Combining Intrusion Detection and Recovery for Enhancing System Dependability

Ajay Nagarajan[1], Quyen Nguyen[1], Robert Banks[1] and Arun Sood[1,2]
International Cyber Center and Department of Computer Science
[1]*George Mason University, Fairfax, VA 22030*
[2]*SCIT Labs, Inc, Clifton, VA 20124*
{*anagara1, qnguyeng, banksr3, asood*}*@gmu.edu*

Abstract – Current cyber defenses are reactive and cannot protect against customized malware and other zero day attacks which persist for many weeks. Using Receiver Operating Characteristic curve analysis and damage cost models, we trade-off the true positive rate and false positive rate to compare alternative architectures. This analysis provides optimal value(s) of Probability of Detection by evaluating the potential damage from a missed intrusion and costs of processing false positives. In this paper, we propose an approach which involves determining the influencing factors of each strategy and studying the impact of their variations within the context of an integrated intrusion defense strategy. Our goal is to manage the intrusion risks by proactively scheduling recovery for dependable networks.

Keywords- Intrusion Tolerance System, Receiver Operating Characteristic

I. INTRODUCTION

The variety and complexity of cyber attacks are increasing, along with the number of successful intrusions to mission and business systems. Recent breach reports like Wyndham Hotels [1] reported system compromise detection in February 2010, whereas the malware had resided in the system since October 2009. So we infer that not only the Intrusion Detection System / Intrusion Prevention System (IDS/IPS) failed to prevent the intrusion, but current systems were not able to detect the presence of the intruder long after the compromise.

Motivated by the above observations, our research focus has been on a method which consists of two important approaches to enhance cyber defense. First, recognizing that intrusion detection is a hard problem, can we shift focus to minimizing losses resulting from intrusions? If this strategy is successful, we anticipate that the reduced demands on the IDS will in turn lead to fewer false positives. Second, our model uses real world data from recent breach reports and their average costs to evaluate the cost reductions that can be achieved by using a combination of intrusion detection and tolerance architectures. Previously, the classical approach to assess architectures has been based on Single Loss Expectancy and Annual Loss Expectancy. More recently decision trees have been used [14]. In the former, many assumptions are required, and in the latter a lot of data have to be collected. These approaches are good for analyzing systems for which past data can be used. But is this useful for architectural decisions for the future? We are proposing the use of ROC (Receiver Operating Characteristic) curve based analysis, which is a powerful tool system administrator can use with enterprise specific data to build economic models and to compare alternate architectures. DARPA funded Lincoln Lab IDS evaluation [2] was a pioneering paper that evaluated many IDS by generating normal traffic similar to that seen on Air force bases. They used ROC curves to present their results. McHugh [3] published a critique of Lincoln Lab's work in 2000 which primarily considered issues associated with Lincoln's experimental dataset. McHugh pointed out the following problems in Lincoln's application of ROC analysis to IDS evaluation, which are a lack of "appropriate units of analysis, bias towards possibly unrealistic detection approaches and questionable presentation of false alarm data" [3]. In Section IV, we treat these issues.

In this paper, we compare an IDS only solution with IDS and SCIT (Self Cleansing Intrusion Tolerance) combination, SCIT being our approach to intrusion tolerance which is classified in the recovery-based category [4]. From this assessment, optimal value(s) of Probability of Detection and other operational parameters can be selected to balance the potential damage from a missed intrusion and the cost of false positive processing. In our approach, we stipulate that providing an upper bound on the time between the compromise and recovery has many advantages since it does not require the assumption that the system will be able to detect either the intrusion attempt or the compromise.

The rest of the paper is organized as follows. In Section II, we develop the motivation for dependability recovery requirements. Section III briefly reviews the intrusion tolerance approach. Sections IV, explains ROC Analysis usefulness to assess IDS architectures. . Sections V, applies a cost model to evaluate how three different cases behave for a set of hypothetical ROC curves. Section VI is the conclusion.

