# An Empirical Study of Connections Between Measurements and Information Security

Rodrigo Sanches Miani*, Michel Cukier†, Bruno Bogaz Zarpelão*, Gean Davis Breda*, Leonardo de Souza Mendes*
*School of Electrical and Computer Engineering
University of Campinas, Campinas, SP, Brazil
Email: {rsmiani,bzarpe,gean,lmendes}@decom.fee.unicamp.br

†A. James Clark School of Engineering
University of Maryland, College Park, USA
Email: mcukier@umd.edu

***Abstract*—This paper presents an investigation of factors that are likely to affect the security of an organization, in particular, the number of security incidents. Using Intrusion Prevention Systems (IPS) data, provided by the University of Maryland, we derive three potential factors (attackers, corrupted computers and attack types) and their respective measurements. Based on empirical studies and information security literature, we examine the effects of selected factors on the number of security incidents. We use a regression model to test the hypotheses empirically and also to study how those factors are affected over time. We found that the number of potential corrupted computers is positively related to the security incidents while the number of potential attackers and range of attack types does not significantly affect the number of security incidents. We also found empirical evidence that factors could significantly change over time.**

***Keywords-Network and Security Management; Security Metrics; Empirical Study; Security Incidents; Intrusion Prevention Systems.***

## I. INTRODUCTION

In information security, questions such as "Is security improving over time?" and "Are we using effective controls?" could be used to derive measurements to facilitate decision making and improve performance and accountability. Researchers have been trying to measure security using models from other disciplines such as economics, risk, reliability engineering and statistics. However, as pointed out by Jansen [1] much of what has been written about security quantification is definitional, aimed at providing guidelines for defining a security metric and specifying criteria for which to achieve.

Currently, there are many suggestions in the security community for what measures organizations should collect in order to construct security measurement models [2], [3] and [4]. However, as noted by Verendel [5], for most cases it is unknown if the proposed models are valid or not in representing security for systems in realistic environments due to the lack of validation and comparison between such methods against empirical data. In other words, little work has been done to determine the value of these measures in real-world operational environments [6]. Research is needed to validate connections between measures and security, and determine and understand possible correlations.

In this paper, we study connections between metrics derived from intrusion prevention system (IPS) alert events and the number of security incidents. The number of security incidents is an important security indicator that in combination with other metrics can indicate the level of threats and effectiveness of security controls.

Since security incidents may be the result of several factors, from a computer infected with a virus to successful attacks against servers, our aim is to investigate some of these factors that could be derived from Intrusion Prevention Systems (IPS) data. We empirically examine, using a multiple linear regression model, the effects of potential attackers, potential corrupted computers and attack signatures on the number of security incidents reported by the Office of Information Technology (OIT) at the University of Maryland [7]. We also analyze how those factors are affected over time.

This paper is structured as follows. Section 2 describes the background on security incidents and intrusion prevention systems. Section 3 introduces our hypotheses about the relationship between factors and the number of security incidents. Section 4 presents the empirical modeling approach and the regression results. Section 5 discusses the threats to validity of our study. We provide conclusions and directions for future work in Section 6.

## II. BACKGROUND

This section describes the background on security incidents, intrusion prevention systems and security quantification that we will use in this paper.

### A. Security Incidents

A security incident, according to the Center for Internet Security [2], results in the actual outcomes of a business pro-

 ISBN: 978-1-61208-209-7 

cess deviating from expected outcomes for confidentiality, integrity, and availability resulting from people, process, or technology deficiencies or failures. There are many interacting factors that affect the occurrence of a computer security incident, for instance, virus, vulnerabilities, characteristics of the population being attacked, distribution and prevalence of vulnerable operating systems and applications and intensity of attacks [8], [9]. Examining the factors that could lead to security incidents is important to improve forecasting and also to identify conditions which may result in the spread of new incident types.

Security incidents may be reported using operational security systems sources, such as anti-malware software and intrusion detection systems (IDS), host logs and also reports from users. In this work we are interested in studying the number of security incidents.

The number of security incidents indicates the number of detected security incidents that the organization has experienced during a time period. However, the use of this metric should be properly examined to avoid misinterpretation of data. In combination with other metrics, the number of security incidents can indicate the level of threats, incident detection capabilities and effectiveness of security controls and can be used as a security indicator of an organization. Understanding which factors are likely to affect the number of security incidents could begin to paint a picture of system security.

### B. Intrusion Prevention Systems

An IPS is considered as an extension of an IDS that monitors malicious activity and reacts in real time by blocking a potential attack [10]. An IDS is a passive device that monitors activity whereas an IPS is an active device that blocks the potential malicious activity.

