

# A Framework for Cost Sensitive Assessment of Intrusion Response Selection

Chris Strasburg      Natalia Stakhanova      Samik Basu      Johnny S. Wong  
*Dept. of Computer Science    Faculty of Computer Science    Dept. of Computer Science    Dept. of Computer Science*  
*Iowa State University    University of New Brunswick    Iowa State University    Iowa State University*  
*Ames, IA, USA      Canada      Ames, IA, USA      Ames, IA, USA*  
*Email: cstras@cs.iastate.edu    Email: natalia@unb.ca    Email: sbasu@cs.iastate.edu    Email: wong@cs.iastate.edu*

**Abstract**—In recent years, cost-sensitive intrusion response has gained significant interest, mainly due to its emphasis on the balance between potential damage incurred by the intrusion and cost of the response. However, one of the challenges in applying this approach is defining a consistent and adaptable measurement of these cost factors on the basis of system requirements and policy. In this paper, we present a host-based framework for the cost-sensitive assessment and selection of intrusion response. Specifically, we introduce a set of measurements that characterize the potential costs associated with the intrusion handling process, and propose an intrusion response evaluation method with respect to the risk of potential intrusion damage, the effectiveness of the response action and the response cost for a system. We provide an implementation of the proposed solution as an IDS-independent plugin tool and demonstrate its advantages on the several attack examples.

**Keywords**—D.4.6. Security and Privacy Protection; D.4.8. Performance; H.1.2. User/Machine Systems; K.6.5. Security and Protection;

## I. INTRODUCTION

The proliferation of complex and fast-spreading intrusions against computer systems has increased the demands on intrusion detection and response, requiring not only advances in detection mechanisms but also the development of sophisticated and automated response systems.

The majority of automated intrusion response systems rely on mapping attacks to pre-defined responses [1], [2]. This approach allows a system administrator to deal with intrusions faster and more efficiently, however it lacks flexibility as few of these systems incorporate intrusion cost factors.

In recent years a trend toward cost-sensitive modeling of response selection has emerged [3], [4], [5], [6]. The aim of this strategy is to balance intrusion damage and response cost to ensure adequate response without sacrificing the normal functionality of the system under attack. However, defining an accurate measurement of these cost factors and ensuring consistent evaluation across varied computing environments are common challenges in using a cost-sensitive approach.

Typically, intrusion response selection is based on an intuitive manual assessment of various factors such as likelihood and severity of intrusion, extent of intrusion damage, effectiveness of response actions, or expected duration of the response.

However, these factors are rarely applied to the automated evaluation of intrusion response in published literature, mainly due to the lack of common interpretation of these factors and the absence of concrete metrics with clear conceptual meaning for measuring them. For instance, most of the existing models supporting automatic response selection introduce response cost as one of the factors in the selection of the suitable response strategy. However, they do not agree on what constitutes the response cost and how it can be measured. Some suggest that response cost includes the labor cost of personnel involved in response deployment and criticality of the attack [3], while others see response cost as a measure of its effectiveness against a detected attack and its disruptiveness to legitimate users [5].

*Solution Methodology:* In this paper we present a host-based framework for the cost-sensitive assessment of intrusion response. We introduce a set of measurements to characterize potential costs associated with the intrusion handling process and propose a method for evaluating intrusion response with respect to potential intrusion damage, response effectiveness and response cost for a system.

The main contributions of this work can be summarized as follows:

- 1) **System aware adaptable model:** the proposed model is adaptable to different environment settings, i.e. systems with varying operational requirements.
- 2) **Consistent approach to response selection:** the proposed evaluation metrics are defined in terms of system resources, bringing a common ground to the selection process.
- 3) **Generalized approach to response selection:** while some existing approaches offer intrusion specific methods of response selection [7], we

establish a general assessment mechanism which supports the analysis of response measures for attacks in the context of the security policies of the given system.

- 4) **Intrusion handling cost measures:** this work provides a more complete cost assessment of the attack handling process, considering not only direct damage caused by the intrusion but also indirect costs that often remain hidden.
- 5) **Implementable in practice:** we provide a detailed assessment process which allows system administrators of a broad range of skill levels to employ the approach.

*Organization:* The remainder of the paper is organized as follows. A brief overview of related work is given in Section II. Section III provides an overview of the proposed intrusion selection framework and presents the details of the intrusion damage and intrusion response cost calculations. The experimental results are given in Section V. Section VI concludes the paper.

## II. RELATED WORK

A number of techniques aimed at enhancing intrusion response automation were proposed and deployed over the last five years. A comprehensive review of this work is given by [8], [9].