II. MOTIVATION

As cyber defense efforts increase, passive efforts such as establishing anti-virus software, firewall protection, or improving password strength and encryption, and the organization's workload are constantly challenged by the need to apply patches immediately. Security researchers are uncovering close to 55,000 new malware samples a day, overwhelming malware analysis resources [5]. Increasingly,

automated analysis technologies are used to keep up with the volume, but they still lack the precision to decipher compressed, encrypted, and obfuscated malware [6]. McAfee recent crash of tens of thousands of PCs globally illustrates the unpredictable system effects after compromise and their collateral damage, which creates even more uncertainty and less dependability for Enterprise Security [7].

The current reactive cyber defense approaches are expensive and inadequate. We expect that, automated recovery and Intrusion Tolerance System (ITS) will be useful in addressing the increasing malware and patch workload, but what are the cost impacts of malicious threats and false positives on dependability and security attributes?

III. INTRUSION TOLERANCE APPROACH

ITS architecture objective is to tolerate unwanted intrusions and restore the system to its normal state. Various ITS approaches are reviewed by Nguyen and Sood [4]. In our paper, we use the recovery-based SCIT (Self-Cleansing Intrusion Tolerance) model [4], which is applicable to servers that are open to the Internet, such as Web, and DNS servers [8]. Using round-robin cleansing, at any point in time, a server in a SCIT cluster can have one of the three states: offline cleansing, offline spare and online transaction processing. The duration that a SCIT server is exposed to the Internet is called its Exposure Time. The architecture is simple, and does not rely on intrusion detection. Implementation of SCIT scheme can be based on virtualization. The interfaces between controller and the group of servers to be protected are trusted.

Another benefit of a recovery-based ITS is to shrink down breach duration, which has the effect of reducing losses and their costs. Indeed, this intrusion tolerance strategy would mitigate the effects of malicious attacks. Intrusion detection is known to be a hard problem, and current cyber defense systems reportedly detect less than half the malware. Still servers and apps account for 98% of the total record compromised. Verizon DBIR 2010 [9] underscores this problem by noting that only 11% of the compromises were detected within minutes or hours. Thus, current cyber defenses cannot protect systems against customized malware and other zero day attacks; once an attack is successful, it can persist for many weeks. This emphasizes the need for a recovery-based Intrusion Tolerance approach since detection triggered ITS might again fall short of the needs.

IV. RECEIVER OPERATING CHARACTERISTIC (ROC)

ROC analysis has been long used in signal detection theory to present the tradeoff between hit-rates and false-positive rates of classifiers. ROC analysis was initially used during World War II in the analysis of radar signals to differentiate signal from noise. It was soon introduced in Psychology to map the perceptual detection of signals [10]. ROC curves are useful for assessing the accuracy of predictions. A ROC curve plots the fraction of true positives (hits) versus the fraction of false positives, and hence has a direct relationship with diagnostic decision making. The ideal prediction method would yield a co-ordinate (0, 1) on the ROC curve. This represents 100 % true positives and zero percent false-positives, and is referred to as the perfect classification.

A. *Using ROC to assess IDS quality.*

The most attractive feature of ROC analysis is the fact that the tradeoff between probability of detection and probability of false positive can be derived directly. This allows a system administrator to instantly determine how well a classifier performs and also to compare two classifiers. We care about false positives in addition to the probability of detection since there is a need to characterize human workload involved in analyzing false positives generated by traffic. According to [2], false positive rates above 100's per day could make IDS almost useless even with high probability of detection since security analysts must spend hours each day investigating false positives.

DARPA funded Lincoln Lab IDS evaluation [2] appears to be the first to perform tests to evaluate many IDS by generating normal traffic similar to that on a government site. McHugh [3] reviews and analyzes the validity and adequacy of artificial data used to estimate real world system performance. In this paper, we present a methodology to compare various IDS's, each of which is represented by a ROC curve. We utilize Verizon's 2010 results representing a cross section of multiple industries. Furthermore, these data validate firsthand real world evidence over a broad five year range from 2004-2009 with the addition of US Secret Service confirmed cases.

