributes out of known alternatives. The results of inception phase together with the initial overall quality design form a kind of sketch of the target system in relation to context. In one of the case studies actors were given a large combined selection of domains and attributes and asked to prioritize them. This was found to be difficult for some informants because of too many items to deal with. Therefore, in the following case studies domain collection was separated from the "palette" and actors were asked to prioritize the domains (step 3) and attributes separately. During overall quality design conflicts may arise and a method for solving them must be found.

Case study observations suggest that after letting actors prioritize domains the most useful and productive step is eliciting positive and negative facts (step 4) indicating existence or the lack of quality inside each prioritized domain. This starts the detailed quality design. In the meta-model these facts are called "indicators" and they constitute an important part of individual quality attribute models. The identification of use cases, scenarios (step 5) and individual quality attributes (step 6) related to the facts can follow thereafter, as well as prioritizing the attributes. Next comes looking for contributors (step 7) and possibly attributing (step 8) the qualities more precisely to certain system constituents. In these steps actors need a lot of support from someone experienced in quality modeling. The assumption that stakeholders are able to understand and communicate present and future needs in a clear way has been recently criticized for example by Holmström and Sawyer (2011, 35). Defining internal contributors is part of system model and consequently requires that developers take part in the process. The steps from 5 to 8 are in practice carried out rather simultaneously than one after another. Unless system component or sub-system specific attribute collections and sets are used, quality characteristics are usually at the beginning of quality modeling attributed to the information system as a whole. More specific attributions grow up during the modeling process, especially in connection with contributor element. All the steps of detailed quality design can potentially affect and cause changes in the overall quality design, i.e. the prioritized domain-attribute set.

After listing indicators and identifying use cases, scenarios and contributors there exists enough material for creating a formal representation each desired relationship between system and context called "model" (step 9). At this stage indicators, model and its verbal description can be used to refine the general definition of the attribute in question in domain-attribute set. The remaining steps are identification of attribute relationships (step 10), especially conflicts, and designing a procedure for measuring (step 11) actual attribute values in connection with testing and operating the target system.

The principles of flexibility and freedom are important in the selection of system specific domain-attribute sets and in quality-driven development in general. There are, however, some core requirements that are commonly acknowledged to be important per se and should always be included in attribute sets. First of all any information system must be feasible (before even trying to create or acquire it), available, accessible and sustainable. In addition, it must be useful and therefore frequently used or, in some cases (e.g. computer games), have an ability to entertain. All the other characteristics, usability in the front line, follow from or affect the before mentioned. They hamper or make it easier to use the information system, make it less accessible etc. Figure 12 depicts this view.

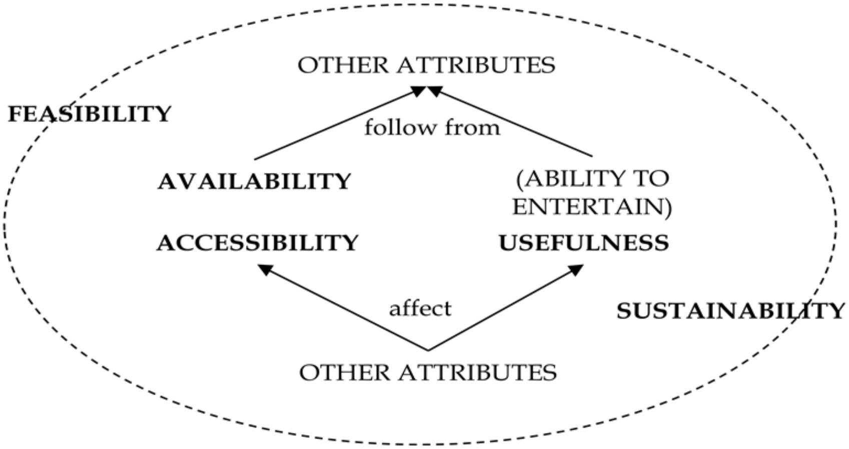


Fig. 12. Core quality requirements

The division into core qualities and other qualities resembles the division into key quality factors and locally defined factors by Khaddaj and Horgan (2005) in their Adaptable Quality Model (AQM). The key factors are required of all products. Locally defined factors, in turn, apply only to the current product being developed. AQM defines in total seven key qualities: maintainability, usability, cost/benefit, security, reliability, timeliness and correctness. Compared to AQM Figure 12 represents even more fundamental requirements. In a recent article Buschmann (2011), in turn, underlines just usefulness (in his terminology "business suitability") and usability as the key requirements for software.

## 5. Conclusion

The above sections have given a definition for quality with respect to information systems as technical artifacts. In addition they have presented a holistic conceptual framework for quality modeling and a way to apply it in the course of system development. The last section finally promoted quality model to arch-model of an information system. As a theory, the quality meta-model does not assert anything testable about the relationships between its elements. The most relevant method of evaluating the framework is assessing the actual quality models created by applying it. These system specific models must be useful, contain elements that represent the real world and be comprehensible. In addition the meta-model itself must be comprehensive, flexible, general enough and applicable to a variety of contexts. This entails repeated case studies. While the framework does not offer testable propositions, it opens a number of questions for future research. How to arrange quality attributes into categories? Can a widely acknowledged division into domains be achieved? Are some actor perspectives and views more informative and productive than others? What are the most relevant axes in the frame of reference that determine the appearance of system to an actor? How is the physical, infrastructural and cultural context reflected in quality mnodels?

## 6. References

Allen D., Karanasios S., Slavova M. (2011). Working With Activity Theory: Context, Technology and Information Behavior. *Journal of the American Society for Information Science and Technology. 62(4),* 776-788.

Alter S. (1999). A General, Yet Useful Theory of Information Systems. *Communications of the Association for Information Systems. 1(3),* Art. 13.

Alter S. (2008). Defining information systems as work systems: implications for the IS field. *European Journal of Information Systems, 17(5),* 448-469.

Buschmann F. (2011). Unusable Software Is Useless, Part1. *IEEE Software, 28 (1)*, 92-94.

Finne A. (2006). Applying a Twofold Quality Model: Producing Groundwork for System Specific Attribute Models. In Hautop H., SutinenE., Duveskog M., Kinshuk, Mkocha A. (eds.) *Proceedings of the Fourth IEEE International Workshop on Technology for Education in Developing Countries, Iringa, Tanzania, July 10-12, 2006,* 3-4.

Finne A. (2011). Towards a Quality Meta-Model for Information Systems. *Software Quality Journal, 19(4),* 663-688.

Homström J., Sawyer S. (2011). Requirements engineering blinders: exploring information systems developers' black-boxing of the emergent character of requirements. *European Journal of Information Systems, 20(1)*, 34-47.

IEEE (Institute of Electrical and Electronics Engineers) Standard Glossary of Software Engineering Terminology (1990). New York: IEEE.

ISO (International Organization for Standardization) 8402 (1994). *Quality management and quality assurance – Vocabulary*. Geneva: ISO/IEC.

ISO (International Organization for Standardization) 9000 (2000). *Quality management systems – Fundamentals and vocabulary*. Geneva: ISO/IEC.

ISO (International Organization for Standardization) 9126 (2001). *Software Engineering – Product Quality – Part 1: Quality Model*. Geneva: ISO/IEC.

ISO (International Organization for Standardization) 12207 (2008). *Systems and software engineering – Software life cycle processes*. Geneva: ISO/IEC.

Jacobson I., Booch G., Rumbaugh J. (1999). *The Unified Software Development Process.* Boston: Addison-Wesley.

Khaddaj S., Horgan G. (2005). A Proposed Adaptable Quality Model for Software Quality Assurance. *Journal of Computer Sciences, 1(4)*, 482-487.

Kotonya G., Sommerville I. (1996). Requirement engineering with viewpoints. *IEEE Software Engineering Journal, 11(1),* 5-18.

Macome E. (2008). On Implementation of an Information System in the Mozambican Context: The EDM Case Viewed Through ANT Lenses. *Information Technology for Development, 14(2),* 154-170.

Muhren W., van den Eede G., van de Walle B. (2008). Sensemaking and Implications for Information Systems Design: Findings From the Democratic Republic of Congo's Ongoing Crisis. *Information Technology for Development, 14(3),* 197-212.

Narasimhaiah G., Lin S-L. (2010). Determinants of software quality: A survey of information systems project managers. *Information and Software Technology 52 (6)*, 602-610.

Reeves C., Bednar D. (1994). Defining quality: Alternatives and implications. *The Academy of Management Review, 19(3),* 419-445.

Savolainen R. (2006). Information use as gap-bridging: The viewpoint of sense-making methodology. *Journal of the American Society for Information Science and Technology, 57(8),* 1116-1125.

Scholl H., Eisenberg M., Dirks L., Carlson T. (2011). The TEDS Framework for Assessing Information Systems From a Human Actor's Perspective: Extending and Repurposing Taylor's Value-Added Model. *Journal of the American Society for Information Science and Technology, 64(4),* 789-804.

Silva L. (2007). Institutionalization Does Not Occur By Degree: Institutional Obstacles in Implementing a Land Administration System in a Developing Country. *Information Technology for Development, 13 (1),* 27-48.

Thanasankit T., Corbitt B. (2000). Cultural Context and its Impact on Requirements Elicitation in Thailand. *The Electronic Journal on Information Systems in Developing Countries, 1(2),* 1-19.

# 8

# Services for the Digital Citizen

Seppo Sirkemaa
*Turku University of Applied Sciences,*
*Life Sciences and Business,*
*Business Information Technology, Turku,*
*Finland*

## 1. Introduction

Internet makes it possible to provide services in a new way, making it possible to create added value to the user. At the same time organizations may re-organize and streamline their processes. Internet is changing the way we purchase products and services. Through Internet we may gather background information on competing products and services, compare and purchase things without the need to leave home. From the business perspective there are several targets when moving activities to the internet, serving customers on a 24 / 7 –basis and global reach are issues that may prompt the development of e-business applications. One of the key drivers in e-business is that internet makes it possible to increase company's efficiency and effectiveness (Rust & Kannan, 2003).

In private sector information technology, internet-based applications and technologies are used widely in e-business applications. Clearly, public sector needs to move from paper to electronic correspondence, and from this toward a self-service model where citizens can get the answers and make transactions through the internet (Atkinson & Leigh, 2003). The concept of service is inextricably linked to e-business applications and the types of services there are in e-business environment (de Ruyter et al., 2001). Here self-service is typical, users have learned to help themselves in finding information and buying products. This is the case also with citizens that are using electronic services provided by public organizations.

The development of electronic services in public sector organizations has been relatively slow (Hasan & Tibbits, 2000; McIvor et al., 2002). This is interesting, because it seems clear that also public sector would benefit from electronic access to services. It is not surprising that there is pressure and an increasing demand for development of e-services in public sector. It is noteworthy, that those who can access internet are in a different position than people who do not have this opportunity (Cullen, 2001). Equal access to internet and services that are made available through it is an issue all over the globe, and it concerns also citizens in the industrialized world.

In this article we look at the challenges that development of electronic services in public sector organizations face. It is an environment which calls for cooperation of various departments and functions, and interaction between service providers, experts and other stakeholders. The question of interest is what makes providing electronic services in public sector so different from development of e-business applications in private companies. In public sector the goal is not only to move services to internet, it is also a question of developing one-stop government solutions (Kubicek & Hagen, 2000; Gouscos et al., 2003). Let us look closer at development of electronic services in public sector.

## 2. Value of electronic services

Electronic services are in this context referred to as e-Services, which relates to services that public organizations provide. The term e-Services is further defined as interactive, content-centred services are accessed through the internet (Rust & Kannan, 2002; Rust & Kannan, 2003). Most e-Services are related to information: the internet is a way to access information independently of time and location. However, there is an increasing demand on interactive and transaction-enabled services through the internet (Ancarani, 2005). Clearly, e-Services need to be integrated into processes and systems of the organization that provides them, especially if the services are transaction-related (de Ruyter et al., 2001).

Travel industry is a good example of an industry, which uses information technology extensively. The customer or the traveller has the possibility to make reservations, get to know hotels, car-rental services and more in the travel destination. The internet gives the traveller services which earlier were possible only through the travel agency. This gives the traveller better control on the travelling experience. The result is a change in the infrastructure of travel industry; today customers are increasingly making reservations by themselves. From the service providers perspective it is critical to have visibility in the internet; the service should be listed when the traveller is planning trip and uses search engines like Google in this. The issue here is that traveller needs to do the planning; technology just shows different options (routes, hotels, fares etc.) to choose from. It is likely that in the future this is not enough, more advanced services will be developed which help in travel planning and adjust to changes in schedules, for example (de Ruyter et al., 2001).

With e-business small- and medium sized companies can compete globally. The most significant benefits of e-business are connected to transactions and communication (Dutta & Roy, 2003). Internet lowers transaction-related costs for both buyers and sellers. Companies can change prices on-line when raw-material costs change, for example. At the same time buyers have access to up-to-date prices directly from their terminals - most online shoppers use comparison-shopping engines (Mulpuru, 2007). Internet allows restructuring of processes which results better profitability - these are important issues for all companies, and are motivators for development in the public sector as well. Even though goals, ethics and values are slightly different business-like performance measurement has been evolving in public sector organizations as well (Van Der Wal et al. 2006; Parhizgari & Gilbert, 2004).

## 3. Electronic services in public sector

Let us look at services in public sector. Services can be described by three dimensions; services can be general in opposite to individualized, separate in opposite to coordinated (integrated) and informative in opposite to performative (Goldkuhl & Persson, 2006).

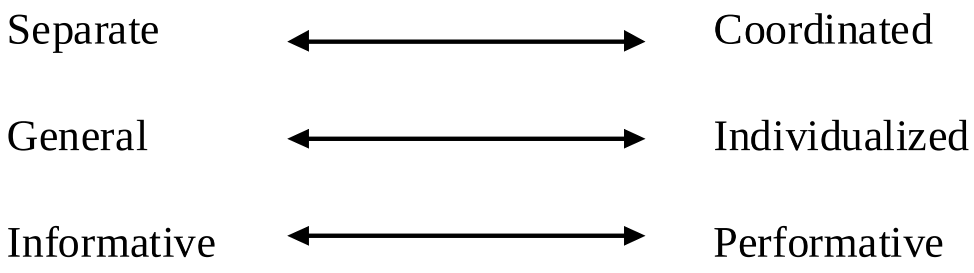


Fig. 1. Three polarities of e-services

The first dimension relates to integration of services. Separate e-services are services from one single agency or office. At the other end there are coordinated e-services which are result of cooperation of several agencies or offices. Some coordinated services are fused together, but there are also services that are merely aligned and may still be separated from each other. Fused services are totally integrated and individual services are not separable. (Atkinson & Leigh, 2003)

The services may be general in a way that they are not designed for any special group of users. Again, services or part of services may be tailored to specific users or user needs. The third dimension of e-service refers to the degree of information and interaction in the service; whether there is interaction and transaction in the services. This dimension is close to seeing development of e-services as a stage model (Atkinson & Leigh, 2003; Asgarkhani, 2005).

In general, the development of electronic services tends to go through different stages, starting with presence on the internet and moving towards transactional services that make it possible to carry out activities right away (Atkinson & Leigh, 2003; Asgarkhani, 2005). The first stage has mostly to do with providing different kind of information to citizens. In the second stage there are often different forms and applications on the website for the user to download and fill. The most enhanced stage is called transactional services. They are result of services that are made transactional. For example, transactional service is when the citizen can fill-in an application for renewing drivers' license, send it and receive acknowledgement. It might also be possible for the citizen to later track the progress of the drivers' license renewal, for example.

## 4. Inter-organizational challenges

Usually development involves cooperation of several people. Especially in development of electronic services like one-stop government services there is a need to combine resources and expertise from different sources. This means that people from various functions, units and locations are brought together, and also outside expertise is needed. Hence, development can be seen as a partnership.

The definition of partnership ranges from working relationships to active transactions and collaboration between organizations. Here the term partnership arrangements include different types of joint ventures, subcontracting, alliances and acquisitions. In this paper the term partnerships refers to inter-organizational cooperation. In a partnership actors learn to know each other in the long run. Often relationships are relatively intensive and even personal. In business relationships competence and goodwill are needed for trust to develop (Blomqvist, 2002). The important issue here is that partnerships are based on commitment to cooperation between different actors.

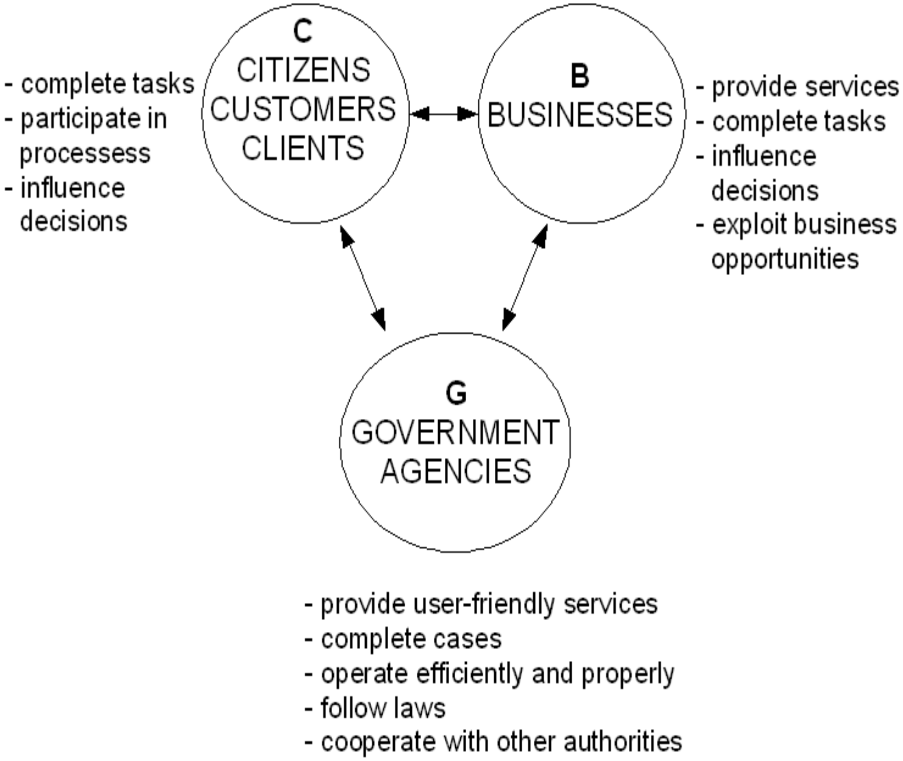


Fig. 2. Actors in public information systems (Sundgren, 2005)

Service providers can be organizations which are "actual" service providers, or they can be outside companies which develop, run or maintain the service in cooperation with the actual service provider. For example, the programming of the application could be done by a specialized software company while the actual service provider concentrates on what is the content of the service. Even though there would be several parties involved in developing and providing the service one should be the service provider who is responsible for the service. This service provider may set rules like minimum requirements or standards.

In development of e-services there are differences among organizations in terms of technical skills, organizational structure and in the attitude towards innovations. This means that not all organizations are ready for changes, and the pace of development differs among organizations (Ancarani, 2005). Thorough planning is therefore needed in development of more sophisticated e-service solutions; this is the case especially in public sector organizations.

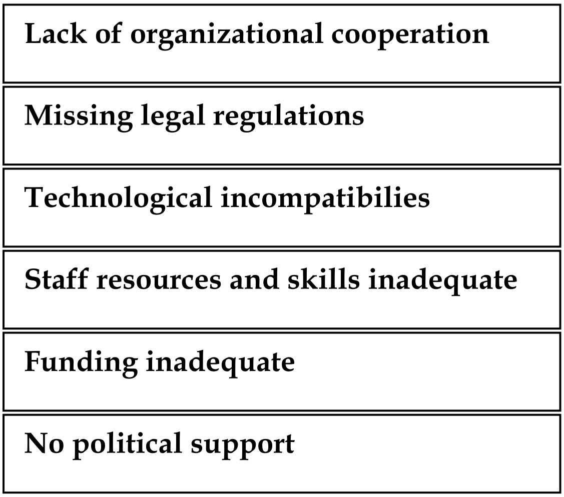

Lack of organizational cooperation
Missing legal regulations
Technological incompatibilies
Staff resources and skills inadequate
Funding inadequate
No political support

Fig. 3. Barriers in development of electronic services

Organizational cooperation is one factor that needs to be addressed in development of electronic services (Kubicek & Hagen, 2000). In public sector initiatives where services cross departmental boundaries this is a challenge. In most projects there are often external organizations, IT expertise and special skills that are needed. Cooperation of several partners, units and stakeholders can become a barrier for projects that involve several organizations. Development is often faced with the fact that stakeholders act too independently, because projects tend to be poorly coordinated (Irani, Love & Montazemi, 2007).

The lack of alignment between organizational goals is put forward as a major factor in the set of organizational and managerial challenges. Furthermore, the size of project and the diversity of users and different organizations involved make the development work more demanding. Dawes and Pardo (2002) also address the existence of multiple and partially conflicting goals in public sector projects involving several stakeholders. In inter-organizational projects there is a built-in delay that is result of inadequate organizational cooperation (Kubicek & Hagen, 2000). They (ibid.) identify six areas which cause failures and delays in development of electronic services. The first key area is lack of organizational cooperation. The second key area is missing legal regulations and the third is that necessary pre-conditions in regard to technology are not met. The fourth key area is human factors, skills and resources. The last barriers are result of inadequate funding and political support.

Gil-García and Pardo (2005) found that challenges to various e-government initiatives are cross disciplinary and may be grouped into five categories: (1) information and data, (2) IT, (3) organizational and managerial, (4) legal and regulatory, and (5) institutional and environmental. Information and data (first category) covers the capturing, management, use, dissemination, and sharing of information. In this category the developers also need to address data quality and data accuracy as well as dynamic, changing information needs. Information technology (second category) refers to issues like technological incompatibility and complexity, security, usability, technical skills and experience, and technological newness which all present challenges for development and use of services. Organizational and managerial issues (third category) are the main challenges to information systems development (Gil-García & Pardo, 2005). It is clear that laws and regulations must be taken into account when developing electronic services (fourth category). The institutional and environmental challenges (last category) are result of the institutional framework in which public organizations. The framework also includes the existing policy environment.

## 5. Focusing on services in public sector

There is an almost infinite potential in development of electronic services in the public sector. Typically, services that are provided through the internet are connected to sharing information. Public sector services are mostly connected to information – and internet is a very efficient way to gather and share it. We argue here that developers of public sector e-services should do more than they have done so far.

In public sector there are numerous electronic, information technology powered services already today. The citizen may use an arbitrary computer connected to the internet and apply for children's day-care, or inform the authorities that the address has changed, for example. These can be done by filling in a form on-line, or downloading it and printing for manual processing. Other typical e-services include seeking information from service providers' internet-pages or requesting further information and advice. It is still clear that many public services are in their early stages; often they are based on existing ways of doing things instead of thinking what citizens need (Howard, 2001). Services need to be integrated which calls for a total re-planning of services in order to better meet the needs of the citizens (Atkinson & Leigh, 2003).

The developers of e-Services need to better understand users of public services. Clearly, the citizens should not have to surf the internet and try to find different services that are spread all over. Better integrated, portal-type sites would make it possible to find relevant information effortlessly. This involves integration of services that are generated in separate offices, departments and units (Atkinson & Leigh, 2003). For example, too often agencies provide information only from their "own" services and activities. Instead, information should be widely available so that users would not have to guess or know what other related information and services there are so that users could better have their problems solved. Services should also include information, advice and links that are not provided by the agency itself. There is a need *"…to approach the Web with a philosophy of helping users solve problems, not merely delivering their same old services through new medium"* (Atkinson & Leigh, 2003).

It is often noticed that people send lot of email to public administration. This is because it is easier to ask than try to figure out what agency to contact and how to proceed. Public administration is full of administrative jargon and official pseudonyms – which are likely to be transferred to web when they have been digitized. The whole structure of the web-sites is based on different agencies, departments and units (stovepipe-structure) rather than integrated portals (Atkinson & Leigh, 2003). In addition, the sites are often relatively unfriendly and there are no comprehensive search-engines that would make it easier to find information from the site, for example. In this way poorly designed internet site can easily increase the burden of the staff in public sector organizations when the amount of incoming email queries go up. This should indicate that existing service through the internet needs to be developed. The solution to the problem is that the services should be more straightforward, easier to understand and include self-service –type of elements.

One suggested method for increasing the amount of self-service over the internet is giving rebates for those using e-services. In e-business it is common that customers who make the reservations over the internet, for example, receive a rebate or discount coupons that they can use when shopping again in the future. The goal is to develop lock-in, and push the customers to using services that are available on the internet. The customers can do e-shopping whenever it best suites them, they don't need to wait on phone, for example. At the same time self-service frees staff from answering customer calls to more productive work. The question here is that why could not public sector organizations use similar techniques in order to promote e-services and ”locking” citizens.

True e-services need to be developed around user needs. Transferring existing papers, files and information from different agencies into web, and placing some hyperlinks between them is not enough. The services should be integrated, enhance self-service and trust so that users see the added value of electronic services. Technologically, users should be able to complete most of their transactions online. Here easy-to-use, robust and trustworthy services are needed so that more users start using e-services in public sector.

## 6. Adoption of e-services

The successfulness of any service depends on whether users start using it or not. This is widely referred to as adoption. Previous research has found several factors that affect the formulation of attitudes and behavior to innovations and leads to their adoption (Rogers, 2003). Relative advantage, compatibility, complexity, triability, communicability and perceived risk are attributes that are important here, they outperform even other types of adoption predictors like situational variables and user characteristics. Rogers (2003) argues that innovation attributes have explained 49 to 87 per cent of the variance in the rate of adoption of various innovations. In predicting adoption the focus has been in adoption of products the factors may be applied to adoption of services which are delivered electronically (de Ruyter et al. 2001). From the array of attributes are perceived risk and relative advantage most relevant in explaining adoption of e-services.

The perceived risk is particularly applicable and important to services as the perceived risk tends to be considerably higher than in case of products. In internet the risk is high because the customers do not know whether the service providing e-business is “big or small, new or established, legitimate or illegitimate” (Hagel & Singer, 1999). It is still noteworthy that organizational reputation has a strong influence on trust, attitude and behavior (de Ruyter et al. 2001). In internet trust plays a critical role because - depending on the service - users have to release personal or financial data to the e-service provider. This takes in an environment where the user may have very little information on the service, the e-service provider and their trustworthiness.

While using internet for users will look for innovations that provide an advantage over current services and products. It becomes operationalized in functionalities and properties as “time-saving”, “range of options” and “ease of use” (de Ruyter et al., 2001). Also convenience is an important issue. For example, there is no need to go to a certain office at a given time, when there is an e-service for doing the task it can be accomplished whenever and wherever using a computer that can connect to the service through the internet. When the innovation provides relative advantage to the user it is seen as a trigger to use of the innovation in question. In e-business relative advantage means that companies should offer better and preferably unique services to the customer if they want to distinguish themselves from other businesses (Tambini, 1999). The e–service must offer benefits over existing services and ways of doing things for attracting potential users and gaining "critical mass" behind the service.