Several works specifically focused on the costs and benefits of intrusion response. The approach proposed by Lee et al. [3] is based on a combination of several factors: the cost of detecting the intrusion, the amount of damage caused by the attack and the operational cost of reaction to the intrusion. Wu et al. [5] rely on attack graphs to identify the actions required to achieve possible attack goals in a distributed system. As opposed to these approaches, models proposed by Toth and Kruegel [10], Balepin et al. [4] and Jahnke [11] not only consider costs and benefits of the responses, but also introduce a link between the cost of responses and the system resources in the network.

One common problem of these models is the lack of consistency. While there is significant conceptual overlap, each model approaches response cost evaluation and selection from a different perspective. Wu et al. [5] measures the response effectiveness against a detected attack based on past experience. Lee et al. [3] relates the response cost to required labor efforts. The works by [10], [4] and [11] consider response cost in terms of system resources, but offer varying evaluation methods. [4] measures the response cost as the sum of manually assigned costs of affected resources. [10] calculates response cost as a function of system capability reduction, while [11] extends the idea in [10] by adding a fine-grained quantification of system resource unavailability.

The emerging theme in these works effectively establishes employing system resources in the evaluation of intrusion response, and we build our approach to response selection based on this promising trend. In this context, our model can be viewed as a generalization of the existing approaches.

## III. INTRUSION RESPONSE SELECTION

Traditionally, triggering an intrusion response is left as part of the administrator's responsibility. Choosing the response measures and their parameters, an administrator essentially weighs the benefits and risks of each response compared to the potential intrusion damage. Thus, the problem of intrusion response selection can be formulated as the problem of minimizing the potential system damage incurred by an attack and a deployed response. Formally, given a set of possible intrusions  $I$  and a system damage function  $SD(i \in I)$ , the goal of the intrusion response selection is to identify the set of responses  $R$  that allow to minimize the impact of intrusions while ensuring the least possible damage from responses' deployment. That is:

$$\text{minimize } \left\{ \sum_{i \in I} SD(i) \right\}$$

To resolve this problem we propose a framework for intrusion response selection based on cost-benefit analysis.

### A. Evaluation metrics

Evaluation of the response actions effectiveness in the context of a specific intrusion requires analysis of several factors such as likelihood and severity of intrusion, extend of the potential intrusion damage, effectiveness of suitable response actions, response cost for the system, etc. These factors can be broadly divided into two groups: *factors associated with intrusion damage* and *factors describing response cost*.

*Intrusion Cost Factors:* The intrusion damage  $D$  caused to a system at any particular time incorporates potential damage from four categories: deploying a response due to false alarms  $D_{FalseAlarm}$ , deploying no response when an intrusion occurs (false negative alarms)  $D_{FalseNegative}$ , deploying a sub-optimal response due to a mis-labeled alarm  $D_{MislabeledAlarm}$  and finally, resulting from a true attack  $D_{true}$ .

$$D = \left\{ \begin{array}{l} D_{true}, D_{FalseAlarm}, \\ D_{FalseNegative}, D_{MislabeledAlarm} \end{array} \right\}$$

The response selection strategy, although aimed at minimizing the overall cost, only partially addresses this damage set.

*False Alarm Damage* results from responding to false alarms. The only damage incurred to the system is cost associated with the deployed response action  $r$ , denoted as  $RC$ .

$$D_{FalseAlarm} = \sum_{i \in FalseAlarms} RC(r, i) \quad (1)$$

*False Negative Damage* is caused by undetected (hence un-responded to) attacks. The damage consists of that caused by an attack. As no response is deployed, the response selection strategy cannot address this damage type.

$$D_{FalseNegative} = \sum_{i \in FalseNegative} D_{true}(i) \quad (2)$$

*Mis-labeled Alarm Damage* is a result of deploying a response to an incorrectly detected intrusion  $\hat{i}$ . Depending on the damage estimate for the true intrusion  $i$ , this may include residual attack damage which can be addressed through follow-up response measures, or the negative effect of the deployed response if this cost exceeds the true intrusion damage.

$$D_{MislabelledAlarm} = \sum_{\hat{i} \in Mislabelled} (D_{true}(i) - D_{true}(\hat{i})) \quad (3)$$

*True Alarm Damage* is caused by a truly occurring attack detected by the IDS. Estimating the actual attack damage is usually a part of the risk assessment process [12]. While apriori measurement of actual damage for an attack is not always possible, approximate values can be estimated. Although these values are useful to show the benefit of security product investments, they present vague guidance for system administrators during the selection of a response for an occurring intrusion.

