Theoretical and Applied Economics
Volume XVII (2010), No. 1(542), pp. 49-68

The Informatics Security Cost of Distributed Applications

Ion IVAN
Bucharest Academy of Economic Studies
ionivan@ase.ro
Dragoş PALAGHIŢĂ
Bucharest Academy of Economic Studies
mail@dragospalaghita.ro

Abstract. *The objective, necessity, means and estimated efficiency of information security cost modeling are presented. The security requirements of distributed informatics applications are determined. Aspects regarding design, development and implementation are established. Influence factors for informatics security are presented and their correlation is analyzed. The costs associated to security processes are studied. Optimal criteria for informatics security are established. The security cost of the informatics application for validating organizational identifiers is determined using theoretical assumptions made for cost models. The conclusions highlight the validity of research results and offer perspectives for future research.*

Keywords: security; costs; models; analysis prediction; efficiency.

JEL Code: C63.
REL Code: 10J.

1. Introduction

The pursued objective is building cost estimation models for informatics security subsystems in distributed informatics applications.

The necessity is given by:

- the increase in complexity in informatics applications(Pendharkar et al., 2008, pp. 1181-1188) which determine measures in database exploitation, usage of executable programs and web site usage;
- the rise in user number and diversity according to (Khansa, Liginlal, 2009, pp. 216-235) which generates measures in accessing the resources of the informatics application;
- data base administration which generates measures in updating, reallocation, and restructuration;
- the development of E-commerce which implies a large volume of information transactions which implies measures[(1)] in maintaining confidentiality, integrity and availability of information;
- the transition to the new economy which generates the virtualization of business processes, which brings, in turn, an increase in the quantity of data streams transferred online and a change in the nature of company assets from physical to virtual, determining measures in protecting them and the related key transactions at an organizational level (Huang, Hu, 2008, pp. 793-804);
- the accelerated development of technologies used for elaboration of software products which imply more complex security measures which are meant to protect a larger area of assets and transactions.

To this purpose:

- distributed informatics applications that are used in electronic commerce, online payments, enterprise resource planning and stock management are considered;
- data that relates to behavior, maintenance, fixing, optimization and restructuration costs are considered (Kumar et al., 2008, pp. 1853-1867);
- data sets regarding the user and application behavior, level of validation and potential security risks are constructed (Anwar et al., 2009, pp. 13-25, Lu et al., 2009, pp. 4617-4625, Aroba et al., 2008, pp. 1944-1950);
- direct and indirect cost influence factors are determined;
- cost model structures based on different influence factors are determined (Lee, Kim, 2009, pp. 453-475);
- by using different mathematical methods the model coefficients are estimated;
- the models are validated using specific methodologies (Jasmine, Vasantha, 2008, pp. 951-954);
- the models are refined by using quantitative methods and genetic algorithms (Ivan et al., 2008, Vişoiu, 2009, pp. 861-866).

In order to complete a gradual approach a literature review of the most important cost estimation models is done. The accent is set towards informatics security and its part in the increase of software development costs and the decrease in current exploit of informatics applications are analyzed.

Influence factors are identified in order to increase the level of informatics security. Vulnerabilities are analyzed and linear and non-linear models are developed. The next step is developing optimization constructions and finally the model validation using data obtained by the current exploitation of the informatics application for validating organizational identifiers.

In order to complete this paper resources made available through contract no. 47/01.10.2008. in the doctoral school of the University of Economics Bucharest, Romania.

2. Literature review

In the papers dedicated to this subject (Vacca, 2009 [VACC09], Tipton, Krause, 2008 [TIKR08], Stamp, 2005 [STAM05], Pfleeger, Pfleeger, 2006 [PFLE06]) basic concepts of information security are presented. Table 1 presents the coverage degree of the mentioned works.

Table 1

Coverage degree of literature work

Contents	[VACC09]	[TIKR08]	[STAM05]	[PFLE06]
Cryptography	X	X	X	X
Application security		X	X	X
Network security	X	X	X	X
Internet security	X	X		
Database security				X
Operating system security	X		X	X
Physical security	X	X		X
Security architectures		X		
Risk analysis	X	X		X
Risk management	X	X		
Security management	X			X
Access control	X	X	X	X
Security protocols	X	X	X	X
Insecurity			X	
The economics of security				X
Confidentiality	X	X	X	X
Legal aspects of information security		X		X

There are scientific journals in the informatics security field like:

- Computers & Security, ISSN: 0167-4048, published by Elsevier Advanced Technology;

- International Journal of Information Security, ISSN: 1615-5270, published by Springer New York;
- Journal of Cryptology, ISSN: 1432-1378, published by Springer New York;
- IEEE Transactions on Information Forensics and Security ISSN: 1556-6013, published by IEEE;
- Cryptologia, ISSN: 1558-1586, published by Taylor & Francis;
- IET Information Security, ISSN: 1751-8709, published by Institution of Engineering and Technology;
- ACM Transactions on Information and System Security, ISSN: 1094-9224, published by ACM
- Security and Privacy, ISSN: 1540-7993, published by IEEE;
- IEEE Transactions on Dependable and Secure Computing, ISSN 1545-5971, published by IEEE.