Relative advantage plays a role in formation of attitudes and behavior towards e-services. Also organizational reputation has an important influence on users. It has a strong positive impact on the customers trust in, quality perception of, and intention to use the e-service (de Ruyter et al., 2001). As a result, the reputation of the services becomes an impediment for successful e-services.

## 7. Discussion

We have looked at development of electronic services and the challenges that this involves. The focus has been on services in the public sector, and they have been mapped against commercial e-business services. In this context interplay of several units, functions and organizations is needed – especially if the provided services are sophisticated, and providing users one-stop government e-services.

Development of electronic services - or information technology in general - requires connecting technologies and applications in order to provide solutions for users. There is a variety of underlying information infrastructures, applications and services that may be owned, maintained or developed by organizations from private or public sector (Ancarani, 2005; Sundgren, 2005). Similarly, development of e-Services is a combination of expertise and effort from people in the organization and from external environment.

The types of electronic services vary greatly in public sector. It is natural to expect that services are integrated into processes and information systems of the organization that provides them. However, in public sector organizations it is common that departments and units provide services to citizens rather independently. Departments have different processes and information systems which are not connected. In many cases information is stored in separate databases. This may be enough when services are oriented to information delivery between the public administration and the citizens. For example, providing downloadable documents and forms is simply offering documents in electronic format and making them accessible through the internet.

Over time more services are developed, more features are added to existing services, and more enhanced, transactional services are developed (de Ruyter et al., 2001; Atkinson & Leigh, 2003; Asgarkhani, 2005). This is challenging as when services become more sophisticated the overall complexity increases. It has been noted that moving to services that are transactional is a big step (Howard, 2001). Transactional services require connectivity, information in other systems and data-bases needs to be accessed, combined and updated from users' interface through the web. This is challenging from the information systems viewpoint as the situation calls for connecting originally separate systems which may be based on different software and database structures.

Cooperation is a challenge for management of the development of electronic services. It is not uncommon that managers find themselves making decisions about technology for which they are unprepared or even ill-equipped (Gil-García & Pardo, 2005). Successful development calls also for top management commitment, linkage to business, technical alignment, knowledgeable personnel and involvement of users (Pardo & Ho, 2004).

It is very important to look at services from user's perspective – whether they are connected to e-business or public services. As long as there are citizens that do not use electronic services organizations must to provide same service electronically and as a traditional service – the result is increased costs instead of cost savings. If the service is based on existing departments, administrative procedures and processes it may not be able to provide added value to the user. There is a need to do things differently, cross boundaries and redesign processes when designing e-Services. Hence, the work of developing and rebuilding government for the digital age is just beginning (Atkinson & Leigh, 2003).

The success of electronic services depends on whether users – digital citizens - find them valuable and start using them. In e-business solutions it has been found that sites need to be both easy to use and add value to the user, these are key attributes that increase the use of services (Igbaria et al. 1995; Lee & Turban, 2001; Lim et al. 2008). The added value lies in properties as "time-saving", "range of options" and "ease of use" (de Ruyter et al., 2001). The web sites should also provide enjoyable experiences, these kind of sites will probably be visited also in the future (Shang et al., 2005).

## 8. References

Atkinson, R.D. & Leigh, A. (2003). Customer-oriented E-Government: Can We Ever Get There? *Journal of Political Marketing*, 2 (¾), pp. 159-181.

Ancarani, A. (2005). Towards quality e-service in the public sector: The evolution of web sites in the local public service sector. *Managing Service Quality*, Vol. 15, No. 1, pp. 6-23.

Asgarkhani, M. (2005). The Effectiveness of e-Service in Local Government: A Case Study, *The Electronic Journal of e-Government*, Vol.3, No 4, pp.157-166.

Blomqvist, K. (2002). *Partnering in the Dynamic Environment: The Role of Trust in Asymmetric Technology Partnership Formation*. Acta Universitatis Lappeenrantaensis 122.

Cullen, R., 2001. Addressing the digital divide. *Online Information Review*, Vol. 25, No. 5, pp.311-320.

Dawes, S.S. & Pardo, T.A. (2002). Building collaborative digital government systems, In: McIver, W.J. & Elmagarmid, A.K. (Eds.) *Advances in digital government. Technology, human factors, and policy.* Kluwer Academic Publishers, Norwell, MA.

Dutta, A. & Roy, R. (2003). Anticipating Internet diffusion. *Communications of the ACM*, (2), pp.66-71.

Gil-García, J.R. & Pardo, T.A. (2005). E-government success factors: Mapping practical tools to theoretical foundations. *Government Information Quarterly* (22): pp. 187-216.

Goldkuhl, G. & Persson, A. (2006). "From e-ladder to e-diamond – re-conceptualising models for public e-services", in *Proceedings of the 14th European Conference on Information Systems (ECIS2006)*, Göteborg.

Gouscos, D., Lambrou, M., Mentzas, G. & Georgiadis, P. (2003). A Methodological Approach for Defining One-Stop e-Government Service Offerings. In: Proceedings of Electronic Government, Second International Conference: pp. 173-176.

Hagel, J. & Singer, M. (1999). *Net Worth: Shaping Markets When Customers make the Rules*. Harvard Business School Press, Boston, MA.

Hasan, H. & Tibbits, H.R. (2000). Strategic management of electronic commerce: an adaptation of the balanced scorecard, *Internet research: Electronic Networking Applications and Policy*, Vol. 10, No. 5, pp.439-450.

Howard, M. (2001). e-Government across the Globe: how Will "e" Change Government? *Government Finance Review*, August.

Igbaria, M., Iivari, J. & Maragahh, H. (1995). Why do individuals use computer technology? *Information and Management*, 29(5), pp. 227-238.

Irani, Z., Love, P.E.D. & Montazemi, A. (2007). E-government: past, present and future. *European Journal of Information Systems.* 16, pp. 103-105.

Kubicek, H. & Hagen, M. (2000). One Stop Government in Europe: An Overview. In: Hagen, M., Kubicek, H. (Eds. 2000). *One Stop Government in Europe. Results from 11 National Surveys.* University of Bremen, Bremen 2000: 1- 36.

Lee, M.K. & Turban, E. (2001). A Trust model of consumer Internet shopping. *International Journal of Electronic Commerce,* 6(1), pp. 75-91.

Lim, K-S., Lim, J-S & Heinrichs, J. H. (2008). Testing an Integrated Model of E-Shopping Web Site Usage. *Journal of Internet Commerce*, Vol 7(3), pp. 291-312.

McIvor, R., Mchugh, M. & Cadden, C. (2002). Internet technologies supporting transparency in the public sector. *The International Journal of Public Sector Management*, Vol. 15, No. 3, pp.170-187.

Mulpuru, S. (2007). Topic overview: US online retail. *Forrester Research*. Available from http://www.forrester.com/go?docid=41752

Parhizgari, A.M. & Gilbert, G.R. (2004). Measures of organizational effectiveness: private and public sector performance. *Omega*, 32(3).

Rogers, E.M. (2003). *Diffusion of Innovations*. 5th Edition. Simon & Schuster.

Rust, R.T., Kannan, P.K. & Peng, N. (2002). The Customer Economics of Internet Privacy. *Journal of the Academy of Marketing Science*, Vol 30, No 4, pp.455-464.

Rust, R.T. & Kannan, P.K. (2003). E-service: A New Paradigm for Business in the Electronic Environment, *Communications of the ACM*, Vol 43, No. 6, pp.37-42.

de Ruyter, K., Wetzels, M. & Kleijnen, M. (2001). Customer adoption of e-service: an experimental study. *International Journal of Service Industry Management*, Vol 12, No 2, pp.184-207.

Shang, R.A., Chen, Y.C. & Shen, L. (2005). Extrinsic versus intrinsic motivations for consumers to shop on-line. *Information and Management*, 42(3), pp. 401-413.

Sundgren, B. (2005). What is a Public Information System? *International Journal of Public Information Systems*, Vol 2005:1, pp.81-99.

Tambini, A.M. (1999). E-shoppers demand e-service. *Discount Store News*, Vol 11, No 28.

Van Der Wal, Z., Huberts, L. Van Den Heuvel, H. & Kolthoff, E. (2006). Central Values of Government and Business: Differences, Similarities and Conflicts. *Public Administration Quarterly*, 30(3).

# The Requirements for the Legal Regulation of Commercial Relations in Cloud Computing

Ivan Pogarcic[1], Marko Pogarcic[2] and Matej Pogarcic[3]
[1]*Polytechnic of Rijeka,*
[2]*Faculty of Law, University of Rijeka,*
[3]*Faculty of Civil and Geodetic, Engineering, University of Ljubljana,*
[1,2]*Croatia*
[3]*Slovenia*

## 1. Introduction: Users and IT (His highness the user and his court jester)

Definitions of Information System are numerous, as well are the seriously written textbooks that consider that subject. In most cases a pragmatic side of Information System is being emphasized. From the pragmatic aspect, Information System is connected to its users since it is made to match their needs. Nevertheless, a definition rarely explicitly addresses user, tough he is indirectly implied. If Information Systems are defined through users' relationships in usage of technical and technological systems, one usually applies to ICT or Information-Communication Technologies. (Kroenke, 2008) It is essential, out of numerous reasons, to differentiate Information Systems from computer solutions and backup to Information Systems and associated business systems. (O'Brien, 2003). Regardless to definition, business system and Information System are determined by user and his needs. Business processes realize those needs and materialize certain benefits, while effective performance of business processes requires timely information. Simultaneously, business activities are followed by routines and repetitions in usage and acquisition of needed information. Computer is an instrument that man releases primarily from wearisome activities that can be figured in automatic sequence of computer orders or complex calculations that would otherwise require much more time. Development of computer sciences has been followed by a constant need for Information Science education of all users. However, velocity of development of technical instruments and technological applicative solutions has frequently increased to higher levels than the educational level of users. Historically observed, Information Science hasn't necessarily separated term user into two basic groups: "material" or real user who applied services of computer applicative solutions and user who applied potentials of technique and technology in order to provide essential benefits and possibilities to a previous group. The latter group is made of experts and Information Scientists. ICT development has blurred border between these two groups. Out of user-user perspective, as time passed the Information Science education moved closer user-user to user-Information Scientist. Their relationship has during that time period been specific to service industry, respectively service, provider and user occurred. Information Sciences have therefore, been wrongly placed within "service industry sciences".

Term "service science" should in this paper be understood as working name, since, naturally classification of science doesn't recognize that term. Why working name? Information science, formed as science, can be considered as relatively young science with strong trend of development and formation. If information systems are put in the centre of information science, that is, on a position of subject of prevailing research, then even the former claim can be put in question. However, the area of information systems is relatively scientifically unfinished, according to its definition of establishment. Incompleteness is obvious, especially in the area of precise definitions of complex terms and frequent identifying with different terms such as: information systems, management information systems, information management, business information systems, business information technologies and alike. Naturally, each of the mentioned terms has indirectly a pragmatic side of definition within. For instance, (Rose, 2000) information systems can be observed as field of computers and communications, but within social context, so they can be placed in class of social and humanistic sciences. However, if area is anticipated as technical, then it is close to natural sciences. In both cases, development of information systems should be connected to humanistic sciences, such as sociology, psychology, but even the newer ones, as telematics or cognitonics. Naturally, this sort of thinking was a subject to criticism. For instance, (Chekland and Holwell, 1998) this area can be classified as "crucial, but confusing", (Banwille and Landry, 1992) as "fragmented adhocracy", or (Whitley, 1984) as area that "produces diffusive, discursive knowledge of commonsense objects". Most critics (King, 1993) of this area's status as scientific entirety usually claim that it misses "intellectual basis", that is actually, borrowed by other "referential field of sciences".

The fact is processes of information systems have numerous ad hoc situations, so the attitude of "adhocracy" of this area can be accepted and characterised if wanted to be projected as level. Appropriate is missing or low level of obstruction to entering the area, weak coordination of research activities, unidentified tasks as reason of "floating" reputation, lack of standards or possibility of their change, even by unprofessional persons, and finally, confusingly defined terminology with many wrong synonyms based more on commonsense then precise definition. Since this area has neither completely, nor precisely, defined subject of research, it should be considered, at least temporarily, independently from other sciences. Still, using it doesn't imply one-way communication with other sciences. Involvement of computers as technique and programme solutions derived by same platforms is unquestionable in all sciences, so it is manifested as service information science.

Each scientific area should have "intellectual basis" to provide ground for defining scientific settings of area, theories, methodologies, research modes, and finally, control of all mentioned through practical realisations and concrete situations.

Precisely in that relationship between science and theory, according to practice, lies one of the basic reasons of unidentified area of information systems. Accelerated development and improvement of technical basis – hardware, demanded an answer in technology and methodology of development. When user is put in the centre of situation, it becomes complicated, but consequences are considerably clearer. Usage of information technologies demands knowledge, skills and competitions by users. It can be achieved by appropriate education and training of users. Traditional systems of education are inert, and so are modifications and adjustments of education. It is controversial that education is one of the latest areas that have introduced computers. That way, user has, according to his level of education, been late in area of development of hardware and software. Still, the complete technique and technology have been created in order to easier and more qualitatively satisfy users' needs. Pragmatic side of this relationship is realised through implementation of applicative solutions in users' environment. However, pragmatism is accompanied by relationships applicable in commerce, since all applicative solution is considered to be a product that requires promotion, marketing and treatment of users, usually as buyers of these products. To satisfy customers'-users' needs within such circumstances implies to make a product more simple, easier and closer to user.

Through period of information sciences' development, approaches and generally paradigms of approaching the users have been frequently modified. Basically from monolithic to applicative solutions, specialised only to certain task so to free user from concern and engagements that additionally burden him through grid and cluster architecture of information systems of paradigms that have been changed from recent form known as cloud computing.

If simplification is understood as certain form of abstracting the volume of user's tasks, then the process hides other traps that can become brakes to performing business. The truth is that relationship between users, architecture and structure of information system, that is professional personnel, can be redefined so to maximally free user of tasks that belongs to information scientists. At the same time, the more precise raster of user's needs has been developed. The usual applicative coverage of users' daily needs has been broadened by covering more demanding exceptions that can emerge among those needs.

Cloud computing has supplemented paradigm that accompanied object oriented approach with paradigm of weak binding or binding software components according to task's needs.

Legal aspect of such relationship has been, during the history of Information Sciences, frequently neglected or observed to belong to someone else who regulates such relationships. Today, when most of human needs have been, one way or another, supported by computers, relationship between users and service providers necessarily demands legal regulation in combination with legal regulation of computer products' usage.

## 2. Overview: From the monolith to the clouds (Le roi s'amuse)

Meaningful usage of computers requires certain programme backup that considers existence of data. Data is being processed according to these programmes respectively algorithms applied in the programmes. Besides programme solutions one requires appropriate communication between user and computer. Realisation of monolith application assumes that programmes and data are physically at the same place – computer, accompanied by user. Although this shape of application is considered to be a beginning of computer backup, the monolith applications are not purely historic since even today specific systems can request such a solution. Specific system has specific users so their position is bonded to special characteristics of system's functions. Legal regulations in such environment are strongly defined since those needs are monolithically connected to such relationship.

Usually, though wrongly, monolithic organisation of application is considered a historical form of organising the information systems (fig 1). Though the majority of monolithic organised systems are historically basic form, that form of organisation exists even today, if system and user are organised in that manner, out of any reason possible. Monolithic assumes coexistence of systems and users in same environment, without more significant

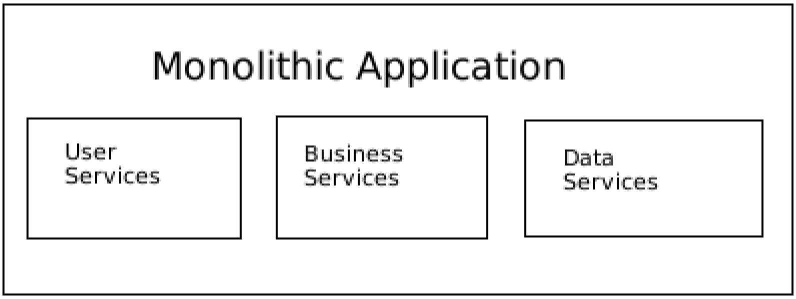


Fig. 1. Monolithic Application. Source: //rubyonrailslink.blogspot.com/2010/09/single-tier-architecture.html

need to communicate with environment. Regulation of rights and obligations in monolithic systems is strictly defined by organisational rules of system to which such information system belongs.

(Batini&Scannapieca, 2006) define area where all information systems are located according to their structure and architecture (fig. 2). Area is defined by coordinates: totally, heterogeneity and distribution. Considering these three characteristics according to measurable intensity, monolithic systems are located in the centre of coordinate system. That is, they are classified as homogenous, not distributive and specialised. Authors have located P2P information systems in a diagonal spot. Systems of type peer to peer are declared as autonomous and independent from computer providers. Such systems are different than usual client-provider architecture. Usually possibilities of all network-connected computers are being used. Scalability is the strongest characteristic of such systems. When user is logged on a network, the complete capacity of a system grows.

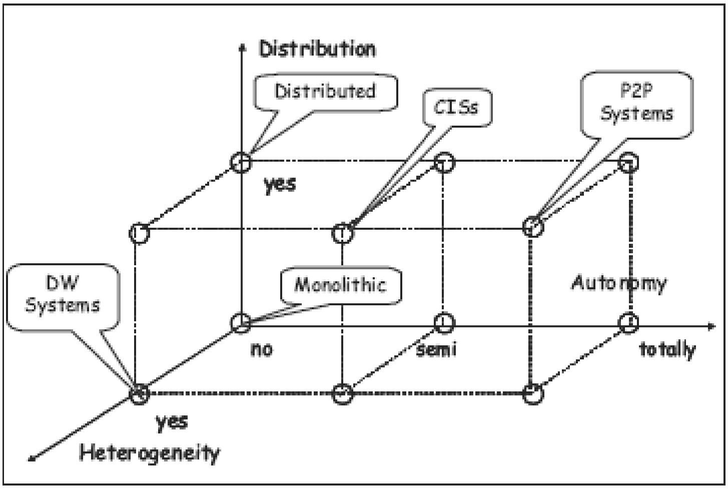


Fig. 2. Type of Information Systems (Batini&Scannapieco, 2006)

Importance of data and need to manage data and to preserve them has led to double and triple-layer organisation of applicative solutions. Indirectly it led to layering a category of users, information scientists and narrower specialization of certain cadres. Specialization of cadres has at the same time meant a diversification of rights derived from obligations given to them. Final or key user of system has in this architecture been extracted, while his approach has been enabled through defined interface. Responsibility for data base has been shifted to skilled and prepared workers, usually information science experts residential to that specific business system. Legal aspects of information system in these architectures have been solved as accompanying problems at level of managerial structures of business systems. Separation of data bases into a special tier has demanded regulation of security mechanisms that necessarily included legal regulation and protection of same. Layering of systems into tiers has more precisely articulated rights that belong to users and information science experts.

Separation of data bases into special tier put users' obligations towards data into frames of concern for accuracy and correctness while organisational and security aspects have resumed in authorisation of IT department. Necessity to manage data and to preserve them led to double-layer and triple-layer organisation of applicative solutions (fig. 3). Indirectly, layering has been made between class of users-information scientists and narrower specialisation of certain cadre. This happened regardless to saving data in a simple file or organising them as more complex form – data basis. When possibility of physical dislocation of users beyond business environment emerged, a need for relationship's organisation known as client-server organisation also emerged. As consequence a need for specialised users who will maintain dataware appeared in line with users who will concern computer networks or working stations – personal computers used by clients. Narrower specialisation of Information Scientists together with increase of consummation possibilities of system led to a need for adjusting system to final user. Graphic interface as solution to communication demands brings more possibilities to user but it imposes need for introduction of the third layer (known as middle tier) which includes logics of applicative solutions. Still, the relationship service user – Information Scientist service provider causes a legal regulation to be solved in general level, usually managerial, by leaving it to personnel outside such a relationship. If the relationships in a prior mentioned architecture of information systems are analysed from time distance, one fact is indisputable and unavoidable. That is relatively low information science education of final users. It also indirectly puts user into inferior position and leaves possibility to information scientists, usually organised in IT department of business system, a freedom that realistically shouldn't be given to them since it can be misinterpreted.

Though only few scientists have concretely analysed that problem, the author can, upon his own experience, confirm frequent conflict between users and information scientists. Sometimes and somewhere the supremacy of information scientist has led to problems in implementation of solutions and left an impression of compulsion by users, with complexes of ignorance and inadequacy to problems. Users' response to such situation was necessary and it was expressed through need for information education and a simpler usage and managing of computer products. Simplification suited to both groups, so it has led to innovation and improvement within area of hardware and software. That trend led to multi-tier architecture and consequently to stronger diversification of information science cadre. (Fig. 4) Data-tier is divided into data tier and data-access-tier. Necessity of implementing the business-tier more intensively includes final user into organisation, though his inclusion into information science tasks remains at level of consummation.

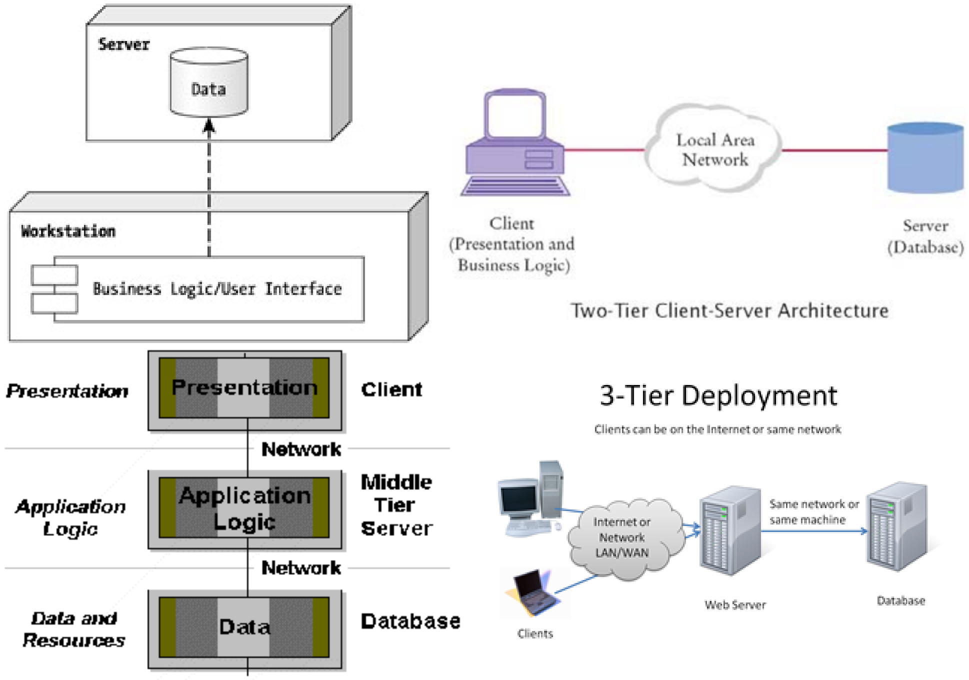
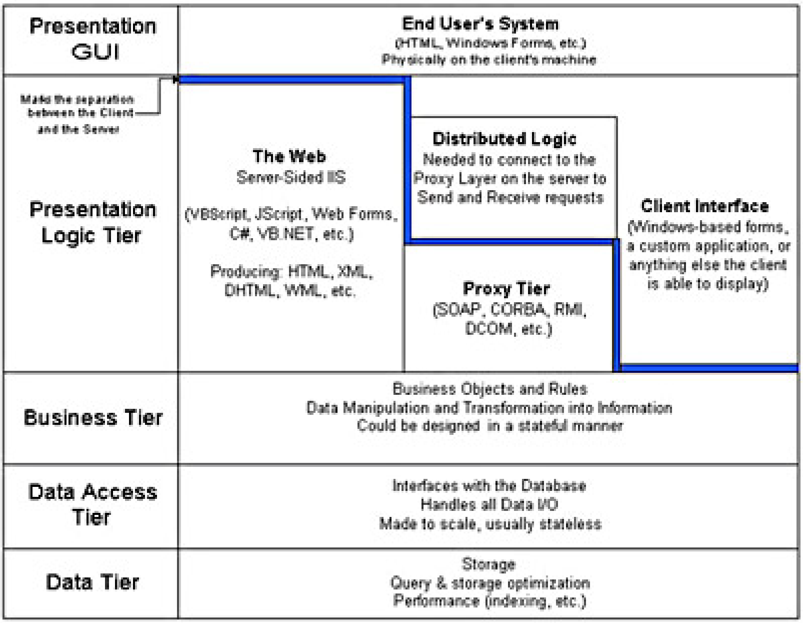


Fig. 3. Two and three tier Architecture (Source: http://www.iro.umontreal.ca/~pift1025/bigjava/Ch27/ch27.html, http://flylib.com/books/en/2.642.1.11/1/)



| | | | |
|---|---|---|---|
| Presentation GUI | **End User's System**<br>(HTML, Windows Forms, etc.)<br>Physically on the client's machine | | |
| Made the separation between the Client and the Server<br>Presentation Logic Tier | **The Web**<br>Server-Sided IIS<br>(VBScript, JScript, Web Forms, C#, VB.NET, etc.)<br>Producing: HTML, XML, DHTML, WML, etc. | **Distributed Logic**<br>Needed to connect to the Proxy Layer on the server to Send and Receive requests<br>**Proxy Tier**<br>(SOAP, CORBA, RMI, DCOM, etc.) | **Client Interface**<br>(Windows-based forms, a custom application, or anything else the client is able to display) |
| Business Tier | Business Objects and Rules<br>Data Manipulation and Transformation into Information<br>Could be designed in a stateful manner | | |
| Data Access Tier | Interfaces with the Database<br>Handles all Data I/O<br>Made to scale, usually stateless | | |
| Data Tier | Storage<br>Query & storage optimization<br>Performance (indexing, etc.) | | |

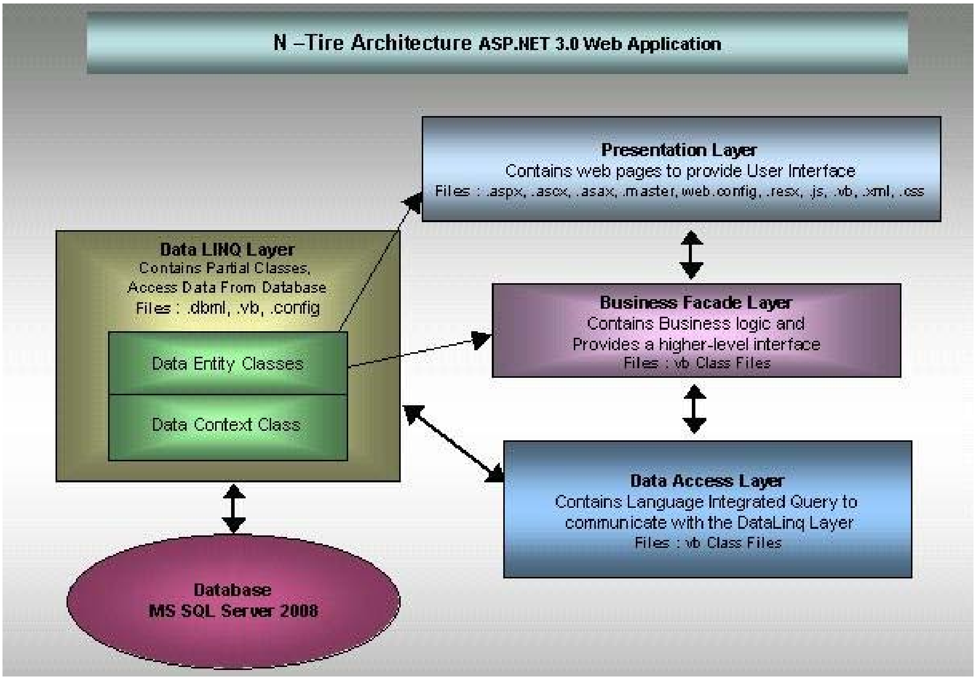

N –Tire Architecture ASP.NET 3.0 Web Application
Presentation Layer
Contains web pages to provide User Interface
Files : .aspx, .ascx, .asax, .master, web.config, .resx, .js, .vb, .xml, .css
Data LINQ Layer
Contains Partial Classes,
Access Data From Database
Files : .dbml, .vb, .config
Data Entity Classes
Data Context Class
Business Facade Layer
Contains Business logic and
Provides a higher-level interface
Files : vb Class Files
Data Access Layer
Contains Language Integrated Query to
communicate with the DataLinq Layer
Files : vb Class Files
Database
MS SQL Server 2008

Fig. 4. Multi tier Architecture (Source: http://www.15seconds.com/issue/011023.htm and http://www.codeproject.com/KB/aspnet/NTierApp_UsingLINQ.aspx)

Simultaneously, Information Science education of system's users has increased while his demands are becoming broader, bigger and frequenter. Technical tools are becoming more sophisticated as technological solutions. Technological solutions as product undergo through all rights and regulations that accompany any other market product.