The investigated IPS device is a signature-based IPS where the blocking decision relies on a set of signatures that are regularly released by the vendor as attacks are newly discovered on the Internet. Basically, when the characteristics of an attack match the ones of a defined signature, the attack is blocked and an alert is recorded.

Based on the data provided by the IPS, it is possible to derive metrics to assess the volume and the nature of the malicious activity. For instance, a set of metrics that includes the number of alerts, number of distinct targets, number of distinct IPS signatures and number of blocked attackers was proposed by [11].

### C. Security Quantification

The ISO/IEC 27004 [12] standard includes specific guidelines about information security measurement such as: measures and measurement development, measurement operation, data analysis and measurement results reporting, and also a template on which to describe a metric.

Jansen [1] provides an overview of the security metrics area and look at possible avenues of research that could be pursued to advance the state of the art. The author states that much of what has been written about security metrics is definitional, aimed at providing guidelines for defining a security metric and specifying criteria for which to achieve. However, relatively little has been reported on actual metrics that have been proven useful in practice.

Condon et al. [8], [9] describe the application of time series models and software reliability models on computer security incidents. They found that certain incidents are caused by well-defined vulnerabilities and might easily patched against but others may exhibit propagation behavior similar to contagious diseases in animals or still can result from economic incentives external to an organization or environment.

Considering other empirical analysis about security measurement, Chrun et al. [11] presents a method that ranks potentially corrupted computers using security metrics derived from imperfect IPS event data. However, nothing was said about the relationship between corrupted computers and computer security incidents.

Cukier and Panjwani [13] conducted an empirical analysis to quantify the link between vulnerabilities and malicious connections. They conclude that a high number of vulnerabilities on services do not necessarily imply a high number of malicious connections or successful attacks.

This study can be used as motivation to investigate another relevant security issues such as: does the high number of attacks on networks imply a high number of security incidents? Does the high number of corrupted computers on networks imply a high number of security incidents?

Our work focuses on extracting empirical relationships between IPSs and computer security incidents datasets and also how to use these results to improve the knowledge about system security.

## III. Research Hypotheses

In this section, we investigate the relationship between the selected factors and their impact on the number of computer security incidents. From this investigation, we formulate hypotheses to guide us in the study.

### A. Attackers

From models that predict the likelihood of burglary or other conventional crimes, it has been demonstrated that parameters such as a motivated offender, a suitable victim and the absence of a motivated guardian can have a major effect on the probability of crime [14].

In information security, the same analogy is hard to achieve. According to Schechter [15], the number of potential attackers is likely to be positively correlated with the rate of security breaches and the resulting security risk. In addition to that, certain variables such as means, motive and

 ISBN: 978-1-61208-209-7 

opportunity to attack, may also influences the number of potential attackers.

As pointed out by Chrun [16], the three major attacker's characteristics which influence on security are motivation, qualities and expertise. For this reason, it is difficult to predict that the number of attackers increases the number of incidents, for example. Besides, an attack could be initiated from an external or internal source and insiders do not generally demonstrate the same attack pattern that external attackers do [17].

With this in mind, we propose the following alternate pairs of exploratory hypotheses: i) H1(a) The number of attackers increases the number of security incidents and ii) H1(b) The number of attackers does not increase the number of security incidents.

### B. Attack signatures

According to Chrun et al. [10], analyzing the range of attack types that target an organization might help the security team identifying popular attack types and making decisions for their network security. With this in mind, it is reasonable to assume that the number of attack signatures is a factor that might be linked to the number of security incidents.

In the simplest case, we can say that an increase in the number of attack signatures might be linked to the number of incidents due to the higher number of attack types that should be handled by the organization. We can also consider the severity of attack signatures. IPS vendors usually classify each attack signature based on the impact of attack on the network. More severe attacks could represent more security incidents. However, attack signatures may have limited impact on the number of security incidents due to the way that organizations react to different attack types. In other words it is difficult to predict that the number of attack signatures increases the number of incidents, for example.

Therefore, we propose the following alternate pairs of hypotheses, also considering the severity of signatures: i) H2(a) The number of attack signatures increases the number of security incidents, ii) H2(b) The number of attack signatures does not increase the number of security incidents, iii) H3(a) The severity of attack signatures increases the number of security incidents and iv) H3(b) The severity of attack signatures does not increase the number of security incidents.

### C. Corrupted computers

A corrupted computer is a potential source of attack inside an organization. Malwares, computer viruses, worms and even unpatched applications are some feasible causes for computer corruption.