There are conferences in informatics security:

- IEEE Intelligence and Security Informatics, which had, in the 2009, sections relating to the distribution of information and data mining, protecting the infrastructure and emergency response, informatics in the context of terrorism;
- IEEE Symposium on Security and Privacy contained in the 2009 edition sessions about attack and defense methods, informatics security, malware code, information loss, confidentiality, formal bases of information security, network security, physical security and web security;
- ACM Conference on Computer and Communications Security which in the 2009 edition contains sections relating to informatics security, secure system design, techniques of ensuring information security, confidentiality, mobile service security, applied cryptography and system security;
- USENIX Security Symposium includes subjects as authentication and authorization, autonomous methods, grid computing, email related security, virus protection methodologies, cybernetic attack defense mechanisms;
- Computer Security Foundations Symposium hosted sections as protocol design, web security, session authorization, session checking, application security analysis;
- Network and Distributed System Security Symposium includes topics as web attacks, cryptography, confidentiality and integrity.

The presented journals and conferences offer a broad coverage area of actual informatics security and highlight the actual research progress made in this field of informatics.

3. Informatics security

Information is defined(2) as facts and ideas that are represented or coded as different data forms.

According to this definition, security is represented(2) by the measures taken to protect a system. Security is also considered like a condition of a system which results from enforcing and maintaining measures to protect assets. Security imposes that the resources of a system must not be the subject of unauthorized access, unauthorized changes or accidental, and destruction or loss of protected assets (Vydrin, 2009, pp. 261-275).

Information security can be defined(3) as the protection of information and informatics systems from unauthorized access, use, divulgation, interference, modification and destruction in order to ensure:

- integrity(3), defined as protection against incorrect information modification or destruction and includes ensuring non-repudiation and information authenticity;
- authenticity(3) is necessary to ensure that data, information or transactions are original; non-repudiation(4) implies the fact that no one is able to deny sending or receiving a transaction; authenticity and non-repudiation are applied in electronic commerce by using digital signatures(3);
- confidentiality(4) is defined as keeping authorized restrictions regarding access and publication, including means to protect private life and personal information;
- availability(3) is represented as insuring timely and trustworthy access to information.

In the mention papers (Vacca, 2009, Tipton, 2008) informatics security addresses problems related to:

- informatics security risk by analyzing general concepts related to risk, vulnerabilities and threats (Alhazmi, Malaiya, 2008, pp. 14-22);
- access control highlighting different authentication modalities and user required protocols; the vulnerabilities of control systems are analyzed considering the main types of attack they are subject to;
- cryptography aspects of security systems pursuing network key management, the most efficient encryption methods are identified and the encryption algorithms are described; advantages and disadvantages for each encryption method are presented;
- application security by detailing the newest security methods and presenting the role of application quality in this process;
- Internet security by describing vulnerabilities and threats in the online environment;
- network security by identifying vulnerabilities of transfer protocols and analyzing the threats of network communication;

- wireless network security, presenting aspects relating to vulnerabilities, threats and recommended security policies;
- cellular network security detailing security protocols for radio transmissions, attack modalities of radio networks and defense strategies applied to radio networks;
- ways of increasing the security level by improving code quality, by identifying and eliminating system physical vulnerabilities, user training, by increasing access control systems quality, by improving password management techniques, by increasing the quality of security policies and clearly differentiating user roles, by developing a quality characteristic system associated to the security system and proceeding to improvement of the individual characteristic quality level in order to obtain a global quality increase of the informatics security system;
- management aspects of information security by clearly detailing the main components of a informatics security management system;
- intrusion detection systems and vulnerability evaluation techniques;
- the identification of legal aspects of informatics security by studying laws and regulations imposed globally for better practices in this domain.

Within informatics security systems quality plays an important role, being a major influence factor in the well being of an informatics application. The quality characteristics are placed in a hierarchy based on:

- source code:
  - homogeneity, which is represented by the nature of source code to have the same characteristics and properties in all of the modules belonging to the security system; the use of operators and operands in a similar fashion is desired as using the same kind of formatting in each of the system modules;
  - intelligibility is defined as the characteristic of source code to be perceived easily by developers that did not have any prior encounter with it before; source code intelligibility in security systems is useful due the resource economy it produces by allowing easy understanding of the implementation logic thus rapidly making improvements or modifications to existing code;
  - testability is the capacity of the source code to undertake the testing process easily by covering all logical paths; a high testability level in security systems ensures a minimization of defect numbers and omissions of the system thus improving the global quality level; testability is ensured by the homogenous development of the security system and using logical internal reporting systems for all operations and events encountered in the security framework;
  - maintainability, which if present in a high level minimizes defect fixing application improvement costs, this characteristic is in tightly related to homogeneity and intelligibility of source code;