## 3. Elaboration: Paradigm shift (Mixing cards with the war for middle-earth)

Necessity for separating computer solutions into layers or multi tier organisation has resulted in more detailed diversification of activities in their development. Changes in approach and consideration of the complete issue moved barycentre from structural system towards system's architecture as a new centre of gravity. Modifications in paradigms are most obvious in the approach to projecting and programming. Object oriented software is the cause and consequence of changes that will result in emergence of opened platforms and multiply usable software. In order to satisfy increased appetites of users, a proclaimed paradigm of inheritance within object oriented approach has provided multiple uses.

However, object oriented approach has also caused a "rearrangement" of role and position of user by regulating approaches to software through mechanisms of encapsulation. That way the system's structure can be observed as a method of its construction or compliance and combination of its parts. Possibility of organising the Information System within conditions of detachment of business system refers to organisation of computer network of different possibilities and purposes. Along with significant improvement and upraise of technical possibilities, computers evidently, by virtue of increasing memory capacities and speed of processing, should be differently organised. The centre of gravity has been moved from structure to architecture. Architecture of computer systems includes structure's moderation together with organisation of technical backbone and applicative solutions so the architecture is both conceptually and realistically broader than structure.

Object-oriented paradigms have primarily, through proclamation of late binding, initiated reorganisation of programming, though object-oriented approach gives advantage to projecting instead of programming. That is understandable, since programme demands are biggest source of misunderstanding among information scientists. When object-oriented approach has emerged and developed, the final users were already well educated in field of information science so their demands represented greater problems to information scientists. Though at the time prices of hardware were dropping and opened platforms of operative systems were intensively used, to each business system investing into information science represents an issue that requires ultimate attention. At that time Sollow paradox became famous: "total income of business system decreases as investments in information science become higher". That should initiate considerations of real usage and need of IT sector as organised fragment of business system. It will also bring IT employees into a considerably different position than the one their forerunners have had.

If historical aspect of information scientists – experts is considered, than, with a considerable attention, conclusion can be made that their position has frequently weakened. From position of unavoidable expert-wizard who was highly respectable his position changed into "someone who is an expert in his field", but doesn't necessarily get paid for his work. Evaluation and assessment of full business contribution of "internal" information scientists has started.

Each individual business process or use care, as it is commonly named, is maybe specific but not to the point where business is made strictly from that phenomena. Case of usage is more a category where concrete realisation of same cases can be placed. Repeating the same cases is measure of need for programming and considering specificities that determine exceptions in same realisation. This attitude is prevailing in initiating, and later on, intensifying intention for outsourcing the IT sector. This helped to pronounce information-science backup as service within business system. Only time will tell how correct that is. It is important to mention that this declaration has a certain delusion within, that is, it makes information-science backup as sporadic need which is absolutely not the case. However, that kind of attitude towards information science has been inherent continuously from the first formation of IT sector within business system up to the moment of excluding IT management from process of making important business decisions. It is especially important that decisions made upon level of business system include CEO of information science, which is often not the case.

The prior paragraph has been given an allegoric subtitle "king is amused" as a result of author's experienced thinking about relation user-information scientist. All misunderstandings upon this line are consequence of dissatisfaction of one or other group of participants of the mutual act. Attempt of upgrading, of any side of participants, will lead to psychological consequences awkward to relationships and business itself.

In an attractive blog title "10 dirty secrets in IT business" Jason Hiner from TechRepublic site mentions a set of "secrets" that confirm the above mentioned. http:// www. techrepublic.com/blog/hiner/10-dirty-little-secrets-you-should-know-about-working-in-it/546. The fact is information science experts are making themselves a huge favour. One shouldn't engage into a serious business to conclude how usage of slang among information scientists can be frustrating to users so to make them think that something is being covered up. Reason can also be found in usage of technology that will help IT to firm its position, but not with actual help to business making. Next, the older IT experts will regularly be averse to introduction of new technologies. All situations mentioned are not amusing to users, especially since user is directly responsible for realisation of business processes and existence of business system. It is normal that detronization of information science in this case becomes inevitable. Psychological state of mind within these frames defines level of satisfaction with users and information scientists. Unsatisfied user will in each situation re-evaluate his obligations and the ones of information scientists as service provider.

Further on, every evaluation of obligations causes re-evaluation of rights that belong to them. In business systems it will regularly lead to, more or less, conflict situations and disturbed personal relationships. The final result is destructive to making a business. When business system has its own IT sector, these conflict situations have internal character so they are internally solved. Legal aspects of those misunderstandings are within domain of specialised business system that is mostly connected to HR sector.

Separating users from information-science assignments provides contribution to both sides under condition that information-science backup is qualitative and timely. At the same time, assignments are delegated more precisely thus making obligations more understandable and better defined. It also expresses better rights of users and information scientists. Object-oriented approach has enabled these relationships and caused further changes of paradigms that will lead to SOA paradigm (fig. 5).

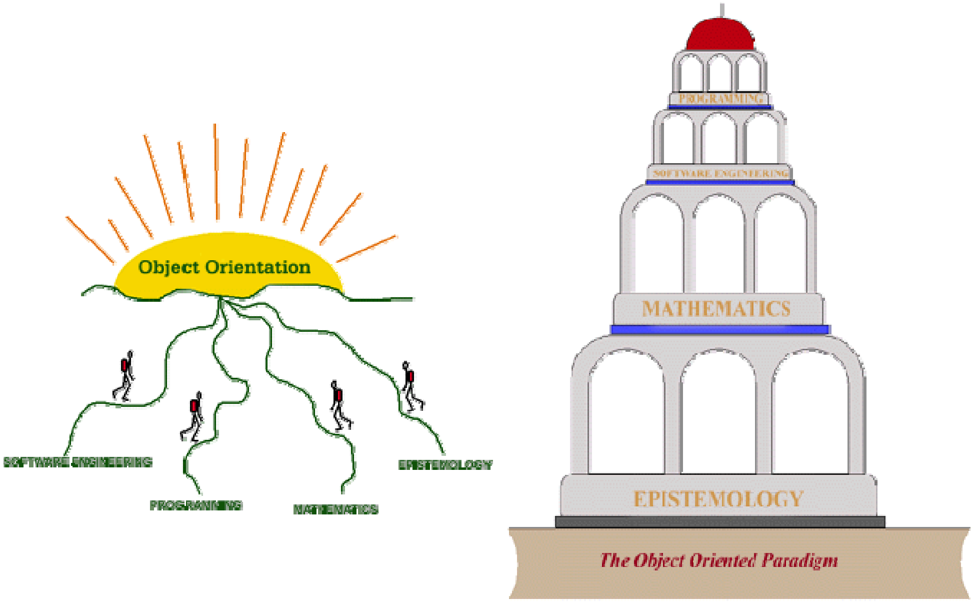


Fig. 5. OO Paradigm (Source: www.hl7labs.gr/pages/ Edsger/Edsger/tower.gif and www.hl7labs.gr/ pages/Edsger/Edsger/oop.htm)

When discussed about architecture, SOA (Service Oriented Architecture) can be observed as politics, practice and frame that enable application of functionalities which have to be provided through collective services. (Liebow, 2005) Services are being offered to the applicants by virtue of standardised interface. (CORE.gov, 2005) In SOA environment, term user has been replaced with benefit user who shares services. Service applicant is user of services provided by system but it is indirectly underlined that user is also driver of activities used for realisation of services. SOA also more pronouncedly affirms paradigm postulates of object oriented approach especially of so-called late binding resources. This turns late binding into weak binding in SOA environment while final operative form of application becomes a dynamic category active only in time of performance. It helps to achieve full effect of application that can be acquitted through optimisation and proper monitoring of costs. IBM (Balzer, 2004) defines principles of basic regulations for development, maintenance and usage of SOA architecture (fig. 6). These are:

- Possibility of multiple usage of programme solutions (reusability)
- Granularity
- Modularity
- Possible usage of composability (a system design principle that deals with the inter-relationships of components)
- Ability to decompose in components
- Interoperability
- Standard-compliance, common and for specific kind of industry
- Possibility of services identification and categorization
- Possibility of ordering, provisioning and delivery
- Monitoring and tracking

SOA architecture is made of functional elements and elements that provide system's quality, and are connected to policy of launching services to market.

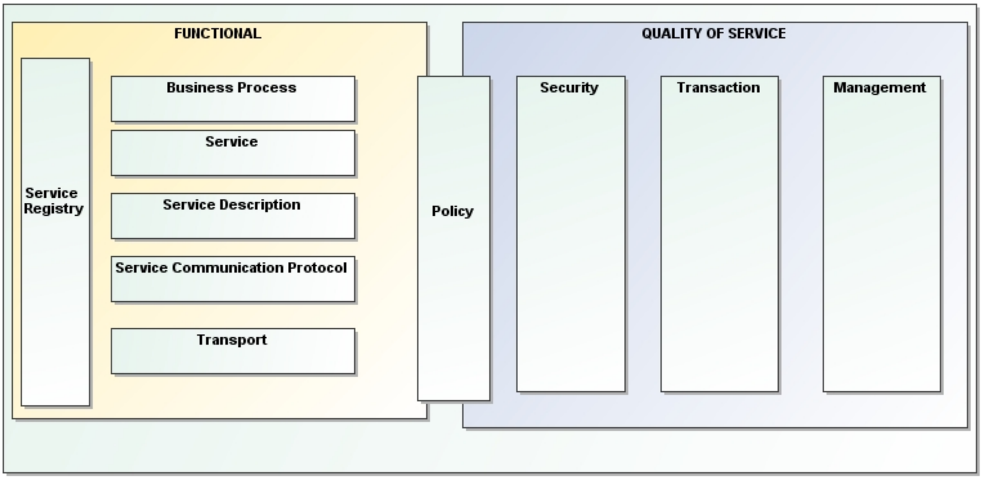


Fig. 6. SOA Architecture (Source: http://www.mondotechnologies.com/en /index.asp ?w=0|0|1)

Functional elements make service register with description of services, business processes, defined communication protocol, defined services and mode of delivery. Elements of insuring the quality are: safety of services, realised transaction and service management (fig. 7). Still, the quality of service and its functionality is ensured by politics that integrate these two parts and present them to user who will eventually benefit from it. Politics of realisation of services can be observed as binding material that connects constructing elements of SOA architecture SOA architecture should provide several attributes to business systems: to propose simple, flexible and efficient system that will provide needed information services with lower costs of providing services and insuring their integration and migration. When settings of SOA architecture are considered, one can assume further development in application of object-oriented paradigms and more expressed implementation of information science architecture. If one analyses terminology, he can notice that some classical terms are replaced with newer.

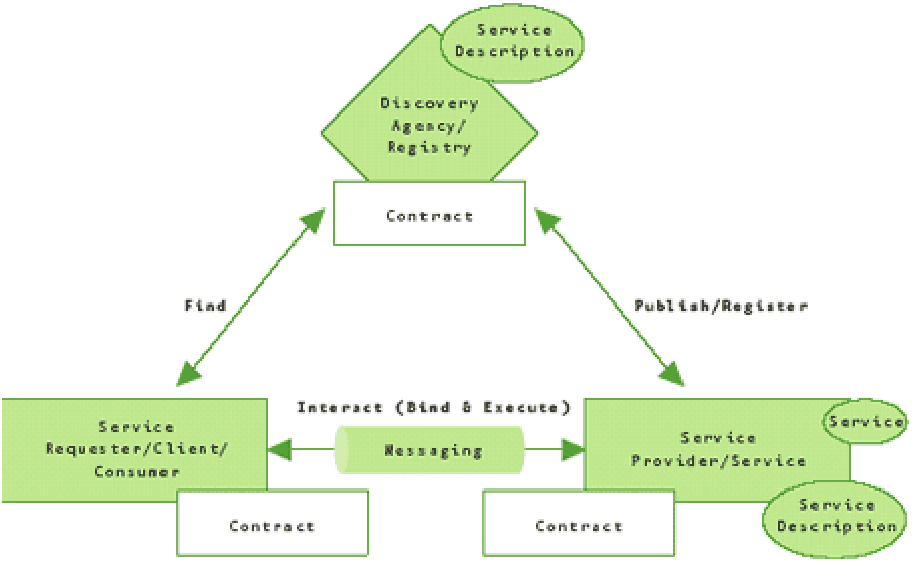


Fig. 7. Contracts in SOA Architecture (Source:http://msdn.microsoft.com/en-us/library/aa480027.aspx#aj2mpsoarch_topic2)

So that user can frequently be called consumer or service client while service creators are named service providers. If consistence in application (and around it) is insisted on, the one that is named information system, then this could lead to semantic mess. User can be asked who is actually a service provider: the one who interferes or the one who actually provides the same services. This kind of relationship makes a good basis for paradigm shift that will end up in SOA evolution to cloud computing. So to conclude, precise definition of terms is necessary. Whoever experienced functioning of SOA architecture or web services (though these are two different things), experienced all advantages of such realisation of information system, but also all disadvantages that can emerge when using services through such systems (fig. 8).

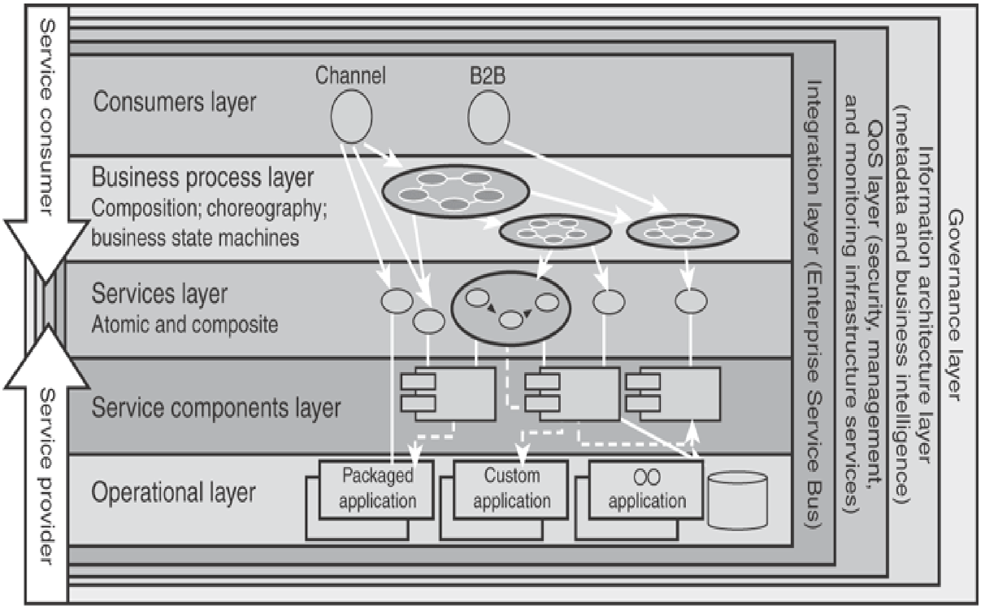

Service consumer
Service provider
Consumers layer
Channel
B2B
Business process layer
Composition; choreography; business state machines
Services layer
Atomic and composite
Service components layer
Operational layer
Packaged application
Custom application
OO application
Integration layer (Enterprise Service Bus)
QoS layer (security, management, and monitoring infrastructure services)
Information architecture layer (metadata and business intelligence)
Governance layer

Fig. 8. Logical view of SOA Reference Architecture (Bieberstein at all, 2008)

Though it may be thought that users-information scientists and active users within business systems are put aside the truth is different. Separation of applicative solutions to its finer parts – services – that are activated accordingly can represent a nightmare for Information Scientists, especially when on the other side are users who cannot articulate their demand in mutually understandable manner. Segment of service providing policies within SOA architecture that can solve this kind of misunderstandings in practice is known as help desk institution. Though as an idea it can be accepted, good realisation of help desk activities is rare in practice. User of help desk regularly receives incomplete instructions that don't help him to solve problem. If help desk is analysed in context of developing relationship user – information scientists and development and upgrade of ICT, one can acquire extremely unpleasant impression. Help desk eliminates not only physical but also virtual communication between final user of application and its professional creator-author. Help desk employee have usually gone through training that should correspond to user with frequent anticipation that they do not understand essence of processes that are information-science supported. User's need for help desk is a sign that his daily business functioning is brought into question through applicative non-functionality.

If help desk employee isn't able to provide adequate help and to eliminate problem that user declared, a situation could turn to conflict. Picture (Fig. 8) indicates that all architecture relies on precise and well defined contracts between actors of providing and consuming the services. Contract is specification of method by which consumer or user of service will communicate to service provider. Contract defines format of demands and mode of answers that are expected through services. Contract of services can demand various conditions and pre-conditions. Pre- and post-conditions describe how service should be performed in order to satisfy specific function. Contract should define quality of service and specify non-functional aspects of services.

It is clear here that realisation of recent forms of information systems should involve more intensively legal profession. Contract is legal category so it should be confined to those who are professional in that field. Exactly those kinds of situations and necessary facts are to be underlined in this paper.

Required legal protections of both users' categories which appear in such environment is essentially different since it returns relationships to a level that existed in time when real users were supposed to have lower level of Information Science knowledge and congruent capability of using the application. At the same time emerged a possibility of so-called Outsourcing or extracting the IT sector from business system as an activity that is not core business activity. It is not a coincidence that simultaneously a need and possibility for outsourcing emerges, that is excluding IT sector from business system as activity which is so-called core business activity.

## 4. Interlude: We are in the cloud; we need some barrister (Hypothetic use case)

Business logic and applicative tier have already, in the position of organising system according to three tiers or multi tier structure, been categorised as middle tier that is "inter-level tier". That also led to excluding the responsibility for contents belonging to specific tier from those to whom it doesn't actually belong. If Fig. 9. is analysed and completed with human component, then on the left side, one can perceive only users, that is, service consumers according to SOA terminology. Suggesting the Cloud computing isn't relevant at the moment. More important are possibilities provided by Internet. Naturally, that is possible if Internet is actually available. On the extreme right side one can perceive data tier. If human component is to be implemented, it becomes clear that it also requires attendance of consumers and providers of services. Traditionally put, all users become involved. However, it is clear that structure, organisation and maintenance of data remains activity within domain of information science experts, while data remain within domain of users. Quality of data is, therefore, responsibility of users, while quantity and frames remain within sphere of information science experts.

However, presence and possibilities offered by Internet have improved attitudes toward proclaiming SOA paradigms towards finding solutions that will be categorised Cloud computing. Cloud computing improves SOA concepts together with object-oriented paradigms.

For instance, (Miller, 2008) according to some authors CC has common characteristics with autonomous computer systems when latter is capable of self-managing, (Reese, 2002), CC inherits important settings of client-server architecture, (Papadopoulos, 2009), CC structure has characteristics of grid computing since it can overtake the form of distributive or parallel computing where computers can be organised through clusters or weakly connected networks or mainframe computers for needs of big organisations with critical applications such as enterprise resource planning, and financial transaction processing. (Vaquero at al., 2009) recognized in cloud computing Utility computing — The "packaging of computing resources, such as computation and storage, as a metered service similar to a traditional public utility, such as electricity. " (Wei&Blake, 2010) recognized structure Peer-to-peer —

Distributed architecture without the need for central coordination, with participants being at the same time both suppliers and consumers of resources (in contrast to the traditional client–server model) and finally Service-oriented computing – software-as-a-service. That way one can turn weak binding into paradigm of loosely binding.

New architecture follows structural diversification of systems. In this sense, CC offers architectural complex structure in three forms: Saas – software as a Service, IaaS – Infrastructure as a service and Paas – Platform as a Service (fig. 9)

Terminology once again has a mild modification. So, besides weak binding, components are no longer in tiers (that basically means binding) but are also organised through layers.

Layers reflect architecture (fig. 10):

- **Client (Cloud clients):** computers and/or computer programmes available to Cloud computing for delivery of applications as the most important part (Malik, 2008)
- **Application (**Cloud applications): applications as services or "Software as a Service" provided as services by Internet, without required installation and initiation of application on client computers combined with simpler maintenance and backup (Mathur&Nishchal, 2010),

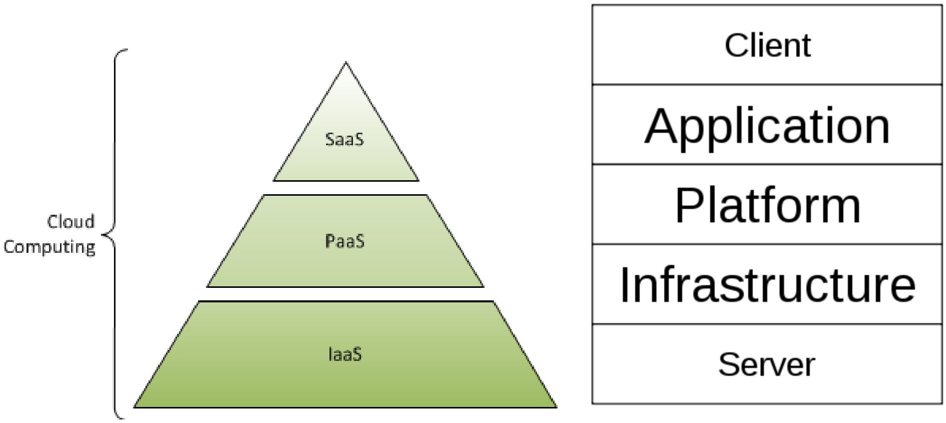


Fig. 9. Cloud Computing architecture( Source: http://www.chades.net/?tag=churchill-club

- **Platform (**Cloud platforms): "Platform as a Service", providing computer platform as service combined with usage of infrastructure and maintenance of applications. It provides implementation of application without costs, obligatory purchase and managing hardware and software on basic level. (Schofield, 2008)
- **Infrastructure (**Cloud infrastructure): Cloud Infrastructure Services or "Infrastructure as a Service", providing computer infrastructure – virtualisation of a platform, environment – as a service, combined with frame for data processing and networking. Instead of buying the provider, software, central memory and network equipment, users buy client versions, that is, complete all external resources as service. Service providers charge upon communal bills, or, price depends upon time of consummation (Pariseau, 2008).

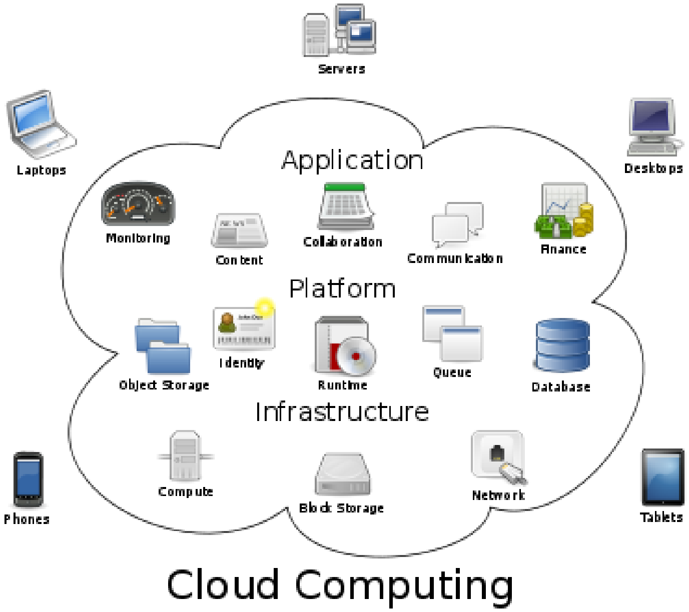


Fig. 10. Cloud Computing (Source: http://en.wikipedia.org/wiki/Cloud_computing)

Server: Provider tier comprises of computers and / or computer programmes especially designed for service delivery, including the multi-core processors, cloud of specific operative systems and combination of offers. (Markoff, 2008)

Diversification to tiers implies that most of activities realised in environment that represents information system supported by computers is being treated as service. Every service of that kind requires carefully prepared contract according to which it will be consumed. However, contract doesn't regulate just the above mentioned services. For service provider sufficient is the list on contract that refers to individual service, which is not the case with service consumer. Consumer should frame his needs within appropriate lines.

To a consumer frame or cloud could be (fig. 12):

- Public cloud: Type of cloud in which services, such as applications and storage are made available to a wider public over the Internet. These are usually located outside the user's offices and provide possibility of reducing risks and costs of providing flexible and temporary broadening of infrastructure (Frank, 2008)
- Community cloud: Cloud of certain organisation. It is usually formed by several business systems with mutual needs (safety, harmonisation, authorisation etc.), regardless to inner or outdoor management. Costs of functioning are lower than costs of public cloud, but higher then ones in private cloud.
- Hybrid cloud: hybrid cloud is comprised of two or more clouds (private, mutual or public) which are unique as community but insures a possibility of multiple implementation of a concrete model.

- Private cloud: infrastructure which exists for individual or individual organisation that can be managed within or outside the organisation (Mell&Grance, 2011)

If named classifications are analysed, it becomes clear that cloud can expressively and efficiently be realised with attentive arrangements and clearly defined contract or contracts. Depending on users /consumer's needs and demands, or what sort of organisation he requires, he'll have to sign the same amount of individual or several combined contracts.

Simultaneously to shifting the paradigms towards definitions of services and modes of their realisation, the importance of regulating performance and ensuring service's results contributed to shaping a need for proper type of contract specific to this area. Basic definition of contract is: Contract is approving declaration of two or more parties directed towards achieving a certain goal. Contract in general form demands:

- Consent and will of parties to sign a contract
- Subject of contract
- Clause or basis for completion
- Other conditions important to signing a contract

Consent and declared will is important condition for signing the contract. However, it doesn't have to be sufficient. Term of sufficiency is completed by defining the subject of contract and clause. Sometimes contract can have a certain form. In that sense, relationships of this kind included, after certain amount of time, form of contract known as SLA (Service Level Agreement).