The Lincoln Lab experiment used ROC for presenting the results of the evaluation. McHugh [3] criticized Lincoln Lab's use of ROC curves primarily on the following grounds. We have attempted to address each of these concerns in our work:

- *Determining appropriate units of analysis.* Unit of analysis is the quantity of input on which a decision is made. Lincoln lab used sessions as the unit of analysis, the problems of which were outlined in [3]. McHugh also emphasized the need for using similar units of analysis across all IDS's to be evaluated. In our case, we consider a simple system and consistently use query / packet as our unit of analysis across all IDS's.
- *Errors per unit time.* In [2], a pseudo-ROC curve with x-axis as False Positives per day instead of Percentage False Positives was used. This led to two incomparable units being used on two axes, and the results in turn became strongly influenced by factors like the data rate that should typically be irrelevant. In this paper, we consistently use probability of detection and that of false positives for all ROC curves. In such a case, given that the distributions of signal and noise are realistic,

McHugh [3] recognizes that the ROC presentation should give a good account of detector performance in similar environments. Given enough characterizations of the signal and noise distributions, McHugh further acknowledges that it is even possible to investigate optimal detectors.

- McHugh [3] criticizes Lincoln Lab's methods of scoring and constructing ROC curves which lead to problems like bias towards unrealistic detection approaches, but not the use of ROC curves itself. In our case, the emphasis is not on constructing ROC curves but on comparing IDS's using our cost-model once we have their respective ROC curves. While there is a need for alternative taxonomies, the scoring method from the attacker's perspective is still utilized for real world incidents.

According to [2], there have been a number of similar efforts. In order to be able to compare multiple IDS systems, the ROC curves should be generated using similar or preferably same test data. According to Orfila et al. [11], if two ROC curves intersect at some point, there is no way of claiming that one is better than the other since some system administrators might want high probability of detection (top right corner of ROC curve) and some might want low probability of false positive (bottom left corner of ROC curve).

Stolfo et al. [12] presents an alternative method to perform evaluation based on cost metrics. Authors help formalize the costs involved in evaluating an IDS into three types: 1) Damage cost, 2) Challenge cost or Response cost and 3) Operational cost.

In [13], Drummond et al. propose the use of cost curves for evaluating classifiers. Cost curves plot expected cost vs. Probability Cost Function (PCF). Here PCF is a function of probability of detection, probability of false positive and its corresponding costs. Although cost curves are good to compare classifiers, the representation does not provide for the system administrator to quickly see the cost trend of operating at different points (P_f, P_d) on the ROC curve. Also [13] does not suggest a way to determine the expected cost of operating at a point on ROC curve.

In [14], Gaffney et al. argued that both ROC analysis and cost analysis methods are incomplete. They used decision analysis techniques and provide an expected cost metric that reflects IDS's ROC curve based on a decision tree approach. This cost model requires a lot of data to be collected and does not reflect the magnitude of actual costs associated with breach events. For this, we propose a cost-model for the calculation of expected cost of operating at any point on the ROC curve.

V. COST MODEL

In this section, we look to overcome each of the shortcomings of earlier approaches by proposing a cost model that consists of two elements:

- A formula for the expected cost of operating at any point on the ROC curve
- Cost metrics derived from published breach investigation reports

A. *Expected Cost calculation.*

The cost of operating IDS at any point on the ROC curve (P_f, P_d) is a combination of the following:

- Operational Costs – Cost involved in operating the IDS and keeping it running.
- Damage Costs – the amount of damage caused by an intruder in case of a successful attack.
- Response Costs – the cost involved in responding to a potential intrusion on detection.

Out of the three costs mentioned above, operational costs and response costs greatly vary from organization to organization based on a number of factors like size of the organization, type of organization etc. Since these two costs are not entirely quantifiable, for the purposes of this paper, we employ the objective function proposed in [15]:

Expected Cost of operating at any point on the ROC curve = Cost of Misses + Cost of False Positives.

Thus, for every point on the ROC curve (P_f, P_d), we have an expected cost:

$$\text{Expected Cost} = (C_m * p * P_m) + (C_f * (1-p) * P_f),$$

where

C_m – Cost of a miss p – Prior probability of Intrusion

C_f – Cost of a false positive P_d – Probability of detection

P_m – Probability of a miss = $(1-P_d)$

P_f – Probability of a false positive

Note that this expected cost is for one incoming query. If there are 'n' incoming queries, the above expected cost must be multiplied by 'n'. The value of metrics used in the cost model is summarized in Table 1.