The relationship between corrupted computers and security incidents could be depicted using computer security data breach reports. A study conducted by Verizon [18] using 141 breach cases, revealed that 38% of studied security incidents were caused by a computer corrupted by malware. Other study conducted by PricewaterhouseCoopers [19] using the survey data of 539 respondent organizations showed that 14% of incidents were caused by viruses or malicious software corruption. Besides that, 23% was caused by systems failure or data corruption which could also involve corrupted computers.

A similar survey [20] based on the responses of 351 computer security practitioners in U.S. corporations, government agencies, financial institutions, medical institutions and universities, revealed that 67.1% of studied incidents were provoked by malware corruption. Despite the differences between the studies, the rate at which corrupted computers affected the investigated attacks could be considered significant.

In face of these findings we propose the following hypothesis: H4(a) The number of corrupted computers increases the number of security incidents.

## IV. Empirical Modeling Approach

This section presents the description of the dataset and our empirical modeling approach.

### A. Data and measurements

The dataset provided by the University of Maryland consisted of over 2615 security incidents and 6.687.874.770 IPS alerts recorded during a period of four years and four months (from September 4, 2006 to December 31, 2010).

The data were grouped in a weekly basis ($t = 226$ weeks), with Monday as the first day of the week. Grouping the data in a time window is a technique to minimize the effects of lag time between occurrence of an event and submission of an incident report. However, due to the human interaction in the incident reporting process, we cannot prove that an incident reported in certain week would have been related to IPS alerts from the same week.

The incidents dataset also included 21 different incident types. Only one type called "nethicsreq" was removed from the dataset. It represents the identification of an illegal use of copyrighted material like music and movies. This kind of incident cannot be detected by the IPS and, therefore, is out of scope of our investigation.

The IPS is located at the edge of the organization so it cannot detect traffic originating inside the organization and targeting computers inside the organization. This dataset includes two cases, represented by Figure 1: 1) where a computer outside the organization is targeting the organization, and 2) where a computer inside the organization is targeting computers outside the organization.

Chrun et al. [11] proposed metrics for both cases. When the organization is the target (case 1), the following metrics were proposed: number of alerts, number of distinct targets,

 ISBN: 978-1-61208-209-7 

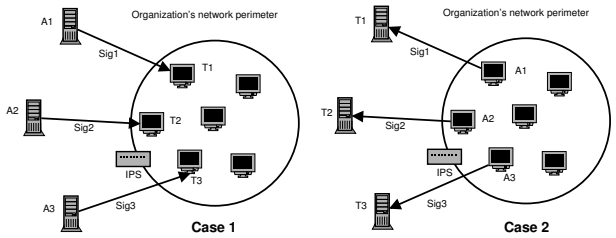


Figure 1. IPS location and attacker's perspective

number of distinct IPS signatures, number of alerts per target and number of attackers per target. When focusing on the traffic originating inside the organization and targeting computers outside the organization (case 2), the following metrics were proposed: number of alerts, number of distinct attackers, number of distinct IPS signatures, number of alerts per attacker and number of targets per attackers. We derive the investigated factors (attackers, corrupted computers and attack signatures) from these two sets of metrics.

We derive the number of security incidents (denoted as $I_t$, $t = 1, \ldots, 226$) from the provided incidents dataset and the number of potential attackers, attack signatures and potential corrupted computers from the provided IPS alerts database.

We consider the number of potential attackers in this study as the number of distinct external blocked attackers (denoted as $A_t$, $t = 1, \ldots, 226$) in the IPS as showed in case 1.

The number of attack signatures is extracted using two measures, one from case 1, the number of distinct signatures from external attackers (denoted as $Ve_t$, $t = 1, \ldots, 226$), and one from case 2, the number of distinct signatures from internal attackers (denoted as $Si_t$, $t = 1, \ldots, 226$). In this way it is possible to estimate the attack types associated with computers that are being exploited by external attackers and also the attack types associated with computers that are being used to launch attacks against external targets. The same applies for the severity of a attack signature where the severity level of a signature is classified by the vendor in three levels: critical (denoted as $SCe_t$ and $SCi_t$), major (denoted as $SMe_t$ and $SMi_t$) and minor (denoted as $SNe_t$ and $SNi_t$).

The number of potential corrupted computers could be measured using two metrics from the IPS dataset: number of distinct targets in case 1 (denoted as $Ce_t$, $t = 1, \ldots, 226$) and the number of distinct internal blocked attackers, showed in case 2 (denoted as $Ci_t$, $t = 1, \ldots, 226$).

In case 1, it is clear that the number of distinct targets reflects the number of targeted computers and thus potentially corrupted computers. The number of distinct internal blocked attackers in case 2 could be seen in two different ways: if a computer inside the organization is launching an attack, it might the result of a willing attacker who launches an attack, or it may be due to an already corrupted computer launching attacks. Therefore, it is reasonable to assume that the number of external targets and the number of internal attackers could be used to approximate the number of corrupted computers on a network. Table I presents the descriptive statistics of these variables.