SLA (Service level agreement) is a contract or part of it that defines services between two parties, where one party is a buyer – demander, while the other party is service provider. Subject of contract is service. As legal institution contract can be used for services or realisations. SLA necessarily has to include:

- Name of service – complete description of package promised by provider
- Way of delivery
- Way of service verification, time of usage and measurement of usability
- Way of treating the mishandling of contract

Qualitatively prepared SLA will help service provider to explain user what, how and in which way he delivers service and gives a guaranty that promised and expected level of service will be delivered.

SLA can be defined in different levels:

- Evaluation upon SLA: arrangement with individual group of buyers in line with covering all services used. For instance, SLA between providers (of IT services) and big service organisation such as financial systems, payment systems, systems of charging, systems of procurement etc.
- Services upon SLA: contract for all buyers who use service is being delivered to service provider.
- Multi-tier SLA: SLA divided into different tiers, with each tier handling different groups of buyers of same service, within same SLA.
- Corporative SLA tier: contract covers all demands of each buyer in a complete business system.

- Evaluation of SLA tier for evaluation of tier: covers all SLM (Service Level Management) questions relevant to specific group of buyers, regardless to services used.
- Service on SLA tier: covers all SLM (Service Level Management) questions relevant to certain services, that is, specific group of buyers.

SLA is treated as contracts of services in which tier of service is formally arranged. In practice term SLA is sometimes used in comparison to contracted deliveries (of services) or performances. SLA is regularly a legal biding formal or informal "contract" (fig. 11).

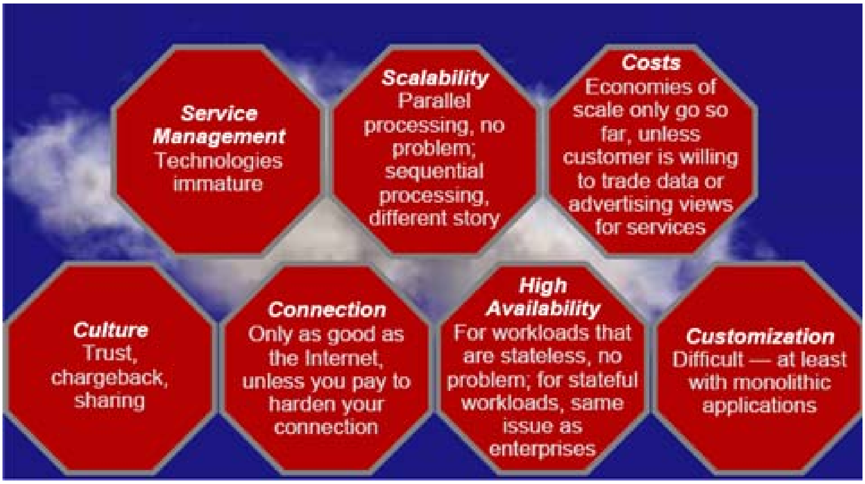

Service Management
Technologies immature
Scalability
Parallel processing, no problem; sequential processing, different story
Costs
Economies of scale only go so far, unless customer is willing to trade data or advertising views for services
Culture
Trust, chargeback, sharing
Connection
Only as good as the Internet, unless you pay to harden your connection
High Availability
For workloads that are stateless, no problem; for stateful workloads, same issue as enterprises
Customization
Difficult — at least with monolithic applications

Fig. 11. What's Holding Cloud Computing Back?
(Source: http://helloanilyadav.blogspot.com/2010/07/cloud-computing.html)

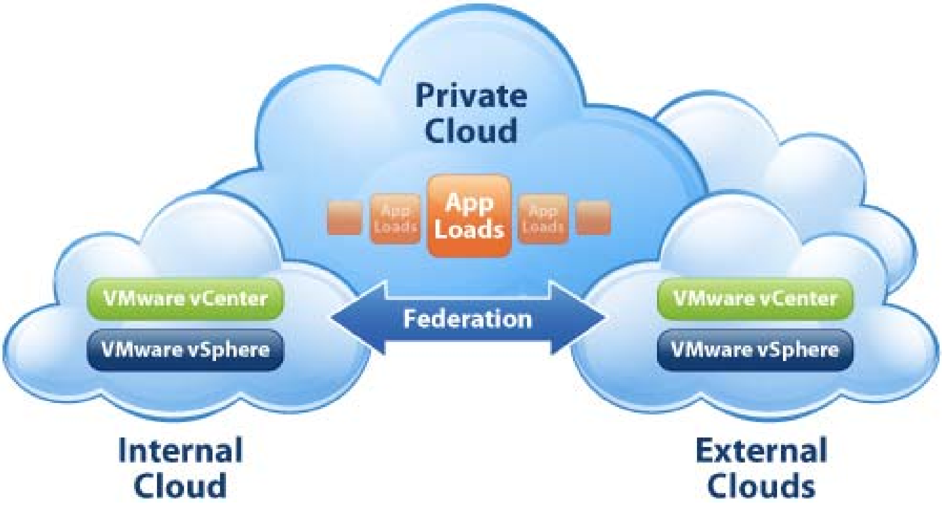


Fig. 12. Type of Clouds (Source: http://computinged.com/business/cloud-computing-facilities-implementations-and-costs-planning/)

However, there is also a wrong situation when contract between service provider and other parties isn't SLA – since the services tier has been defined by (main) buyer. Every field or range of services should have defined "tier of services". SLA can determine level of availability, usability, efficiency, performance or other characteristics and services, such as mode of charge. "Tier of service" can also be mentioned as goal and minimum which enable users to be informed about expected minimum that ensures measurable, usually average aimed value that implies level of performance. Some contracts can arrange penalties in case of not respecting SLA. In practice SLA is sometimes used to define relations for contracted deliveries of services or performances.

### 4.1 Postlude: A hypothetical use case, if the cloud computing is guilty?

In everyday life more and more people are using services of which they don't know the precise definitions, or they haven't been timely explained to them, so they don't know the possible consequences. For instance, widely used net-banking represents a service which is simply defined web service. User in this kind of situations usually doesn't know where the service provider is, that is, where data are located, though he's not introduced with complete infrastructure, and there's no need of that. Further on, mobile phones made these services free of time and geographic location. Advantages are impressive, risk considerably lower, but still possible. Provider of such services is obliged to ensure user with proper information of all possible outcomes during the service providing and afterwards, especially if there is a possibility for user to suffer from any sort of damage.

Among generally accepted services one could count the GPS (Global Positioning System) navigator for automobiles. GPS navigator is using spatial information about location and time anywhere on the Earth, or near it, GIS (Geographic Information System) based upon Global Navigation Satellite System (GNSS) in all weather conditions, with undisturbed line of views from four or more GPS satellites. It is maintained by USA Government and is freely available to all using GPS receiver with some technical restraints that can be removed only for military users. (http://en.wikipedia.org/wiki/Global_Positioning_System). Hence, to most GPS users GIS service represents cloud service or it can be treated that way. GIS is then proclaimed as a system that represents integrated hardware and software with data for collecting, managing, analysing and presenting all modes of geographic information. The purpose of GIS is to enable insight, understanding, making enquiries with interpretation and visualisation of data in form of maps, graphs, reports and charts. Procedure of formation and models of reports are maximally simplified so to be available and understandable to all users. Still, the basic advantage of GIS technology is a possibility of its integration into any form of information system with purpose of ensuring the most qualitative business.

In this part of paper authors analyse hypothetical case in which automobile accident has happened due to mistakes in navigation caused by wrong data provided at the same moment. Emphasis has been given to possible legal consequences of such case, but primarily to legal shortage of defined relations between service provider and service user.

**Hypothetical case:**

Owner of an expensive automobile equipped with equally expensive devices that include GPS is a male person in his thirties. During a longer journey he has been using services of

GIS system that is implemented into latest GPS device. During the journey he followed instructions given by GPS device, so he faced situation and circumstances on highway which mostly haven't correspond to information provided by GPS device. In that specific situation the automobile has been completely ruined, while the owner has been experienced severe body contusions which haven't threatened his life, but could contribute to consequences of minor body disability.

**Epilogue:**

Afore mentioned hypothetical case has all characteristics required to fall a dispute and to claim indemnification of a victim. For purpose of this paper it is important to determine an essential need for legal definition of using services, which are, in this situation, cause of a possible accident.

- Though it is disputable, for this hypothetical case and paper it is needed to determine an important fact: how is GIS treated. Is that Cloud computing service or not? If it is, in which form of Cloud computing a victim could be introduced? These are primary conditions that could be used in eventual dispute with purpose of indemnification.

There is a version of web GIS concept according to which web GIS is every GIS using web technology, or, in narrower sense, it is every GIS that uses web to connect system components. (Fun Sun, 2010). Every such service should necessarily have its own SLA – contract of usage.

Some Service Level Agreement (SLA) ere enforceable as contract but many are really agreements that are more along the lines on operating level agreements (OLA) and may not have the force of law. It is good to have an attorney review this document before you make a major commitment to a Cloud provider. (Sosinsky, 2011).

Indemnification of a victim will necessarily depend upon quality of SLA, its content and form of declaration, if such exists. Since usage of GPS devices is an independent fact in this case, arbiter, seller of GPS devise and selling of devices emerge as factor that additionally complicates described situation. However, accident hasn't been caused by device but instead by data and information provided by device. Within these circumstances responsibility is thrown to provider of data basis and mode of its delivery since responsibility for data (or data basis in information science terms) service provider responds for consequences when data haven't been used properly.

This hypothetical case is only an attempt of directing attention to an urge for more intensive legal treatment of relations that can emerge between service providers and service users within Cloud computing. Service provider should be prepared to provide SLA palette that will correspond to implemented form of Cloud computing. Public cloud carries within dangers similar to outcomes named in a prior hypothetical case. Service provider is usually, in practice, somehow prudishly concealed behind institution of help desk that usually doesn't provide sufficient help, but only smooth communication. Therefore a transparency of relations and their strict regulation is most required.

## 5. Epilogue: Who really need IT and informatics? (The empire strikes back)

Motto "I am not interested how you do the job, just do it" has been transferred from object oriented environment to SOA environment by simultaneous modification of late binding paradigm into a more declared weak binding paradigm. Graphic interface and object oriented paradigms have broadened possibilities for all users' categories. Programmers as user category have been significantly disburdened in their work by possibility of applying the package and libraries of final programme modes which, combined with minor modification and change, can be used in every situation imposed by concrete problem in business environment. Basic paradigms from object oriented environment have been further improved and applied in SOA environment. SOA environment as paradigm is characterised by assembly of mutually connected services. Services have minimal interdependence and constant distinction. Should all services be united in one portfolio of services and should they be made disposable to the user, all his needs should be satisfied. Projecting based on analysis of use case permits such a granularity of software solutions. It is important that user is introduced with a possibility that he can use such solutions when need emerges. That way a possibility is offered which necessarily doesn't have to be consummated, but it is important to know that it exists. Extraction of IT sector as non-core business out of business system's frame demands a different attitude towards the ICT experts so it presents them in different light.

Application of all accomplishments of good practice from the above mentioned modes of organising computer supported Information Systems through integrating all possibilities offered by Internet and Web, appreciation of users' needs and a possibility of their modification have been united in a new approach known as cloud computing. (Korri, 2006) Basic advantages of cloud computing are: reductions in duration of performance and response, lower costs of approach, decrease of risks in field of infrastructure, maintaining the level of innovations, reliability, possibility of development, safety and sustainability. Since cloud computing is proclaimed by slogan pay-for-what-you-use the fact is it represents a model similar to the one of spending electricity, fuel and water. Cloud can be public or private. Public cloud sells services unlimited respectively those services that are available with fee to anyone on the Internet. Private cloud is usually an owner of private network and data that provide services to a limited number of users. In such circumstances Information Science necessarily demands appropriate legal treatment of every service individually. Both categories of users should be maximally protected but so should be every service as a product, tool or an instrument used by both sides. Legal aspect has been maximally underlined in all levels up to a measure that it requires a special legal regulation. Service as product can lead to unwanted results that can cause certain damage to user or service applicant. Necessity of protection has been articulated through forms known as SLA and the beginning has been denoted in Cloud that has a prospect of becoming a dominant form of computer backup to business in a long term period.

## 6. Conclusion: What can we do without IT? (Adam, why are you naked?)

User or requestor of Information System's services has, during the time flow, been put in different relationships and different positions. Primal need of users' high specialisation in knowledge and skills within Information Science has been lost. Business systems that insisted on computer backup, approached to organisation of private IT sectors that employed such experts. Reference to such category of users has changed with the time flow. Difference in basic knowledge of Information Science among experts users and users themselves has disappeared while cognition of own possibilities has grown and user increased his knowledge of own possibilities. Final user has in such circumstances become aware of himself and his possibilities but has also fallen in trap of overestimating the same. On the other hand, underestimating the role of information scientists and wrong interpretation of need for such personnel in business brought to outsourcing IT sector and leaving the computer backup service such as Cloud computing.

Insecurity, non-transparency, indecisiveness and many similar negative characteristics doesn't have to accompany cloud solutions but they remain as possible side effects. User coming from business system has perceived Cloud as vulnerable and "stripped" to the level where it has to provide information services precisely ensured and regulated to a level securing the legal strength. Business system's user has in Cloud realised that he has to have Information Science services completely protected by law. Only that way will they be able to reinforce themselves with required tools and backup in accelerated business and fiercer market competition. Information Science asks for law to be its companion. Information science requires law as constant associate in business making and providing information to business needs. Development of information system will demand a person in team professional enough to perform that assignment. Actually a qualitative synergy between law and information science is obligatory.

## 7. References

Balzer , Y. (2004), Improve your SOA project plans, IBM,

Batini,C. ,Scannapieco, M. ( 2006), Data Quality: Concepts, Methodologies and Techniques (Data-Centric Systems and Applications), Springer, ISBN-13: 978-3540331728

Frank, G.,(2008), *Defining "Cloud Services" and "Cloud Computing"*". IDC. 2008-09-23. Retrieved 2010-08-22

Fu, P., Sun, J., (2010), *Web GIS: Principles and Applications*, ESRI Press, ISBN-13: 978-1589482456

Korri, T. (2010), *Cloud computing: utility computing over the Internet*, Helsinki University of Technology,
www.cse.tkk.fi/en/publications/B/5/papers/Korri_final.pdf/(accessed: 10.5.2010.)

Kroenke, D M. (2008). *Experiencing MIS*. Prentice-Hall, Upper Saddle River

Liebow, M., (2005), *Do customers really want SOA?*, IBM, TechRepublic, ZDNet News

Malik, O., (2008), *What Makes a Cloud Computer?*, Gigaom.com. 2008-06-22. Retrieved 2010-08-22.)

Markoff, J. (2008). "*Microsoft Plans 'Cloud' Operating System*". Nytimes.com. Retrieved 2011-08-20.)

Mathur, P. ,Nishchal, N., (2010), *Cloud computing: New challenge to the entire computer industry, (in Parallel Distributed and Grid Computing (PDGC), 1st International Conference, pp. 223 – 228, ISBN: 978-1-4244-7675-6 )* Eccentex.com. Retrieved 2010-08-22

Mell, P., Grance, T. (2011) "The NIST Definition of Cloud Computing (Draft)", *(Recommendations of the National Institute of Standards and Technology)*. National Institute of Science and Technology., Retrieved 2011-07-24.)

Miller, R., (2008), *What's In A Name? Utility vs. Cloud vs Grid*, Datacenterknowledge.com. Retrieved 2010-08-22.)

O'Brien, J A. (2003). *Introduction to information systems: essentials for the e-business enterprise*. McGraw-Hill, Boston, MA

Papadopoulos, G., (2009), *Sun CTO: Cloud computing is like the mainframe,* Itknowledgeexchange.techtarget.com. 2009-03-11. Retrieved 2010-08-22

Pariseau, B., (2008) , *EMC buys Pi and forms a cloud computing group*, Searchstorage. techtarget.com. 2008-02-21. Retrieved 2010-08-22

Reese, G.(edit.), (2002), *Database Programming with JDBC and Java*, O'Reilly & Associates.

Schofield, J. (2008), *Google angles for business users with 'platform as a service'*, London: Guardian.

Sosinsky, B., (2011), *Cloud Computing Bible*, Wiley, 2011, ISBN-13: 978-0470903568

Vaquero, L.M. et al.,(2009) *A break in the clouds: towards a cloud definition*, Newsletter ACM SIGCOMM Computer Communication Review, Volume 39 Issue 1, TechPluto. Retrieved 2010-09-14

Wei,Y., Blake, M.B., (2010) *Service-Oriented Computing and Cloud Computing: Challenges and Opportunities*". IEEE Internet Computing, vol. 14 no. 6, pp. 72-75 Retrieved 2010-12-04

# Developing a Theoretical Framework for the Adoption of Biometrics in M-Government Applications Using Grounded Theory

Thamer Alhussain[1] and Steve Drew[2]
*[1]King Faisal University,*
*[2]Griffith University,*
*[1]Saudi Arabia*
*[2]Australia*

## 1. Introduction

Mobile devices have become the world's most common means of interpersonal communication; and, the growing marketplace for new software, or "apps", enriches an already burgeoning array of purposes to which mobile technology can be lent. We are thus witnessing the advent of conditions for a range of mobile technology enabled information systems. According to the latest statistics produced by the Central Intelligence Agency (CIA), there were 5.3 billion mobile subscriptions worldwide in 2010 out of a world population of about 7 billion people (World Fact Book 2011). With the advancements in mobile technologies, several governments have started looking to provide their services via wireless and mobile devices. Mobile government (m-government) is a new delivery channel using Information and Communication Technology to deliver and improve government services that complements current e-government (Antovski and Gusev 2005). Currently, a number of m-government applications exist in several countries around the world. With the growth of m-government services, the importance of security for its acceptance and adoption has been noted in many studies (NECCC 2001; Al-khamayseh et al. 2006; Clarke and Furnell 2005, 2007). Requirements for user acceptance lead to a greater need for user and government authentication to protect data, services, and the promotion of public trust. The negative security perception is a serious issue that citizens have regarding the use of mobile services which may affect their adoption of the technology for critical applications (Chang and Kannan 2002).

This chapter will describe an enquiry into how biometric technology, which can provide reliable user authentication, can play an integral role in providing secure m-government services. We use Grounded Theory methodology to understand reality from the point of view of the participants including mobile users, service providers, and network operators in order to develop a substantive theory for the adoption of biometric authentication in m-government security. In the field of information systems, Urquhart et al. (2009) indicated that Grounded Theory has been proved to be extremely useful in this field which led them to recommend its application to help generate theories in information systems.

This chapter provides unique perspective on investigating the adoption of biometric authentication in the context of mobile government applications, taking into account requirements and opinions of the people involved in m-government including mobile users, service providers, and network operators. This chapter addresses a gap in the literature regarding the factors influencing the adoption of biometric authentication in m-government security. The main contribution of this chapter is the development of a new substantive theory that provides a theoretical framework for the factors influencing this technology's adoption. Thus, it provides rich insights and increased understanding of the concerns and perceptions of the abovementioned stakeholders regarding the application of biometric authentication to mobile devices for government services. Moreover, this chapter provides a new example of the application of Grounded Theory methodology to qualitative information systems research.

This chapter is structured as follows. It begins with a brief background relating to the information security and mobile government. Next, the chapter discusses the adoption of biometric technology within the context of electronic and mobile government. The chapter then explains and justifies the methodological choices along with the description of Grounded Theory methodology. The chapter also explains the context of the study presented in this chapter in addition to the data collection procedures. The application of Grounded Theory is then detailed and described. Finally, the paper concludes by developing a new theoretical framework for factors influencing the adoption of biometric authentication in m-government security and providing several considerations for the adoption of biometrics in m-government applications.

## 2. Information security and m-government

The primary entities of m-government are mobile phone users, government agencies as service providers, and the network operators. Although they have several different requirements, they share security as one of the most important system requirements. As mentioned above, security is the most important issue facing m-government applications and it is a basic feature of the mobile communication infrastructure. Specifically, security has five features that need to be considered, which are user authentication, data integrity, service availability, information confidentiality, and non-repudiation of user participation in transactions. A biometric system enhances the identification, authentication and non-repudiation of the information's user to support facets of information security. It can help "to provide identity-based access control and to authenticate integrity of information with respect to subject involved" (Vielhauer 2006, p. 18).

### 2.1 Authentication strategies

There are three general categories of authentication as follows:

- Something the user knows (e.g. PIN or password).
- Something the user has (e.g. cards or tokens).
- Something the user is (e.g. biometrics).

The Personal Identification Number (PIN) is a secret-knowledge authentication method and consequently relies upon knowledge that only the authorized user has. Although the PIN and password are the most commonly used methods for authentication in information systems (Scott et al. 2005), such secret-knowledge approaches unfortunately have long-established problems, with weaknesses often being introduced by the authorized users themselves. These are most clearly documented in relation to passwords, with bad practices including the selection of weak and easily guessable strings, sharing passwords with other people, writing them down where others can find them, and never changing them (Clarke and Furnell 2005). Consequently, these approaches are the easiest targets for hackers.

A security token is a physical entity or item that an individual possesses to establish personal identification, such as a passport, ID card, or credit card (Jain et al. 2000). This token based approach is approximately similar to the secret knowledge approach, as it basically relies upon the user remembering to bring along something to ensure security whereby the token needs to be physically present (Clarke and Furnell 2007). Therefore, secret knowledge and token based authentication approaches are unsatisfactory methods of achieving the security requirements of information systems, as they are unable to differentiate between an authorized and an unauthorized person who fraudulently acquires the knowledge or token of the authorized person (Jain et al. 2000). On the other hand, biometric authentication relies upon the unique physiological and behavioural characteristics of an individual; hence, it cannot be forgotten, lost or stolen.

### 2.2 The current authentication system in m-government

The current security method in mobile phone based m-government applications is based on the use of 4 to 8 digit Personal Identification Numbers (PINs). This method can be applied to both the mobile device and the user's Subscriber Identity Module (SIM) which is a removable token containing the cryptographic keys required for network authentication. As mentioned above, the PIN is an approach providing low level authentication, as it is based on something the user knows. However, the existing SIM card, a token based approach, can be physically removed from the mobile device when not in use; however, users usually leave it inside the mobile device for convenience as well as to avoid loss or damage (Clarke and Furnell 2007). Thus, the PIN and SIM card approached carry the risk of loss or theft which can compromise the security of information, especially with the inclusion of sensitive personal information which confirms the need of advanced approach for ensuring and enhancing the security of data in mobile devices.

Providers of second generation (2G) and third generation (3G) mobile networks deliver smartcards with pre-installed symmetric keys which are used by the network to authenticate the mobile device and, in the 3G case, for the mobile device to authenticate the access network. The authentication system is based on the trust relationship that exists between the access network provider and the service provider via a roaming agreement, and between the user and the service provider via the service subscription. The symmetric session keys for data confidentiality and integrity sent over the airwaves are derived during the authentication process. However, data confidentiality and integrity extending over the whole path between the communicating parties is not provided by the access network security of second and third generation systems which has to be provided on the network at application levels for end-to-end security (Dankers et al. 2004). With this in mind, Public Key Infrastructure (PKI) combined with biometric authentication may present a suitable integrated solution to achieve end-to-end m-government security.

### 2.3 Biometrics and m-government

Integrating biometric authentication into mobile devices can be done in two different ways. The first technique is to store the biometric template in an external database (Giarimi and Magnusson 2002). In this case, the biometric data have to be sent over the network every time the user wants to be verified and, during that process, the data are encrypted, which forms the external database for storage rather than security. The problem is that the users have no control over their own biometric pattern once it leaves the device. Furthermore, it can potentially take a long time to perform verification when data are being sent over the mobile network due to traffic overload and the number and size of the files in transit. However, it does not take up much memory in the mobile device. The second technique is to store the biometric template in the device or particularly on the smart card which will enable users to control their biometric pattern (Giarimi and Magnusson 2002). The biometric verification should take place when the users want to log in to their mobile device and when they want to perform a government service. Moreover, this can be integrated with the Public Key Infrastructure, as mentioned earlier, to provide a more secure authentication system.

## 3. Adoption of biometric technology

With the advantage of reliable authentication of biometric technology, many security applications around the world have adopted and implemented biometric technology. Currently, biometric technology has been adopted in many applications such as access control, national identity, immigration, proving attendance, military identification, e-government, and e-commerce applications.

With the application of biometric technology, e-government aims to give its citizens improved services and better access to information as it can provide reliable identification of individuals as well as the ability for controlling and protecting the integrity of sensitive data stored in information systems. Many researchers such as Ashbourn (2004), Bonsor and Johnson (2008), Scott et al. (2005), and Wayman et al. (2005) argue that a wider use of biometric technology can be applied to e-government projects. With variations on attendance registration mentioned above, biometric technology is used for e-voting to ensure that voters do not vote twice. With biometric technology, governments prevent fraud during elections. Moreover, biometric technology can be used to ensure correct working times are recorded and that only authorized personnel have access to government property and resources.

Biometric technology can also be used by e-governments for business. For instance, many banks use facial recognition systems to minimise chances of theft. For example, photos are taken on the bank slips which are stored on computer software. As a result, this has avoided the issue of fraudulent bank slips when withdrawing money, since ATMs are a quick method of withdrawing money. This has helped the government to conduct its activities effectively (Bonsor and Johnson 2008).

In business, there is frequently the need for full identification of employees to ensure that, in case of any problem in that firm, the management is in a position to identify the person responsible for that act. Commercial applications may also require full identification capability, digital certificates, human interface, and one or more authentication devices to ensure that the business can run well. People are also in a position to do their business properly and invest in any organisation as long as that organisation has an identity as an effective company (Ashbourn 2004).

Biometric technology is also used in the identification of citizens by e-governments. If they choose, every nation should ethically be able to identify its citizens and non-citizens by using national identification cards, visas, and passports. As a result, e-governments are in a position to identify its citizens in the production of these documents, hence reducing the issue of illegal immigration. A good example is the United States whereby, since the events of September 11 2001, it has widely adopted biometric technology. Two laws, relating to identification of transport workers and to immigrants, were made in the United States triggering a mass deployment of biometrics. Now, seven million transportation employees in the United States have biometrics incorporated into their ID cards. Moreover, in order to closely control visitors who enter and leave the country, all foreign visitors are required to present valid passports with biometric data; consequently, over 500 million U.S. visitors have to carry border-crossing documents which incorporate biometrics (Ashbourn 2004). Several European governments have also started to implement the use of biometrics. The U.K. government has established issuing asylum seekers with identification smart cards storing two fingerprints. General plans have also been made to extend the use of biometric technology throughout the visa system in the U.K. as well as in France, Germany and Italy (Scott et al. 2005).

E-governments use the various types of biometric identification in order to control certain illegal behaviour. For example, the Japanese government plans to use biometric technology in passports to tackle illegal immigration and to enable tighter controls on terrorists. This will be applied within a computer chip which can store biometric features like fingerprints and facial recognition data (Scott et al. 2005).

Other e-governments are using the biometric technology to secure access to certain defence bases and similarly secure areas. Biometrics can also provide potential for security cost savings. For instance, hand recognition has been used at the Scott Air Force Base to save more than $400,000 in manpower costs through their metro-link biometric access gate (Frees 2008).

