The Role of Weighted Entropy in Security Quantification

Constantinos Patsakis, Dimitrios Mermigas, Sotirios Pirounias, and Gregory Chondrokoukis

***Abstract*—There is no doubt that an Information System faces various challenges every day and security is only one of them. It takes a lot of planning and even more work to keep it up and running. Faulty wiring, hardware problems, overloading, new versions, backups are just some of the possible failing factors. Yet, the probability of their occurrence is almost impossible to predict. The aim of this paper is to study the "behavioral model" that the technical risk factors which affect the security level of an Information System follow and suggest possible failing points. We will attempt to incorporate a more practical approach, by using real life scenarios that have occurred in many corporations, rather than "manufacturing" a theoretical approach that suits the needs of our findings. To this end we use data that are publicly available through open source databases. The analysis proposed, of the data collected, performed using weighted entropy methodology. Finally, the output of this research is going to be used as an input to the proposed model of our research group for quantifying security using Stochastic Processes[1].**

***Index Terms*—Security, entropy, quantification, risk assessment.**

I. Introduction

We live in an era of extreme uncertainties. The markets are unstable, stock exchange has become more volatile than ever, marketing is experiencing difficulties in persuading new customers to buy more products or services, competition between corporations is fierce and yet in these conditions, managers are called to undertake more risks, in order to produce better results that will increase the trust of the stakeholders and attract new investments. However, corporations have found some comfort with the adoption of complex Information Systems (IS) that led them to expand their boundaries of geographic coverage and thus be able to address to a much bigger amount of consumers. Enterprise networks have become essential to the operation of enterprises of any kind and the internet economy is constantly expanding.

In this context of growing technology dependency and increasing security risk, better measurement strategies are essential for two basic reasons: New vulnerabilities are identified with thrifty frequency. The time-lag in the vulnerability life-cycle between the discovery and exploitation phase is becoming narrower. As the number of vulnerabilities is increasing exponentially, we investigated a risk assessment methodology that groups vulnerabilities with an intuitive manner. To this end we define an IS technical risk factor as every discernible element of the system (e.g. the operation system platform used, the antivirus program used etc.). The risk factors analyzed on this paper are strictly technical and it is the topic of further research to estimate the risk factors emanating from other dimensions of an IS. We will not study the risk factors related directly to people, like physical security, or factors related to the operational dimension of a company/organization. In order for a model to fully evaluate the security risk level of an IS, it must incorporate all those three dimensions.

Our intention is to create a methodology for technical risk factor assessment using data which is easily obtainable and provides a metric that is unbiased and feasible to measure. That is the basic reason we relied the data accumulation for our research on an open source database such as the Open Source Vulnerability Database (OSVDB). Unbiasedness is observed also with the provision of a methodology that is not tied to a specific risk platform as is the case with Information Technology (IT) risk analysis tools.

It is agreed that IS security cannot be managed if it cannot be measured adequately with objectivity and unbiasedness. Objective measurement of risk can provide answers to important questions, such as "What is the current security level", "What is the current security status of an IS?", "When is an IS more vulnerable?". In this study we investigate answers to the second question that will ground further research for answering the first and most crucial one.

With the final target of objectivity we use weighted entropy methodology as a means for IS technical risk factor assessment. The output of our proposed model leads to risk factor prioritization and provides crucial inputs for the model the authors proposed on [1].

In our study we did not followed a theoretical approach but instead we utilized a coherent statistical analysis of "literally all" the technical risk factors that have damaged or disabled IS throughout a time period that spans over twenty (20) years. By the end of our analysis we will be able to provide a new methodology for assessing the security level of a risk factor that, in turn, will help managers safeguard company's business continuity.

In what follows,we firstly discuss the previous work and the method used for data acquisition. Afterwards, we present how weighted entropy could be used to quantify security, using the method in [1] and we close with some conclusions and ideas for future work.

II. Previous Work

The area of vulnerabilities taxonomy is still under debate among the researchers. Even the more universally recognized taxonomy provided by the National Vulnerability Database (NVD) of the National Institute Standards and Technology (NIST) has problems and cannot be considered as totally unambiguous [2]. This is one of the reasons we attempted to avoid analysis of IS risk factors based on vulnerabilities

Manuscript received August 15, 2012; revised October 19, 2012.

The authors are with the Department of Informatics, University of Piraeus, 80 Karaoli & Dimitriou str, 18534 Piraeus, Greece (e-mail: kpatsak@gmail.com, dmermigas@gmail.com, spirounias@yahoo.com).

categorization.