For a deployable framework, we need to consider practical measurements that can be easily provided by system administrators. Thus, we define system damage caused by an attack using three components<sup>1</sup>:

- *System resources affected by intrusion.* System resources, denoted by  $SR = (sr_1, \dots)$  can be broadly viewed as services provided by the system (e.g., FTP, HTTP, etc.) which contribute to its value. The value or weight of each  $sr_j$ , denoted  $W(sr_j)$  is assigned according to its impact on the security policy of the system, and computed as a combination of the resource importance for system confidentiality  $F_C(sr_j)$ , availability  $F_A(sr_j)$  and integrity  $F_I(sr_j)$  and the weight of these factors on the security policy of system.

<sup>1</sup>In practice, the evaluation of potential attacks can be performed through the analysis of the enabled signatures in the employed intrusion detection system.

$$W_{sr_j} = F_C(sr_j) \times w(C) + F_I(sr_j) \times w(I) + F_A(sr_j) \times w(A) \quad (4)$$

- *Intrusion impact on system resources.* The intrusion impact on system resources is defined in terms of the loss in confidentiality, availability and integrity. The set of considered intrusions is denoted  $I = (i_1, \dots)$ , where  $i_k$  is associated with the corresponding intrusion signature. The intrusion system impact  $SI(i_k)$  is computed by assessing the possible effect of  $i_k$  on each system resource  $sr_j$  and characterizes a potential degradation of the system state under intrusion  $i_k$ . The effect of  $i_k$  on system resource  $sr_j$  is denoted by  $E(i_k, sr_j)$  and is assessed in three dimensions: confidentiality  $c$ , availability  $a$  and integrity  $i$ . As such:

$$E(i_k, sr_j) = i_k(C) + i_k(I) + i_k(A) \quad (5)$$

To address the uncertainty in estimating the intrusion impact in advance, the effect on each category is a binary value indicating a possibility of resource damage. For instance, in a denial of service (DoS) attack, the confidentiality of stored data will not be affected, so file resource confidentiality impacts will be 0, but availability impact will be 1. Thus  $E(DoS, file\_resource) = 0 + 0 + 1$ .

$$SI(i_k) = \sum_{sr_j \in SR} E(i_k, sr_j) \times W_{sr_j} \quad (6)$$

- *Operational cost* is the baseline cost present for an intrusion, regardless of system damage caused. Costs in this category include mandatory reporting requirements, forensic analysis, etc. In the risk-assessment field these costs are usually expressed as monetary values, however, in other areas a more effective solution appears to be the application of relative measurements [3]. In our context, we also choose the latter approach and assign a relative measure to the operational cost of an intrusion  $i_k$ , denoted by  $OC(i_k)$ , with the value in the range  $[0, 1]$ .

The potential system damage caused by a true intrusion  $i_k$  is given as follows:

$$D_{true}(i_k) = SI(i_k) + OC(i_k) \quad (7)$$

*Response Cost Factors:* A challenge in assessing response cost is accurately defining numeric values. Assigning monetary values, although providing a concrete metric, is not always possible. A more effective solution is the use of relative measurements based on system-specific policies.

We employ the methodology for response cost assessment originally presented in [13]. Response cost,  $RC$ ,

is essentially the price of deploying a response  $r \in R$  on a given system. The cost estimate is composed of three quantities:

- *the operational cost (OC)* of a response in a given environment. This measures various aspects of the response associated with its daily maintenance.
- *the response goodness (RG)* with respect to intrusion(s) the response is able to mitigate. The response goodness represents the ability of a response to mitigate damage caused by the intrusion to the system resources.
- *the response impact on the system (RSI)*. Finally, the impact of a response on the system quantifies the damage caused to system resources and is estimated independently from the effectiveness in countering the intrusion(s).

Intuitively, the combination of *OC* and *RSI* constitutes the penalty associated with the response, while *RG* is the benefit of this response measure in terms of suspected intrusions. Thus, the general measure of response cost *RC* in the context of the suspected attacks *I* is given as follows:

$$RC(r, I) = OC + RSI - RG \quad (8)$$

The *RC* measure does not serve as indication of the response cost in a particular situation, but reflects the general level of risk associated with its deployment. Thus while invoking the response solely based on the *RC* measure ensures the minimum expected damage caused by a response, it will not necessarily yield the optimal result for the system, e.g. in the case of a false positive, or insignificant intrusion damage.

