December 2003

Configuration of Intrusion Detection Systems: A Comparison of Decision and Game Theoretic Approaches

Huseyin Cavusoglu
Tulane University

Srinivasan Raghunathan
University of Texas at Dallas

CONFIGURATION OF INTRUSION DETECTION SYSTEMS: A COMPARISON OF DECISION AND GAME THEORETIC APPROACHES

Huseyin Cavusoglu
Tulane University
New Orleans, LA USA
huseyin@tulane.edu

Srinivasan Raghunathan
The University of Texas at Dallas
Richardson, TX USA
sraghu@utdallas.edu

Abstract

Intrusion detection systems (IDSs) have become a core component of a firm's IT security architecture. While IDSs enable real time detection of intrusions, a common criticism has been the frequency of false alarms, which undermines their effectiveness. A fundamental problem with IDSs for intrusion detection is achieving the optimal balance between detection rate and false positive and false negative rates. Many firms use decision theoretic approaches to deal with the IDS configuration problem. While decision theoretic approaches are appropriate for configuring many types of machine learning and classification software that suffer from false positive and false negative errors, we argue that decision theoretic approaches have fundamental limitations for configuring IDSs. Decision theoretic approaches are based on the presumption that configuration does not influence the behavior of hackers. Game theoretic approaches recognize the fact that hackers do modify their strategies in response to firms' actions. In this paper, we compare the decision and game theoretic approaches to the IDS configuration problem when firms are faced with strategic hackers. We find that under most circumstances firms incur lower costs when they use game theory as opposed to decision theory because decision theory approach frequently either over- or under-configures the IDS. However, firms incur the same or lower cost under decision theory approach compared to the game theory approach if configurations under decision theory and game theory are sufficiently close. A limitation of the game theory approach is that it requires user specific utility parameters, which are difficult to estimate. Decision theory, in contrast to game theory, requires the attack probability estimate, which is more easily obtained.

Introduction

Intrusion detection systems (IDSs) have become a core component of a firm's IT security architecture. IDSs embed pattern recognition intelligence to classify an event as normal or intrusive. Although IDSs enable real time detection of intrusions, some security experts criticize that false alarms undermine their effectiveness (Axelsson 2000). A fundamental problem with IDSs for intrusion detection is achieving the optimal balance between detection rate and false positive and false negative rates. False positives occur when the system classifies a normal event as an intrusion. False negatives occur when the system classifies an intrusion as a normal event. While it is desirable to have low false positive and low false negative rates, a reduction in one type of error is often accompanied by an increase in the other type. The goal of configuration of is to balance the two error rates in order to minimize the firm's cost. In this paper, we investigate the problem of optimal configuration of intrusion detection systems using decision and game theoretic approaches.

Decision theory is often used to analyze the risk associated with configuration problem. Recently Gaffney and Ulvila (2001) proposed a decision theory based approach to configure IDSs. Calibration of algorithms used in many classification programs is also based on decision theoretic approaches, which simply compare the costs associated with false positive and false negative errors. While the decision theory based approach can provide a useful starting point for managing risk in settings where potential for intrusion exists, we argue that this method is incomplete because of the problem's strategic nature. The reason for the limitation of the decision theory approach can be stated as one simple proposition: it does not allow a firm's decisions to influence the behavior of hackers. Researchers and practitioners have long recognized the behavioral influences of a firm's actions on hackers (Cavusoglu et al. 2002). For example, it has been pointed out that security should be viewed as a cat-and-mouse game played by firms and hackers (Jajodia and Miller 1993). Hackers do not randomly select their targets. They rationally make their choices based on how much effort will be required to succeed in hacking, the probability of getting caught, and penalty. Such strategic interactions between a firm's decisions and hackers have to be captured in the model used to determine the IDS configuration. Because decision theory is designed to analyze decision making under uncertainty where nature is the only opponent, it is fundamentally inadequate to address the intrusion detection problem where firms deal with strategic adversaries.[1] Modeling the interaction between firm and user decisions requires game theory.

Traditional decision theory assumes that the firm exogenously estimates the probability of intrusion in choosing the configuration. Although the firm can perform sensitivity analysis with respect to the estimated attack probability, the model still provides only partial solutions. In the game theoretical model, both the firm's configuration and the attack probability are endogenously determined. In this paper, we derive the optimal configuration under decision theory and game theory and compare the resulting costs. We find that game theory results in a lower cost except when the configuration under decision theory is sufficiently close to the configuration under game theory. The results of our model caution firms that ignoring the reactions of their strategic adversaries can cause significant harm to firms.

Related Work

The paper that comes closest to ours is Gaffney and Ulvila (2001). They analyze the IDS configuration problem from a decision theory perspective and determine the best operating point of the IDS for a given environment. Their study integrates the costs of dealing with false positive and false negative errors and the quality profile of the IDS as indicated by its receiving operating characteristics (ROC) curve, which relates rates of these two errors. Although Gaffney and Ulvila consider hostility of operating environment as one of the factors that determine the configuration, they assume a fixed attack level. Our study assumes that the realized attack level is determined by operating environment (firm and hacker specific factors) and configuration and is not necessarily the same as estimated attack level. In contrast to Gaffney and Ulvila, one of the goals of this paper is to develop a framework that explicitly models the strategic aspect related to user behavior in response to the IDS configuration and the firm's strategy in a cost-minimizing framework. Another goal is to compare the decision theoretic approach with the game theory approach for the IDS configuration.

The broader area of this paper relates to configuration management and performance evaluation of software. Guidelines from commercial classification software manufacturers as well as research institutes emphasize the need for proper configuration of detection systems. For example, CERT's guidelines (CERT 2001) on installing security software cautions firms against accepting the default settings automatically and advises appropriate configuration to balance security and operational requirements. Similar observations have been made for other detection software such as explosives detection systems used by airports (NMAB 1998). According to a federal report, between November 2001 and February 2002, security screeners missed 70 percent of knives, 60 percent of simulated explosive devices, and 30 percent of guns. Commenting about this, the former FAA security chief Billie Vincent noted, "The current metal detectors won't do the job. If you turn it high enough to detect that much metal, you will have an alarm on every person going through."[2] This effectively implies hand screening of all bags and passengers, which will balloon the cost of security at the airports.