### B. Empirical Modeling

In order to investigate how the variables defined in the previous section might be linked to the number of security incidents, we built a multiple linear regression model. Our dependent variable is the weekly number of incidents $I_t$ and the independent variables are $A_t$, $Ce_t$, $Ci_t$, $Se_t$, $SCe_t$, $SMe_t$, $SNe_t$, $Si_t$, $SCi_t$, $SMi_t$ and $SNi_t$. The general notation of the multiple regression functions can be written as:

$$I_t = \beta_0 + \sum_{k=1}^{j} \beta_k X_k + \epsilon_t \quad (1)$$

where $\beta_0$ is the constant term, $\beta_1$ to $\beta_j$ are the coefficients on the $j^{th}$ independent variable, $j$ is the total number of independent variables, $\epsilon_t$ the error term and $X_k$ represents the set of independent variables that could include $A_t$, $Ce_t$, $Ci_t$, $Se_t$, $SCe_t$, $SMe_t$, $SNe_t$, $Si_t$, $SCi_t$, $SMi_t$, $SNi_t$. In a multiple linear regression model, we assume that the relationship between the variables is linear and the error terms are distributed normally.

With a multiple linear regression model, it is possible to detect the effect of the independent variables on the dependent variable using a variable selection approach, that is, the screening of the candidate variables to obtain a regression model that contains the 'best" subset of independent variables [21]. The main idea is to select the independent variables, run the regression model and study its significance through the p-value obtained in each variable. Further information about variable selection in multiple linear regression models can be found in [22] and [23].

Our goal is to investigate the effects of attackers, attack signatures and corrupted computers on the number of security incidents using several variables. Therefore, we propose five different regression models in order to investigate the differences between using the number of attack signatures or the severity of attack signatures, and also analyze the impact of each single severity level.

In the first model, we analyze the number of attackers, attack signatures and corrupted computers. In the second model we exclude the number of signatures, in order to analyze the signature severity levels. In models 3, 4, 5 we analyze the impact of each severity level. Model 3 includes only the critical attack signatures, model 4 includes only the major attack signatures and the model 5 includes only the minor attack signatures. Table II summarizes the regression models and results.

We noticed that in all proposed regression models, the weekly number of attackers, $A_t$, cannot significantly affect

 ISBN: 978-1-61208-209-7 

Table I
DESCRIPTIVE STATISTICS OF VARIABLES (226 OBSERVATIONS)

| **Variables** | Mean | Std. Dev | Min | 1st Quartile | Median | 3rd Quartile | Max |
|---|---|---|---|---|---|---|---|
| $I_t$ Incidents | 9.80 | 12.236 | 0 | 3.00 | 6.00 | 11.25 | 64 |
| $A_t$ External attackers | 8674.53 | 57686.163 | 31 | 795.00 | 1424.00 | 2193.50 | 822698 |
| $Ce_t$ External targets | 1532.70 | 767.803 | 78 | 1030.50 | 1262.00 | 1801.50 | 6244 |
| $Ci_t$ Internal attackers | 103.87 | 76.991 | 11 | 56.00 | 83.50 | 131.25 | 504 |
| $Ve_t$ Ext_Signatures | 86.40 | 124.119 | 15 | 62.00 | 72.00 | 82.25 | 1347 |
| $SCe_t$ Ext_Signatures_critical | 58.25 | 81.797 | 10 | 43.00 | 50.00 | 56.00 | 1074 |
| $SMe_t$ Ext_Signatures_major | 14.21 | 15.266 | 2 | 7.00 | 10.00 | 17.00 | 151 |
| $SNe_t$ Ext_Signatures_minor | 14.04 | 36.583 | 0 | 8.00 | 11.00 | 14.00 | 535 |
| $Si_t$ Int_Signatures | 29.00 | 32.132 | 8 | 20.00 | 24.00 | 29.25 | 426 |
| $SCi_t$ Int_Signatures_critical | 16.22 | 5.794 | 4 | 12.00 | 15.00 | 19.00 | 40 |
| $SMi_t$ Int_Signatures_major | 7.38 | 27.896 | 0 | 2.00 | 3.00 | 5.00 | 377 |
| $SNi_t$ Int_Signatures_minor | 5.40 | 3.810 | 0 | 4.00 | 5.00 | 6.00 | 39 |