#### IV. INTRUSION RESPONSE EFFECTIVENESS

Intuitively, the expected effectiveness of a response measure is the difference between the potential intrusion damage the response might prevent and the response cost.

However, in practice this evaluation is bound to the specific attack context and thus also depends on several factors: (1) the likelihood of the intrusion, (2) the probability that we are considering the correct intrusion in the case of multiple possible attacks, and (3) the potential ability of the response to counter a suspected attack.

The first two factors depend on the *effectiveness of the intrusion detection system*  $E(IDS)$ , in other words, its ability to correctly identify the attack in a timely fashion. An evaluation of IDS effectiveness is based on a variety of factors, the most basic being false alarm rate, detection rate, IDS capability, expected cost, etc. [14], [15]. These factors relate to IDS alert confidence and allow one to probabilistically estimate the potential extent of intrusion damage.

The third factor, the potential ability of response  $r$  to counter an attack  $i$ , denoted as  $SF_{r,i}$ , relates to the expected response performance and indicates how successful the response was in countering this intrusion in the past. Let  $Pr_{success}(r, i)$  be the percentage of successful response deployments of  $r$  against intrusion  $i$  and  $S_{level}$  be the *success level*, indicating the degree of success in handling  $i$ . Then  $SF_{r,i}$  can be computed as follows:

$$SF_{r,i} = Pr_{success}(r, i) \times S_{level} \quad (9)$$

A common approach to response success quantification utilizes the distinction of two response outcomes: *response success*, if the deployed response achieves the expected result (e.g., blocked intrusion, collected the data), and *response failure* otherwise. A limitation of this approach is the inability of the response system to distinguish intrusion steps disabled as a result of the response, which may later result in an additional response needlessly deployed to counter neutralized intrusion behavior.

Another limitation lies with the response, as it is often represented by a set of multi-targeted actions. Labeling such a strategy as failed essentially indicates that none of the response actions succeeded, which underestimates the value of the deployed response.

While for the purpose of this work we adopt a common view of the response outcome, considering more sophisticated strategies to resolve partial response success is a promising research direction. As such we consider *success level* as binary variable which takes a value of 0 in case of the response failure and a value of 1 if the response succeeds.

Based on these factors, the expected value of response  $r$  for a given intrusion set  $I_s$  can be estimated as follows

$$RV(r, I_s) = \sum_{i \in I_s} RC(r, i) - \langle D(i) \times SF_{r,i} \times p_i \times E(IDS) \rangle \quad (10)$$

where  $p_i$  is the likelihood of intrusion  $i$  and  $D(i)$  corresponds to one of the intrusion types: true positive, false positive or mis-labeled intrusion. Since accurately identifying the specific intrusion type is often not possible in advance, we rely on  $E(IDS)$  to adjust the calculations dynamically during the response evaluation process.

As the primary aim of response cost analysis is to minimize the potential system damage  $SD$  incurred by an attack given a suspected set of IDS intrusions  $I_s$ , the goal of the response engine is to select a response  $r$  which minimizes the impact of  $I_s$  while ensuring the least possible damage from response deployment. More

formally:

$$SD(I) = \min_{R'' \subseteq R} \{RV(r, I)\}$$

## V. INTRUSION RESPONSE SELECTION ASSESSMENT

To evaluate the effectiveness of the proposed framework, we implemented it as a plugin tool for an intrusion detection system (IDS) and performed a series of experiments focusing on two issues: the cost-benefit advantage of the approach and its scalability.

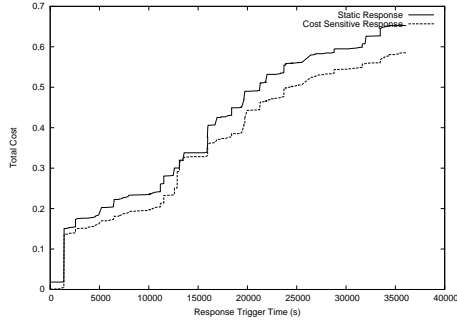


Figure 1. Evaluation of the response on a system with high availability requirements (public web server).

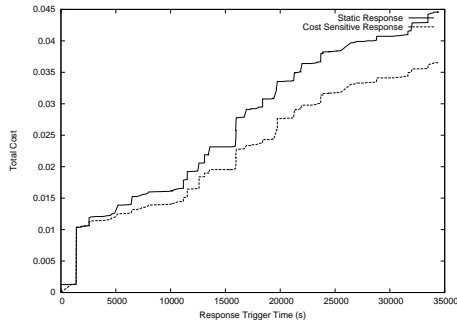


Figure 2. Evaluation of the response on a system with high confidentiality requirements (medical data input system).