In order to support and manage software configuration, a whole industry that develops configuration management tools has evolved.[3] The overriding goal of these efforts has been the performance of the software as measured by its classification accuracy. Evaluation of software such as intrusion detection systems, machine learning systems, and other classification programs has relied on false positive and false negative rates (Durst et al. 1999; McHugh 2000; Sarkar and Sriram 2001). Modeling the accuracy of classification software is a well-established area with many known models and measures such as lift, response ratio, L-Quality,

[1] Fellingham and Newman (1985) make the same observation in the auditing context.

[2] **www.cnn.com/2002/US/03/25/airport.security/?related**.

[3] Configuration management tools are much broader in scope than the configuration task we consider in this paper. They support related tasks such as version control support and configuration process management support.

and others (Shapiro and Masand 1999, Shapiro and Steingold 2000, Steingold et al. 2001). All these models use Bayesian decision theory. Our model also relies on Bayesian decision theory to generate the quality profile of the IDS. However, unlike previous models, our model for configuration does not use classification accuracy exclusively. Firms are interested in not only raw performance measures but also the overall cost of the detection process. Recently, researchers have recognized the importance of costs of misclassification in measuring IDS performance. Lee et al. (2002) developed a detection model that incorporates these costs in the classification algorithm itself in order to minimize costs. Our model for the IDS configuration also minimizes a firm's cost. The innovation in our analysis, absent in prior work, is the explicit modeling of the strategic aspect related to user behavior in response to IDS quality and a firm's response in a cost minimizing framework.

The need for incorporating user behavior in software configuration has been recognized in the IT security context in order to develop better security breach prevention and detection software. Johsson and Olovsson (1997) pointed out that the common criteria used to evaluate security software "reflect static design properties and do not incorporate the interaction with the environment in a probabilistic way." Using an experiment, they modeled attacker behavior and concluded that the attacking process can be split into learning, standard attack, and innovative attack phases. In a related work, Ortalo et al. (1999) proposed measures using a Markovian model that estimates the effort an attacker might expend to exploit system vulnerabilities. While these models capture the dynamic aspects of user behavior, we use a static model that captures attacker behavior through the attack probability.

The rest of this paper is organized as follows. The next section summarizes the statistical decision theory that underlies IDSs. We then introduce our model framework to address the IDS configuration problem. The optimal configurations under decision theory and game theory approaches are derived. The results under these two approaches are compared. Finally, we present our conclusions.

IDSs and ROC Curves

The principles underlying IDSs are grounded in classical decision theory. The basic components of a decision theory problem, as applied to the intrusion detection scenario, are shown in Figure 1. The first is a source that generates the inputs to the IDS. The source is the users interacting with the system that the IDS is designed to protect. In the simplest case, there are two types of users: normal (H_0) and hacker (H_1). The normal user generates legal or authorized transactions. The hacker generates illegal transactions or intrusions. The probabilistic transition mechanism controls the relative frequencies of legal transactions and intrusions. In a typical real life detection scenario, a large percentage of transactions are legal. The skewed nature of the frequency distribution makes detection of illegal transactions difficult. The IDS observes the transaction but does not know whether it is a legal transaction or an intrusion. The goal of the IDS is to classify each transaction as legal or illegal. Two types of errors can occur in this classification: classification of an illegal transaction as a legal transaction (false negative) and classification of a legal transaction as an illegal transaction (false positive).

We define

Probability of detection = P_D = Pr(classify into $H_1|H_1$ is true), or
Probability of false negatve = 1-P_D
Probability of false positive = P_F = Pr(classify into $H_1|H_0$ is true)

In general, we would like to have P_D as large as possible and P_F as small as possible in an IDS. However, it is not always possible to increase P_D and decrease P_F simultaneously. This is because of variability associated with the data of legal and illegal transactions and imprecision of algorithms and models used by IDSs. Many IDSs classify transactions based on whether a numerical score computed from transaction data exceeds a threshold value and/or whether the transaction data satisfy a rule. The quality parameters P_D and P_F of an IDS can be configured, although not independently, by setting its threshold value or relaxing or tightening the rules. Consequently, the quality profile of an IDS is characterized by a curve that relates its P_D and P_F, known as the receiver operating characteristics (ROC) curve (Van Trees 2001).

The ROC curve of an IDS can be derived experimentally or analytically (Durst et al. 1999, Lippman et al. 2000, McHugh 2000). The analytical procedure proceeds as follows. Consider an IDS that uses a numerical score x computed from transaction data and a threshold value t to detect intrusions. Let the IDS classify a transaction as illegal if $x > t$ for that transaction; otherwise as legal. It follows that

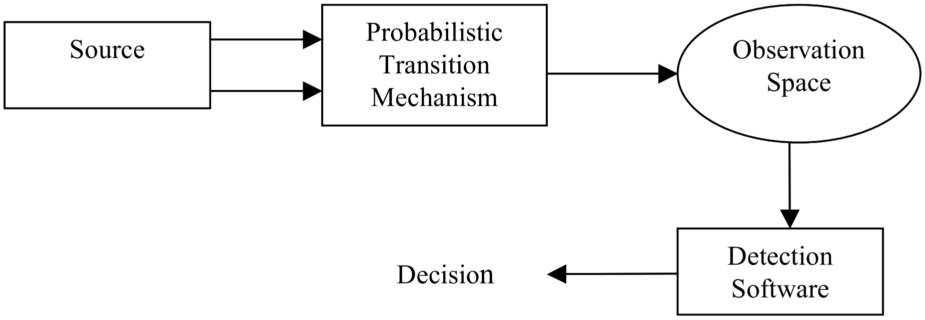

Figure 1. Components of a Detection Problem

$$p_D = \int_t^{\infty} f_F(x)dx \quad and \quad p_F = \int_t^{\infty} f_N(x)dx$$

where $f_N(x)$ and $f_F(x)$ are the probability density functions of x for legal and illegal transactions respectively. Figure 2 illustrates these probability calculations.