Table 1- Metrics values used in the Cost Model

Metrics	Value	Explanation	Ref
Median number of records lost per breach (M)	1,082	Removes outliers. Better estimate of the "typical value"	[9]
Average cost of compromised record (D)	$ 204	Direct Cost: $ 60 + Indirect Cost: $144	[16]
Cost of a Miss (C_m)	$220,000	M * D = 1082 * $ 204	[9], [16]
Cost of a False Positive (C_f)	$ 400	Assumption: Labor Cost + Overhead Cost = $ 400	
Median Compromise Duration per breach	14 days	Compromise to Discovery time + Discovery to Containment time	[9]

In this paper, the probability of detection P_d and that of a false positive P_f will constitute the operational parameters.

We use the median number of records lost for assessing damage. In many cases, the outliers in breach data can skew the data, because most of the losses come from only a few

breaches. Therefore, the Mean becomes highly skewed and is not a good estimate of the typical number of records lost per breach. Median is a better estimate of the typical value [16].

B. *Evaluating classifiers using our Cost Model.*

For the purposes of this paper, we do not address how the ROC curves are constructed. Proper construction and use of ROC curves in Intrusion / Anomaly detection have been addressed in [17]. We just show how the cost model can be implemented once they are constructed. Figure 1 gives a family of hypothetical ROC curves, each representing a classifier. We will implement our cost model on these ROC curves in three different cases to evaluate the classifiers' behaviors:

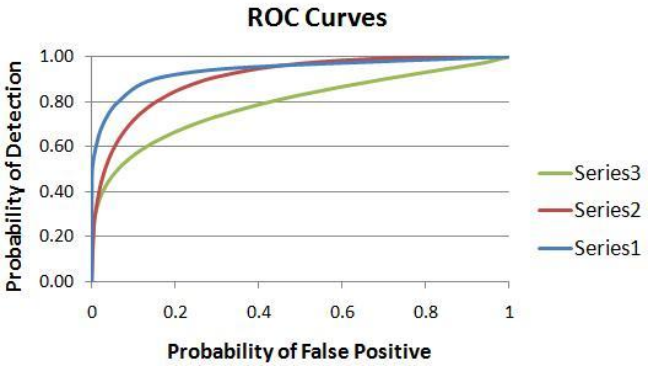

Figure 1 - Receiver Operating Curves

Table 2 provides the values of the parameters used in the cost model in each of the three cases. Within each case, the value of 'p' remains the same for both IDS and SCIT+IDS. Therefore, the number of intrusions that occur in each of these architectures are the same since Number of intrusions = [Number of incoming queries * Prior probability of intrusion (p)]. The baseline IDS and SCIT+IDS scenarios are provided for Case 1. Case 2 and Case 3 help investigate the impact of 'C_m' and 'p' on system cost and security. Figures 2 through 7 illustrate this. It is noted that the y-axis scale is different in Figure 6.

CASE 1a. *IDS:* (Figure 2)

This is a stand-alone IDS system. The cost keeps decreasing as Probability of Detection (P_d) is increasing. As P_d increases, number of misses decrease along with the significant associated costs. However, after a threshold, if we keep increasing the value of P_d, the expected cost stops decreasing and starts increasing rapidly. At this point, the cost of False Positives exceeds the cost of misses and so the gains from containing misses start diminishing. This point is known as the "minimal cost point on the ROC curve (MCP)". For e.g., in Case 1a, the MCP for Series 1 is 70 and it occurs at (P_f, P_d) = (0.20, 0.85). MCP for each series of every case we evaluated is tabulated in Table 3.

CASE 1b. *SCIT + IDS:* (Figure 3)

Now we add SCIT to existing IDS and evaluate the system using our Cost Model. We assume that the exposure time of SCIT is 4 hours[1]. This reduces the compromise duration of the system from 14 days to 4 hours. We assume that data is ex-filtrated uniformly over time. Since the cost of a miss was $220,000 earlier with compromise duration of 14 days, now it significantly reduces to $2,620 for compromise duration of 4 hours.

CASE 2. (Figures 4 & 5)

Assumption: As compared to the baseline (Case 1), IDS cost of a miss is reduced from $220,000 to $60,000.

CASE 3. (Figures 6 & 7)

Prior Probability of Intrusion is increased fivefold from p = 0.001 to p = 0.005.