Table II
REGRESSION RESULTS - ENTIRE DATASET

| Variables | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| $A_t$ | -0.00001 (0.000014) | -0.00001 (0.000014) | -0.00001 (0.000014) | -0.00001 (0.000014) | -0.00001 (0.000014) |
| $Ce_t$ | 0.002650* (0.00106) | 0.00259* (0.01257) | 0.00273* (0.001) | 0.00266* (0.001) | 0.002698* (0.001) |
| $Ci_t$ | 0.022947* (0.01083) | 0.0157 (0.01257) | 0.0121 (0.0117) | 0.02394* (0.01) | 0.0256* (0.0111) |
| $Se_t$ | -0.00717 (0.00657) | - | - | - | - |
| $SCe_t$ | - | -0.01432 (0.0185) | -0.00939 (0.0098) | - | - |
| $SMe_t$ | - | 0.075 (0.1162) | - | -0.03446 (0.055) | - |
| $SNe_t$ | - | -0.0211 (0.03569) | - | - | -0.0254 (0.0222) |
| $Sit$ | 0.006619 (0.02578) | - | - | - | - |
| $SCi_t$ | - | 0.3259* (0.1634) | 0.3363* (0.1564) | - | - |
| $VMit$ | - | -0.00023 (0.036) | - | -0.0003 (0.0292) | - |
| $VNit$ | - | -0.16 (0.27) | - | - | -0.1378 (0.2227) |

* coefficient significant at 0.05 ** coefficient significant at 0.01 ( ): standard error

the number of security incidents, supporting H1(b). According to our previous discussion, attacker's characteristics such as motivation, expertise and qualities influence the number of potential attackers. Our results suggest that we cannot use the number of attackers as an indicator of security incidents. Besides that, we may also need to develop metrics to help characterize the attacker in order to refine the attacker's dataset. It may be necessary to study characteristics such as attacker geographical origin, time of attack, targets and type of attack to create a subset of possible relevant attackers.

According to Table II, the Model 1 shows that the number of attack signatures, represented by $Si_t$ and $Se_t$, cannot significantly affect the number of security incidents, supporting H2(b) and also consistent with the results presented in [13]. While the number of attack signatures may be used to find corrupted computers [10] and reveals the range of attack types that target the organization, they are not related to the number of security incidents.

This fact can be due to the limitation of signature-based devices. The detection of new attacks in such devices depend on how fast the vendor will provide new signatures for them. In other words, new attacks will not be detected nor blocked and such unblocked attacks may be the origin of some security incidents, decreasing the impact of attack signatures on security incidents.

Regarding the severity of signatures, none of the distinct signatures from external attackers ($SCe_t$, $SMe_t$, $SNe_t$) significantly affect the number of security incidents.

The same thing occur with the distinct signatures from internal attackers ($SMi_t$, $SNi_t$), except with the critical signatures from internal attackers, $SCi_t$, that significantly affect the number of security incidents, as seen in models 2 and 3. Thus, we conclude that only critical signatures associated with computers from the organization that are being used to launch attacks against external attackers impact the number of security incidents, supporting H3(a), that is, the severity of attack signatures increases the number of security incidents.

Across all these five models, we found that the number of external targets, $Ce_t$, significantly impacts the number of security incidents. The other metric related to corrupted computers, number of internal attackers, $Ci_t$, significantly impacts the number of security incidents in models 1, 4 and 5, according to the severity of attack signatures.

In cases where the number of internal critical signatures is considered, the internal attackers are not significantly. Thus, hypothesis H4(a) is supported. In other words, the more corrupted computers, the more frequent the security incidents. Thus, in our case, the number of corrupted computers could be used as an indicator of security incidents.

The results also reveal that the impact of internal threats, such as the number of internal attackers and the number of internal critical signatures ($Ci_t$ and $SCi_t$), on the number of security incidents are more significant than the external threats, such as external attackers and signatures ($A_t$, $SCe_t$, $SMe_t$, $SNe_t$). This result may indicate that external threats

 ISBN: 978-1-61208-209-7 

do not impact the number of security incidents or that the reported security incidents are not reflecting the actual external threats.

Using the same datasets, we can also study the behavior of the factors over time. Since attacks could change over time, it is reasonable to assume that the metrics should also change in order to follow the new trends.

As pointed out by [6], cyber technology is so dynamic that the meaning of metrics changes over time. There may be additional factors that influence the significance of the measure, as well as different relative importance for the existing factors [6]. Therefore, we investigate the variables over two periods of time: 2007-2008 and 2009-2010. We decide to exclude the four months of 2006 in order to preserve the entire year. We also chose the aggregation of two years in each analysis to maintain an acceptable number of observations per dataset (105 per dataset). Tables III and IV summarize the regression results.