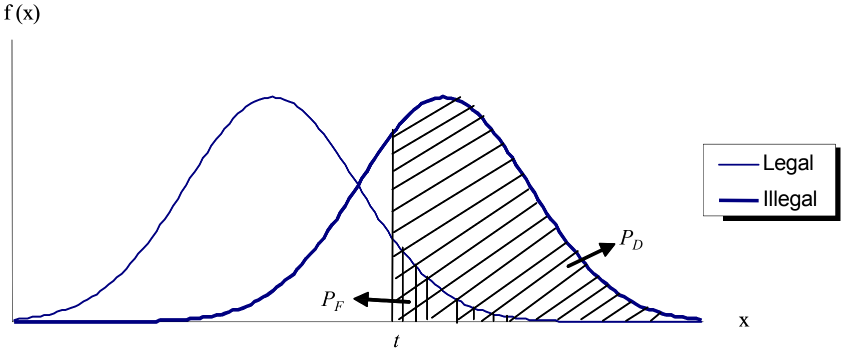

Figure 2. Computation of P_D and P_F

The shape of the ROC curve depends on the probability density functions of x_N and x_F. We assume that the numerical score used to distinguish normal from illegal transactions follows an exponential distribution. Exponential distributions, besides being analytically tractable, capture the skewed nature of transaction data very well.[4] Let the numerical scores for the normal and illegal transactions follow exponential distributions with parameters λ_N and λ_F, $\lambda_N > \lambda_F$, respectively. Then we can write P_D and P_F as

$$P_D = \int_t^{\infty} \lambda_F e^{-(\lambda_F x)} dx = e^{-\lambda_F t} \tag{1}$$

[4] We performed the analysis for the case when the numerical scores are normally distributed. Although closed form analytical expressions are unavailable, numerical solutions give qualitatively same insights.

$$P_F = \int_t^{\infty} \lambda_N e^{-(\lambda_N x)} dx = e^{-\lambda_N t} \quad (2)$$

Thus P_D can be expressed as a function of P_F as

$$P_D = P_F^{\,r} \quad (3)$$

where $r = \frac{\lambda_F}{\lambda_N}$ is between 0 and 1. Lower values of result r in steeper ROC curves.

Figure 3 shows sample ROC curves for various values of r.

Since IDSs are imperfect, they are only used as decision support tools in many commercial organizations. A security expert further investigates transactions classified as illegal by the IDS. Similarly, the expert may also randomly investigate transactions classified as legal by the IDS. The frequency of such random investigations depends on various costs including the cost of performing the investigation and the cost of misclassification.

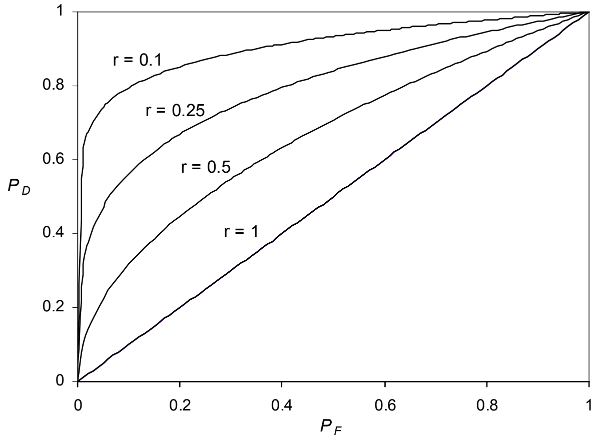

Figure 3. ROC Curves

The Model Framework for Configuration the IDS

We consider a firm that uses an IDS to detect intrusions. Our model of IDS is similar to that described in the previous section. A user can generate legal or illegal transactions. Users may intrude depending on factors such as the benefit they derive from break-in, the penalty they will receive if they are caught, and the likelihood that they will be caught. We assume that a user committing the intrusion derives a benefit of μ. If the intrusion is detected, the user incurs a penalty of β. The penalty can take different forms depending on the nature of intrusion. It can be the cost from legal prosecution or social humiliation. We denote the probability that a user commits an intrusion by ψ.

The IDS analyzes each of the transactions. If the program deems a transaction to be illegal, it generates a signal. The firm then decides whether to investigate or not investigate the transaction. The firm makes a decision about whether to investigate or not based on the state (the signal or the no signal) it is in. However, when the program generates a signal, the firm does not know with certainty whether there has been a true attack or whether the program generated a false alarm. Similar uncertainty exists also when the program does not generate a signal.

The firm supplements the IDS with manual investigation by a human expert. The human expert may investigate only a proportion (ρ_1) of transactions that generated signals. Furthermore, the expert may investigate a proportion (ρ_2) of transactions that did not generate signals. The firm incurs a cost of c each time the human expert performs a manual investigation. We assume that manual investigation always detects intrusions.[5] If the firm detects the intrusion, the firm does not incur any loss,[6] other than the cost of manual investigation. When an intrusion is undetected, the firm incurs a damage of *d*. Most companies estimate these possible damages in the risk assessment phase prior to implementing and configuring the IDS. The quality profile of the detector is modeled through its ROC curve.

Optimal Configurations under Decision Theory and Game Theory Approaches

Our analysis for deriving the optimal configuration proceeds using backward induction follows. First, for a given configuration, i.e., P_D and P_F, we determine the optimal ρ_1 and ρ_2 as a function of P_D and P_F. Then, we determine the optimal P_D and P_F by minimizing the firm's expected cost subject to the ROC curve constraint. The firm can use decision theoretic or game theoretic approach in determining the optimal decisions. We derive the optimal configuration for the decision theory approach first followed by the game theory approach.

Decision Theory Approach

A decision theoretic approach determines optimal configuration by minimizing its cost for a given risk environment. Firms assess the risk before applying the decision theoretic model. This method does not consider the behavioral implications of the configuration on users. The decision theoretic model for a given configuration is shown in Figure 4.