### 3.1 Technology adoption factors among empirical studies

Empirical studies related to the acceptance and adoption of mobile phones and electronic services via the Internet have mostly applied models based on the use of Diffusion of Innovation (Rogers 1995), the Technology Acceptance Model (Davis 1989), or the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al. 2003). For instance, Jahangir and Begum (2008) introduced a conceptual framework that considered perceived usefulness, ease of use, as well as security and privacy as important factors that influence users' acceptance and adoption of electronic banking services. Another study by Tassabehji and Elliman (2006) highlighted trust and security as major factors affecting e-government adoption. Moreover, AlShihi (2007) indicated that trust has a wide impact on m-government acceptance. Lee et al. (2002) found that social influence and self-efficacy variables significantly affect perceived usefulness and perceived ease of use for user acceptance of the mobile Internet. Moreover, Teo and Pok (2003) found that social factors including perceptions of relative advantage play a significant role in influencing intentions for the adoption of Wireless Application Protocol WAP-enabled mobile phones amongst Internet users. Kaasinen (2007) found that perceived value, ease of use, trust and ease of adoption are important factors that influence user acceptance of mobile Internet services. AlGhamdi et al. (2011) pointed out that the provision of trustworthy and secure online payment options is a critical key determining the decision for online customers to accept/reject buying online from a specific retailer.

Thus, highly similar acceptance factors appear under various theories and models covering innovation acceptance and adoption. Moreover, the set of factors that proposed in TAM variants and UTAUT correspond closely with factors identified in DOI theory. For instance, Moore and Benbasat (1991) indicated that while developing an instrument based on DOI concepts to determine an individual's perceptions regarding the acceptance and adoption of an information technology innovation recognised the similarity between the construct of perceived usefulness with perceived relative advantage, and between perceived ease of use with perceived complexity.

### 3.2 Biometric adoption among empirical studies

Although there are a lack of academic studies concentrating on the factors that influence the adoption of biometric authentication systems, most of the published papers in this area (Harris and Yen 2002; Kleist et al. 2005; Lease 2005; Uzoka and Ndzinge 2009) identified different factors that are quite dissimilar to those discussed in the technology adoption theories and models outlined in previous section. For example, Lease (2005) found that managers' positive perceptions of security effectiveness, need, reliability, and cost-effectiveness correlate with their willingness to recommend the use of biometric technology, while Uzoka and Ndzinge (2009) indicated that ease of use, communication, and size and type of organisation are the most important factors affecting the intention to adopt biometric technology in organisations. However, Harris and Yen (2002) stated that the adoption of biometric systems can be influenced by managerial, economical, operational and process-related factors. Kleist et al. (2005) also indicated different affecting factors of biometric systems including users, administration, environment, infrastructure, cost, communication system, as well as security needs and requirements.

As a result, the review of the relevant literature on the technology adoption factors did not lead to any hypotheses, but rather helped to enhance awareness of the existing factors and to identify the gap in relevant knowledge. The case in this chapter discusses the adoption of biometric authentication in the m-government context and in particular, in Saudi Arabia, which adds some specificity to the area of biometric technology adoption.

## 4. Methodological choices

A review of the extant literature on theories relating to authentication and mobile government security also did not lead to any hypotheses per se, but broadened the considerations relevant to this work. What was needed was a method for determining the full range of stakeholder considerations that would influence the adoption of m-government with biometric security. A review of methods available for an interpretive study with the main purpose of creating a substantive theory to guide development, indicated that Grounded Theory methodology (Strauss and Corbin 1990) was the ideal vehicle for investigating actualities from the real world to generate or discover theory grounded in context specific data that has been systematically gathered and analysed (Creswell, 1998).

The investigation part of this study was carried out by the use of questionnaire and semi-structured interviews for the data collection. By conducting the interviews and questionnaires, we explored the factors influencing the adoption of biometrics in m-government through the concerns and perceptions of mobile communication users', service providers', and network operators' about applying biometric authentication into mobile devices for government services. Data were analysed following Strauss and Corbin's (1990) approach of Grounded Theory. The use of Grounded Theory helped to develop a substantive theory that identifies and describes the factors influencing the adoption of biometric authentication in m-government in Saudi Arabia.

### 4.1 Grounded theory methodology

Grounded Theory is one of the most widely used methodologies in qualitative research (McLeod 1999). It originated in nursing research by Glaser and Strauss (1967) and then has been adopted in several areas of research such as sociology, business, management, and information systems (Mansourian 2006). More specifically, Grounded Theory was first developed by Barney Glaser and Anselm Strauss in 1967 in their book "The Discovery of Grounded Theory". They defined Grounded Theory as "the discovery of theory from data – systematically obtained and analysed in social research" (Glaser and Strauss 1967, p. 1).

In subsequent years two different approaches of Grounded Theory have emerged, one by Glaser and the other by Strauss and Corbin. These two approaches became more visible by the publication of Strauss and Corbin's book in 1990. The Grounded Theory approach, according to Strauss and Corbin (1990), is a "qualitative research method that uses a systematic set of procedures to develop an inductively derived Grounded Theory about a phenomenon" (p. 24). However, Glaser (1992) clarified that "Grounded Theory is based on the systematic generating of theory from data, that itself is systematically obtained from social research" (p. 2).

Glaser (1992) thought that the research should allow the theory to emerge during the observation of the codes and data analysis. Glaser's approach concerns with a classic philosophy emphasizing an inductive emergence of the theory as well as the researcher's role within that process (Heath and Cowley, 2004). Glaser (1992) focuses on the importance of letting the theory emerge from the data by allowing the data to speak for itself and avoiding imprinting preconceived ideas onto the theory (Creswell 2008). By contrast, Strauss and Corbin's (1990) perspective emphasised more on a systematic approach involving validity and verification (Heath and Cowley, 2004). Strauss and Corbin's (1990) approach indicated that Grounded Theory should be inductively derived from the study of the phenomenon it represents. It should be discovered, developed, and verified through systematic data collection and analysis of the data that pertaining to the phenomenon.

It emerges that the Strauss and Corbin (1990) approach is significantly more prescriptive in specifying the steps to be done during the coding and data analysis. More specifically, Strauss and Corbin (1994) identified Grounded Theory as "a general methodology for developing theory that is grounded in data systematically gathered and analysed. Theory evolves during actual research, and it does this through continuous interplay between analysis and data collection" (Strauss and Corbin 1994, p. 273).

Glaser (1992) and Strauss and Corbin (1990) differed on the role of the literature review as an influence on the methodology. Glaser (1992) believed that specific reading related to the area under study before or during data collection could strongly influence the emerging theory, thus, it should not be reviewed until the theory begins to emerge. While Strauss and Corbin (1990) believed that the researcher will come to the research area with a background about the relevant literature which is a basis of professional knowledge and it is important to acknowledge and use it, as will be discussed in the next section. They believe that some understanding of the research area through the literature review will enhance the theoretical sensitivity of the researcher when generating theory.

However, Grounded Theory has been presented as a general methodology applicable for both qualitative and quantitative studies (Strauss and Corbin 1994). Strauss and Corbin (1998, p.27) stated that "briefly, we maintain that the aim of theorizing is to develop useful theories. So, any methodology, whether qualitative or quantitative, is only a means for accomplishing that aim. We do not believe in the primacy of either mode of doing research"

The study related in this chapter adopted Grounded Theory methodology to develop a substantive theory for the adoption of biometric authentication in m-government security in Saudi Arabia. In particular, this study followed Strauss and Corbin's (1990) approach as it allows researchers to take into account previous relevant theories and literatures to help gain insights into the data. It also provides extensive guidance and a comprehensive framework for researchers, while, Glaser's approach is much less structured. Further justifications for the use of Grounded Theory are provided in the following section.

### 4.2 Justification for using grounded theory

According to Goulding (2002), the usefulness of the application of Grounded Theory appears where there is a lack of integrated theory in the literature. From the initial literature review provided in the earlier in the chapter, it can be noted that there was a lack existing theories regarding the utilization of biometric authentication and mobile government security, especially which might be applied in developing countries such as Saudi Arabia. Combining this finding with the main application of Grounded Theory for investigating actualities in the real world, the researchers use Grounded Theory to develop a substantive theory that describes how biometric authentication can play an integral role in providing secure m-government services by investigating the phenomenon within the real world entities involved in m-government which are mobile users, service providers, and network operators.

Moreover, comparing with other qualitative analysis methods, Grounded Theory provides systematic method of analysis including open, axial, and selective coding that helps to develop a theory that is grounded in data. This is consistent with Charmaz's (2006) indication that the main strength of Grounded Theory is that it provides means for the analyzing processes including specific steps for developing concepts, categories, and theory. Piantanada et al. (2002) point out the usefulness of the Grounded Theory in such interpretive research. They note "the procedures of Grounded Theory provide interpretive researchers with a disciplined process, not simply for generating concepts, but more importantly for coming to see possible and plausible relationships between them" (p. 3).

Urquhart et al. (2009) indicated that Grounded Theory has been proved to be extremely useful in the field of information systems, which led them to recommend its application to help generate theories in that research area. Furthermore, Urquhart and Fernandez (2006) stated that the value of Grounded Theory in the field of information system has become widely acknowledged in the research community.

Accordingly, the Grounded Theory approach fits the purpose of this study, which should lead to the development of a substantive theory for the adoption of biometric authentication in m-government security in KSA. Through the interviews and questionnaires, the researchers explore the factors that influence the adoption of biometrics in m-government through users', service providers', and network operators' concerns and perceptions regarding applying biometric authentication into mobile devices for government services.

## 5. Context of the study

This study was supported by Saudi government and data collection primarily took place in the Kingdom of Saudi Arabia. Therefore, Saudi and Islamic cultural issues needed to be considered throughout this study. The Kingdom of Saudi Arabia is located in the south-eastern part of the Asian continent. It occupies 2,240,000 sq km (about 865,000 sq mi). The total population reached 28.5 million in mid-2009 with an annual growth rate of 2.9 percent; however, it is estimated that approximately 5.5 million of the population are non-Saudis (World Fact Book 2011).

The need for the services, means and methods of e-government in Saudi Arabia has emerged by responding to the developments and changes of the modern world in all fields. Saudi Arabia, like other countries, is seeking to make use of the great technological advancements in communication means and information due to their importance in providing services which are better, faster, more accurate, and with stricter controls. Particular attention has therefore been given to e-government as an international approach and a general trend that requires a response to and the use of modern technology as a means for its success. Based on this, the e-government program was introduced in 2005 and was called "Yesser", an Arabic word meaning "facilitator". This program was set as a result of the execution of the communications and information technology national plan through the support of electronic transactions and applications by government organisations. It plays the role of the enabler/facilitator of the implementation of e-government in the public sector. Moreover, it aims to raise the public sector's efficiency and effectiveness, offer better and faster government services, and ensure availability of the required information in a timely and accurate fashion (E-government program "Yesser" 2011).

Due to the enormous significance of e-government applications, there are now more than 180 electronic services being offered by 50 different organisations. An example of a most successful e-government service is the payment system called "Sadad". Sadad was implemented by the Saudi Arabian Monetary Agency in order to facilitate and streamline the bill payment transactions of end consumers via all banking channels, including bank branches, ATMs, telephone banking, and Internet banking. In 2008, the number of transactions conducted by Sadad exceeded 5 million transactions per month, with a monthly growth rate of 22% (Sadad 2008).

The use of mobile devices is rapidly increasing among the people in the KSA. According to a recent report in 2010 by Communications and Information Technology Commission (CITC) in Saudi Arabia, the latest statistics in 2010 indicated that there were 4.3 million telephones (fixed lines) in use. By comparison, the total number of mobile subscriptions is 47 million with average annual growth rate for the last eight years at around 43%. This CITC report also stated that mobile penetration in Saudi Arabia stood at 172% which is higher than the world average of 67%, the developing countries average of 57% and the developed countries average of 114%. However, the CITC report indicates an estimated 11 million Internet users with an average annual growth about 33% over the eight years period (2001-2009).

Therefore, as the number of mobile phone users is higher than that of Internet users, the Saudi government is concentrating on developing delivery of its services through mobile devices. Currently however, m-government applications in the KSA are at an early stage and most are based on the use of SMS. For instance, the Ministry of Education has been sending final exam results to the final level high school students via mobile phones since 2003. In the process of this service, the Ministry of Education provides a soft copy of the students' final exam results to the Saudi Telephone Company (STC) and students are required to send an SMS message containing a student number to the STC to receive a text message containing their results. The main disadvantage of this service is the lack of privacy where anyone who knows a student's number can get that student's results without their permission (Abanumy and Mayhew 2005).

The Ministry of Interior also started to provide several services via mobile devices through its different sectors, such as the General Directorate of Passports and General Department of Traffic. For example, drivers can inquire from the General Department of Traffic about their fines via their mobile devices. A driver can send an SMS message containing their ID number and then will receive a text message containing the result.

Another m-government application is weather notifications. Mobile users can get an SMS message containing weather conditions from the weather forecasting authority. Moreover, a number of hospitals have started an appointment reminder application that reminds the patients of their appointments by sending an SMS message containing the date, time and clinic location.

### 5.1 The use of biometrics in the KSA

As mentioned earlier, several governments have implemented biometric authentication in various types of applications. The Kingdom of Saudi Arabia, as other countries, has implemented biometrics in several places as follows.

Fingerprint technology has been applied for registering employees' attendance in several government agencies such as Ministry of Interior, Ministry of Foreign Affairs, the General Organisation for Technical Education and Vocational Training, the Royal Commission for Jubail and Yanbu, and Supreme Commission for Tourism. Furthermore, a number of agencies such as the Ministry of the Interior, the Ministry of Foreign Affairs, and the Saudi Monetary Fund have implemented biometrics to authenticate employees in special security cases like entering via some doors in their buildings.

Recently, the Ministry of Interior started to require citizens submit their biometrics when they issue or renew their ID national card as well as residents' biometrics when they issue or renew their residential cards. More specifically, the Directorate General of Passports has implemented fingerprint technology in several cities in the Kingdom for foreign people. This system now has the biometrics for about 7 million Saudi residents.

### 5.2 Conduct of the study

As stated above, this study used both questionnaire and semi-structured interviews for the data collection. In particular, eleven face-to-face semi-structured interviews were conducted in the Kingdom of Saudi Arabia with the managers of online services and IT security managers of mobile e-government service providers including the Ministry of Interior, National Information Center, General Directorate of Passports, The Saudi E-Government Program (Yesser), National Centre for Digital Certification, Al-Elm Information Security Company, and Sadad Payment System. Four semi-structured interviews were also conducted with managers and IT security providers in mobile communication network services including the Saudi Telecom Company (STC) and Etihad Etisalat (Mobily).

Theoretical sampling guided by Grounded Theory methodology was applied in this study, and refers to the selection of participants based on criteria specified by the researcher and according to preliminary findings (Glaser and Strauss 1967). The early stages of continuous data analysis pointed out matters that need further exploration, therefore, the process of sampling was directed by the on-going theory development and interviews were conducted until theoretical saturation was reached.

The interview questions were of an exploratory nature. More specifically, open-ended questions were designed to help identify the factors influencing the successful implementation of biometric authentication in m-government security. These comprised questions on benefits, challenges, barriers, and concerns about this application of biometrics taking into account the different roles of the target organisations. Furthermore, the data collected from the first interviews helped to modify the questions for the subsequent interviews. This was as intended and followed the guidelines of Grounded Theory methodology.

A survey questionnaire was presented to mobile communications users to explore their concerns and perceptions regarding applying biometric authentication in their mobile devices for government services. Users from both genders were chosen as participants from a range of relevant age groups and education levels. The questionnaire sought responses from a selection of choices under the basic headings of “Background Information”, “ICT Experience”, “Mobile Devices and Government Services”, “Mobile Device Security” and “Biometrics and Mobile Government Services”. It was also designed to give opportunity for the participants to make comments after each question. 420 questionnaires were distributed and 330 were returned from the participants. Nineteen of the 330 were excluded from the study because they were deemed incomplete. Thus, a total sample of 311 questionnaires was included in the analysis.

It is noteworthy that interviews rather than questionnaires were conducted with the much smaller number of government service providers and network operators in order to more fully explore their individual perspectives. The questionnaire by comparison, was distributed to mobile users in order to collect larger amounts of data about mobile communication users’ concerns and perceptions in a shorter time scale than would have been possible with interviewing.

## 6. Application of the grounded theory

Collected data was subjected to analysis using Grounded Theory methodology which was executed by carefully following Strauss and Corbin’s (1990) approach. As mentioned earlier, this study incorporates the suggested techniques by Strauss and Corbin (1990) including sampling, coding, memo writing, reviewing of literature, and making constant comparisons to analyse the data and enhance theoretical sensitivity. More details about the application of these techniques are provided in the following sections.

### 6.1 The Use of the literature

"All kinds of literature can be used before a research study is begun: both in thinking about and getting the study off the ground. They can also be used during the study itself, contributing to its forward thrust" (Strauss and Corbin, 1990, p.56). However, Strauss and Corbin (1990) distinguish between different types of literature which are technical and nontechnical literature and they argue that both are of equal usefulness, and can be used at the same points in Grounded Theory analysis procedures.

Technical literature refers to theoretical and philosophical papers as well as other research studies which characterize the writing of a professional discipline and it can be used as background material for comparison against the findings of Grounded Theory. Strauss and Corbin (1990) stated several reasons for the use of technical literature early in the study. For instance, it can be used in order to stimulate theoretical sensitivity by providing concepts and relationships for comparison against the data, therefore, previous theories can be modified, extended, or amended depending on the situation. Moreover, it can be used to stimulate questions for interviews or the other data collection techniques and can also be used to help direct theoretical sampling. Another reason for using technical literature is to provide a secondary source of data that can help by providing supplementary, externally sourced, validity to the research findings (Strauss and Corbin 1990).

In contrast, nontechnical literature refers to the use of other materials including reports, records, and manuscripts (Strauss and Corbin 1990). It can be used either as "primary data or to supplement interviews and field observations in Grounded Theory studies" (Strauss & Corbin, 1990, p.48). However, some researchers believe that the initial review of the literature is important as it helps in enabling readers to classify the researcher's perspective as the research begins as well as providing justification for applying the Grounded Theory study (Antle 1986).

In the study related here, a review of the relevant literature and previous theories established current thinking in the areas of mobile government and security. The main objective of this literature review was to enhance awareness of the existing knowledge and to identify the gap. In addition, technical literature was used as background material for comparison against the findings for the development of the substantive theory of this study. Nontechnical literature such as newspapers and government reports were used as well in order to support several emerging issues resulting from the empirical study.

### 6.2 Memos

Memos are "written records of analysis related to the formulation of theory" (Strauss and Corbin 1990, p.197). They are written continuously through the research process in order to reflect upon and explain meanings and processes, including identifying relationships between codes and categories, as well as providing a depth of understanding of the concepts (Strauss and Corbin 1990). In this study, memos were written to help describe and explain the data analysis, as well as the relationships among concepts and categories. It further helped to explore data, and to group concepts and codes into categories. For example, during the development of the "system requirements" and "procedural issues" categories, one of the memos written stated that "Organisational and users' factors influence the identified system requirements. System requirements relate to the system itself. Procedural issues which emerged from entities factors include current and future authentication system issues that are related to the system as well". This memo helped in developing a core category - "system factors" - which combined system requirements and procedural issues along with their sub-factors.

### 6.3 Constant comparison

According to Charmaz (2006), constant comparison is described as being core to Grounded Theory. It refers to the process of constantly comparing data set to data set and its coding in order to refine the development of theory. Strauss and Corbin (1990) indicated that constant comparison reminds the researcher to constantly return to the data which can help in verifying the emerging categories as well as examining and comparing concepts for similarities and differences.

In this study, constant comparison was employed by comparing incoming data with the previous data to find out whether the same concepts appear and are relevant for the new cases and whether the codes were placed in correct category and were reliable and truly represent the empirical data.

### 6.4 Coding procedures

Strauss and Corbin (1990) defined three coding procedures in Grounded Theory which are open, axial, and selective coding.

#### 6.4.1 Open coding

An open code refers to "the analytic process through which concepts are identified and their properties and dimensions are discovered in data" (Strauss and Corbin 1998, p. 101). In this study, open coding was considered in the initial phase of the analysis process. A total of 115 open codes were created based on 15 interviews and 311 questionnaires. During this stage, the analysis was done by using phrase-by-phrase coding. In order to capture what has been said in the interview, conceptual labels were appended to almost every phrase. These labels were mostly too close to the exact words and context of the interview. Phrases with the same idea were attached with the same open code; otherwise, a new open code was created if the existing one did not fit. As the coding process continued and the researchers became well focused and confident in the process, the codings were revisited and refined so that they were reasonably understandable and provided more meaningful concepts, taking into account that they truly represented the empirical data.

However, it is important to note that some sentences represent only one concept, while others represent more the one concept. For example, the following sentence represents the code, financial benefits: "Also, it may have financial benefits later". Whereas the next sentence represents four concepts: lack of m-government services, lack of e-government services, enquiry services, and target users: "The most current m-services as well as e-services are enquiry services for both citizens and residents".

The aim of this procedure is to begin the unrestricted labelling of all data and to allocate representational and conceptual codes for all incidents highlighted within the data.

### 6.4.2 Axial coding

Axial coding refers to the process of making connections and links between the categories (Strauss and Corbin 1998). In this stage, the open codes were put together in new ways by making connections and relationships between and among codes and concepts in order to develop core codes. Figure 1 below shows the emergent categories and the relationships between them; however, further developing was made as will be discussed in the following section.

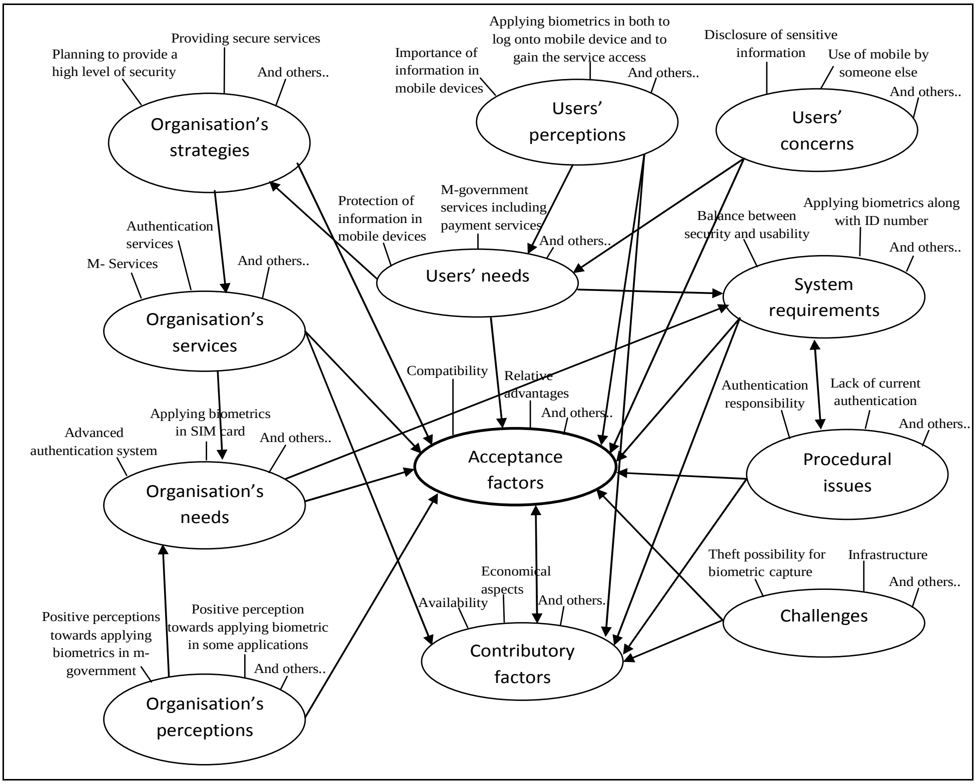

Planning to provide a high level of security
Providing secure services
And others..
Organisation's strategies
Importance of information in mobile devices
Applying biometrics in both to log onto mobile device and to gain the service access
And others..
Users' perceptions
Disclosure of sensitive information
Use of mobile by someone else
And others..
Users' concerns
Authentication services
M- Services
And others..
Organisation's services
Protection of information in mobile devices
M-government services including payment services
And others..
Users' needs
Balance between security and usability
Applying biometrics along with ID number
And others..
System requirements
Compatibility
Relative advantages
And others..
Applying biometrics in SIM card
Advanced authentication system
And others..
Organisation's needs
Acceptance factors
Authentication responsibility
Lack of current authentication
And others..
Procedural issues
Economical aspects
Availability
And others.
Contributory factors
Theft possibility for biometric capture
Infrastructure
And others..
Challenges
Positive perceptions towards applying biometrics in m-government
Positive perception towards applying biometric in some applications
And others..
Organisation's perceptions

Fig. 1. Categories and relationships

The most closely interrelated open codes were aggregated under core categories. This is was done taking into account constant comparison of the open codes listed under that axial code with each other by asking questions such as: "How do these axial codes show connections and explain factors that influence the adoption of biometric authentication in m-government security?"

Through the process of axial coding, the following twelve categories were created: acceptance factors, contributory factors, challenges, system requirements, procedural issues, users' needs, users' perceptions, users' concerns, organisation's strategies, organisation's services, organisation's needs, and organisation's perceptions. However, it is important to note that the categories of users' needs, users' perceptions, and users' concerns were developed based on the questionnaire results.

### 6.4.3 Selective coding

Selective coding can be identified as "the process of integrating and refining the theory" (Strauss and Corbin 1998, p.143). According to Strauss and Corbin (1990), the procedure of selective coding requires a selection of a focal core category which is the central phenomenon that has emerged from the axial codes and relating it to other categories in addition to validating those relationships. In this study, the acceptance factors code is the core category as illustrated in Figure 2.

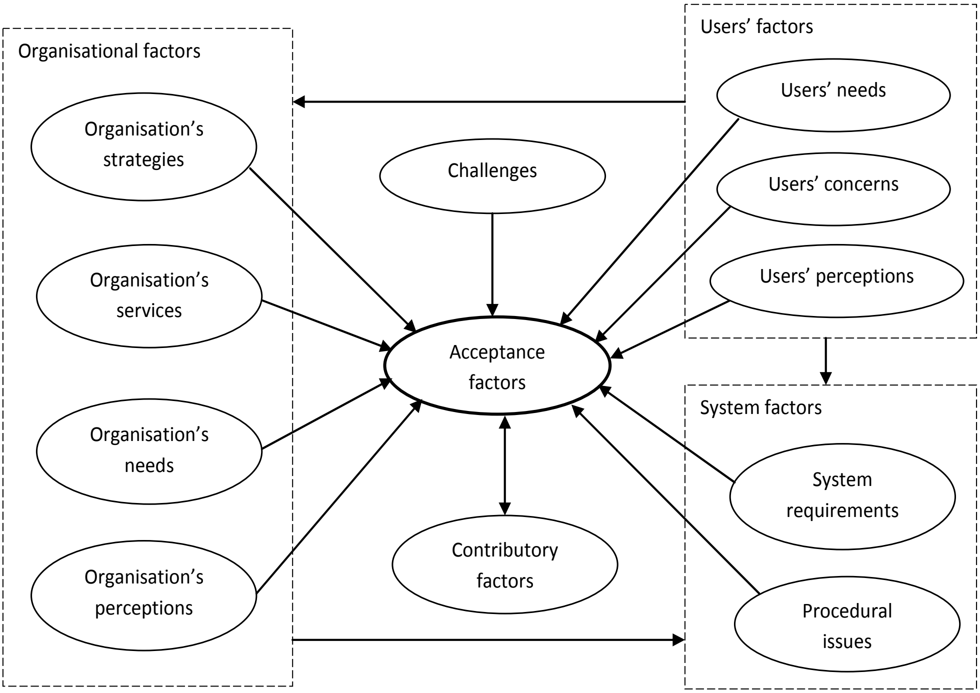


Fig. 2. Core category and relationships

The identification of the core category informs substantive theory that identifies the factors that influence the adoption of biometric authentication in m-government security. However, after becoming more familiar with the area and being well focused according to the suggestive categories, some of the categories were combined and incorporated with others, as justified in Charmaz (2006).