- data type veridicity thus pursuing the elimination of buffer overflow which leads to the security system corruption; lower and upper limiting of buffers is needed in order to ensure the system is protected from this type failure; enforcing buffer limitation as a standard for all input received by the application is a proven technique for avoiding memory corruption;
- error perception is a characteristic which oversees the level on which the security system reports errors and interprets them correctly; a high level of perception in the security system reduces the costs relating to restoring the informatics application, paying compensations and the cost of reengineering the security system altogether;
- using secret phrases in source code is a known security issue mostly when the phrases take the form of access tokens and are hardcoded for future use; this practice is not recommended as it raises several security hazards that are hard to control;

▪ interaction with the informatics application:
  - compatibility, meaning using common communication protocols thus maintaining communication between the two entities in the best conditions possible;
  - coexistence, defined as security system's ability to work at optimal parameters within the informatics application; a high level of coexistence increases the reliability degree of the informatics application and minimizes the maintenance costs associated to the security system;
  - accuracy is represented by the exact nature with which the signals emitted by the informatics application are perceived by the security system;
  - securing information is characterized by the security system's ability to protect and ensure the confidentiality of data used and processed in the informatics application;

▪ user interaction:
  - transfer security in the informatics application by implementing efficient authentication and authorization systems thus ensuring a correct user authentication effort, minimizing identity theft cases and costs with:
    - restarting the security system of the informatics application after a breach;
    - damage evaluation provoked by the unauthorized access in the informatics application;
    - paying compensations due to compromising protected assets;
  - the validity of data inputted by the user; by ensuring a high level of which the attack opportunities are minimized and human-machine interaction is improved; data validity is ensured by implementing validation controls and procedures to prevent the most common and dangerous informatics attacks to which the application is exposed.

Informatics security is a necessary requirement for large distributed systems (Dudin et al., 2009, pp. 234-240). It is imperative to develop secure systems in the conditions of an increasing number of threats and threat agents.

4. Influence factors for informatics security

Informatics applications are complex constructions used in defined social and economical contexts. The influence factors are numerous and have diverse effects.

Direct influence factors consist of:

- the target group which is defined as all the individuals that form the collectivity which uses the informatics product; the target group influences security directly through:
  - structural diversity, a collectivity structural analysis being necessary to determine behavioral patterns differenced based on age, sex and education in order for the security system to register user actions and assign a behavioral pattern to application users such that an adaptive security policy system is used to grant or deny privileges to them;
  - dimension such that the security system is correlated to the number of individuals that access the application; this way the security system will work at optimal parameters;
  - the social status in the collectivity, thus if it proves to be true that certain individuals in it are against actions or thoughts that the application owner sees as favorable, a greater amount of effort must be made to ensure an increase in physical and logical security of the application;
- the development process quality has a direct influence on informatics security because:
  - a high level of quality leads to the minimizing the number of defects which in turn reduces the informatics security risk;
  - a low level of quality increases the number of vulnerabilities in the application thus increasing the informatics security risk
- in the development cycle of the security system fixed quality objectives should be followed:
  - homogeneity of source code by developing modules and procedures which integrate totally in the security system;
  - intelligibility of implemented procedures in order to minimize testing, optimization and maintenance time of the security system associated to the informatics application;
  - flexibility of network communication and reporting systems in n order to function with an extended set of report formats thus assuring a high compatibility degree with intrusion detection systems;
  - scalability of components in order to easily increase the adaptability of the security system;
- used development technologies represent an important aspect because they influence the level of informatics security by:

- quality transfer, if the instruments used in the development stage have a high quality level then by using them the developed security system will benefit of a high quality level;
- the degree to which the development assistance tools help the developer make good decisions by providing useful observations at development time;
- the novelty degree of used instruments and tools and their coverage level of the newest informatics attacks thus allowing the developer to bring the performances of the security system to the highest standards;

- the environment in which the informatics product is used and in which the security system activates influences the level of security by the degree of provided physical security;
- hardware elements have a direct influence on the security system by their wear resistance and reliability considering they have to work continuously; performance is another key issue for hardware equipment being necessary to ensure a small response time for each event in the security system;
- dynamic elements of the problem that the informatics application needs to solve, this implies an increased flexibility level to handle new and unforeseen events generated by structural or logical changes in informational transfers required by modifications in the problem structure.

The indirect factors that determine security are:

- complexity, which has an important effect over informatics security (Pocatilu, 2004), as the software product's complexity grows so does the number of defects thus decreasing the level of informatics security. Complexity is defined using the following models:
  - Halstead (1977), a model which is characterized by the following equations:
  - the length of code *N* which is represented by the sum of the operator number N_1 and operands number N_2:

$$N = n_1 \times \log_2 n_1 + n_2 \times \log_2 n_2$$

where:

n_1 represents the number of distinct operators;

n_2 represents the number of distinct operands;

- volume is the product of the code length with the minimum number of bits needed to store operators and operands:

$$V = N \times \log_2 n$$

where:

$$N = N_1 + N_2$$

$$n = n_1 + n_2$$

- difficulty is defined as:

$$D = \frac{n_1 \times N_2}{2 \times n_2}$$

- the effort for implementing the program is computed using:
  $E = D \times V$
- cyclomatic is defined by:
  $C = m - n + 2$

where:

m is the number of arcs in the graph associated to the program;
n is the number of nodes of the graph associated to the program.