Table III
REGRESSION RESULTS - 2007-2008

| Variables | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| $A_t$ | -0.000005 ($7 \cdot 10^{-7}$) | -0.000005 ($7 \cdot 10^{-7}$) | -0.000005 ($7 \cdot 10^{-7}$) | -0.000005 ($7 \cdot 10^{-7}$) | -0.000005 ($7 \cdot 10^{-7}$) |
| $Ce_t$ | 0.0049** (0.000892) | 0.0042** (0.000963) | 0.004487** (0.0009) | 0.0044** (0.0009) | 0.00467** (0.00092) |
| $Ci_t$ | 0.022947 (0.01083) | 0.00554 (0.011) | -0.003311 (0.01) | -0.001997 (0.01) | -0.00112 (0.01025) |
| $Se_t$ | -0.003317 (0.003433) | - | - | - | - |
| $SCe_t$ | - | -0.02552* (0.0103) | -0.0053 (0.0051) | - | - |
| $SMe_t$ | - | 0.215934** (0.076) | - | 0.0012 (0.03) | - |
| $SNe_t$ | - | -0.039 (0.01962) | - | - | -0.01152 (0.011) |
| $Si_t$ | 0.01318 (0.1048) | - | - | - | - |
| $SCi_t$ | - | -0.01896 (0.1698) | 0.0243 (0.1511) | - | - |
| $SMi_t$ | - | -0.02117 (0.32511) | - | 0.1421 (0.2845) | - |
| $SNi_t$ | - | -0.36126 (0.3692) | - | - | -0.2217 (0.3711) |
| | * coefficient significant at 0.05 | ** coefficient significant at 0.01 | ( ): standard error | | |

Analyzing the Tables III and IV, we can note that in both datasets the number of attackers and the number of signatures cannot significantly affect the number of security incidents, the same as in the previous analysis.

Regarding the differences between the two periods, in 2007-2008 the number of external targets, reflecting the potentially number of corrupted computers, significantly impacts the number of security incidents. However, the other metric related to corrupted computers, number of internal attackers, does not significantly affect the number of security incidents as showed in the analysis with the entire dataset.

This metric became significant only in the next period, 2009-2010, showing that the effects of measures changes over time. This may be due to several factors as the increase or decrease of these metrics or changes in the reported type of security incidents. In other words, the measurement context is very important to improve the accuracy of measures and metrics and also to develop new metrics.

Table IV
REGRESSION RESULTS - 2009-2010

| Variables | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| $A_t$ | -0.000006 (0.00009) | 0.000022 (0.00009) | 0.000016 (0.00009) | 0.000005 (0.00009) | 0.00001 (0.000014) |
| $Ce_t$ | -0.0004 (0.0039) | -0.0026 (0.0043) | -0.00085 (0.0039) | -0.00088 (0.00375) | -0.001 (0.004) |
| $Ci_t$ | 0.04247* (0.0146) | 0.02879 (0.01743) | 0.0349* (0.01659) | 0.0442** (0.0141) | 0.037* (0.015) |
| $Se_t$ | -0.04219 (0.07) | - | - | - | - |
| $SCe_t$ | - | 0.08362 (0.1817) | -0.06287 (0.11221) | - | - |
| $SMe_t$ | - | -0.3715 (0.33279) | - | -0.2019 (0.21867) | - |
| $SNe_t$ | - | 0.0117 (0.618) | - | - | -0.07624 (0.4617) |
| $Si_t$ | -0.00353 (0.0281) | - | - | - | - |
| $SCi_t$ | - | 0.24 (0.229) | 0.1978 (0.22121) | - | - |
| $SMi_t$ | - | -0.0736 (0.0418) | - | -0.0133 (0.031) | - |
| $SNi_t$ | - | 0.7165* (0.3534) | - | - | 0.3471 (0.27187) |
| | * coefficient significant at 0.05 | ** coefficient significant at 0.01 | ( ): standard error | | |

### C. Discussions

Currently, there are many recommendations in the security metrics literature for what measures organizations should gather. However, as noted by Black [6], little work has been done to determine the value of these measures in real-world environments, including which measures are most supportive of particular metrics.

Our findings show that, in our case, IPS metrics might be linked to the number of security incidents. Based on the investigated hypothesis the following results were achieved: i) the number of attackers does not increase the number of security incidents, ii) the number of attack signatures does not increase the number of security incidents, iii) the severity of attack signatures increases the number of security incidents and iv) the number of corrupted computers increases the number of security incidents.

The number of attackers, number of attack signatures and severity of external attackers does not significantly affect the number of security incidents. Since there are some characteristics of attackers that might influence the rate at which a system is attacked, finding metrics that characterize attackers is a possible avenue to research, in order to understand the type of attackers that affects the number of security incidents.