More specifically, system requirements and procedural issues were combined to be in the category of 'system factors', because they were all mentioned by the interviewees as part of the system factors. Similarly, the categories of organisation's strategies, organisation's services, organisation's needs, and organisation's perceptions were combined to be 'organisational factors' as all of them related to the organisation and were mentioned as the organisation's viewpoints. Acceptance factors were also combined with contributory factors to be called 'enabling factors' as all the factors in this category were mentioned as factors that would enable the achievement of successful adoption of biometric authentication in m-government security.

It is noteworthy that the presented codes and categories in the above sections are the final set of the re-coding process. As suggested by (Miles and Huberman 1994), in order to ensure that codes and categories are applied consistently, it is significant for the researchers to verify all codes and categories that are assigned to the data. The benefit of re-coding process consisted of reconfirming and refining the codes and categories. Moreover, a comparison between newly and previously assigned codes and categories assisted us as well to check whether the codes and categories were reliable and truly represented the empirical data. The initial codes were mostly too close to the exact words of the data while the final codes provide more meaningful concepts.

As evidenced from the data analysis, one concept that was frequently stressed by the participants was the idea of acceptance. It had been introduced as a category involving: relative advantages, compatibility, ease of use, trialability, observability, trust, and privacy; and had been promoted to the "core" category. This, as known in Grounded Theory methodology, was because it linked up all of the other categories.

Another important category is the category of contributory factors which include availability, awareness, legislation, economical aspects, as well as social and cultural aspects. During the data analysis, it is found that both acceptance factors and contributory factors are involved in promoting the adoption of biometric authentication in m-government security.

Furthermore, the emerged categories show that organisational factors including organisation's strategies, services, needs, and perceptions influenced by users' factors which consist of users' needs, concerns, and perceptions. At the same time, both organisational and users' factors influence acceptance factors in different manners. For instance, applying biometric authentication along with a public key infrastructure (PKI) in m-government services is an important need of service providers due to the security relative advantage of this combination between biometrics and the PKI. This application also meets the users' need for protection of their personal and sensitive information through the use of m-government services. This combination of factors can consequently be seen to positively influence the acceptance of biometrics in m-government security among both users and service providers.

Similarly, system factors including system requirements and other related system issues such as the authentication responsibility and user registration at a website influence the acceptance factors and are at the same time influenced by users and organisational factors. This is because the majority of the system requirements, for example, have basically emerged due to organisations' and users' needs and perceptions.

It was also found that several challenges influence the adoption of biometric authentication in m-government security. These challenges include technical and related challenges such as the biometric registration and enrolment process and the current lack of research into security and m-government in the Kingdom of Saudi Arabia. These challenges, which directly influence the acceptance factors, need to be considered before starting the implementation of biometrics in m-government in order to enhance the acceptance factors.

## 7. The development of the theoretical framework

Figure 3, below, illustrates the findings, focusing on the categories that have emerged from the open, axial, and selective coding phases. Based on the application of Grounded Theory, this theoretical framework, pictured, encompasses and organizes the concepts that form the factors influencing the adoption of biometric authentication.

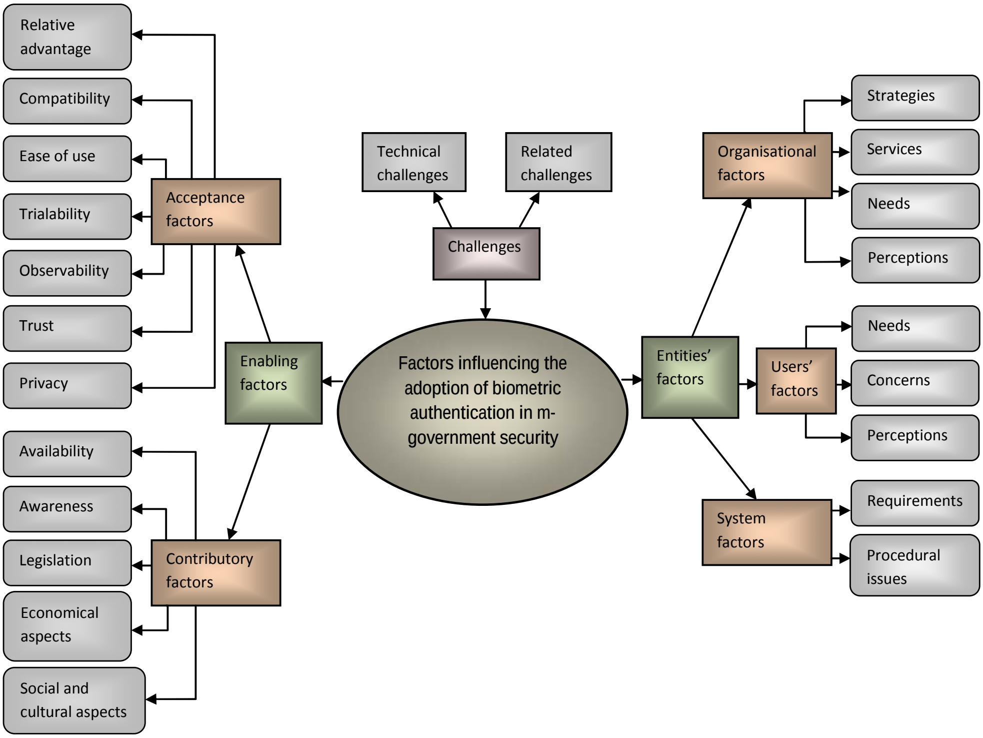


Fig. 3. Theoretical framework for the adoption of biometrics in m-government security in the KSA

Figure 3 depicts a new theoretical framework for the factors influencing adoption of biometric authentication in m-government security in Saudi Arabia, which was derived by the use of Grounded Theory as described above. Analysis and discussion of the results indicated that "entities factors", which include users', organisational, and system factors, as well as enabling factors involving acceptance and contributory factors, influence the adoption of biometric authentication in m-government security. As reported earlier in the previous section, this will also be influenced by responses to technical and non-technical challenges.

It is noteworthy that acceptance factors including relative advantage, compatibility, ease of use, trialability, observability, trust, and privacy, as well as contributory factors involving availability, awareness, legislation, economical aspects, and social and cultural aspects, are the most important factors that will enable the KSA in achieving the adoption of biometric authentication in m-government security.

The results of this study presented in this chapter are supported by a number of findings reported in literature. In particular, the findings among the acceptance factors, for example, are close to existing theories such as Diffusion of Innovation (Rogers 1995) and the Technology Acceptance Model (Davis 1989). Relative advantage, for instance, is similar to Rogers' theory of Diffusion of Innovation (1995). However, by comparing this concept with "perceived usefulness" in the Technology Acceptance Model (TAM) (Davis 1989), it seems that "relative advantage" is more accurate in representing users', service providers', and network operators' perceptions regarding the application of biometrics in m-government security. Furthermore, it is noticeable that the set of factors proposed in TAM variants and Unified Theory of Acceptance and Use of Technology (UTAUT) correspond closely with factors identified in DOI theory. While in the Technology Acceptance Model, two specific variables of perceived usefulness and perceived ease of use are hypothesized to be fundamental determinants of user acceptance (Davis 1989); the Diffusion of Innovation theory concentrates on five concepts - relative advantage, compatibility, complexity, trialability, and observability - as the five factors that affect the rate of innovation diffusion. However, the acceptance factors among the developed framework include the five concepts of DOI theory in addition to the concepts of trust and privacy as emerged from the data analysis.

Empirical studies related to the acceptance and adoption of mobile phones and electronic services via the Internet mostly indicate similar factors. For instance, Jahangir and Begum (2008) introduced a conceptual framework that considered perceived usefulness and ease of use, as well as security and privacy, as important factors that influence users' acceptance and adoption of electronic banking services. Another study by Tassabehji and Elliman (2006) highlighted trust and security as major factors in e-government adoption. AlShihi (2007) also indicated that trust has a wide impact on m-government acceptance. Kaasinen (2007) found that perceived value, ease of use, trust, and ease of adoption are important factors that influence user acceptance of mobile Internet services. However, while highly similar acceptance factors appear under various theories and models covering innovation acceptance and adoption, the developed theoretical framework in this chapter is more comprehensive. This is because it identified acceptance factors along with other factors, such as contributory and organisational and users' factors that influence adoption. Although the concepts of the developed theoretical framework in this study presented in this chapter are more comprehensive, it was created to accurately represent what different participants in Saudi Arabia including users, service providers, and network operators meant by their views, allowing it to be applied by individuals as well as organisations.

In addition, some of the identified contributory factors can be related to other findings in the literature. For example, while "availability" in this study indicates the availability of mobile devices with biometric attachment as well as m-services, Quantz (1984) identified availability of new technology as an important factor for the adoption of a new technology. Lee et al. (2002) found that social influence and self-efficacy variables significantly affect perceived usefulness and perceived ease of use for user acceptance of the mobile Internet. Teo and Pok (2003) also found that social factors, including perceptions of relative advantage, play a significant role in influencing intentions for the adoption of Wireless Application Protocol (WAP)-enabled mobile phones amongst Internet users. Furthermore, while the developed theoretical framework in this chapter classified social and cultural aspects as contributory factors that influence the adoption of biometric authentication systems in m-government security, Myers et al. (2002) stated culture as a factor that influenced users' decision to accept and adopt a particular system.

The developed theoretical framework in this chapter identified several organisational factors as contributing to m-government acceptance. Feng (2003) and Alharbi (2006) confirmed organisational issues and culture as important factors that need to be considered when applying new technology. More specifically, Feng (2003) stated that e-government projects are not a technical issue, but rather an organisational issue. A result of the study presented in this chapter indicated a lack of organisations with clear implementation strategies, and only a few organisations following Yesser's strategies regarding the implementation of e-government applications. This supports a literature finding where Al-Shehry (2008) pointed out some doubts about transforming Yesser's ideas into reality. He mentioned that some organisations failed to follow the general standards set by Yesser. He also indicated some issues that have not been adequately addressed in Yesser's strategy, such as organisational readiness, awareness, and the re-engineering of business processes. Moreover, Sahraoui et al. (2006) indicated that there is a lack of clear vision and strategy for the deployment of e-government services in Saudi Arabia. Consideration towards applying advanced levels of authentication in electronic and mobile services were also apparent among the organisational factors; which supports the finding that reveals that successful e-government strategies have to include effective security controls for the processes and systems of the government, and to ensure privacy for personal information (OMB 2002). According to Satyanarayana (2004), an e-government strategy should identify the infrastructure needs, required process transformations within government, and the technical framework, along with an indicative timeline.

Another important category among organisational factors is addressing an organisation's needs, which in this study, represent the needs of both government organisations and mobile network companies regarding the authentication of m-government security in Saudi Arabia. A number of researchers such as (Bergstrom 1987; Putnam 1987; Roberts and Pick 2004) mentioned organisational needs as a factor that affects the adoption of new technology. According to Bergstrom (1987), organisational needs influence the decision processes involving new technologies. Putnam (1987) identified organisational needs as a critical factor that may impact the success of a modernisation project in organisations where new technology is involved. Such a determined need of advanced authentication system reflects the organisations' perceptions towards the importance of security in m-government applications, which supports the literature findings of (Al-Khamayseh et al. 2006; Chang and Kannan 2002). More specifically, Al-Khamayseh et al. (2006) indicate that the security of m-government services is considered the hallmark of successful m-government, while a study by Clarke and Furnell (2005) found that additional and advanced authentication systems are required for mobile devices. Furthermore, this result relates with the literature finding where Nanavati et al. (2002) confirmed that increased security is one of the main benefits of adopting biometric authentication compared with traditional authentication methods such as PINs and tokens. Nanavati et al. (2002) found that the need for high levels of security frequently plays an important role in an enterprise's decision to deploy biometrics. A reliable authentication of the user accessing an agency’s Website is a basic requirement, since the lack of user authentication may cause serious threats through unauthorized access (Department of Commerce 2003). Moreover, while an organisation in the developed framework indicates a need to apply biometric authentication in only some advanced m-services, Nanavati et al. (2002) emphasized that the application of biometric systems should have a limited scope. Another identified need is to store biometric capture on the SIM card, which is consistent with the literature finding that suggests storing the biometric template on the smart card of the mobile device to enable users to control their biometric pattern (Giarimi and Magnusson 2002).

The theoretical framework in this chapter indicates the importance of taking into account organisations’ perceptions in order to adopt biometrics in m-government. Several studies by Putnam (1987), Ettlie (2000), and Roberts and Pick (2004) mentioned that perceptions of a specific security technology are one of the important elements in the decision to recommend the technology to an organisation. Beatty et al. (2001) stated that the more likely organisations were to perceive an innovation as consistent with their perceptions, the more likely they were to adopt it. Positive perceptions emerged, for example, towards the application of fingerprinting, which is consistent with literature findings (ORC 2002) that fingerprint scanning is the most commonly experienced technique, followed by the use of signature dynamics. This also may agree with Giesing’s (2003) study where most employees pointed to fingerprint technology as their preferred biometric.

The developed theoretical framework brought to light several users’ factors that influence the adoption of biometric authentication in m-government security. This framework identified users’ factors including their needs, perceptions and concerns, while the most recent literature mentioned only the importance of users’ factors. For example, Ashbourn (2004) stated that users can have a direct impact on the operational performance of biometric systems. They can be an essential factor in the successful implementation of biometrics (Ashbourn 2004; Giesing 2003; Scott et al. 2005). User adoption and perception problems related to the implementation of the new technology have been clarified by Giesing (2003) as a factor that would prevent an organisation from adopting biometric technology. Thus, the biometrics research community is well-advised to study the users’ side regarding the use of biometrics (Bolle et al. 2004).

Wayman et al. (2005) highlighted the importance of understanding system requirements, procedures, and other related issues, including systems management and user psychology, in order to gain successful integration of biometric systems. The developed theoretical framework in this chapter identified a number of system factors including requirements and procedural issues that need to be considered to adopt biometric authentication in m-government security in Saudi Arabia. According to Kanellis and Paul (2005), in order to have a good chance of project success for such an ICT system, system requirements need to be considered before the implementation commences.

Finally, as illustrated in Figure 3, the developed theoretical framework indicates several challenges that would influence the adoption of biometric authentication in m-government applications in Saudi Arabia. In contrast, most of the literature in the area of m-government (Al-khamayseh et al. 2006; NECCC 2001) mentioned only security and privacy as challenges of the implementation of m-government, while Lallana (2008) stated cost issues. However, there are some identified challenges that support some of the literature findings. For example, registration and enrolment processes of people's biometric credentials has been identified as one of the biggest challenges facing the adoption of biometrics in m-government, and this supports the literature finding where Hirst (2005) stated that the ease of enrolment is a determining factor for the successful implementation and use of a biometric system.

To summarise, while the developed theoretical framework (Figure 3) in this chapter supports a number of findings reported in existing theories and literature, it is unique and more comprehensive than other related existing theories such as Diffusion of Innovation (1995) and the Technology Acceptance Model (Davis 1989). It includes factors influencing the adoption of biometric authentication in m-government among mobile communication users, service providers and network operators, adding further dimensions such as contributory aspects, organisational aspects, user aspects and system aspects.

## 8. Considerations for the adoption of biometrics in m-government

The theoretical framework proposed in Figure 3 can be used to understand the factors influencing the adoption of biometric authentication in m-government security. Based on the findings presented in this chapter, several considerations can be suggested for those who are involved the adoption of biometrics in m-government applications.

First, the findings of this study revealed that, to be effective, there is a need to provide mobile communication users and service providers with an advanced authentication system for m-government services; therefore, government and decision makers should consider this need in order to enhance the adoption and implementation of m-government services. Based on the viewpoints of the participants, the application of biometric authentication would play an integral role in enhancing the security of m-government. However, it will be important for decision makers to take into account that such legislation needs to be carefully crafted to safeguard the rights of the involved entities, and the people involved will need to be aware of the new laws and regulations. Similarly, it will be important that legislation is enforced and that all parties involved in the application are well informed. Technical support will be required to make mobile devices with biometric attachments available around the country at a reasonable cost on the normal users' level.

It would be appropriate that the application of biometric authentication is implemented along with the use of a Public Key Infrastructure (PKI), and that service providers apply biometrics with only the advanced m-services to enhance usability and user acceptance. It is also important to consider the security of the templates and databases as well as the management of biometric data during enrolment, transmission, storage and authentication. The enrolment process, which essentially introduces the user to the authentication system, needs to be considered as it would affect the adoption of biometrics in m-government. A high level of cooperation between government agencies will be required to effectively introduce biometrics in m-government services. Biometric standards and systems evaluation could effectively take place before and during the implementation process.

Furthermore, to encourage mobile users to adopt and use biometrics in their mobile devices for government services, it would be strategically advantageous for awareness programs to take place before the technology is introduced. In fact, education programs for both mobile users and service providers will help to ensure awareness of the new technology, the purpose of its implementation, and its benefits. It is also important to note that privacy, relative advantage, ease of use; trialability, compatibility, and observability, play a significant role in promoting the adoption of biometric authentication in m-government applications.

## 9. Conclusions

There is a dearth of empirical studies in the adoption of biometric authentication, certainly within the context of m-government in developing countries. This chapter contributes rich insights into people's perceptions on the application of biometrics in this context, which adds weight to existing opinions on how to best apply biometrics in m-government security. The empirical data identifies several issues including the users' factors, organisational factors, system factors, acceptance factors, contributory factors, and challenges that influence the adoption of biometric authentication in m-government in the KSA context. The data also reveals that biometric authentication might best be implemented with the use of PKI in a smartcard installed in mobile devices, and preferably only for advanced government services. This is as opposed to consistently applying biometric authentication to the mobile device itself, and applying it for all online government services.

A main finding and contribution of this work is the use of Grounded Theory in the development of a new substantive theory for the adoption of biometric authentication in m-government security in the KSA. This chapter contributes to the literature by identifying and describing the factors that influence the adoption of biometric authentication in m-government applications. It further contributes to theory by providing rich insights and increased understanding of the concerns and perceptions of mobile phone users, service providers and network operators regarding the application of biometric authentication to mobile devices for government services.

This chapter adds a methodological contribution by providing a new example of the application of Grounded Theory coding procedures to develop a theoretical framework for the adoption of biometric authentication in mobile services applications in a developing country. This chapter described the application of Grounded Theory, including the development of the concepts and categories, and then the generation of the substantive theory, providing a unique insight into this qualitative information system research.

## 10. References

Abanumy, A and Mayhew, P 2005, M-government Implications For E-Government In Developing Countries: The Case Of Saudi Arabia, EURO mGOV 2005, Brighton, UK, pp. 1-6.

Alharbi, S 2006, *Perceptions of Faculty and Students toward the Obstacles of Implementing E-Government in Educational Institutions in Saudi Arabia*, a PhD thesis, West Virginia University, USA.

Al-khamayseh, S, Lawrence, E, and Zmijewska, A 2006, Towards Understanding Success Factors in Interactive Mobile Government*, Second European Conference on Mobile Government*, Brighton, UK.

Al-Shehry, A 2008, *Transformation towards e-government in the Kingdom of Saudi Arabia: technological and organisational perspectives*, a PhD thesis, De Montfort University, UK.

AlShihi, H 2007, M-government Services in Oman: Success and Failure Factors, viewed on 5th of January 2009 at
www.csdms.in/mserve/2007/fullpapers/HafedhAlShihi.pdf

Antle, K 1986, *Writing and Evaluating the Grounded Theory Research Report. In From Practice to Grounded Theory* (Chenitz W.C. & Swanson J.M., eds), Addison-Wesley, Mill Valley, California, pp. 146–154.

Antovski, L and Gusev, M 2005, M-Government Framework, *Proceedings of the First European Conference on Mobile Government*, 10-12 July, University Sussex, Brighton, UK.

Ashbourn, J 2004, *Practical biometric from aspiration to implementation*, London: Springer.

Beatty, R, Shim, J, and Jones, M 2001, Factors influencing corporate web site adoption: A time-based assessment. *Information & Management*, vol. 38, pp. 337-354.

Bergstrom, R 1987, Critical issues in CIM implementation. *CIM Technology*, pp. 5-6.

Bolle, R, Connell, J, Pankanti, S, Ratha, N and Senior, A 2004, *Guide to Biometrics*, Springer, New York.

Bonsor, K and Johnson, R, *How Facial Recognition Systems Work, How Stuff Works*, viewed on 2nd March 2008 at available at
http://computer.howstuffworks.com/facialrecognition.htm

Chang, A and Kannan, P 2002, *Preparing for Wireless and Mobile Technologies in Government*, IBM Endowment for the Business of Government.

Charmaz, K 2006, *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*, Sage, Thousand Oaks, California.

Clarke, N and Furnell, S 2005, Authentication of users on mobile telephones – A survey of attitudes and practices, *Computers & Security*, vol. 24, no. 7, pp. 519-527.

Clarke, N and Furnell, S 2007, Advanced User Authentication for Mobile Devices, *Computers & Security*, vol. 26, no. 2, pp. 109-119.

Creswell, J 1998, *Qualitative inquiry and research design: Choosing among five traditions*, Thousand Oaks, Calif.: Sage Publications.

Creswell, J 2008, *Educational research: planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, N.J., Pearson/Merrill Prentice Hall.

Dankers, J, Garefalakis, R, Schaffelhofer, R and Wright, T 2004, PKI in mobile systems, *Security for Mobility*, IEE Telecommunications 51, Mitchell, C. (ed.), The Institution of Electrical Engineers, UK, pp. 11-33, 2004.

Davis, F 1989, Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly*, 13(3): 319-339.

Department of Commerce (DC) 2003, *Information Security Guideline for NSW Government - Part 2 Examples of Threats and Vulnerabilities*, the Office of Information and Communications Technology, Australia.

E-government Program (Yesser) 2011, The Ministry of Communications and Information Technology, available at http://www.yesser.gov.sa

Ettlie, J 2000, *Managing technological innovation*, New York, John Wiley and Sons.

Feng, L 2003, Implementing E-government Strategy is Scotland: Current Situation and Emerging Issues, *Journal of Electronic Commerce in Organizations*, vol. 1, no. 2, pp. 44-65.

Frees, R 2008, Biometric technology improves identification security, U.S. Air Force, viewed on 3rd December 2008 at http://www.af.mil/news/story.asp?id=123084564

Giarimi, S and Magnusson, H 2002, *Investigation of User Acceptance for Biometric Verification/Identification Methods in Mobile Units*, Department of Computer and Systems Sciences, Stockholm University, Sweden.

Giesing, I 2003, *User Perceptions Related to Identification Through Biometrics within Electronic Business*, University of Pretoria.

Glaser, B 1992, *Emergence vs Forcing: Basics of Grounded Theory Analysis*, Sociology Press, Mill Valley, CA.

Glaser B and Strauss A 1967, The *Discovery of Grounded Theory: Strategies for Qualitative Research*, Aldine, Chicago.

Goulding, C 2002, *Grounded Theory: A Practical Guide for Management*, Business and Market Researchers London, Sage.

Harris, A and Yen, D 2002, Biometric authentication: Assuring access to information, *Information Management and Computer Security*, vol. 10, no. 1, pp. 12-19.

Heath, H and Cowley, S 2004, Developing a Grounded Theory approach: a comparison of Glaser and Strauss, *International Journal of Nursing Studies*, vol. 41, pp. 141-150.

Hirst, C 2005, *A Primer on Biometric Technologies*, Gartner Research.

Jahangir, N and Begum, N 2008, The role of perceived usefulness, perceived ease of use, security and privacy, and customer attitude to engender customer adaptation in the context of electronic banking, *African Journal of Business Management*, vol. 2 (2), pp. 032-040.

Jain, A, Hong, L and Pankanti, S 2000, 'Biometric identification', *Communications of the ACM*, vol. 43, no. 2, pp. 90-98.

Kaasinen, E 2007, 'User acceptance of mobile Internet services', In Workshop on Mobile Internet User Experience, *Mobile HCI 2007 Conference*, September 9, 2007, Singapore.

Kanellis, P and Paul, R 2005, User behaving badly: Phenoma and paradoxes from an investigation into Information systems misfit, *Journal of Organisational and End-use Computing*, vol. 17, no. 2, pp. 264-291.

Kleist, V, Riley, R and Pearson, T 2005, Evaluating biometrics as internal control solutions to organizational risk, *Journal of American Academy of Business*, vol. 6, no. 2, pp. 339–343.

Lallana, E 2008, *mGovernment: Mobile/Wireless Applications in Government, eGovernment for Development Information Exchange*, University of Manchester's Institute for Development Policy and Management, UK, viewed on 2nd February 2009 at Hhttp://www.egov4dev.org

Lease, D 2005, *Factors influencing the adoption of biometric security technologies by decision making information technology and security managers*, a PhD thesis, Capella University, USA.

Lee, W, Kim, T and Chung, J 2002, *User acceptance of the mobile Internet*, In M-Business 2002, Athens, Greece.

Mansourian, Y 2006, *Adoption of Grounded Theory in LIS research*, New Library World, vol. 107, no. 9/10, pp. 386-399.

McLeod, J 1999, *Practitioner Research in Counselling*, London: Sage.

Miles, M and Huberman, A 1994, *An Expanded Sourcebook - Qualitative Data Analysis*, 2nd ed. Thousand Oaks, California, SAGE Publications, Inc.

Moore, G and Benbasat, I 1991, 'Development of an instrument to measure the perceptions of adopting an information technology innovation', *Information Systems Research*, vol. 2, no. 3, pp. 192-222.

Myers, M and Tan, F 2002, Beyond Models of National Culture in Information Systems Research, *Journal of Global Information Management*, vol. 10, no. 1, pp. 24-32.

Nanavati, S, Thieme, M, and Nanavati, R 2002, *Biometrics: Identity verification in a networked world*, New York, John Wiley and Sons, Inc.

National Electronic Commerce Coordinating Council (NECCC) 2001, *M-Government: The Convergence of Wireless Technologies and e-Government*, Research and Development Workgroup.

OMB. (2002). *E-government Strategy: Executive Office of the President Office of Management and Budget*, Washington, D. C.

ORC, Opinion Research Corporation 2002, *Public Attitudes Toward the Uses of Biometric Identification Technologies by Government and the Private Sector*, Summary of Survey Findings.