TABLE I: THE WEIGHTS OF SEVERAL MICROSOFT PRODUCTS

Vendor	Product	Total Vulnerabilities	Impact			Total Entropy Log base 10	Taking into account the time factor
			None	Partial	Complete		
Microsoft	Windows XP	3836	616	1520	1700	2,172	0.22
Microsoft	Windows 2003	2496	300	890	1306	2,062	0.35
Microsoft	Office XP	210	9	90	111	2,00	0.13

Patriciu *et al.* present an open source framework based on the Common Vulnerability Scoring System (CVSS v2) in order to evaluate security metrics for IS risk assessment[3]. The authors provided a risk prioritization method where vulnerabilities are ranked according to severity and risk. Our research is also based on an open source framework. We also investigate a method of objective risk prioritization considering risk factors which contain groups of vulnerabilities.

The CVSS is the most widespread Vulnerability Scoring System and is supported by the Forum of Incident Reporting and Security Teams (FIRST). This system analyzes and evaluates vulnerabilities based on a fairly unbiased methodology and is open to public use [4].

Frei et. al. suggest a methodology for the assessment of the risk exposure of an IS statistically analyzing the different phases of the lifecycle of each vulnerability[5]. More specifically they quantify the time lag between the exploit phase and the solution release phase as an indicator of the security level of a system. As analyzed in the following section of our study, we use the same kind of data sources as the authors do but we suggest a different kind of methodology in order to quantify security.

Alhazmi et. al. analyze vulnerabilities by category and try to connect each category to a security level so that risk analysis will be focused on vulnerabilities with higher risk[6]. Our focus on the contrary is not vulnerability-specific but risk factor-specific.

Homer et. al. provide an objective quantification of security risk using attack graphs and CVSS metrics[7]. They provide a likehood estimation for the successful utilization of a vulnerability by an attacker. Again the proposed methodology for security risk quantification is vulnerability-specific.

Peotta et. al. propose a risk analysis framework using internal and external data for vulnerability analysis which is performed using CVSS v2[8]. The main idea is the constant assessment of the security level of an IS in order to encounter the vulnerabilities generation which is due to the rapid technological advancements. This agrees to our basic concept that security is not static and must be dealed on an ongoing manner.

III. DATA ACQUISITION

It is very common as well as useful to present an example in order to support an argument. However, when it comes to security risk factors, things are a little bit different. What should be considered the starting point and where should we end the research for risk factors? The number of risk factors that have temporarily or permanently impaired an IS cannot be easily deducted. Fortunately, there are many organizations that provide publicly available information like that of the Open Source Vulnerability Database (OSVDB) [9]. The OSVDB is the type of database that are called "reporting vulnerability mechanisms". Those are maintained by organizations such as the Open Security Foundation, the Security Focus, the National Vulnerabilities Database , the MITRE Corporation and individual software developers. The databases also record some of the significant attributes of the vulnerabilities[5].

OSVDB synopsis states that "OSVDB is an independent and open source database created by and for the security community. The goal of the project is to provide accurate, detailed, current and unbiased technical information on security vulnerabilities"[9]. Therefore, in order to study the risk factors for an IS we decided to use this database which holds more than twenty (20) years of real data. Our decision was based primarily upon the need to include all the important risk factors that can possibly impair an IS and also to achieve unbiasedness in the investigation of different IS.

Let us take an example of two researchers where one of them is affiliated to Windows Operating Systems (OS) while the other to Unix-like OS. It is our human nature that would lead the first to overlook –up to a certain extend – some security risks of Windows OS while the second to argue in favor of Unix-like OS. That is why, in order to overcome our human nature, manage to stay unbiased and perform a thorough research on security risk factors,we used the OSVDB.

The data that we used were actual data, that is, real cases that actually happened. From these, we managed to draw conclusions as to how often do we have a security vulnerability for a certain operating system, how severe it is as well as when was it patched. Based on the above conclusions we can determine the patterns of the offered security from certain software in the pass of time.

The data of the OSVDB contained various software vendors which have built software applications. Each application may have more than one versions and each vendor may have more than one application. A example of that is Microsoft Corporation which has distributed over time different applications, such as different versions of OS, different versions of Microsoft Office or even different versions of Microsoft SQL Server. Within the database various records of vulnerabilities can be found for affecting the above applications as well as the impact for each one of them.