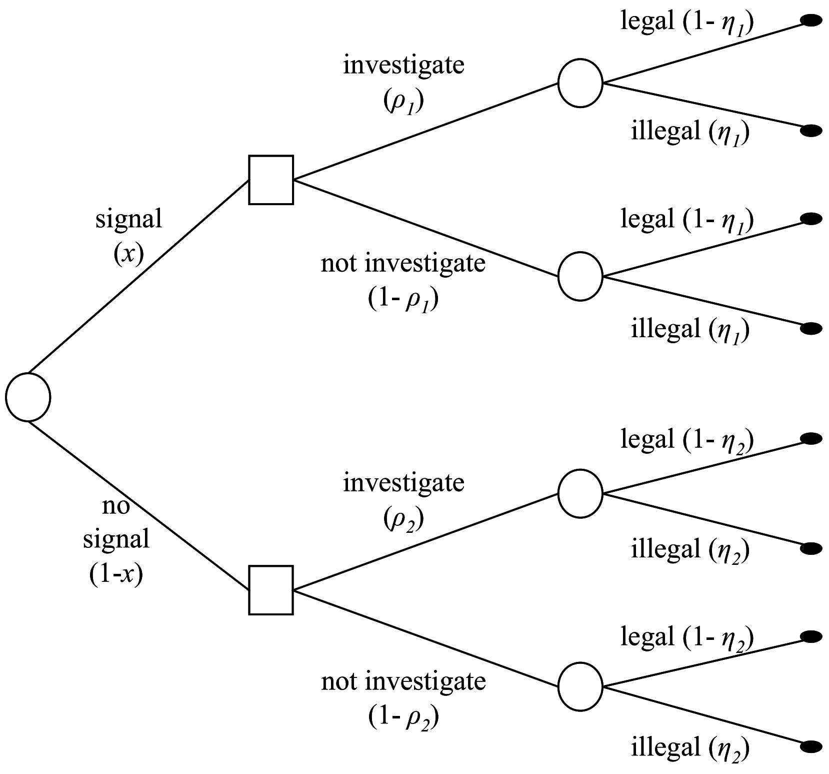

Figure 4. The Decision Theoretic Model for Configuring an IDS

[5] We can extend the model easily to the case when manual investigation is not 100 percent effective. The results will not change qualitatively.

[6] Again, our model can easily be extended to the case in which the firm recovers only a portion of the damage that has already been inflicted. The qualitative nature of results under this more general model is identical to that presented in this paper.

When the firm observes a signal or no signal from the IDS for a transaction, it updates its belief about the transaction type using Bayes' rule. If the firm gets a signal, it determines the posterior probability of intrusion as

$$\begin{aligned}\eta_1 &= \text{P(intrusion|signal)} = \frac{\text{P(signal|intrusion)P(intrusion)}}{\text{P(signal|intrusion)P(intrusion)+P(signal|no-intrusion)P(no-intrusion)}} \\ &= \frac{P_D\psi}{P_D\psi + P_F(1-\psi)}\end{aligned} \tag{4}$$

Similarly, when the firm does not get any signal, it calculates the posterior probability as

$$\begin{aligned}\eta_2 &= \text{P(intrusion|no-signal)} = \frac{\text{P(no-signal|intrusion)P(intrusion)}}{\text{P(no-signal|intrusion)P(intrusion)} + \text{P(no-signal|no-intrusion)P(no-intrusion)}} \\ &= \frac{(1-P_D)\psi}{(1-P_D)\psi + (1-P_F)(1-\psi)}\end{aligned} \tag{5}$$

The probability of the firm being in the signal state can be computed to be $x = (P_F + \psi(P_D - P_F))$.

The expected cost for each action in each of the decision nodes is computed to be the following:

$Cost\ (investigate \mid signal) = \eta_1 c + (1 - \eta_1)c = c$
$Cost\ (investigate \mid signal) = \eta_1(d) + (1 - \eta_1)0 = d\eta_1$
$Cost\ (investigate \mid no\ signal) = \eta_2 c + (1 - \eta_2)c = c$
$Cost\ (investigate \mid no\ signal) = \eta_2(d) + (1 - \eta_2)0 = d\eta_2$

The firm will decide to investigate or not investigate in a state (signal and the no signal) by choosing the action that yields the lower expected cost. The following result shows the firm's optimal strategies in the signal and no signal states.

Result 1. *The optimal manual investigation frequencies for a given IDS configuration in the decision theory framework are as follows.*

$\frac{c}{d} < \eta_2$ ①	$\eta_2 < \frac{c}{d} < \eta_1$ ②	$\frac{c}{d} > \eta_1$ ③
$\rho_1 = 1, \rho_2 = 1$	$\rho_1 = 1, \rho_2 = 0$	$\rho_1 = 0, \rho_2 = 0$

The above result is consistent with our intuition. Choosing to investigate every transaction irrespective of whether the program generates a signal or not is the best strategy for the firm if the ratio (cost/benefit) of investigation is sufficiently low. On the contrary, it is optimal for the firm not to investigate any transaction if the ratio is sufficiently high. If the ratio is moderate, the firm should investigate all and only those transactions that generated signals. Substitution of the above optimal investigation strategies in the firm's expected cost expressions gives the firm's expected cost for different parameter regions, as given in Table 1.

Table 1. Firm's Expected Costs under Decision Theory Approach

Region	Firm's Expected Cost
1	c
2	$d\psi + (c - d)\psi P_D + c(1 - \psi)P_F$
3	$d\psi$

An inspection of the expected cost expressions reveals that configuration is relevant only in region 2 because the expected costs in other regions are independent of P_D and P_F.