In Figure 1 the graphical representation of the influence of direct and indirect factors over informatics security is presented.

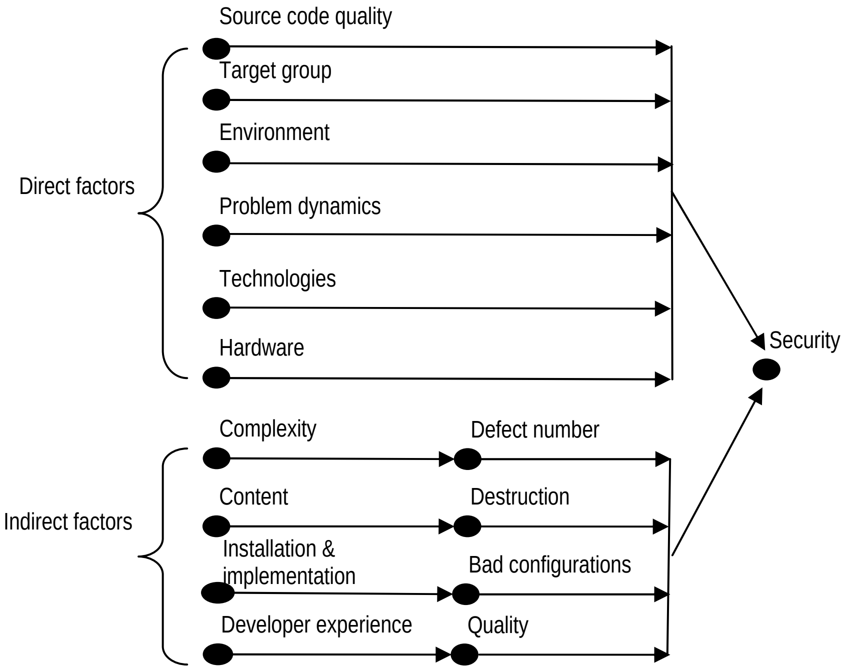

Figure 1. *Graphical representation of the security influence factors*

The influence factors are an important element in informatics security analysis and in establishing cost model coefficients.

5. Informatics security cost models

A model is a mathematical expression used to describe an economical process using a set of variables and operators in order to quantify the relation between them[(6)]

In order to build models:

- the way data is collected is analyzed;
- data collecting procedures are established;
- automated data acquisition procedures are developed;
- influence factors are placed in correspondence with variables;
- a specific technology is developed.

Economical models according to some authors (Ivan, Vişoiu 2005) consist of:

- exogenous variables that are associated to factors influencing a process or another variable that results from the model;
- endogenous variables or result variables which are associated to the pursued objective by completing the model;
- data sets collected manually or automatically by using specific procedures of data acquisition;
- the model coefficients which are a result of applying parameter estimation models based on the data sets;
- operators which are used to develop expressions and marking connections between the component factors of the economical model;
- elementary mathematical functions that are part of nonlinear models;
- composite mathematical functions that are used for the elaboration of more complex models.

There are several types of models associated with cost estimation:

- linear models, where mathematical expressions are defined in which the relations between exogenous and endogenous variables has a form like:

$$Ct = \sum_{i=1}^{NF} Ch_i$$

where:

Ct – total cost of implementing the security system;
Ch_i – expense no, i;
NF – number of endogenous variables;

- non-linear models, which contain multiplications, divisions, logarithmic expressions, integrals, square roots and complex mathematical expressions, the analytical form of the nonlinear cost model is:
  Ct = f(CM,CDS)

where:

Ct – the total cost of implementing the security system;
CM – model complexity;
CDS – the cost of developing the security system.

The following cost model is defined[(7)]:

$$Ct = a_i \times KLoC^{b_i} \times EAF$$

where:

a_i, b_i – table coefficients attributed according to the project type;

KloC – number of source code line upon delivery;

EAF – estimated effort coefficient based on data collected using a specific methodology.

In order to develop cost models the variables that influence cost must be determined and separate them into independent and dependent variables; Table 2 presents the correlation between time frames and values of identified factors.