Our research area is mainly focused on a certain software vendor that is Microsoft, due to the great variety of products as well as versions that it has produced over time. The large number of software products and versions that have been

released produced, in turn, a substantial number of vulnerabilities. By measuring the vulnerability occurrence of each product, we managed to calculate their possibility,p_i using the following formula:

$$p_i = \frac{Numberofvulnerabilitiesofimpacti}{Totalnumberofvulnerabilities}, \quad (1)$$

where *i* is the index number of an impact scale. Later on in this paper, we will use these p_i to measure the security level of an IS.

IV. Measuring Security Level

Almost every day a new vulnerability is found when it comes to IS security. The exact day that each new security risk will threat a specific IS cannot be predicted, thus we ought to think security as more random and less static. In[1], the authors proposed a new method for quantifying the security status of a service or a system. Now, towards the same direction, we will try to measure the security level of an IS by studying its behavioral model using weighted entropy to determine the weight of each risk factor.

Shannon [10] is considered to be the father of information theory and was the first that incorporated the term information entropy in an IS for measuring the uncertainty associated with a random variable. Following his steps, we measured the volume of entropy in conjunction with the random factor, that of the "vulnerability appearance". The definition of the information entropy is, however, quite general, and is expressed in terms of a discrete set of probabilities p_i:

$$H(X) = -\sum_{i=1}^{n} p_i log(p_i) \quad (2)$$

In our study we consider risk factors to be elements of the system. As Frei *et al.*[5] state, the security level of an IS is determined by the software vulnerabilities of its elements and their interaction. For example, the operation system that is decided to be used on an IS is a certain risk factor. Measuring the entropy related to the vulnerabilities that a specific risk factor has produced we manage to achieve an unbiased quantitative measurement of its risk level.

Since we decided to use the OSVDB database, finding the p_i was not actual an issue. Nevertheless, assigning the same probability to any risk factor –or vulnerability- that was recorded throughout the years was not going to produce accurate results. If we were to do that, then we would have had overlooked a substantial piece of information that is, the perceptual quality of each risk factor. Stated in other words, each risk factor is not expected to compromise a system at the same degree. Hence, we studied the impact of each risk factor in our database. The aim was to determine a weight of each factor, not according to the total number of vulnerabilities, but according to their severity over the years. It is obvious that the widely used operating systems, should be more prone to attacks, hence more vulnerabilities would be found and published about them. This should not change the weight of the risk factor, unless most of them appear to be severe.

In the used database, we found out that there were three kinds of impact that each risk factor could have to an IS: (a) none, (b) partial and (c) complete. Using this qualitative distinction, we assigned different weights to each kind of impact: 1, 3 and 10 respectively.

Using the weighting coefficients proposed above in the volume calculation of the information entropy we result to the following revised formula:

$$W(X) = W(w_1, w_2, \dots w_n; p_1, p_2, \dots p_n) = -\sum_{i=1}^{n} w_i p_i log(p_i), \quad (3)$$

where w_i is the weight of each security factor impact.

Putting it all together, we used the OSVDB database which gave us unbiased information about all risk factors recorded for vendors as well as their products. We used the revised formula for calculating the volume of information entropy. We assigned different weight according to the impact of each risk factor and finally we calculated the desired entropy for a product. The results can be seen in .

If we assume that we have an ISin which we have identified all its risk factors, in order to measure its security level using the formula proposed in[1], we have to achieve a more accurate determination of f_i, so we should assign a weight to each one of them and set it as its exponent. We demand that if one of the risk factors is zeroed then the total security level is zeroed, something that implies a product. If we assume that each factor is independent of all the others, then in order to apply the needed weights, the proposed formula in [1]becomes:

$$\int \prod f_i^{\frac{1}{c_i}}(t)dt \quad , \quad (4)$$

where c_i is the weight of each factor. We set c_i as calculated weighted entropy of risk factor i, thus

$$c_i = -\sum_{j=1}^{n} w_j p_j log(p_j), \quad (5)$$

where n is equal to the number of different impacts we have and w_i is their weight respectively. In this way, we determine how much is the product affected by these factors, using the weighted entropy of the severity of their vulnerabilities as the unbiased measure. Moreover, functions f_i are only subject to time variable.

Since $0 \leq f_i \leq 1$ we have that if it increases then the corresponding weight increases as well.Similarly, when one decreases, the same happens for its weight. In order to balance the whole quantity, we raise it to the power equal to the number of years that passed from the discovery of a vulnerability and we define it as k.

Using the above methodology we apply the same weight for vulnerabilities of previous years, with the ones from the current year. Of course we want to take into consideration the continuity of the software developers, yet our measure should apply a penalty to the products that are more prone to recent

vulnerabilities. Therefore we have to tamper the c_is.