Result 2: *IDS is valuable and configuration is important only if (cost/benefit) ratio of manual investigation, ($\frac{c}{d}$), is greater than η_1 and less than η_2 .*

Since configuration is relevant only in region 2, firms only in that region will use the IDS to supplement their manual investigations. We focus only on this region to determine the optimal configuration. That is, we assume that the firm's cost parameters are such that it operates in region 2. Writing P_F as a function of P_D (using equation. 3), we get the firm's expected cost in region 2 as $c\sqrt[r]{P_D}(1-\psi)+d\psi+P_D(c-d)\psi$. Minimizing this over P_D gives

$$P_D^* = \left[\frac{c(1-\psi)}{r(d-c)}\right]^{\frac{r}{r-1}} \tag{6}$$

$$P_F^* = \left[\frac{c(1-\psi)}{r(d-c)}\right]^{\frac{1}{r-1}} \tag{7}$$

Substituting these optimal configuration points into the expected cost expression gives the expected cost of $d\psi+(r-\psi)(d-c)\left[\frac{c(1-\psi)}{r(d-c)}\right]^{\frac{r}{r-1}}$.

The decision theory approach uses an estimate of the probability of attack while solving the configuration problem. However, in reality the probability that a user commits an intrusion need not necessarily be the same as that estimated by the firm. The realized probability of attack depends critically on how the user determines his/her strategy.

Game Theory Approach

The game theory approach differs from the decision theory approach in one important respect. In the game theory approach, the firm makes its decisions by anticipating the behavior of the user in response to its actions. Thus, while the decision theory approach assumes that the probability of intrusion y is unaffected by P_D, P_F, ρ_1 , and ρ_2, the game theory approach utilizes the fact that the user will change his/her strategy based on the firm's decisions. We consider the case when the user and the firm make their decisions simultaneously. We identify a Nash equilibrium. In a Nash equilibrium, neither player has an incentive to deviate from the equilibrium as long as the other player does not deviate.

In order to analyze the strategic interactions between the firm employing the IDS and the user, we enhance the decision theory model to include the user's and firm's strategies. The game theoretical model for the firm is shown in Figure 5.

Node 1 represents a user interacting with the IDS. Node 2 represents a normal transaction whereas node 3 represents an illegal transaction. Nodes 6 and 7 represent transactions that generated a signal from the IDS. Node 7 represents a true attack (true positive) whereas node 6 represents a false alarm (false positive). Nodes 4 and 5 represent transactions that did not generate a signal from the IDS. Node 5 denotes a missed attack (false negative) whereas node 4 implies that the IDS did not give any false alarm (true negative). The firm does not know whether the transaction is in node 6 or node 7 when there is a signal. Similarly, when the program does not generate a signal for a transaction, the firm does not know whether the transaction is in node 4 or node 5. The dashed curves in Figure 5 enclose the information sets of the firm. There are two information sets. The one in which there is a signal, and the one in which there is no signal from the program. The firm must make decisions about whether to conduct a manual investigation of a transaction without knowing exactly which node it is in within the information set.

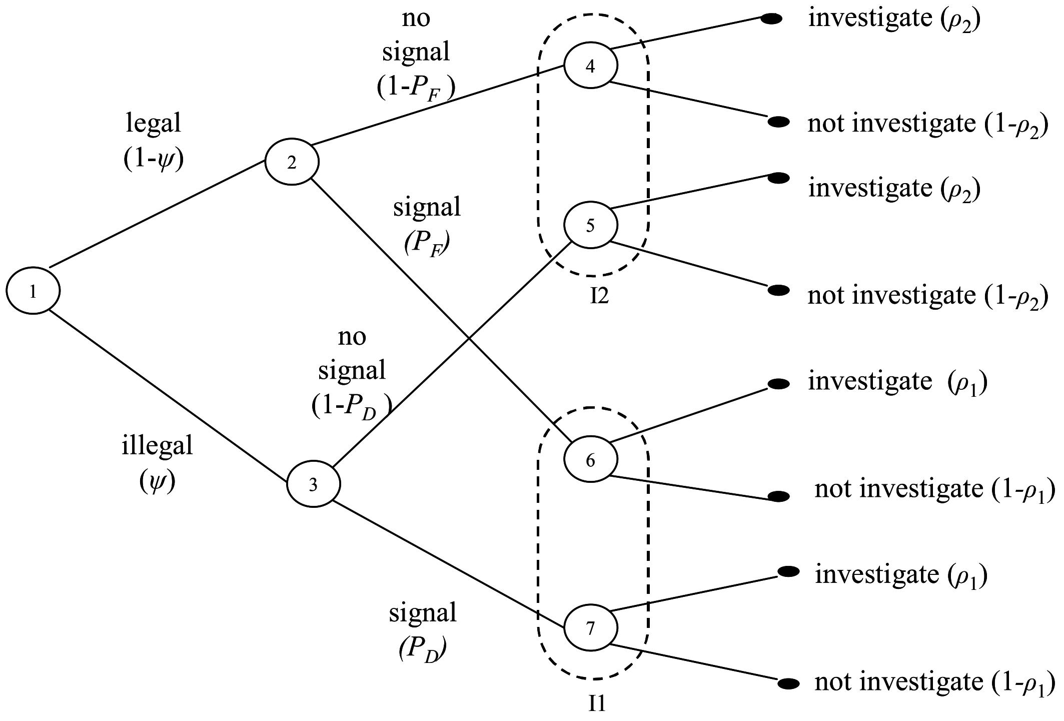

Figure 5. The Game Theoretic Model for Configuring an IDS

The strategy set of a user is $S^h \in$ {*no-intrusion, intrusion*}. The strategy set of the firm is $S^f \in$ {*investigate|no-signal, do-not-investigate|no-signal, investigate|signal, do-not-investigate|signal*} where *investigate|no-signal* means that the firm chooses to investigate the transaction even though the IDS does not generate a signal.