Table 2 – Parameter values used in the cost model

	P	C_m	C_f	Compromise Duration
Case 1a: IDS	0.001	$220,000	$400	14 days
Case 1b: IDS+SCIT	0.001	$2,620	$400	4 hours
Case 2a: IDS	0.001	$60,000	$400	14 days
Case 2b: IDS+SCIT	0.001	$715	$400	4 hours
Case 3a: IDS	0.005	$220,000	$400	14 days
Case 3b: IDS+SCIT	0.005	$2620	$400	4 hours

C. *Results: Comparison of IDS's.*

Figure 8 compares the MCP's of 3 IDS' whose performances are indicated by the ROC curves in Figure 1.

- Series 1 IDS clearly outperforms all the other IDS' in all three cases.
- It is most expensive to operate the IDS' in case 3 since prior probability of intrusion is high which in turn leads to more misses.

D. *Results: Comparison of SCIT + IDS's*

Figure 8 also presents the minimal cost points for IDS + SCIT. We have used an exposure time of 4 hours. We note that as compared to the IDS only case, the costs are much lower. The minimal cost points are achieved using a much lower value of Probability of Detection which in turn leads to a lower Probability of False Positive. We conclude that this makes the IDS design much easier and the system easier to operate. The reliability of the IDS results also increase.

[1] The SCIT servers tested in our lab and independently tested at Lockheed Martin and Northrop Grumman have Exposure Times of 1 or 2 minutes. Here, we use larger values of Exposure Time to emphasize the advantage of the concept.