A practical implication of the results could be the development of automated tools to analyze the relationship between IPS data and security measurements. Such tools would be used by the security team, as a first step towards understanding the organization's overall security posture. Another implication is the study of IPS metrics to build security

 ISBN: 978-1-61208-209-7 

incident prediction models. For example, evaluating whether the number of attackers, signatures and corrupted computers are predictive of security incidents. If so, security analysts can use this prediction to prioritize security inspection and to implement preventive measures.

Our study indicates that more than one significant IPS metric might be derived from the dataset. This finding suggests the use of different metrics instead of finding a single security metric. In other words, as noted by [24], since security is a set of attributes, we should also use a set of metrics for measuring it. In combination, metrics could begin to paint a picture of system security.

The results also provide indications that metrics behavior change over time. Given the dynamic nature of information security, it is not certain that events of the past will provide a trustworthy prediction to the future, since attackers actively work to change the threat environment. Therefore, security metrics must be able to reflect significant changes in the underlying assumptions about how the system changes over time.

Our findings might be restricted to networks like those of universities: with nodes that are not fully controlled by the IT department. Private organizations, for instance, have different concerns about information security. The way that the security perimeter of a university is secured is completely different from a private company. Therefore, a similar study, when conducted in such organizations, could show different results.

Finally, we studied the relationship between IPS data and security incidents, but, similar research could be conducted between another security data, for instance, security incidents and firewall logs, intrusion detection systems and network flow data and so on.

## V. Threats to validity

The incidents used in our study were reported based on three sources of events: i) an IDS, ii) reports from users and iii) reports from other system administrators. Since recorded incidents led to the blocking of the suspected computer's IP address, the University Office of Information Technology (OIT) verified the authenticity of each incident. As a result, all incidents obtained from these sources were manually reviewed. OIT launched port scans and packet captures to validate the suspicious behavior of identified hosts. Because of the method used by OIT to validate the incidents, we can assume that all incidents used in our work are real. Thus, there are no false positives among the incidents reported. However, we cannot quantify the number of undetected attacks and intrusions that did not lead to a security incident.

The main issue with IPS event data is that the collected data are not perfect [10]. In other words, collected data might contain false positives and might not detect some malicious activity (false negatives). Moreover, since the IPS is a signature-based device, new attacks will not be detected nor blocked.

We have not evaluated the IPS and thus do not know how many false positives and false negatives the IPS produces. Besides, we cannot prove that a blocked attack would have been damaging to the targeted computer. In particular, for an attack to be successful, the targeted computer should have the associated vulnerability. We have scanned several computers for which an IPS alert was raised and noticed that in many cases the vulnerability associated with the alert was not present. This means that even without the IPS, the attack would not have been successful. This also indicates that the IPS identifies and detects an attack in its early stage preferring to block attacks that would not have been successful instead of not blocking a potentially successful attack.

As with all empirical studies, our results are limited to the datasets we investigated. In order to generalize our observations from this study to other environments, further studies should be performed.

## VI. Conclusion and Future Work

In this paper, we investigated some factors that might be linked to the rate at which security is successfully breached and empirically examine how attackers, attack signatures and corrupted computers affect the number of security incidents of an organization. We use two datasets, security incidents and IPS alerts, provided by the University of Maryland to derive our measurements.

Our results reveal that from the set of 11 investigated variables, 3 of them are positively related to the number of security incidents: the number of external targets, number of internal attackers, and the number of critical signatures of attacks launched from computers inside the organization. Since the number of external targets and internal attackers are related to the number of potential corrupted computers in an organization, the following hypotheses were supported:

- The number of attackers does not increase the number of security incidents;
- The number of attack signatures does not increase the number of security incidents;
- The severity of attack signatures increases the number of security incidents;
- The number of corrupted computers increases the number of security incidents.

We also found empirical evidence that relevant metrics changes over time. These findings are consistent with the idea that the security of the overall system cannot be ensured using only a single metric. Since security is a set of attributes, measuring security implies the usage of a set of metrics.

Future research should be conducted to compare the results presented in this work. It would be useful to repeat

 ISBN: 978-1-61208-209-7 

the analysis for some other datasets and investigate the differences between them. For instance, since attack signatures could be associated with certain security vulnerabilities, it would be interesting to investigate whether severity of security vulnerabilities follows the pattern found in our study.

Additional research may be also conducted to evaluate the impact of security factors over other variables, such as network topology and certain security incidents categories. Understanding which factors are likely to affect the security of a network can help network security analysts extract relevant information about the organization security.