The expected cost for the firm depends on the state it is in. The firm's expected cost for the signal and the no signal states respectively are as follows.

$$F_S(\rho_1,\psi) = \rho_1 c + \eta_1(1-\rho_1)d \tag{8}$$

$$F_N(\rho_2,\psi) = \rho_2 c + \eta_2(1-\rho_2)d \tag{9}$$

The firm's overall expected cost for a transaction is given by

$$F(\rho_1,\rho_2,\psi) = (P_F + \psi(P_D - P_F))F_S(\rho_1,\psi) + (1 - P_F - \psi(P_D - P_F))F_N(\rho_2,\psi) \tag{10}$$

The user's expected benefit is

$$H(\rho_1,\rho_2,\psi) = \psi\mu - \psi\beta(\rho_1 P_D + \rho_2(1-P_D)) \tag{11}$$

The firm minimizes $F_S(\rho_1,\psi)$ when it gets a signal from the IDS and $F_N(\rho_2,\psi)$ when it does not get a signal from the IDS. The user maximizes $H(\rho_1,\rho_2,\psi)$. The following result holds in our model (*All proofs are available from the authors).*

Result 3: *The probability of manual investigation when there is a signal is greater than or equal to the manual investigation probability when there is no signal (i.e., $\rho_1 \geq \rho_2$). In addition, both probabilities cannot be positive and less than one at the same time.*

Result 3 is consistent with the intuition that the firm will investigate a larger fraction of the cases that generated signals from the IDS compared to those that did not generate signals. Now we can present the optimal frequency of manual investigations.

Result 4. *The optimal frequencies for manual investigation for a given configuration of the IDS using game theoretic framework are as follows.*

Region	$\frac{c}{d} < 1$	$\frac{c}{d} > 1$
$\frac{\mu}{\beta} > 1$	$\psi = 1$ ③ $\rho_1 = 1, \rho_2 = 1$	$\psi = 1$ ④ $\rho_1 = 0, \rho_2 = 1$
$P_D < \frac{\mu}{\beta} < 1$	$\psi = \frac{c(1-P_F)}{c(P_D - P_F) + (1-P_D)d}$ ② $\rho_1 = 1 , \rho_2 = \frac{\mu - P_D\beta}{(1-P_D)\beta}$	
$\frac{\mu}{\beta} < P_D$	$\psi = \frac{cP_F}{P_D d - c(P_D - P_F)}$ ① $\rho_1 = \frac{\mu}{P_D\beta} , \rho_2 = 0$	

The equilibrium policies given in result 4 are characterized by the ratios $\frac{c}{d}$ and $\frac{\mu}{\beta}$. We can observe that, in general, the probability of investigation decreases and probability of intrusion increases when $\frac{c}{d}$ increases for a given $\frac{\mu}{\beta}$, and the probability of intrusion as well as probability of investigation increases with $\frac{\mu}{\beta}$ for a given $\frac{c}{d}$. Table 2 shows the firm's expected cost under the optimal manual investigation policy for the different regions.

Table 2. Firm's Expected Costs

Region	**Firm's Expected Cost**	**Region**	**Firm's Expected Cost**
1	$\frac{cP_F d}{P_D d - c(P_D - P_F)}$	3	c
2	$\frac{c(c(P_D - P_F) + (1-P_F)d - (P_D - P_F)d)}{c(P_D - P_F) + (1-P_D)d}$	4	d

The following result is apparent from Table 2.

Result 5: IDS is valuable and as a result configuration is important only if benefit to cost ratio of intrusion for the user ($\frac{\mu}{\beta}$) and cost to benefit ratio of manual investigation for the firm ($\frac{c}{d}$) are both less than one.

Configuration is important only in region 1 and region 2. Through configuration, the firm can choose which region to lie in by specifying P_D and P_F. Subtracting the expected cost in region 1 from that in region 2 gives $\frac{c(P_D-P_F)(c-d)(d+c(P_D-P_F)-dP_D)}{(c(P_F-P_D)-d(1-P_D))(P_D(d-c)+cP_F)} \geq 0$. Hence the firm will choose its configuration so that the equilibrium point lies in region 1. Next the firm should decide where to lie within region 1. Writing the cost expression in region 1 as a function of P_D gives $\frac{cd\sqrt[r]{P_D}}{c\sqrt[r]{P_D}+P_D(d-c)}$.

$$\frac{\partial()}{\partial P_D}=\frac{cd\sqrt[r]{P_D}(1-r)(d-c)}{r[c\sqrt[r]{P_D}+P_D(d-c)]^2}\geq 0 \tag{12}$$

This derivative implies that the firm will choose to set P_D as small as possible. Since the firm wants to be in region 1, the firm sets P_D of its detection system to $\frac{\mu}{\beta}$.[7] That is, the optimal configuration for the IDS is

$$P_D^*=\frac{\mu}{\beta} \tag{13}$$

$$P_F^*=[\frac{\mu}{\beta}]^{\frac{1}{r}} \tag{14}$$

Substituting the above optimal configuration point into the cost expression gives an expected cost of $\frac{d}{1+\left(\frac{\mu}{\beta}\right)^{1-\frac{1}{r}}(\frac{d}{c}-1)}$.

Comparison of Game Theoretic and Decision Theoretic Approaches

Having derived the optimal configurations under the decision theoretic and game theoretic approaches, we can now compare the firm's realized costs under these approaches. We use the following definitions for our comparisons.

P_D^D ≡ optimal probability of detection under decision theory approach

P_F^D ≡ optimal probability of false alarm under decision theory approach

P_D^G ≡ optimal probability of detection under the game theory approach

P_F^G ≡ optimal probability of false alarm under game theory approach

ψ_R ≡ realized probability of intrusion

The realized cost at the optimal configuration under the decision theory approach is given by $d\psi_R+(c-d)\psi_R P_D^D+c(1-\psi_R)P_F^D$. We need to determine the realized probability of intrusion in order to compute the realized cost under the decision theory. Since the user is strategic, he/she will adjust his/her strategy depending on the strategy

[7] Actually, the firm will set P_D to $\frac{\mu}{\beta}+e, e>0$, where *e* is an infinitesimally small number.

used by the firm. That is, the user will choose ψ_R based on his/her utility.[8] From the user utility function, we find that the user will intrude, i.e., $\psi_R = 1$, if $P_D^D < \frac{\mu}{\beta}$ and will not intrude, i.e., $\psi_R = 0$ otherwise. Thus, the realized cost under decision theory approach is computed to be the following.