In order to apply the time penalty, we change formula (5)as

$$c_i = -\sum_{j=1}^{n}\sum_{k=0}^{m} e^{-k} w_j p_{jk} log(p_{jk}), \qquad (6)$$

where *m* is the number of years that product i is in the market and p_{jk} is the probability that product i has a vulnerability of severity class j and it was k years ago.

The weight could be furthered improved by inserting the percentage of time used. If for example a certain application is prone to vulnerabilities, yet it is not used all the time then in order to objectively determine its impact on the total security of the IS, we might have to take into consideration the percentage of time this application is used by the IS. Thus the formula becomes

$$c_i = -\sum_{j=1}^{n}\sum_{k=0}^{m} t_{ik} e^{-k} w_j p_{jk} log(p_{jk}), \qquad (7)$$

where t_{ik} represents the percentage of time application i is used, k years ago. Having calculated the weights with any of the above scenarios, all that is left to calculate the security level is to approximate the stochastic functions.

V. Conclusions

Since security is a significance milestone for any business, it is very crucial to measure and quantify it. Unfortunatly, this work embeds the human factor and the perception of someone over how exposed he feels to the attacker.

This work denotes the necessary steps that one should take when applying an unbiased method for measuring security. However, applying a weight to each category of risk factors,involves a certain level of objectivity, yet it makes some steps forward to a more unified, unbiased and objective quantification of security. Moreover, it shows that if we are able to approximate the stochastic functions that the risk factors follow, then we will be able to quantify security, not only for the current state, but make an approximation of the security level in the future.

References

[1] C. Patsakis, D. Mermigas, S. Pirounias, N. Alexandris, and E. Fountas, "Towards a formalistic measuring of security using stochastic calculus," in *2010 3rd IEEE International conference on computer science and information technology*, 2010.

[2] National Vulnerability Database. [Online]. Available: http://ndv.gov/com

[3] V. Patriciu, I. Priescu, and S. Nicolaescu, "Security metrics for enterprise information systems," *Journal of applied quantitative information systems*, vol. 1, no. 2, pp. 151-159.

[4] P. Mell, K. Scarfone, and S. Romanosky, "A complete guide to the common vulnerability scoring system version 2.0," 2007.

[5] S. Frei, M. May, U. Fiedler, and B. Plattner, "Large-scale vulnerability analysis," *Computer engineering and networks labaratory*.

[6] H. O. Alhazmi, S. W. Woo, and Y. K. Malaiya, "Security vulnerability categories in major software systems."

[7] J. Homer, X. Ou, and D. Schmidt, "A sound and practical approach to quantifying security risk in enterprise networks,"

[8] L. Peotta and P. Gondim, "A framework for risk assessment of information technology in the corporate environment," *International Journal of Forensic Computer Science*, vol. 1, pp. 75-88.

[9] Open Source Vulnerability Database. [Online]. Available: http://osvdb.org/com

[10] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, 1948.

Constantinos Patsakis was born in 1979 at Marousi, Athens, currently he is post-doc researcher at Rovira IVirgili University and waiting his appointment as Lecturer at the Department of Informatics, University of Piraeus. He holds a BSc in Mathematics Department, University of Athens. MSc from Royal Holloway, University of London in Information security and PhD from Department of Informatics, University of Piraeus. His main areas of interest are Cryptography, Cryptanalysis, Cryptocirology and Computer Security. Dr Patsakis has been teaching since 2004 at the University of Piraeus and two years at the Technical Institute of Kalamata.

Mr. Dimitrios Mermigas was born in 1971 at Kallithea, Athens and currently is working at Bank of Greece at the Internal Audit Department, Section of Information Systems Audit. He received his first degree from the Informatics Department, Technological Institute of Athens. Mr. Mermigas then received his Masters' degree on Business Administration from Leicester University on United Kingdom and currently he is a PHD candidate at the Department of Informatics, University of Piraeus. He has a sound knowledge of software development and project management as he has been a software developer and then architect specializing at ERP solutions, for several years.

Mr. Sotirios Pirounias was born in 1978 at Athens, Greece. He received his first degree from the Department of Maritime Economics, University of Piraeus. Mr. Pirounias then received his Masters' degree on Finance and Banking Administration from the University of Birmingham and currently he is a PHD candidate at the Department of Informatics, University of Piraeus. His main fields of study are Risk Management and Computer Security. He is working at Ecological SA, a company that is a leader in facilities services in Greece, as a Financial Director since 2003. His current research interests are concentrated on the assessment of the security of information systems and on the cost appraisal of security incidents.