Table 2

Time frame – factors correlation

T	Ch_{Tsec}	Ch_P	Ch_{RBD}	Ch_D	Ch_{PSS}	Ch_T	Ch_{DESP}	Ch_{OPT}	Ch_{VUL}
t_1	$Ch_{Tsec\,1}$	Ch_{P1}	Ch_{RBD1}	Ch_{D1}	Ch_{PSS1}	Ch_{T1}	Ch_{DESP1}	Ch_{OPT1}	Ch_{VUL1}
t_2	$Ch_{Tsec\,2}$	Ch_{P2}	Ch_{RBD2}	Ch_{D2}	Ch_{PSS2}	Ch_{T2}	Ch_{DESP2}	Ch_{OPT2}	Ch_{VUL2}
...	...	...	...	...	...	...	...	...	...
t_i	$Ch_{Tsec\,i}$	Ch_{Pi}	Ch_{RBDi}	Ch_{Di}	Ch_{PSSi}	Ch_{Ti}	Ch_{DESPi}	Ch_{OPTi}	Ch_{VULi}
...	...	...	....	....	...	...	...	...	...
t_n	$Ch_{Tsec\,n}$	Ch_{Pn}	Ch_{RBDn}	Ch_{Dn}	Ch_{PSSn}	Ch_{Tn}	Ch_{DESPn}	Ch_{OPTn}	Ch_{VULn}

where:

Ch_{Tsec} – total expenses with the security system;

Ch_P – expenses with the development team;

Ch_{RBD} – expenses caused by restoring the data base;

Ch_D – security system maintenance caused expenses;

Ch_{PSS} – expenses with designing the security system;

Ch_T – testing activities expenses;

Ch_{DESP} – compensation related expenses;

Ch_{OPT} – optimization related expenses;

Ch_{VUL} – vulnerability minimization expenses.

Security costs are represented by these expenses leading to the expense model:

$$Ch_{T\sec} = Ch_P + Ch_{RBD} + Ch_D + CH_{PSS} + Ch_T + Ch_{DESP} + Ch_{OPT} + Ch_{VUL}$$

According to Andersen and Choobineh (2008), the informatics security expenses have an important impact in choosing asset security strategies such that cost estimation becomes an essential analysis in selecting the best security system considering environmental conditions faced by the organization.

6. Informatics security optimization

Optimization as a selection process from a set is meant to improve. Considering the set of time frames $M_1, M_2, ..., M_k$ and the set of options $V_1, V_2, ..., V_k$ the optimal variant is:

$$V(M_h) = \min_{1 \le i \le k}\{V_i\}$$

where:

$V(M_h)$ – optimal variant from $M_1, M_2, ..., M_k$.

In order to perform optimization in best possible circumstances it is necessary to define optimal criteria through which the improvement of certain characteristics of the informatics security system associated to the informatics application is pursued (Ivan et al., 2008, pp. 39-56). Thus for informatics security the following optimal criteria are defined:

- minimizing execution time of procedures that make up the informatics security system by data processing algorithm optimization, by eliminating unnecessary source code, by analyzing execution threads and directing them to use the shortest path possible;
- maximizing the application efficiency by defining benchmarks for all operations thus identifying critical areas in the security system;
- maximizing the correctness in the informatics security system in order to identify more clearly the threats and attack attempts that the system faces on secured assets;
- minimizing the fixing duration associated to the development work done on defects of the security system.

Source code optimization is meant to improve the quality of procedures and execution time. By optimizing source code the minimization of defect numbers is also pursued thus improving overall system performance through the improvement of the system's quality characteristics. Optimization through increasing the level of the informatics security system's quality is based on:

- testing through which problems in system quality are observed and by reconsidering segments of the development cycle in which they were generated such that solutions for solving them are implemented;
- internal reporting of the security system through which based on test data simulations lower than expected levels of quality characteristics are unveiled;
- running periodical reports in order to monitor the application evolution in current usage conditions thus highlighting the most frequent problems that appear in the informatics security framework;
- the appearance of new technologies which placed in current usage transfer quality to the security system making it more efficient;
- developing new information security techniques that through implementation will improve the overall performance of the informatics security system;
- developer certification in order to find new work techniques better than the ones used before this leads to efficient and less error prone code development;
- software development company certification in quality systems in order to install software development techniques that follow the best patterns and practices in the field throughout the development life cycle.

Increasing the quality of the informatics security system leads to obtaining an improved version of the system that corresponds to a rise in the individual levels of quality characteristics. The new version holds a net superiority over the old one considering its quality characteristics.

Cost optimization according to [WEB05] represents practices, abilities and behavior adopted by an organization in order to reduce expenses, minimize costs, maintain software system quality at a constant level and maintains the growth potential of the company on an ascending trend. Cost optimization at organizational level, at a development team level and at source code level is pursued. The cost optimization process at the organizational level must identify administrative and production areas of the company that register the highest level of expenses and assume plans for minimizing or eliminating the following costs with:

- auxiliary personal that don't contribute to profit generating activities;
- procuring materials and equipment which are not necessary to ongoing business processes and software development processes;
- employee transportation to this purpose identifying ways of minimizing business travel and using alternative Internet based solutions for business communication;
- external suppliers by making auctions and obtaining a better price for needed supplies and equipment.