$$\begin{cases} cP_F^D & \text{if } \frac{\mu}{\beta} < P_D = \left(\frac{c(1-\psi_D)}{r(d-c)}\right)^{\frac{r}{r-1}} \quad (i.e.\ \psi_R = 0) \\ cP_D^D + d(1-P_D^D) & \text{if } \frac{\mu}{\beta} > P_D = \left(\frac{c(1-\psi_D)}{r(d-c)}\right)^{\frac{r}{r-1}} \quad (i.e.\ \psi_R = 1) \end{cases}$$

Under game theory, the realized cost is identical to the expected costs at the equilibrium because neither party deviates from the equilibrium. Consequently, the realized cost when the firm uses the game theory approach is $\frac{d}{1+\left(\frac{\mu}{\beta}\right)^{1-\frac{1}{r}}(\frac{d}{c}-1)}$. A comparison of the realized costs under the decision theory and game theory approaches gives the following result.

Result 6: *Configuration using game theory results in lower cost than configuration using decision theory unless*

$$1-\frac{r(d-c)}{c}\left(\frac{\mu}{\beta}\right)^{\frac{r-1}{r}} < \psi_D < 1-\frac{r(d-c)}{c}\left(\frac{\mu}{\beta}\right)^{\frac{r-1}{r}}\left(\left(\frac{c}{d}\right)\left(\frac{\mu}{\beta}\right)^{\frac{1}{r}}+\left(\frac{\mu}{\beta}\right)\left(1-\frac{c}{d}\right)\right)^{1-r}$$

The above result can be illustrated using Figure 6, which plots the realized costs under two approaches for different values of ψ_D. It also shows the realized values of intrusion probability in different regions. Game theory approach results in a lower cost to the firm in all cases except when the firm uses a attack probability that lies in the interval (*a*,*b*), where *a* and *b* are respectively the lower and upper limits for ψ_D given in result 6, under decision theory approach.

Decision theory performs worse than game theory if the firm does not deter the user from committing the intrusion. In other words, if the firm estimates a low probability of attack and configures the IDS with a low probability of detection, the user ends up committing an intrusion. In this scenario, the firm incurs a higher cost than under game theory in which the probability of attack is less than one. If the firm is successful in deterring the user from committing an intrusion under decision theory, then its cost under decision theory is lower than game theory only when it does not "over deter" the user, i.e., it does not configure the program with a very high probability of detection. When the firm's estimate of the attack probability is very high, it will configure the program to have a high probability of detection, which will deter the user. While the firm is successful in deterring the user, it also incurs a high cost from the high rate of false signals from the software. Consequently, the firm realizes a lower cost under game theory even though the game theory solution does not deter the user from committing an intrusion.

The result that decision theory can perform better than game theory, although only when the condition stated in result 6 is satisfied, is counter-intuitive. One would expect that the firm will be better off when it makes the decisions by anticipating the reaction of the user, as in game theory, as compared to when it does not take the user reaction into consideration, as in decision theory. The counter-intuitive result can be explained as follows. We can show that when $\psi_D = 1-\frac{r(d-c)}{c}\left(\frac{\mu}{\beta}\right)^{\frac{r-1}{r}}$, i.e., when ψ_D is equal to the lower limit of the condition stated in result 6, the firm uses the same configuration under decision theory as game theory. Under this scenario, though the final outcomes in terms of configuration are identical under the decision theory and game theory methods, the models used to arrive at these outcomes are different in these methods. If the firm is lucky enough to con-

[8] In the simultaneous game, the user will not be able to observe the firm's decision if the firm also uses game theory. We can interpret the equilibrium derived in this paper as the steady state that will be reached after several alternating moves by the firm and the user. However, if the firm is nonstrategic and uses the decision theory approach, the user will adjust his/her strategy in the second period based on what he/she observes in the first period. Since the firm does not adjust its strategy, the equilibrium will be reached in the second period itself.

figure the software correctly under decision theory approach, then the firm will incur the lowest possible cost because essentially the firm behaves as though it knows the realized attack probability, which leads to a first mover advantage for the firm.[9] Neither party has a first mover advantage in game theory, and thus the firm incurs a higher cost under game than decision theory approach. This first mover advantage to the firm under decision theory approach exists as long as the firm does not over-configure (i.e., set a high detection/false positive rate). If the firm over-configures, the first mover advantage is offset by the higher costs associated with more frequent false alarms.