Piantanida, M, Tananis, C, and Grubs, R 2002, Claiming Grounded Theory for practice-based dissertation research: A think piece, *Conference on Interdisciplinary Qualitative Studies*, Athens, Georgia.

Putnam, R 1987, *Selling modernization within your company*, Commline, 13.

Quantz, P 1984, *CIM planning: The future-factory foundation,* CIM Review, 38.

Roberts, G and Pick, J. 2004, Technology factors in corporate adoption of mobile cell phones: A case study analysis. *Proceedings of the IEEE 37th Annual Hawaii International Conference on System Sciences*, 9(9), 90287-90296.

Rogers, E 1995, *Diffusion of Innovations*, The Free Press, New York.

Sadad 2008, payments newsletter, SADAD Payment System, vol.2, issue 10, available online at http://www.sadad.com/Arabic/

Sahraoui, S, Gharaibeh, G, and Al-Jboori, A 2006, Government in Saudi Arabia can it overcome its challenges?, Brunel University, London, paper presented at Government Workshop '06 (eGOV06).

Satyanarayana, J 2004, e-Government - *The Science of the Possible*, New Delhi, Prentice-Hall of India.

Scott, M, Acton, T and Hughes, M 2005, An assessment of biometric identities as a standard for e-government services, *Services and Standards*, vol. 1, no. 3, 2005, pp. 271-286.

Strauss, A and Corbin, J 1990, *Basics of qualitative research: Grounded Theory procedures and techniques*, Newbury Park, CA: Sage Publications, Inc.

Strauss, A and Corbin, J 1994, Grounded Theory Methodology: An Overview, In Denzin, N. and Lincoln, Y (Eds.). *Handbook of Qualitative Research*. Thousand Oaks, Sage Publications.

Strauss, A and Corbin, J 1998 *Basics of Qualitative Research: Techniques and Procedures for Developing Theory*, 2nd ed., Sage, Thousand Oaks, CA.

Tassabehji, R and Elliman, T 2006, Generating Citizen Trust in E-Government Using a Trust Verification Agent: A Research Note, *European and Mediterranean Conference on Information Systems (EMCIS)*, Costa Blanca, Alicante, Spain

Teo, T and Pok, S 2003, Adoption of WAP-enabled mobile phones among Internet users. Omega: *The International Journal of Management Sciences*, vol. 31, no. 6, p. 483-498.

Urquhart, C and Fernandez, W 2006, Grounded Theory Method: The Researcher as Blank Slate and Other Myths, *In Twenty-Seventh International Conference on Information Systems*, pp 457-464, Milwaukee.

Urquhart, C Lehmann, H and Myers, M 2009, Putting the theory back into Grounded Theory: guidelines for Grounded Theory studies in information systems, *Information systems journal*, Blackwell Publishing Ltd.

Uzoka, F.M. and Ndzinge, T 2009, An Investigation of Factors Affecting Biometric Technology Adoption in a Developing Country Context, *International Journal of Biometrics*, vol. 1, no. 3, pp. 307-328.

Venkatesh, V, Morris, M, David, G and Davis, F 2003, 'User acceptance of information technology: toward a unified view', *MIS Quarterly*, vol. 27, no. 3, pp. 425-78.

Vielhauer, C 2006, *Biometric User Authentication for IT Security: From Fundamentals to Handwriting*, Springer, New York.

Wayman, J, Jain, D, Maltoni, H and Maio, D 2005, *Biometric Systems: Technology, Design and Performance Evaluation*, Springer, New York.

World Fact Book 2011, Central Intelligence Agency, available at https://www.cia.gov/index.html

11

# Building Expert Profiles Models Applying Semantic Web Technologies

Valentina Janev and Sanja Vraneš
*University of Belgrade, Mihajlo Pupin Institute*
*Serbia*

## 1. Introduction

Semantic Web (SW) is an emerging research field that has application in different domains such as e-government services, richly interlinked library systems, Web search engines, enterprise knowledge stores, and other. The term "Semantic Web" refers to the World Wide Web Consortium's (W3C) vision of the Web of linked data (called also the Web of Data) as "…an extension of the current Web in which information is given a well-defined meaning, better enabling computers and people to work in cooperation" (Berners-Lee, Hendler, & Lassila, 2001). Since then, many specifications, guidelines, languages, and tools have been developed that facilitate software development, improve performance and create new business opportunities.

This Chapter investigates the existing standard models for representing Human Resource (HR) data and especially in the context of competence management and discusses the process of building and publishing ontology-based expert profiles models. The expert profiles can be used for deploying advanced HR management services for in-house purposes, as well as integration and search of experts in the Linked Open Data (LOD) cloud. During the past 3 years, the Linked Data paradigm, promoted by Tim Berners-Lee (2006) as a part of the Semantic Web roadmap, evolved from a practical research idea into a very promising candidate for addressing one of the biggest challenges in the area of intelligent information management: the exploitation of the Web as a platform for data and information integration in addition to document search (Auer & Lehmann, 2010; Janev & Vraneš, 2011b). According to the statistics from March 2012 (see http://stats.lod2.eu/), the LOD cloud contains over 1,7 billion triples, integrates over 400000 namespaces and registers over 1,7 million instances of type "person".

The Chapter is organized as follows. After introducing the main requirements for managing competence data in enterprises in Section 2, we point to existing and emerging HR approaches and data representation standards (Section 3). Then, follows an example process for building an ontology-based enterprise knowledge stores (Section 4) and two examples for searching enterprise knowledge stores with Semantic Web tools (Section 5).

## 2. HR management requirements from business and technical perspective

Human resource management (HRM) is one of the basic business processes that consists of a wide range of administrative, organizational and employee acquisition and development activities. Administrative activities include management of different employee records (personal data, qualifications, holidays, business trips) and legal procedures of hiring / dismissal, as well as payment processing. Organizational activities cover strategic issues of enterprise organizations, systematization of working places, planning of team work, team formation and development, etc. Employee acquisition and development activities are directed towards definition of requirements, and standards that employees have to fulfil prior to employment, planning of necessary resources, education and development of employees, and employee performance measurement.

Herein, we would like to discuss the requirement for a comprehensive knowledge model for competence management from business and technical perspective.

### 2.1 Competence management business requirements

Competence management (CM) is an important research object in the more general area of human resource management. The idea of "competency" into the HR literature was introduced by the Harvard's psychologist David McClelland in early seventies of the last century (McClelland, 1973) and since then, development and use of competency based approaches within the corporate environment has been rapid (Draganidis & Mentzas, 2006).

#### 2.1.1 Competence management on company level

Companies adopt different competency models and start competence and skills management initiatives in order to create a setting for the empowerment of their workforce and thus increase competitive advantage, innovation, and effectiveness (Houtzagers, 1999). A competency model is a list of competencies which are derived from observing satisfactory or exceptional performance for a specific occupation or task. Related to in-house competence management mainly aimed at building individual competence models are the following requirements:

- building central repositories which define competencies for certain communities;
- building services for identifying experts and finding out and continually recording what people ("experts") in an organization know ("expertise");
- making expertise available to users so they can answer questions or solve problems that exceed personal or workgroup capabilities;
- expertise gap analysis;
- planning the expertise development paths; etc.

#### 2.1.2 Competence management for cooperation and integration of activities with partners on national and international level

In order to be competitive in the global knowledge economy, companies organize themselves in partner networks or even virtual enterprises that require interlinking of activities, or even existing information systems. Business processes in such networks often spawn different specific tasks that are to be solved by the network members. Therefore, it is essential that partner organizations prove themselves with complementary competencies both on an expert and an organizational level. Developing and maintaining competence profiles of all the relevant parties associated with specific task and topic can significantly improve the performance of the partner network or the virtual organization. Related to inter-enterprise cooperation, interoperability and integration are the following requirements:

- standard description of occupations and competences ;
- using multi-lingual dictionaries for building expert profiles;
- interoperability of knowledge models with similar schemas on the Web; etc.

### 2.2 Technical requirements

In order to achieve transparency and comparability of expertise, organizations need tools and technologies to express the core competencies and talents of employees in a standardized, machine processable and understandable format. Based on the competence management business requirements briefly introduced above, we can distinguish three types of expertise management services, namely (1) expert profiling and search, (2) organization profiling and search and (3) knowledge items search and retrieval. Technologies that play a role in implementation of these services originate from the fields of open systems architecture, Web services, information retrieval, data and text mining, clustering, natural language processing, ontology building, etc.

## 3. State-of-the-art analysis of HR standards and literature

### 3.1 Analysis of classical approaches to expertise management

The actual HRM solutions mainly focus on the integration of the distributed legacy databases, typically in the form of the data warehouse where the fact data (i.e. employee data) is arranged in order to answer the analytical queries efficiently. Personal profiles here usually rely on the self declared expertise. Employees keep track of their areas of expertise manually by maintaining a list of keywords or phrases and this list of key qualifications is being defined in the HR sector. This approach is error-prone since users are typically subjective/biased and reluctant to update the file regularly. Also, manually created lists cannot be an exhaustive description of the person's expertise areas. In addition, content based approaches (Sim et al., 2006) to expertise extraction, profiling and finding have been introduced lately that focus on the automatical identification of the expertise entities in the semi-structured and unstructured documents containing the expertise information as well as on the annotation of the identified expertise entities with the semantic mark-up. The input documents are: (1) curricula vitae and résumé that have been published in formats such as text, PDF, DOC and HTML; (2) publications and other legacy documents (Balog et al., 2006; Balog & de Rijke, 2008); (3) e-mails, blog sites and other online social networking related context (Aleman-Meza et al., 2007; Schäfermeier & Paschke, 2011). The expertise extraction and profiling is based on the linguistic analysis, statistical and machine learning classification methods as well as on the inductive logic programming techniques to discover rules for extracting fields from documents (Fang & Xiang Zhai, 2007; Petkova & Bruce Croft, 2006; Jung et al., 2007). Inspired by different research fields such as expert finding, competency management, terminology extraction, keyword extraction and concept extraction (Bordea, 2010), Bordea and Buitelaar (2010) proposed a hybrid approach and the Saffron system for expert profiling and finding.

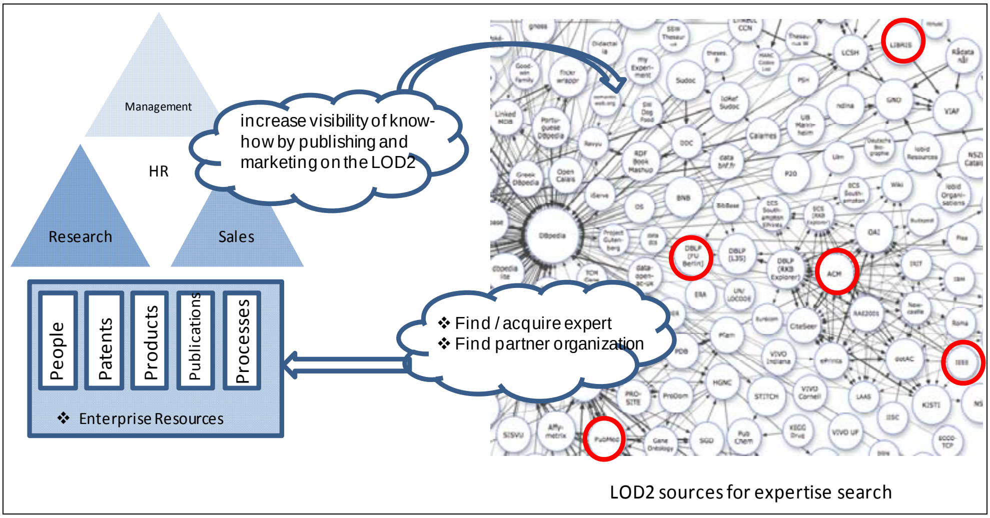

Management
HR
Research
Sales
increase visibility of know-how by publishing and marketing on the LOD2
People
Patents
Products
Publications
Processes
❖ Enterprise Resources
❖ Find / acquire expert
❖ Find partner organization
LOD2 sources for expertise search

Fig. 1. Linking enterprise resources to the LOD cloud.

### 3.2 Review of standards and literature for ontology-based competence management

European Union, through its chief instruments for funding research (FP5 - The Fifth, FP6 – The Sixth and FP7 – The Seventh Framework Programs), has financed several projects that focused on ontology-based competency management. As a result of these projects, several prototype systems have been developed (Bizer et al., 2005; Draganidis et al., 2006) and few ontologies were made publicly available. The developed HR ontologies (Bizer et al., 2005; Müller-Riedlhuber, 2009) are based on widespread used standards and classifications of job profiles and industry sectors such as SOC (Standard Occupational Classification System, www.bls.gov/soc/), NAICS (North American Industry Classification System, see http://www.census.gov/epcd/www/naics.html), NACE (Statistical Classification of Economic Activities in the European Community, see http: //ec. europa. eu/ eurostat/ ramon/), HR-XML (HR-XML Consortium, www.hr-xml.org) and other.

Schmidt & Kunzmann (2006) developed the Professional Learning Ontology that formalizes competencies as a bridge between human resource development, competence and knowledge management as well as technology-enhanced learning. In (Bizer et al., 2005), Bizer developed a HR ontology, an e-recruitment prototype and argued that using Semantic Web technologies in the domain of online recruitment could substantially increase market transparency, lower the transaction costs for employers, and change the business models of the intermediaries involved. In (Paquette, 2007), the author presented a competency ontology and the *TELOS Software Framework for Competency Modelling and Management*.

Furthermore, the research work in the competence management domain in the last decade had a positive impact on several European Public Employment Services, e.g. see DISCO project (Müller-Riedlhuber, 2009). Some of them have already introduced (e.g. Germany, Norway) or are at the moment working on improvements of their matching (vacancies and job seekers) processes by shifting more emphasis to competences.

| Acronym | HR Initiative |
|---|---|
| HR-XML | HR-XML Consortium Competencies Schema, http://ns.hr-xml.org/ |
| SOC | The 2010 Standard Occupational Classification (SOC, www.bls.gov/soc) system is used by Federal statistical agencies to classify workers into occupational categories for the purpose of collecting, calculating, or disseminating data. All workers are classified into one of 840 detailed occupations according to their occupational definition. To facilitate classification, detailed occupations are combined to form 461 broad occupations, 97 minor groups, and 23 major groups. |
| O*NET | The Occupational Information Network (O*NET, http://www.onetcenter.org, based on SOC) is designed to be the nation's most comprehensive resource of occupational information, with a database system that includes 275 descriptors about each occupation. |
| DISCO | European Dictionary of Skills and Competencies, financed by EU Leonardo da Vinci programme & the Austrian Federal Ministry for Education, the Arts and Culture, http://www.skills-translator.net/ |
| e-CF | European e-Competence Framework, a reference framework of 32 ICT competences, http://www.ecompetences.eu/. |
| ESCO | The European Skills, Competences and Occupations taxonomy (under development). A partial classification is already in use in the European job mobility portal EURES (http://ec.europa.eu/eures/). It exists in 22 languages and currently contains around 6000 skill descriptions and 5000 job titles. |

Table 1. International HR initiatives

## 4. Explicit representation of an HR knowledge store

To represent information on the Web and to ensure interoperability between applications that exchange machine-understandable information, the Semantic Web uses the Resource Description Framework (RDF) as a general-purpose language. RDF describes information in terms of objects ("resources") and the relations between them via the RDF Schema, which serves as a meta-language or vocabulary to define properties and classes of RDF resources. The next layer on top of the RDF/RDFS data model serves to formally define domain models as shared conceptualizations, also often called ontologies (Gruber, 1993). Ontologies are nowadays very often used for building integrated inter- and intra-organization business services, and to make the search and retrieval both efficient and meaningful. In this Section we will use the RDF and OWL languages to introduce the most important concepts and relations between concepts relevant for building an expert profile (see Figure 2).

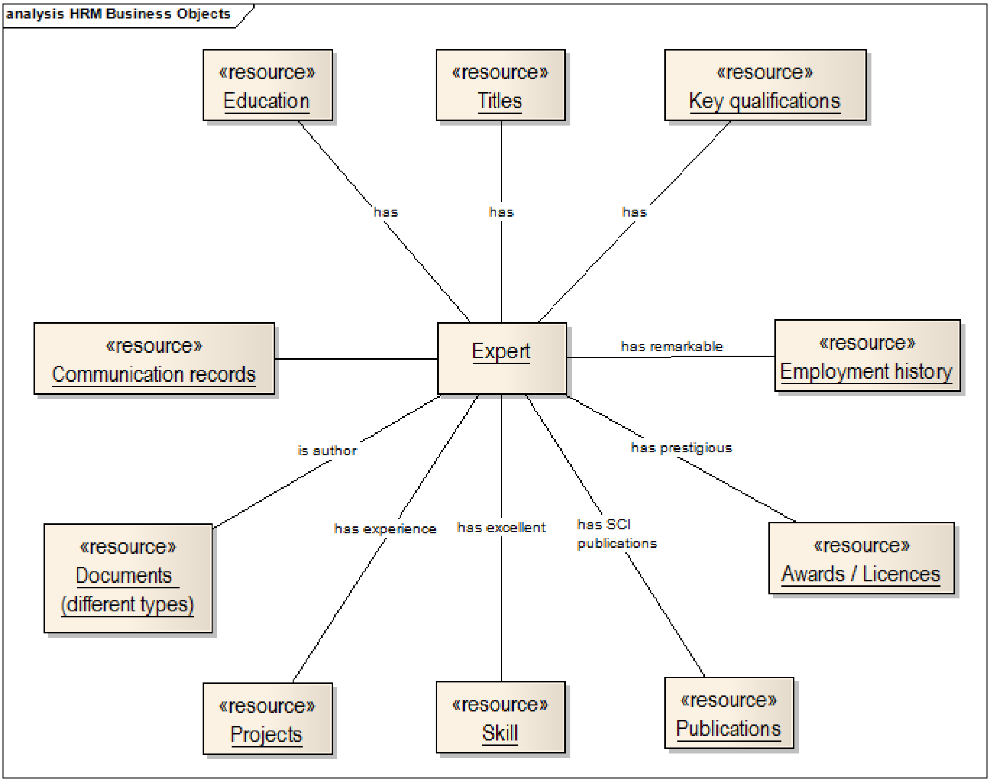


Fig. 2. UML representation of the concept Expert.

### 4.1 Creating a new ontological model

Unlike conventional object-oriented conceptual models like UML where attributes are bound to a specific class, classes and properties (as main entities of ontology) are equally important for ontology building. Therefore, prior to making a decision about the knowledge model design and structure, one have to enumerate the important terms that will be used, e.g. for the HR domain these are *Person*, *Organization*, *Document*, *Project*, *Publication*, *Author*, *Competence*, *Experience*, etc. After that, separate generalization hierarchies for classes and properties are designed. There are several possible approaches in developing a class hierarchy: the top-down, the bottom-up and the combination development process. The top-down development process starts with the definition of the most general concepts in the domain and subsequent specialization of the concepts. The bottom-up development process starts with the definition of the most specific classes, the leaves of the hierarchy, with subsequent grouping of these classes into more general concepts. The combination development process is a combination of the top-down and bottom-up approaches.

In a top-down manner, we can define the most general concepts/properties as subclasses / subproperties of entities from the public vocabularies FOAF, DOAC, and BibTeX (Aleman-Meza et al., 2007)and assign them meaning identical with the existing commonly used classes in the Semantic Web (see Table 2). In that way, the main "components" are defined as subclasses of the public concepts (*foaf:Person*, *foaf:Organisation*, *foaf:Document*, *foaf:PersonalProfileDocument*, *doac:Education*, *doac:Skill*, *doac:Experience, bibtex:Entry*), while links/relations between the components are defined as sub-properties of *foaf:interest*, *foaf:made/maker*, *foaf:topic*, *foaf:primaryTopic*, *foaf:homepage*, etc. Additional classes and properties specific to the domain of interest (e.g. in the ICT domain) can be defined manually with elements from the RDF Schema (www.w3.org/TR/rdf-schema/) or defined automatically in bottom-up manner e.g. using D2RQ server, http://www4.wiwiss.fu-berlin.de/bizer/d2r-server.

| Acronym | RDF model |
|---|---|
| SIOC | The SIOC initiative (Semantically-Interlinked Online Communities, http://www.w3.org/Submission/2007/02/) aims to enable the integration of the online community information. SIOC provides the Semantic Web ontology for representing rich data from the Social Web in RDF. |
| FOAF | The FOAF (Friend of a Friend, http://www.foaf-project.org/) project is about creating a Web of machine-readable pages describing people, the links between them and the things they create and do. |
| DOAC | DOAC (Description Of A Career, DOAC Vocabulary specification, http://ramonantonio.net/doac/0.1/) is a vocabulary used for describing professional capabilities of a worker. It was designed to be compatible with the Europeans Curriculum so that those can be generated from a FOAF+DOAC file. |
| DOAP | DOAP (Description of a Project, DOAP Vocabulary specification (http://trac.usefulinc.com/doap/ ) is a RDF schema and XML vocabulary used for describing software projects and, in particular, open source. |
| Dublin Core | The Dublin Core Metadata Initiative (http://dublincore.org/) is an open organization engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models. |

Table 2. RDF models

For example, the Mihajlo Pupin Institute ontology (MPI) uses concepts from the DOAC+FOAF vocabulary and extends them with new concepts and properties defined in the *imp* and *skills* namespace as follows:

- general description of an expert (*imp:Person rdfs:subClassOf foaf:Person*);

- general description of an organization (*imp:Organization rdfs:subClassOf foaf: Organization*) and a community (*foaf:Group*);
- *imp:PersonalProfileDocument*, based on *foaf:PersonalProfileDocument* for expertise data integration on employee level;
- *imp:RnDProfile*, a disjoint concept of the *foaf:PersonalProfileDocument* for MPI core competences integration on organizational level;
- description of education (*doac:Education*) and skills (*skills: ComputerSkill, skills: LanguageSkill, skills: EngineeringSkills, skills: OrganizationalSkill, doac: SocialSkill*);
- general description of a document (*foaf:Document*);
- personal profile document (*imp:PersonalProfileDocument*);
- R&D profile document (*imp:RnDProfile*);
- various kinds of experience (*imp:WorkingExperience, imp:ScientificExperience rdfs:subClassOf doac:Experience*)
- relations between a person and his/her profile documents and expertise (*foaf:primaryTopic, foaf:topic, imp:topic_interest_project, imp:topic_interest_reference, imp:keyQualifications, imp:responsibilities OnProjects, imp:hasScientificRecord*);
- relations between a organization and its profile document (*foaf:primaryTopic*, *foaf:topic*, *foaf:homepage*);
- relations between a person and his/her expertise (*imp:degree, imp:graduationTitle, imp:useDBMS, imp:useModellingTool, imp:useProgrammingLanguage*);
- relations between a person and the document base (*foaf:workInfoHomepage, foaf:workplaceHomepage*); etc.

### 4.2 From implicit to explicit data representation

After the ontological knowledge base is designed, the next step is to populate the ontology i.e. import data into the ontology and create instances. Manually creating of ontologies is a time consuming task. Semantic Web community has delivered many high-quality open-source tools that can be used for automatic or semiautomatic ontology population i.e. to convert the facts trapped in the legacy systems or business documents into information understandable both for machines and people.

Professional HRM systems, e.g. the SAP Human Capital Management solution, cover the whole life-circle of an employee from her/his recruitment, training, development, and deployment to retirement. They enable tracking of employee movements and adequate tracking of changes in organizational structure. Furthermore, standard SAP HCM processes support skill management and give managers and HR professionals reporting and analysis options that provide a real-time insight into employee qualifications. As a result, the underlying (implicit) data base model is highly normalized and quite complex. Customizing the predefined SAP HCM functionalities or extending them with new client tailored functionalities require SAP consultancy efforts. Therefore, extracting the HR data in explicit format and enriching them with semantic information will make the data easily accessable and processable in other business applications. Table 3 gives an example how specific groups of data, called “infotypes” in SAP terminology, can be mapped to public or in-house defined domain classes.

| | | |
|---|---|---|
| SAP HCM - Personnel Administration | | |
| Organizational Data | | |
| IT-0001 | Organizational Assignment | *imp:inOrganization* (*imp:Organization*) |
| IT-0034 | Corporate Function | *imp:EmploymentType* |
| IT-0016 | Contract Elements | *imp:Document* |
| Personal data | | |
| IT-0002 | Personal Data | *imp:Person* |
| IT-0006 | Addresses | *imp:Address* |
| IT-0009 | Bank Details | *imp:Bank* |
| IT-0021 | Family / Related Person | *imp:hasFamilyMember* (*imp:Person*) |
| IT-0022 | Education | *doac:education* (*doac:Education*) |
| IT-0023 | Other/Previous Employers | *imp:PartTimeEmployment, imp:referer* |
| IT-0024 | Qualifications | *imp:Skill, imp:LanguageSkill, imp:refer* |
| IT-0105 | Communication | *foaf:holdsAccount* (*foaf:OnlineAccount*)<br>*foaf:phone*<br>*foaf:homepage* (*foaf:Document*) |
| IT-0185 | Personal ID | *imp:globalID* |
| SAP HCM - Organizational Management | | |
| P010 | Organization | *imp:Organization* |
| P013 | Position | *imp:JobPosition* |
| The Researcher file | | |
| IT-9110 | MPI scientific titles | *imp:hasScientificRecord* (*imp: ScientificExperience*)<br>*imp:graduationTitle xsd:string* |
| IT-9120 | Postgraduates studies - details | *doac:education* (*doac:Education*) |
| IT-9130 | Key qualifications and Areas of Expertise | *imp:keyQualification xsd:string* |
| IT-9140 | Memberships in scientific organizations | *imp:isMemberOf* (*foaf:Organization*) |
| IT-9150 | Awards, Appreciations | *imp:hasAward* |
| IT-9160 | Projects | *imp:responsibilitiesOnProjects* (*imp:ProjectReference*) |
| IT-9170 | References | *imp:ScientificPaper* |

Table 3. Establishing correspondence between implicit and explicit data representation

Figure 3 represent a screenshoot of mapping the facts from RDBMS tables to instances explicitly represented in the Institute „Mihajlo Pupin“ knowledge store (Janev & Vraneš, 2011a).