Cost optimization is meant to minimize cost generating activities and maximize the efficiency of profit generating ones meaning finding alternative, more efficient means of handling business process to existing ones.

7. The cost of informatics security in the organizational identifier validation application

The application for organizational identifiers validation analyses the orthogonality of company names in order to eliminate situations when two companies share very similar names.

In the orthogonality analysis application a vocabulary VOCDEN = {D_1, D_2, …, D_k} is developed which contains all organization names inputted in the data base.

The following situations are identified:

a) The entity has its name formed from a single word, the analysis is completed by comparing the name of the entity with all other names stored that have only one word, the orthogonality is studied at word level and it is computed for the CI and CB words using:

$$ORTOC(CI,CB) = 1 - \frac{Len(SMC)}{\max\{Len(CI), Len(CB)\}}$$

where:

Len(SMC) – length of the maximum common substring;
Len(CI) – CI length;
Len(CB) – CB length.

In order to obtain the SMC string the maximum common substring extraction operator ⊗ from the two words CI and CB is defined as:

$SMC = CI \otimes CB$

where:

$Len(SMC) \leq \max\{Len(CI), Len(CB)\}$

so $ORTOC(CI, CB) \in [0,1]$.

The words „test” and „rest” are considered

SMC = „test” ⊗ „rest”

SMC = „est”.

For the defined words and the computed SMC the ortogonality is:

ORTOC(„test”, „rest”) = $1 - \frac{3}{4} = 0.25$

Orthogonality in the organizational identifier validation application is computed using a vocabulary extracted from VOCDEN, VOC = $\{C_1, C_2, ..., C_n\}$ which contains company names. The following formula is defined for orthogonality analysis:

$$ORTOTOT(CI, VOC) = \min_{i=1,n}\{ORTOC(CI, VOC_i)\}$$

If ORTOTOT(CI,V) is less than 0.75 the name must be reformulated.

b) The entity has a name formed from mere words the orthogonality is established by vocabulary and word analysis.

Vocabulary level analysis is established first in order to determine the correspondence between texts and validation necessities of the application for orthogonality computation.

In order to analyze the orthogonality of two texts T_1 and T_2 two vocabularies V_1 and V_2 are defined based on the alphabetical sort of the words that compose the two texts. The vocabularies are defined the sorted words sets $V_1 = \{C_{11}, C_{12},..., C_{1n}\}$ and $V_2 = \{C_{21}, C_{22}, ..., C_{2n}\}$ where C_{1i} corresponds to the word on position i in vocabulary V_1 and C_{2j} corresponds to the word j of V_2. For the computations done on the two sets the operator \ defined in the set theory is used defined as below on sets A and B.

$$B \setminus A = \{x \in B \mid x \notin A\}$$

If after computing the text orthogonality one of the following equations is true:

$V_1 \setminus CC=\{\varnothing\}$

$V_2 \setminus CC=\{\varnothing\}$

where:

CC – set of common words.

Then the orthogonality is represented by the formula:

$$ORTOT(V_1, V_2) = 1 - \frac{NCC}{\max(NrCV_1, NrCV_2)}$$

where:

NCC – number of common words;

$NrCV_1$ – number of words in vocabulary V_1;

$NrCV_2$ – number of words in vocabulary V_2.

If both equations are true it results that the vocabularies are identical and the orthogonality is 0.

If both equations are below are true:

$$V_1 \backslash CC \neq \{\varnothing\}$$

$$V_2 \backslash CC \neq \{\varnothing\}$$

where:

CC = $\{CC_1, CC_2, ..., CC_k\}$ the subset of common words.

The individual orthogonality of words contained in the vocabularies V'_1 and V'_2 is computed. The vocabularies V'_1 and V'_2 are computed as:

$$V'_1 = V_1 \backslash CC$$

$$V'_2 = V_2 \backslash CC$$

Each word C_{1i} in vocabulary V'_1 is compared to each word C_{2j} contained in vocabulary V'_2 in order to determine the maximum common substring SMC. The orthogonality is computed using the following method:

$$ORTOC(C_{1i}, C_{2j}) = 1 - \frac{Len(SMC_{ij})}{\max\{Len(C_{1i}), Len(C_{2j})\}}$$

where:

Len(SMC) – the length of the maximum common substring;

$Len(C_{1i})$ – the length of word i in vocabulary V'_1;

$Len(C_{2j})$ – the length of word j in vocabulary V'.

Thus the values presented in Table 3 represent the individual orthogonality of the words.