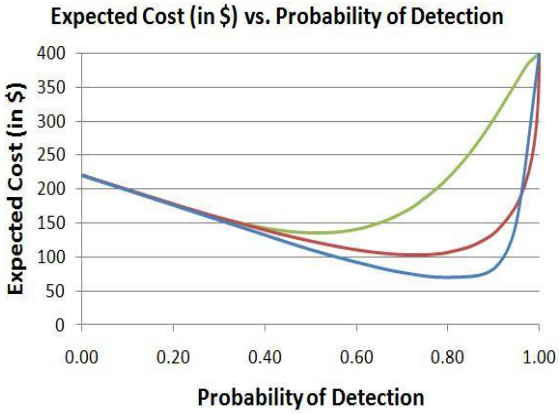

Figure 2 – IDS Case 1a

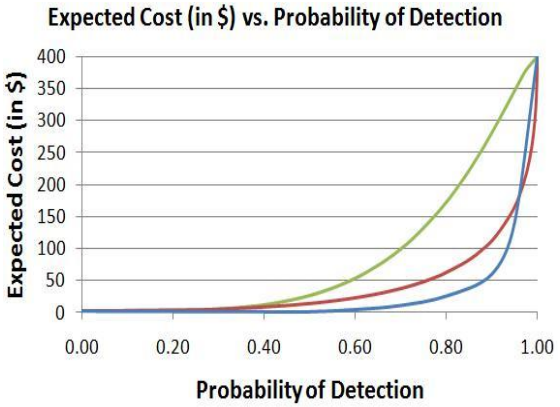

Figure 3 - SCIT + IDS Case 1b

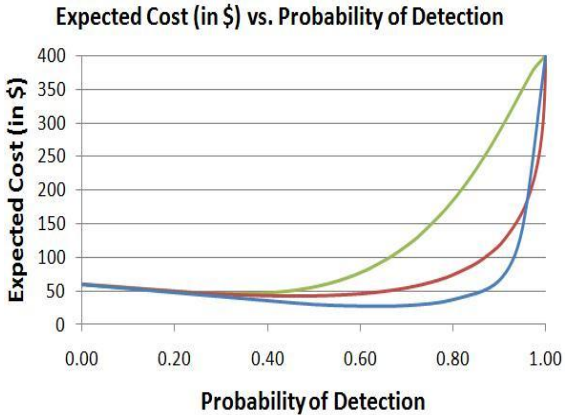

Figure 4 - IDS Case 2a

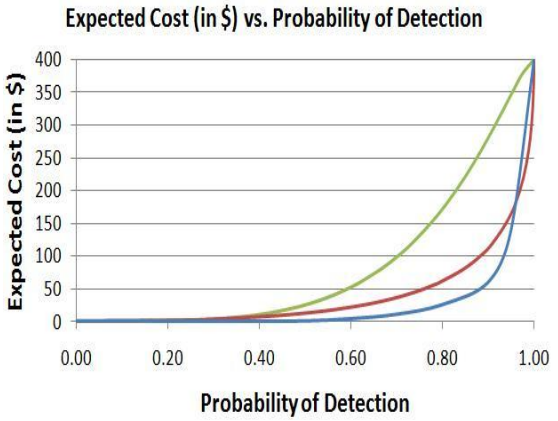

Figure 5- SCIT + IDS Case 2b

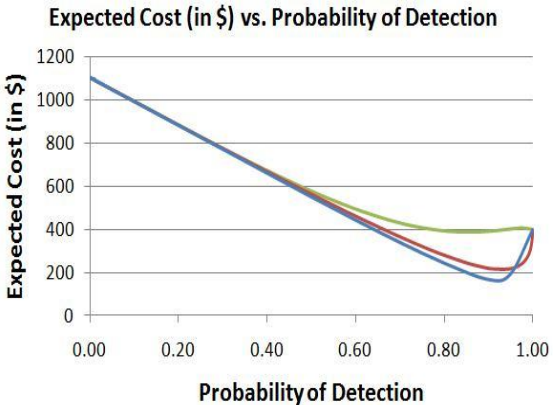

Figure 6 - IDS Case 3a

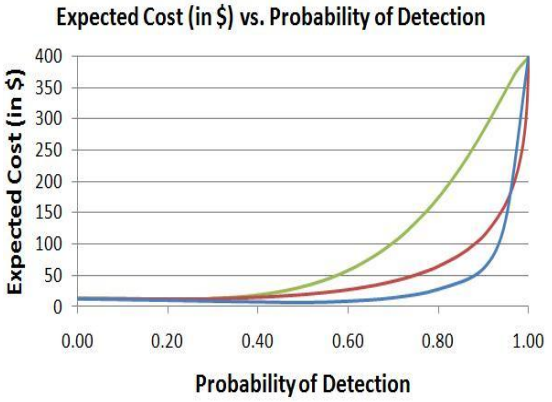

Figure 7 – SCIT + IDS Case 3b

From the results, we can see that the benefits of adding SCIT are as follows:

- Cost of a miss is greatly reduced. As the compromise duration / exposure time of SCIT is reduced, cost of a miss further reduces.
- We can tolerate a larger number of misses now that the cost of a miss is reduced.

E. General Observations (IDS and SCIT + IDS)

- As the cost of miss decreases, we can tolerate more misses and so probability of detection for achieving minimal cost point can now take lower values.
- As C_m decreases, C_f has a greater influence on the expected cost and so there is an increased need to contain false positives. Note that the Probability of

False Positives for achieving minimal cost point now decreases.

As prior probability of intrusion 'p' increases:

- The total number of misses' increases and so does the expected cost.
- To combat this, probability of Detection for achieving minimal cost point increases thus reducing the number of misses. (Note: Number of misses = Number of incoming queries * p * P_m).

Table 3: Minimal Cost Point values

CASES	Minimal Cost Point for Figure 1 ROC Curves - Cost ($)					
	SERIES 1		SERIES 2		SERIES 3	
	IDS Only	IDS + SCIT (ET=4hrs)	IDS only	IDS + SCIT (ET=4hrs)	IDS Only	IDS + SCIT (ET=4hrs)
CASE 1	70	2	102	3	135	3
CASE 2	28	0.5	43	1	45	1
CASE 3	170	7	218	12	386	12