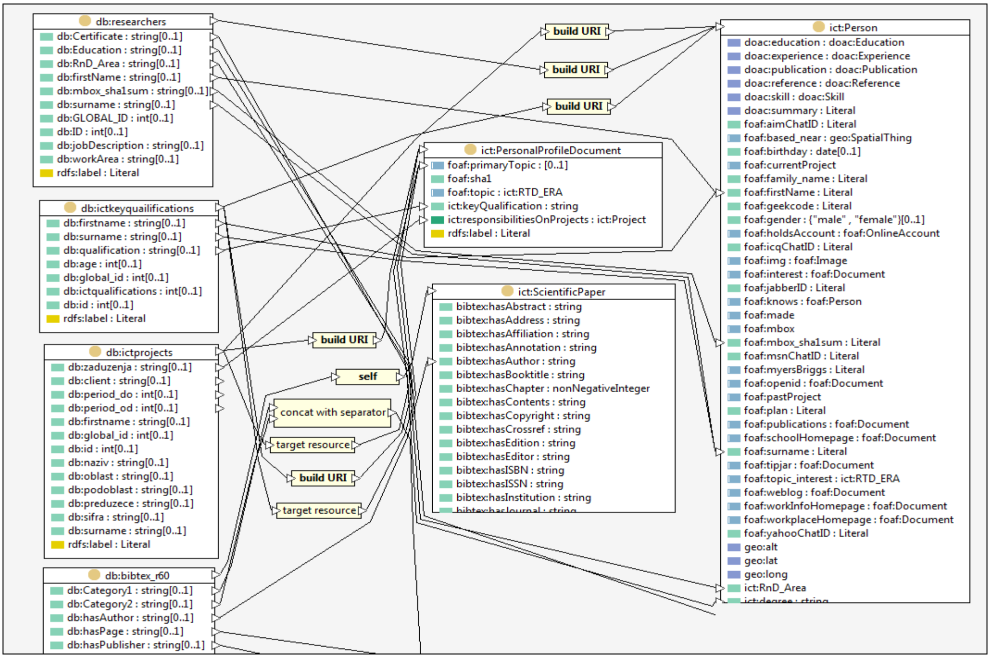

db:researchers
db:Certificate : string[0..1]
db:Education : string[0..1]
db:RnD_Area : string[0..1]
db:firstName : string[0..1]
db:mbox_sha1sum : string[0..1]
db:surname : string[0..1]
db:GLOBAL_ID : int[0..1]
db:ID : int[0..1]
db:jobDescription : string[0..1]
db:workArea : string[0..1]
rdfs:label : Literal
db:ictkeyqualifications
db:firstname : string[0..1]
db:surname : string[0..1]
db:qualification : string[0..1]
db:age : int[0..1]
db:global_id : int[0..1]
db:ictqualifications : int[0..1]
db:id : int[0..1]
rdfs:label : Literal
db:ictprojects
db:zaduzenja : string[0..1]
db:client : string[0..1]
db:period_do : int[0..1]
db:period_od : int[0..1]
db:firstname : string[0..1]
db:global_id : int[0..1]
db:id : int[0..1]
db:naziv : string[0..1]
db:oblast : string[0..1]
db:podoblast : string[0..1]
db:preduzece : string[0..1]
db:sifra : string[0..1]
db:surname : string[0..1]
rdfs:label : Literal
db:bibtex_r60
db:Category1 : string[0..1]
db:Category2 : string[0..1]
db:hasAuthor : string[0..1]
db:hasPage : string[0..1]
db:hasPublisher : string[0..1]
build URI
build URI
build URI
build URI
self
concat with separator
target resource
build URI
target resource
ict:PersonalProfileDocument
foaf:primaryTopic : [0..1]
foaf:sha1
foaf:topic : ict:RTD_ERA
ict:keyQualification : string
ict:responsibilitiesOnProjects : ict:Project
rdfs:label : Literal
ict:ScientificPaper
bibtex:hasAbstract : string
bibtex:hasAddress : string
bibtex:hasAffiliation : string
bibtex:hasAnnotation : string
bibtex:hasAuthor : string
bibtex:hasBooktitle : string
bibtex:hasChapter : nonNegativeInteger
bibtex:hasContents : string
bibtex:hasCopyright : string
bibtex:hasCrossref : string
bibtex:hasEdition : string
bibtex:hasEditor : string
bibtex:hasISBN : string
bibtex:hasISSN : string
bibtex:hasInstitution : string
bibtex:hasJournal : string
ict:Person
doac:education : doac:Education
doac:experience : doac:Experience
doac:publication : doac:Publication
doac:reference : doac:Reference
doac:skill : doac:Skill
doac:summary : Literal
foaf:aimChatID : Literal
foaf:based_near : geo:SpatialThing
foaf:birthday : date[0..1]
foaf:currentProject
foaf:family_name : Literal
foaf:firstName : Literal
foaf:geekcode : Literal
foaf:gender : {"male" , "female"}[0..1]
foaf:holdsAccount : foaf:OnlineAccount
foaf:icqChatID : Literal
foaf:img : foaf:Image
foaf:interest : foaf:Document
foaf:jabberID : Literal
foaf:knows : foaf:Person
foaf:made
foaf:mbox
foaf:mbox_sha1sum : Literal
foaf:msnChatID : Literal
foaf:myersBriggs : Literal
foaf:openid : foaf:Document
foaf:pastProject
foaf:plan : Literal
foaf:publications : foaf:Document
foaf:schoolHomepage : foaf:Document
foaf:surname : Literal
foaf:tipjar : foaf:Document
foaf:topic_interest : ict:RTD_ERA
foaf:weblog : foaf:Document
foaf:workInfoHomepage : foaf:Document
foaf:workplaceHomepage : foaf:Document
foaf:yahooChatID : Literal
geo:alt
geo:lat
geo:long
ict:RnD_Area
ict:degree : string

Fig. 3. Defining mapping rules with TopBraid SPINMap SPARQL-based language.

## 5. Publishing and searching enterprise knowledge stores with SW tools

Once represented as an RDF data store, the HR data can be linked in the (LOD) cloud and become available for further exploitation. Herein, we would like to discuss two possibilities for searching RDF models:

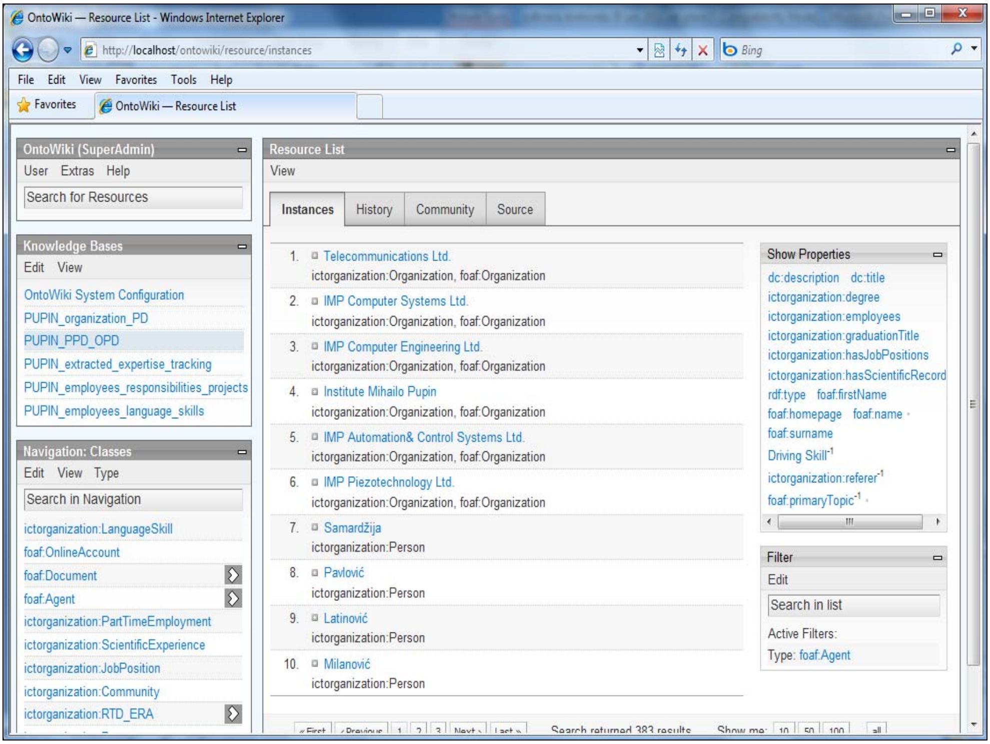
1. using OntoWiki tool (Auer, 2007);
2. using Sig.ma, a service and an end user application to access the Web of Data (Tummarello et al., 2010).

### 5.1 Navigating and querying the semantic knowledge models with *OntoWiki*

In order to publish the developed ontological models on the Web, maintain, search and retrieval in efficient and meaningful way the OntoWiki Knowledge Engineering open-source tool can be used (see ontowiki.net). The main goal of the *OntoWiki* is to facilitate the visual presentation of a knowledge base as an information map, with different views on instance data. *OntoWiki* provides a generic user interface for arbitrary RDF knowledge bases. Each node at the information map, e.g. RDF resource *foaf:Person*, is represented as a Web accessible page and interlinked to related digital resources, e.g. using the *rdfs:subClassOf* semantic property to other RDF resource *foaf:Agent*.

Selection opportunities include (see Fig.4):

- Semantically Enhanced Full-text Search (see the "Search" panel in the upper left corner); A semantic search has significant advantages compared to conventional full-text searches. By detecting classes and properties that contain the matched keywords, the semantic search delivers important feedback to the user how the search may be successfully refined;
- Browsing using semantic relations (see the "Navigation:Classes" panel in the lower left corner);
- Searching using faceted navigation method (see the "Filter" panel in the right most side). *OntoWiki* enables users to select objects according to certain facets i.e. all property values (facets) of a set of selected instances. If for a certain property the instances have only a limited set of values, those values are offered to restrict the instance selection further. Hence, this way of navigation through data will never lead to empty results;


OntoWiki — Resource List - Windows Internet Explorer
http://localhost/ontowiki/resource/instances
Bing
File Edit View Favorites Tools Help
Favorites
OntoWiki — Resource List
OntoWiki (SuperAdmin)
User Extras Help
Search for Resources
Knowledge Bases
Edit View
OntoWiki System Configuration
PUPIN_organization_PD
PUPIN_PPD_OPD
PUPIN_extracted_expertise_tracking
PUPIN_employees_responsibilities_projects
PUPIN_employees_language_skills
Navigation: Classes
Edit View Type
Search in Navigation
ictorganization:LanguageSkill
foaf:OnlineAccount
foaf:Document
foaf:Agent
ictorganization:PartTimeEmployment
ictorganization:ScientificExperience
ictorganization:JobPosition
ictorganization:Community
ictorganization:RTD_ERA
Resource List
View
Instances
History
Community
Source
Show Properties
dc:description dc:title
ictorganization:degree
ictorganization:employees
ictorganization:graduationTitle
ictorganization:hasJobPositions
ictorganization:hasScientificRecord
rdf:type foaf:firstName
foaf:homepage foaf:name
foaf:surname
Driving Skill-1
ictorganization:referer-1
foaf:primaryTopic-1
Filter
Edit
Search in list
Active Filters:
Type: foaf:Agent
« First ‹ Previous 1 2 3 Next › Last »
Search returned 383 results
Show me: 10 50 100 all

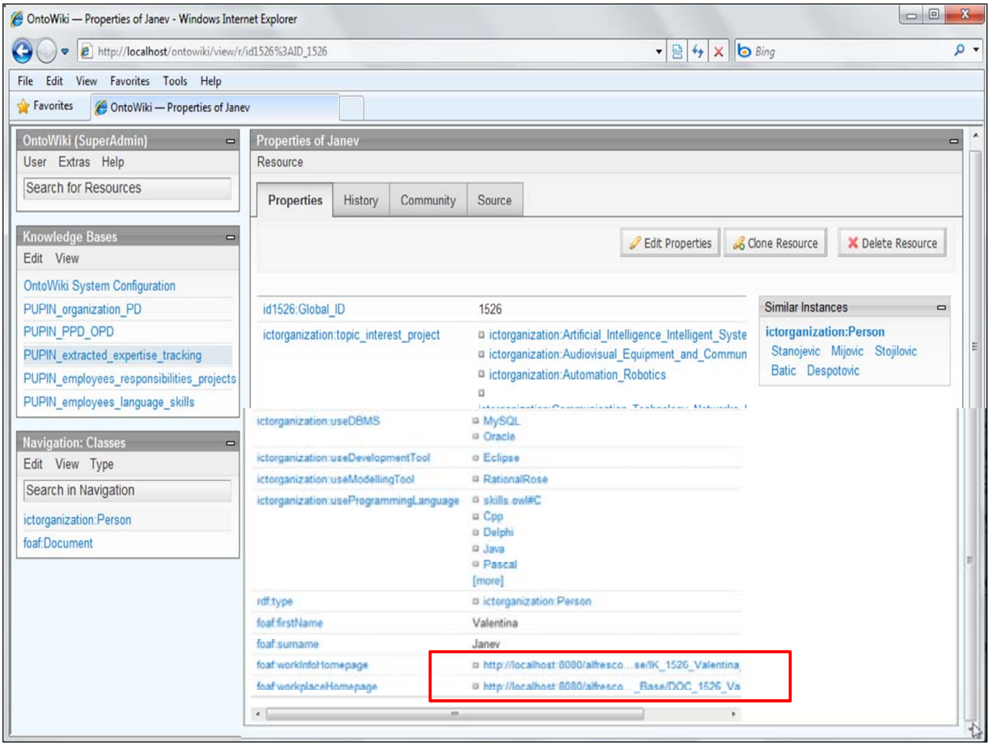
1. Telecommunications Ltd.
   ictorganization:Organization, foaf:Organization
2. IMP Computer Systems Ltd.
   ictorganization:Organization, foaf:Organization
3. IMP Computer Engineering Ltd.
   ictorganization:Organization, foaf:Organization
4. Institute Mihailo Pupin
   ictorganization:Organization, foaf:Organization
5. IMP Automation& Control Systems Ltd.
   ictorganization:Organization, foaf:Organization
6. IMP Piezotechnology Ltd.
   ictorganization:Organization, foaf:Organization
7. Samardžija
   ictorganization:Person
8. Pavlović
   ictorganization:Person
9. Latinović
   ictorganization:Person
10. Milanović
   ictorganization:Person

Fig. 4. Expertise search with OntoWiki.

Once a selection is made, the main content section will arrange matching content in a list view linking to individual views for individual instances. The right sidebar offers tools and complementary information specific to the selected content.

The main steps in the process of navigating and querying of a semantic model can be summarized as follows.

1. Select a knowledge base e.g. Organization and Personal profiles;
2. Select a semantic concept e.g. *foaf:Agent*;
3. Filter the entities using a semantic relation e.g. *rdf:type* in order to retrieve all instances of type *foaf:Person*;
4. Filter the entities with the faceted navigation filter e.g. retrieve the personal data for a person with a surname *Janev;*
5. After reviewing the results, the user may wish to continue navigating the information space by following relations between instances e.g. *foaf:PrimaryTopic*$^{-1}$ can be selected to link the instance *Janev* with its personal profile document *1526-PPD* (see Fig. 5). Links to the MPI document base that stores the publications and other documents created in the MPI working process are framed in red (Janev et al., 2010).

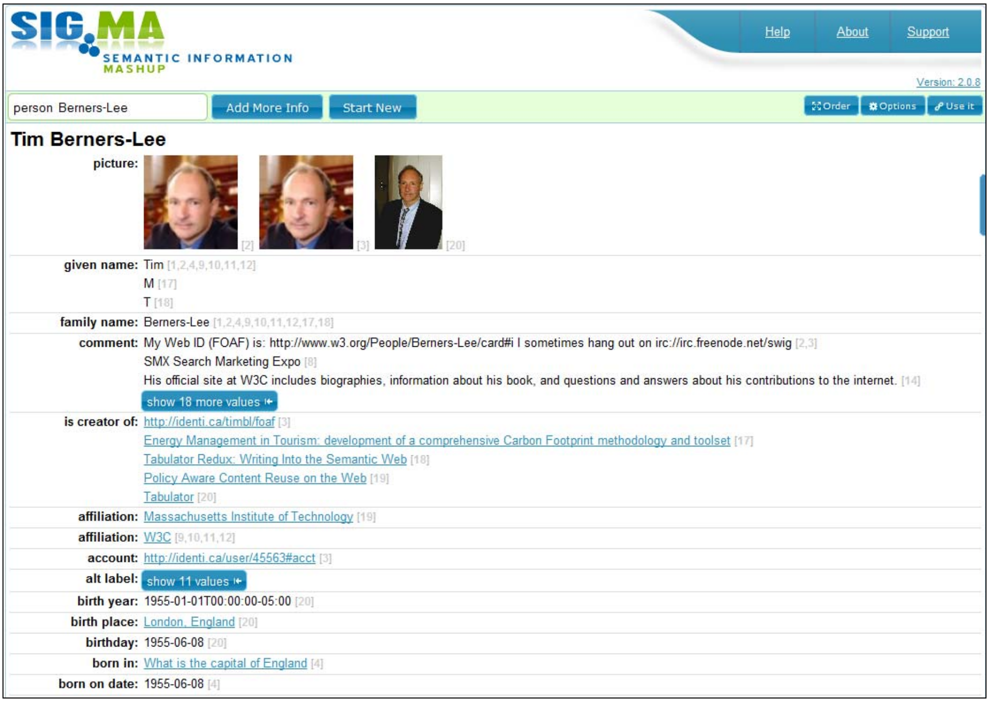


| | |
|---|---|
| id1526:Global_ID | 1526 |
| ictorganization:topic_interest_project | ictorganization:Artificial_Intelligence_Intelligent_Syste<br>ictorganization:Audiovisual_Equipment_and_Commun<br>ictorganization:Automation_Robotics |
| ictorganization:useDBMS | MySQL<br>Oracle |
| ictorganization:useDevelopmentTool | Eclipse |
| ictorganization:useModellingTool | RationalRose |
| ictorganization:useProgrammingLanguage | skills.owl#C<br>Cpp<br>Delphi<br>Java<br>Pascal<br>[more] |
| rdf:type | ictorganization:Person |
| foaf:firstName | Valentina |
| foaf:surname | Janev |
| foaf:workInfoHomepage | http://localhost:8080/alfresco...se/IK_1526_Valentina |
| foaf:workplaceHomepage | http://localhost:8080/alfresco..._Base/DOC_1526_Va |

Fig. 5. Personal profile document of instance Janev-PPD.

## 5.2 Searching Web of Data using Sig.Ma

Once available on the Web, expert profiles can be searched with Semantic search engines, e.g. Sig.Ma.



person Berners-Lee

**Tim Berners-Lee**

**picture:** [2] [3] [20]

**given name:** Tim [1,2,4,9,10,11,12]
M [17]
T [18]

**family name:** Berners-Lee [1,2,4,9,10,11,12,17,18]

**comment:** My Web ID (FOAF) is: http://www.w3.org/People/Berners-Lee/card#i I sometimes hang out on irc://irc.freenode.net/swig [2,3]
SMX Search Marketing Expo [8]
His official site at W3C includes biographies, information about his book, and questions and answers about his contributions to the internet. [14]

**is creator of:** http://identi.ca/timbl/foaf [3]
Energy Management in Tourism: development of a comprehensive Carbon Footprint methodology and toolset [17]
Tabulator Redux: Writing Into the Semantic Web [18]
Policy Aware Content Reuse on the Web [19]
Tabulator [20]

**affiliation:** Massachusetts Institute of Technology [19]

**affiliation:** W3C [9,10,11,12]

**account:** http://identi.ca/user/45563#acct [3]

**alt label:**

**birth year:** 1955-01-01T00:00:00-05:00 [20]

**birth place:** London, England [20]

**birthday:** 1955-06-08 [20]

**born in:** What is the capital of England [4]

**born on date:** 1955-06-08 [4]

Fig. 6. An example of available personal profile document in the LOD cloud.

## 6. Conclusion

Taking into account the new trends in the design and implementation of enterprise information systems (based on adaptable, flexible, and open IT architecture, using open standards and emerging technologies), this Chapter introduced new insight into expertise management and proposed the Semantic Web-based approach to HR data representation, integration and retrieval.

**Ontology-based approach to competency management**: The proposed ontology-based approach to competency management includes establishment of a modular knowledge base of expert profiles and population of the knowledge base with information extracted from different HR related sources. The proposed approach based on emerging technologies and tools does not complement the existing information - integration approach (e.g. integrating the expert data in a form of a database) or the content management approach (e.g. integrating the experts' documents in a form of a document base), but it rather extends, enhances and integrates them with the aim to obtain a complete picture of the available resources.

**Explicit, standard format of expert profile that facilitates data interoperability and expert search:** As the interoperability between different knowledge organization schemas is one of the major Linked Open Data issues, the design of the semantic knowledge model in this Chapter was based on public vocabularies such as FOAF, DOAC, SIOC, DOAC, BibTeX, as well as common vocabularies for modelling case study specific data and relations such as DC, RDF, RDFS, and OWL.

**Enhancing self-declared expertise with competences automatically and objectively extracted using text analysis:** Taking into consideration that self declared expertise cannot be an exhaustive description of the person's expertise areas, the use of text analysis tools for updating the semantic expert profiles with uncovered *latent* knowledge should be considered.

**Meaningful search and retrieval of expertise:** Recently, a new search approach has emerged. It has been named faceted search that combines the navigational search paradigm and the direct, keyword search paradigm. Faceted search methods augment and improve traditional search results by using not just words, but concepts and logical relationships that are components of an ontology. Faceted navigation techniques and semantic relations shorten the search time, improve the relevance of search results, and deliver high-quality search services. This Chapter demonstrates the use of these methods in practice.

## 7. Acknowledgement

This work was supported by a grant from the European Union's 7th Framework Programme (2007-2013) provided for the project LOD2 (GA no. 257943), as well as by the Ministry of Education and Science of the Republic of Serbia under grant agreement no TR-32010 (SOFIA – Software Framework for Intelligent Adaptive Management of Complex Facilities).

## 8. References

Aleman-Meza, B. et al. (2007). Combining RDF vocabularies for expert finding. *The Semantic Web: Research and Applications*, Lecture Notes in Computer Science, Vol. 4519 (pp. 235-250). Berlin / Heidelberg: Springer

Auer, S., Dietzold, S., Lehmann, J., & Riechert, T. (2007). OntoWiki: A Tool for Social, Semantic Collaboration. CKC 2007

Auer S., Lehmann S. (2010). Creating knowledge out of interlinked data. Semantic Web Journal 1 (pp. 97–104)

Balog, K., & de Rijke, M. (2008). Associating people and documents. In C. Macdonald et al. (Eds.), *Proceedings of the 30th European Conference on Information Retrieval (ECIR 2008), Lecture Notes in Computer Science, Vol. 4956* (pp. 296-308). Berlin / Heidelberg: Springer

Balog, K., Azzopardi, L., & de Rijke, M. (2006). Formal models for expert finding in enterprise corpora. In S. Dumais, E.N. Efthimiadis, D. Hawking, and K. Järvelin (Eds.), *Proceedings of the 29th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval* (pp. 43-50). ACM

Berners-Lee, T. (2006). *Linked Data*. Retrieved March 15, 2012 from http://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American, May 2001*. Retrieved January 15, 2007, from http://www.sciam.com/article.cfm?id=the-semantic-web

Bordea, G., (2010). Concept Extraction Applied to the Task of Expert Finding. *The Semantic Web: Research and Applications, Lecture Notes in Computer Science*, 2010, Volume 6089/2010, 451-456, DOI: 10.1007/978-3-642-13489-0_42

Bordea, G., & Buitelaar, P. (2010). Expertise Mining. In *Proceedings of the 21st National Conference on Artificial Intelligence and Cognitive Science,* Galway, Ireland, 2010

Bizer, C., Heese, R., Mochol, M., Oldakowski, R, Tolksdorf, R, Eckstein, R. (2005). The Impact of Semantic Web technologies on job recruitment processes. *International Conference Wirtschaftsinformatik (WI 2005), Bamberg, Germany*

Draganidis, F., Chamopoulou, P., & Mentzas, G. (2006). An ontology-based tool for competency management and learning paths. In *Proc. I-KNOW '06, 6th International Conference on Knowledge Management, Special track on Integrating Working and Learning, 6th September 2006*, Graz , Austria

Draganidis F., & Mentzas, G. (2006). Competency based management: a review of systems and approaches. *Information Management and Computer Security 14(1)*: 51 – 64

Fang, H., & Xiang Zhai, C. (2007). Probabilistic models for expert finding. In *Advances in Information Retrieval, Lecture Notes in Computer Science, Vol. 4425* (pp. 418-430). Berlin / Heidelberg: Springer

Gruber, T.R. (1993). A translation approach to portable ontology specification. *Knowledge acquisition, 5(2):* 199-220

Houtzagers, G. (1999). Empowerment, using skills and competence management. *Participation & Empowerment: An International Journal (2):*27-32

Janev, V., & Vraneš, S. (2011a). *Semantic Web Tools and Technologies for Competence Management: The Case Study of R&D Organization*. LAP LAMBERT Academic Publishing GmbH & Co. KG, 2011. ISBN: 978-3-8454-4166-5

Janev, V., & Vraneš, S. (2011b). Applicability assessment of Semantic Web technologies. *Information Processing & Management*, 47:507–517, doi:10.1016/j.ipm.2010.11.002

Janev, V., Mijović, V., & Vraneš, S. (2010). Automatic extraction of ICT competences from unstructured sources. In J.E. Quintela Varajǎo et al. (Eds.), *Proceedings of the CENTERIS 2010 - International Conference on ENTERprise Information Systems*, Part II, CCIS 110 (pp. 391-400). Berlin / Heidelberg: Springer

Jung, H., Lee, M., Kang, I.-S., Lee, S.-W., & Sung, W.-K. (2007). Finding topic-centric identified experts based on full text analysis. In A.V. Zhdanova, L.J.B. Nixon, M. Mochol, J. G. Breslin (Eds.), *Finding Experts on the Web with Semantics 2007, Proceedings of the 2nd Intl. ISWC+ASWC ExpertFinder Workshop (FEWS'07), Busan, Korea, November, 2007.* Retrieved August 5, 2008 from CEUR-WS.org/Vol-290/

McClelland, D. (1973). Testing for competence rather than for intelligence. *American Psychologist 20*:321-33

Müller-Riedlhuber, H. (2009). The European Dictionary of Skills and Competences (DISCO): an example of usage scenarios for ontologies. In *Proceedings of I-KNOW '09 and I-SEMANTICS '09, 2-4 September 2009, Graz, Austria* (pp. 467 – 479)

Paquette, G. (2007). An Ontology and a Software Framework for Competency Modelling and Management. *Educational Technology & Society 10* (3): 1-21

Petkova, D., & Bruce Croft, W. (2006). Hierarchical language models for expert finding in enterprise corpora. In *Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence* (pp. 599 – 608). IEEE Computer Society

Sim, Y. W., Crowder, R. M., & Wills, G. B. (2006). Expert finding by capturing organizational knowledge from legacy documents. In *Proceedings of the IEEE International Conference on Computer & Communication Engineering (ICCCE '06), 9-11 May 2006, Kuala Lumpur, Malaysia*

Schmidt, A, & Kunzmann, C. (2006). Towards a Human Resource development ontology for combining competence management and technology-enhanced workplace learning. In R. Meersman and Z. Tahiri and P. Herero (Eds.), *On The Move to Meaningful Internet Systems 2006: OTM 2006 Workshops. Part I. 1st Workshop on Ontology Content and Evaluation in Enterprise (OntoContent 2006), Lecture Notes in Computer Science vol. 4278* (pp. 1078—1087)

Schäfermeier, R., & Paschke, A. (2011). Using Domain Ontologies for Finding Experts in Corporate Wikis. In C. Ghidini et al. (Eds.), *Proceedings of the I-SEMANTICS '11, 7-9 September 2011* (pp. 63 – 70). Graz, Austria: J.UCS

Tummarello, G., Cyganiak, R., Catasta, M., Danielczyk, S, Delbru, S., & Decker, S. (2010). Sig.ma: Live views on the Web of Data. *Web Semantics: Science, Services and Agents on the World Wide Web 8(4):* 355-364