Table 3

ORTOC (V'_1,V'_2) orthogonality

	C_{11}	C_{12}	...	C_{1n}
C_{21}	x_{11}	x_{12}	...	x_{1n}
C_{22}	x_{21}	x_{22}	...	x_{2n}
....	...	...	...	...
C_{2n}	x_{n1}	x_{n2}	...	x_{nn}

where:

$$ORTOC(V'_i, V'_j) = x_{ij}$$

In order to keep the word set representative the following sets are defined:

$$Max = \max_{\substack{Len(V'_j) \\ 1 \le i \le nrCV'_i \\ j=1,2}} \{V'_j[i]\}$$

And

$$Min = \min_{\substack{Len(V'j) \\ 1 \le i \le nrCV'_i \\ j=1,2}} \{V'_j[i]\}$$

where:

Max – the vocabulary with the maximum length;
Min – the vocabulary with the minimum length;
Max_i – the word on position i of the maximum length vocabulary;
Min_j – the word on position j of the minimum length vocabulary;
$Len(V'_j)$ – the number of words in the V'_j vocabulary;
$nrCV'_i$ – the number of words in the V' vocabulary;
$V'_j[i]$ – the word on position i in the V'_j vocabulary.

In Table 4 the Max-Min orthogonality matrix is presented.

Table 4

Max-Min orthogonality matrix

	Min_1	Min_2	…	Min_n
Max_1	x'_{11}	x'_{12}	…	x'_{1n}
Max_2	x'_{21}	x_{22}	…	x'_{2n}
….	…	…	…	…
Max_n	x'_{n1}	x'_{n2}	…	x'_{nn}

The formula for computing orthogonality becomes:

$$ORTOC(Max_i, Min_j) =$$
$$= 1 - \frac{Len(SMC_{ij})}{\max\{Len(Max_i), Len(Min_j)\}} * \frac{\max\{Len(Max_i), Len(Min_j)\}}{Len(Max)} =$$
$$= 1 - \frac{Len(SMC_{ij})}{Len(Max)}$$

where:

len(Max) – number of characters from the Max word set.

$$ORTOC(V'_1, V'_2) = \sum_{i=1}^{n} \min_{j=1,n}\{ORTO(Max_i, Min_j)\}$$

In order to define orthogonality for the whole text including text level and vocabulary level orthogonality the following formula is used:

$$ORTO(V_1, V_2) = ORTOT(V_1, V_2) \times ORTOC(V'_1, V'_2)$$

By multiplying the two orthogonality measures the degree of phrase level representativeness of the word othogonality measure is obtained such that a complete analysis of the two texts is developed.

In order to compute the orthogonality for all names that have the same number of words as the inputted name an extraction is made from the VOCDEN vocabulary into VOC = {C_1, C_2, C_l} vocabulary using:

$$ORTOTOT(V_1, VOC) = \min_{i=1,n}\{ORTO(V_1, VOC_i)\}$$

If ORTOTOT(V_1, VOC) is smaller than 0.75 then the name must be reintroduced by the user.

In order to estimate the cost of implementing the security system in the informatics application for validating organizational identifiers the main factors that make up the linear cost model:

$$Ct = \sum_{i=1}^{NF} Ch_i$$

are considered. In Table 5 the parameters associated the auxiliary costs and their estimative values are displayed.

Table 5

Auxiliary costs

Name	Cost	Use duration
Laptop	4000	28
Wireless router	400	28
Software development assist tools	300	28
Software development tools	1500	28
Office space rent	600	28

Table 6 presents the expenses related to software development activities and application testing.

Table 6

Software development costs

Activity	Duration(days)	Cost per day	Total cost
Documentation	4	200	800
Developing security software	3	200	600
Testing security software	3	150	450
Fixing security software	2	200	400
Implementation and installation of the security software	4	150	600

The linear cost model is:

Ct=CA+Cd+Cds+Cts+Cdes+Ciis=6800+800+600+450+400+600=9650

where:

CA – total auxiliary cost;

Cd – documentation costs;

Cds – security software development costs;

Cts – security software testing costs;

Cdes – security software fixing costs;

Ciis – costs relating to implementation and installation of the security software modules.

8. Conclusions

In order to develop security systems in optimal costs conditions a thorough informatics security influence factors analysis is needed. A quality characteristic system must be developed in order to ensure the correlation between quality and influence factors for informatics security.

The research results are represented by efficient cost models that are tested on real world applications. Cost models are presented with regard to the actual necessities of performing software development for informatics security.

The software product for organizational identifier validation is destined for choosing organization names as clear and as different as possible from the ones already stored in the data base. Orthogonality measures are implemented and tested and a complete analysis of organizational identifiers is done. The application has been tested using real world organization names from the Romanian Chamber of Commerce for which orthogonality levels were established. The results show that the application is performing and the implemented models work as expected.

Cost estimation models were applied to the historical project data regarding the development stages of the security system associated to the application. The experimental results show the correctness of the used influence factors and their correlation with actual costs.