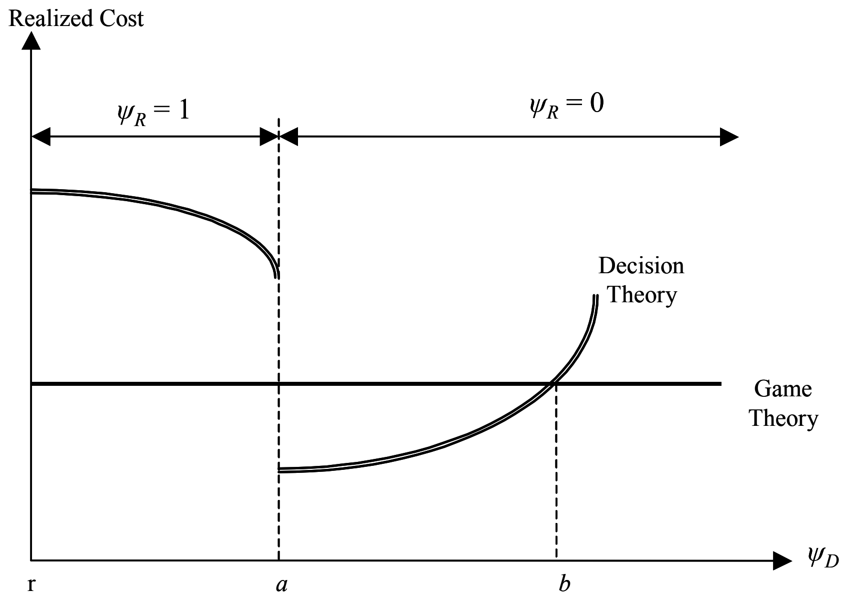

Figure 6. The Game Theoretic Model for Configuring a Detection Program

Our results clearly show that under most circumstances firms realize lower cost when they use game theory as opposed to decision theory to configure IDSs when firms are faced with a strategic adversary. However, there seems to a dichotomy between our results and current business practices, which seem to favor decision theory approach. We hypothesize several reasons for the dichotomy. One reason could be that firms are truly unaware of the potential benefits of applying the game theoretical approach for the intrusion detection problem. We believe that this paper provides insights to firms on how and why game theory performs better than decision theory. Another reason could be that firms view decision theory as a simplification of the more complex game theory approach. The decision theoretic and game theoretic models require estimation of several parameters. In some sense, the game theoretic model requires deeper user-specific parameters that are more difficult to obtain. For instance, game theory requires that the firm knows the user utility and penalty parameters, whereas decision theory requires only the final outcome of users' decisions in terms of attack probabilities. We believe that attack probabilities are easier to obtain compared to utility because firms can use historical log records to estimate attack probability. We conjecture that the difficulty in estimation of user specific parameters is one reason why firms may prefer to use decision theory instead of game theory.

Conclusions

The performance of IDSs is severely limited by their false positive and false negative errors. Achieving a high detection rate is typically accompanied by a high false signal rate. Consequently, firms need to configure the IDS carefully to achieve a balance between these rates. In this paper we presented two models, the first based on decision theory and the second based on game theory, to assist firms in the configuration of IDSs. In the decision theory approach, the firm estimates the attack probability

[9] Recall that when the firm uses decision theory approach, it makes its decisions once based on its estimate of the attack probability. The strategic user adjusts his/her strategy based on the firm's decision. Thus, the firm acts as a leader and the user acts as a follower under decision theory approach.

exogenously and assumes that its actions do not alter users' behavior. In the game theory approach, the firm makes its decisions by assuming that the firm's decisions alter user behavior. We found that firms incur lower cost under most situations when they use game theory as opposed to decision theory. The decision theory approach results in a lower cost only when the firm neither under-configures nor over-configures the IDS by a significant amount. The results of our model caution firms that use of the decision theory approach for the IDS configuration that ignores the reactions of their strategic adversaries can cause significant harm to firms. Traditionally IDS configuration is viewed as a firm's internal problem that affects only the firm. While this is true of systems that deal with operational problems such as transaction processing systems, strategic applications require modeling the strategic interactions. Consequently, the design and configuration of such systems need to take into account the effect of configuration on user behavior.

References

Axelsson, S. "The Base-Rate Fallacy and the Difficulty of Intrusion Detection," *ACM Transactions on Information and System Security* (3:3), 2000, pp. 186-205.

Cavusoglu, H., Raghunathan, S., and Mishra, B. "Optimal Design of Information Technology (IT) Security Architecture," in *Proceedings of the Twenty-Second International Conference on Information Systems*, L. Applegate, R. Galliers, and J. I. DeGross (eds.), Barcelona, Spain, 2002.

CERT Coordination Center. *Security for Information Technology Service Contracts*, CERT Security Improvement Modules, 2001.

Durst, R., Champion, T., Witten, B., Miller, E., and Spagnuolo, L. "Testing and Evaluating Computer Intrusion Detection Systems," *Communications of the ACM* (42:7), 1999, pp. 53-61.

Fellingham, J., and Newman, P. "Strategic Considerations in Auditing," *The Accounting Review* (60), October 1985, pp. 634-650.

Gaffney, J. E., and Ulvila, J. W. "Evaluation of Intrusion Detectors: A Decision Theory Approach," *IEEE Symposium on Security and Privacy*, 2001, pp. 50-61.

Jajodia, S., and Miller, J. "Editor's Preface," *Journal of Computer Security* (16:4), 1993, pp. 43-53.

Jonsson, E., and Olovsson, T. "A Quantitative Model of Security Intrusion Process Based on Attacker Behavior," *IEEE Transactions on Software Engineering* (23:4), 1997, pp. 235-245.

Lee, W., Fan, W., Miller, M., Stolfo, S., and Zadokm E. "Toward Cost-Sensitive Modeling for Intrusion Detection and Response," *Journal of Computer Security* (10:1/2), 2002, pp. 5-22.

Lippmann, R., Haines, J. W., Fried, D. J., Korba, J., and Das, K. "The 1999 DARPA Off-Line Intrusion Detection Evaluation," *Computer Networks* (34:4), 2000, pp. 579-595.

McHugh, J. "Testing Intrusion Detection Systems: A Critique of the 1998 and 1999 DARPA Intrusion Detection System Evaluations as Performed by Lincoln Laboratory," *ACM Transactions on Information and System Security* (3:4), 2000, pp. 262-294.

NMAB. *Configuration Management and Performance Verification of Explosives-Detection Systems*, Publication NMAB-482-3, National Academy Press, Washington, DC, 1998.

Ortalo, R., Deswarte, Y,m and Kaaniche, M. "Experimenting with Quantitative Evaluation Tools for Monitoring Operational Security," *IEEE Transactions on Software Engineering* (25:5), 1999, pp. 633-650.

Sarkar, S., and Sriram, R. "Bayesian Models for Early Warnings of Bank Failures," *Management Science* (47:11), 2001, pp. 1457-1475.

Shapiro, G. P., and Masand, B. "Estimating Campaign Benefits and Modeling Lift," in *Proceedings of KDD-99*, ACM Press, New York, 1999, pp. 185-193.

Shapiro, G. P., and Steingold, S. "Measuring Lift Quality in Database Marketing," *SIGKDD Explorations* (2:2), 2000, pp. 81-86.

Steingold, S., Wherry, R., and Shapiro, G. P. "Measuring Real-Time Predictive Models," in *Proceedings of IEEE International Conference on Data Mining*, 2001, pp. 649-650.

Van Trees, H. *Detection, Estimation and Modulation Theory-Part I.* John Wiley, New York 2001.