## ACKNOWLEDGMENTS

The authors would like to thank the State of São Paulo Research Foundation (FAPESP) and the Coordination for the Improvement of Higher Education Staff (CAPES) that supports this work. We also thank Gerry Sneeringer and the Division of Information Technology at the University of Maryland for allowing and supporting the described research.

## REFERENCES

[1] W. Jansen, "Directions in security metrics research," National Institute of Standards and Technology (NIST), Tech. Rep., 2009.

[2] The Center for Internet Security, "The cis security metrics v1.1.0," http://www.cisecurity.org/, November 2010, retrieved: 07, 2012.

[3] A. Jaquith, *Security metrics: replacing fear, uncertainty, and doubt*. Addison-Wesley Professional, 2007.

[4] M. Swanson, N. Bartol, J. Sabato, J. Hash, and L. Graffo, "Performance measurement guide for information security," NIST Special Publication 800-55, Tech. Rep., 2003.

[5] V. Verendel, "Quantified security is a weak hypothesis: a critical survey of results and assumptions," in *NSPW '09: Proceedings of the 2009 workshop on New security paradigms workshop*. New York, NY, USA: ACM, 2009, pp. 37–50.

[6] P. Black, K. Scarfone, and M. Souppaya, "Cyber security metrics and measures," 2008.

[7] "Division of information technology," http://www.oit.umd.edu/, retrieved: 07, 2012.

[8] E. Condon, A. He, and M. Cukier, "Analysis of computer security incident data using time series models," in *19th International Symposium on Software Reliability Engineering (ISSRE)*, 2008, pp. 77–86.

[9] E. Condon, M. Cukier, and T. He, "Applying software reliability models on security incidents," in *18th IEEE International Symposium on Software Reliability (ISSRE)*, 2007, pp. 159–168.

[10] D. Chrun, M. Cukier, and G. Sneeringer, "Finding corrupted computers using imperfect intrusion prevention system event data," in *Proceedings of the 27th international conference on Computer Safety, Reliability, and Security*, ser. SAFECOMP '08. Springer-Verlag, 2008, pp. 221–234.

[11] ——, "On the use of security metrics based on intrusion prevention system event data: An empirical analysis," in *HASE '08: Proceedings of the 2008 11th IEEE High Assurance Systems Engineering Symposium*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 49–58.

[12] "Iso/iec 27004," http://www.iso27001security.com/, retrieved: 07, 2012.

[13] M. Cukier and S. Panjwani, "Prioritizing vulnerability remediation by determining attacker-targeted vulnerabilities," *Security & Privacy, IEEE*, vol. 7, no. 1, pp. 42–48, 2009.

[14] E. Mustaine and R. Tewksbury, "Predicting risks of larceny theft victimization: A routine activity analysis using refined lifestyle measures," *Criminology*, vol. 36, no. 4, pp. 829–858, 1998.

[15] S. Schechter, "Toward econometric models of the security risk from remote attacks," *Security & Privacy, IEEE*, vol. 3, no. 1, pp. 40–44, 2005.

[16] D. Chrun, M. Cukier, A. Mosleh, and G. Sneeringer, "Investigating the impact of humans in information technology security: A case study at the university of maryland," in *10th International Probabilistic Safety Assessment and Management Conference (PSAM)*, 2010.

[17] E. Schultz, "A framework for understanding and predicting insider attacks," *Computers & Security*, vol. 21, no. 6, pp. 526–531, 2002.

[18] Verizon, "2012 data breach investigations report," http://www.verizonbusiness.com/us/about/events/2012dbir/, 2012, retrieved: 07, 2012.

[19] PriceWaterHouseCoopers, "Global state of information security survey 2012," http://www.pwc.co.uk/audit-assurance/publications/uk-information-security-breaches-survey-results-2012.jhtml, 2012, retrieved: 07, 2012.

[20] R. Richardson, "Csi computer crime and security survey," *Computer Security Institute*, pp. 1–44, 2010.

[21] R. Walpole, R. Myers, S. Myers, and K. Ye, *Probability and statistics for engineers and scientists*. Macmillan New York, 1989.

[22] A. Arora, R. Krishnan, A. Nandkumar, R. Telang, and Y. Yang, "Impact of vulnerability disclosure and patch availability-an empirical analysis," in *Third Workshop on the Economics of Information Security*. Citeseer, 2004.

[23] J. Wang, N. Xiao, and H. Rao, "Drivers of information security search behavior: An investigation of network attacks and vulnerability disclosures," *ACM Transactions on Management Information Systems (TMIS)*, vol. 1, no. 1, p. 3, 2010.

[24] S. Pfleeger, "Useful cybersecurity metrics," *IT Professional*, pp. 38–45, 2009.

 ISBN: 978-1-61208-209-7 