Notes

[1] See http://www.sans.org/reading_room/whitepapers/awareness/ the_need_for_information_security_in_todays_economy_916

[2] See http://www.ietf.org/rfc/rfc2828.txt

[3] See http://csrc.nist.gov/publications/nistir/NISTIR-7298_Glossary_Key _Infor_Security_Terms.pdf

[4] See http://en.wikipedia.org/wiki/Information_security

[5] See http://www.gartner.com/DisplayDocument?doc_cd=166713

[6] See http://en.wikipedia.org/wiki/Economic_model

[7] See http://en.wikipedia.org/wiki/COCOMO

References

Alhazmi, O.H., Malaiya, Y.K., „Application of vulnerability discovery models to major operating systems", IEEE Transactions on Reliability, vol. 57, issue 1, 2008, ISSN: 0018-9529

Anderson, E.E., Choobineh, J., „Enterprise information security strategies", *Computers & Security*, vol. 27, issue 1, 2008, ISSN: 0167-4048

Anwar, Z., Montanari, M., Gutierrez, A., Campbell, R.H., „Budget constrained optimal security hardening of control networks for critical cyber-infrastructures", *International Journal of Critical Infrastructure Protection,* vol. 2, issue 1, 2009, ISSN 1874-5482

Aroba, J., Cuadrado-Gallego, J.J., Sicilia, M.-A., Ramos, Isabel, Garcia-Barriocanal, Elena, „Segmented software cost estimation models based on fuzzy clustering", *Journal of Systems and Software*, vol. 81, issue 11, 2008, ISSN: 0164-1212

Chen, Z., Ji, C., „An Information-Theoretic View of Network-Aware Malware Attacks", IEEE *Transactions on Information Forensics and Security*, vol. 4, issue 3, 2009, ISSN: 1556-6013

Dudin, E.B., Zhlyabinkova, I.A., Zakharova, E.G., Smetanin, Yu.G., „Information security in distributed computing systems. A review", *Automatic Documentation and Mathematical Linguistics*, vol. 43, no. 4, 2009, ISSN: 1934-8371

Halstead, M. (1977). *Elements of Software Science, Operating, and Programming Systems Series*, Volume 7, New York, NY, Elsevier, Bucureşti

Huang, D.C., Hu, Q., „An economic analysis of the optimal information security investment in the case of a risk-averse firm", *International Journal of Production Economics*, vol. 114, issue 2, 2008, ISSN: 0925-5273

Ivan, I. Vişoiu, A. (2005). *Baza de modele economice*, Editura ASE, ISBN: 973-594-570-4, Bucureşti

Ivan, I., Doinea, M., Palaghiță, D., „Optimization of authentication processes in distributed applications", *Theoretical and Applied Economics*, 2008, nr. 6, ISSN 1841 – 8678, Bucureşti

Ivan, I., Vişoiu, A., Ciurea, C., Palaghiță, D., „Model bases and software quality metrics refinement", The 4th International Conference *Economy and Transformation Management*, Timişoara, 2008

Jasmine, K.S., Vasantha, R., „Cost estimation model for reuse based software products", *IMECS 2008: International Multiconference of Engineers And Computer Scientists*, 2008, ISBN: 978-988-98671-8-8

Khansa, L., Liginlal, D., „Valuing the flexibility of investing in security process innovations", *European Journal of Operational Research*, vol. 192, issue 1, 2009, ISSN: 0377-2217

Kumar, V.K., Carr, M., Kiran, R.N., „Software development cost estimation using wavelet neural networks", *Journal of Systems and Software*, vol. 81, issue 11, 2008, ISSN: 0164-1212

Lee, J., Kim, C.-Ki, „Software architecture evaluation methods based on cost benefit analysis and quantitative decision making", *Empirical Software Engineering,* vol. 14, issue 4, 2009, ISSN: 1382-3256

Lu, J., Bai, C., Zhang, G., „Cost-benefit factor analysis in e-services using bayesian networks", *Expert Systems with Applications*, vol. 36, issue 3, 2009, ISSN 0957-4174

Pendharkar, P.C., Rodger, J.A., Subramanian, G.H., „An empirical study of the Cobb-Douglas production function properties of software development effort", *Information and Software Technology*, vol. 50, Issue 12, 2008, ISSN 0950-5849

Pfleeger, Ch.P., Pfleeger, S.L. (2006). *Security in Computing, 4th Edition*, ISBN: 978-0132390774, Prentice Hall

Pocatilu, P. (2004). *Costurile testării software*, ASE Publishing House, ISBN: 973-594-549-5

Stamp, M. (2005). *Information Security: Principles and Practice*, ISBN: 978-0-471-73848-0, Wiley-Interscience

Tipton, H.F., Krause, M. (2008). *Information Security Management Handbook, Sixth Edition*, 456 pages, ISBN: 978-1420067088, Auerbach Publications

Vacca, J.R. (2009). *Computer and Information Security Handbook*, ISBN: 978-0123743541, Morgan Kaufmann

Vişoiu, A., „Refinement methods for security metrics", *Economic Informatics Conference 2009*, ISBN: 978-606-505-172-2

Vydrin, Al.S., „Theoretical aspects of information security", *Journal of Mathematical Sciences*, vol. 156, no. 2, 2009, ISSN: 1573-8795