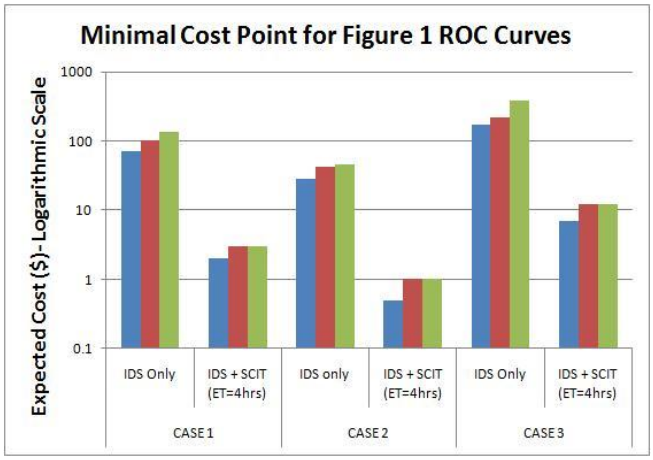

Figure 8 – Minimal Cost Point Comparison

VI. CONCLUSION

Intrusion detection is a hard problem, making intrusions inevitable. Consequently, containing losses by an upper bound on the time between compromise and recovery shows many advantages. ROC analysis, supplemented with cost analysis using median of lost records and average cost of compromised records per breach, reveals tradeoff between high probability of detection, and low probability of false positive. Our approach reduces the cost of a miss; and tolerating a larger number of misses' leads to lower false positive costs.

The SCIT architecture provides a robust security mechanism that guarantees certain security properties by limiting the exposure time. In addition, SCIT does not generate false positives and thus reduces the intrusion alerts management costs. Thus SCIT also provides administrative and economic benefits which make it a reasonable choice to be included in security architecture. In particular, this is expected to be of interest in environments where technical skills are limited. The analysis presented suggests that a combination of IDS with SCIT on host servers provides a robust architectural solution in the face of new attacks.

REFERENCES

[1] Hotchkiss, Kirsten. http://www.wyndhamworldwide.com/customer_care/data-claim.cfm. Jun. 2010.

[2] R. Lippmann, et al "Evaluating Intrusion Detection Systems: The 1998 DARPA Off-line Intrusion Detection Evaluation" Proceedings of DISCEX 2000, Los Alamitos, CA. 2000.

[3] McHugh, John (2000) "Testing intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory" TISSEC, Vol 3, Issue 4

[4] Nguyen, Quyen and Sood, Arun. "Comparative Analysis of Intrusion-Tolerant System Architectures". IEEE Security and Privacy – Volume: PP , Issue: 99 , 2010.

[5] McAfee Labs. "McAfee Threats Report: Second Quarter 2010".http://www.mcafee.com/us/local_content/reports/q22010_threats_report_en.pdf. pg 11.

[6] Bejtlich, Richard. "The Tao of network security monitoring: beyond intrusion detection", Pearson Education, Inc. 2005.

[7] Kravets, David. "McAfee Probing Bungle That Sparked Global PC Crash".Threat Level. http://www.wired.com/threatlevel/2010/04/mcafeebungle/. 2010.

[8] Anantha K. Bangalore and Arun K Sood. "Securing Web Servers Using Self Cleansing Intrusion Tolerance (SCIT)", *DEPEND 2009*, Athens, Greece. 2009.

[9] Verizon Business Data Breach Investigations Report 2010.

[10] Swets. John A. "Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers".

[11] Orfila, Augustin. Carbo, Javier. and Ribagardo, Arturo. "Advances in Data Mining, volume 4065, chapter Effectiveness Evaluation of Data Mining based IDS, pages 377-388. Springer Berlin Heidelberg. 2006.

[12] Stolfo, S. Fan,W. Lee, W. Prodromidis, A. and Chan, P. "Cost-based modeling for Fraud and Intrusion Detection: Results from the JAM Project" Proceedings of DISCEX 2000, Los Alamitos, CA. 2000.

[13] Drummond, Chris. Holte, Robert C. "What ROC Curves Can't do and Cost curves can". 2004.

[14] Gaffney, John E. Jr. Ulvila, Jacob W. (2001). "Evaluation of Intrusion Detectors: A Decision Theory Approach" Security and Privacy.

[15] J. Hancock and P. Wintz. Signal Detection Theory. McGraw-Hill. New York 1966

[16] Widup, Suzanne. (2010, Jul). "The Leaking Vault – Five years of data breaches" – Digital Forensics Association.

[17] R.A. Maxion and R.R. Roberts. "Proper use of ROC curves in Intrusion/ Anomaly Detection" Technical Report, University of Newcastle Nov 2004

[18] 2009 Annual Study: Cost of a Data Breach, Ponemon Institute LLC.