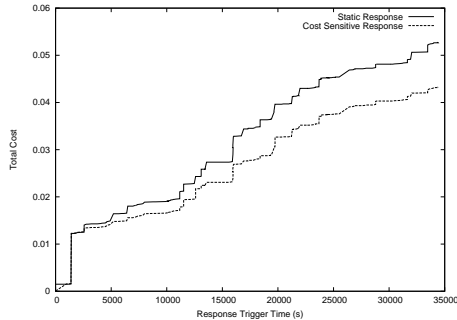


Figure 3. Evaluation of the response on a system with high integrity requirements (central file repository).

**Results.** We evaluated our model using the 1998 DARPA/Lincoln Lab offline evaluation data, in particular week 6, and an open-source, signature-based Snort IDS.

In these experiments we focused on the accuracy of the cost-sensitive selection on systems with different security policies. The characteristics of these systems are given in Table I. We define the *cumulative response cost* metric as our primary criteria, showing the cumulative value of all responses deployed on the system. The results of the experiments in comparison to a traditional system equipped with a static intrusion response mechanism are given in Figures 1, 2, 3.

The proposed cost-sensitive approach clearly outperforms the static response. Although initially the difference between the cumulative response value of our cost-sensitive approach and the system with static response is small, as the number of the deployed responses increases, this difference becomes significant.

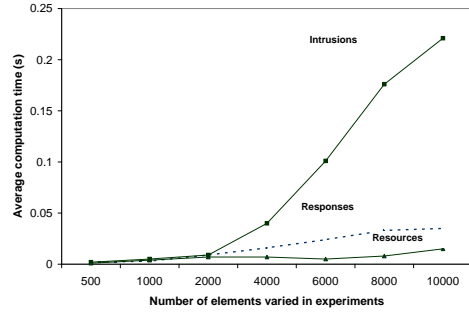


Figure 4. Processing time evaluation

**Performance.** We conducted a set of experiments to measure the performance of our algorithm on artificial data sets. The experiments were run on Intel(R) Core(TM)2 Duo CPU U7700 with CPU of 1.33GHz and 1 GB of RAM and were performed on three variables: number of system resources, number of responses available in the system and number of suspected intrusions. During the experimentation with each variable, the other two variables remained fixed at value 100. The results, given in Figure 4, show that even with a significantly large number of resources to consider, the computation time is only 0.015s. As Figure 4 shows that the highest impact on performance is the number of suspected intrusions simultaneously considered during the response selection process, but even so, it took 0.221s to assess the available responses for 10000 suspected intrusions. However, such large set of intrusions is rarely, if ever, found in practice. Normally there are only a few possible intrusions to consider at a time.

These results show that our approach has reasonable performance requirements, suitable for the efficient analysis of response selection in an automatic setting,

|  |  |
|--|--|
| System legend: <i>public web server providing remote access for affiliates and public information</i><br>System security priorities: low confidentiality 0.1, moderate integrity 0.5, high availability 0.9.   |  |
| System legend: <i>central file repository for distributed collaboration</i><br>System security priorities: moderate confidentiality 0.5, high integrity 0.9, low availability 0.1.   |  |
| System legend: <i>Medical data input system for a hospital</i><br>System security priorities: high confidentiality 0.9, high integrity 0.9, low availability 0.1.  |  |
| System Resources:<br>Web server (providing web access to files for users)<br>FTP server (providing FTP access to get and post files)<br>RPC server (providing remote procedures used by some of the web services and possibly clients)<br>SNMP server (used for system monitoring and management by the administrator) | System responses:<br>Block Attacker's address in .htaccess<br>Block Attacker's IP Address in Firewall<br>Restart Service (Web, FTP, RPC, or SNMP)<br>Reboot System<br>Stop Service<br>Disable User Logins<br>Disable specific user account |

Table I  
THE CHARACTERISTICS OF THE HYPOTHETICAL SYSTEMS USED IN THE EXPERIMENTS.

as well as in support of manual response assessment during an administrator's daily routine.

## VI. CONCLUSION AND FUTURE WORK

In this paper we presented a host-based framework for the cost-sensitive assessment and selection of intrusion response. We incorporate a set of evaluation metrics for the practical assessment of costs and benefits associated with intrusion response, and introduce a balanced strategy for response selection according to the security policy of the given system.

In the future, we intend to investigate the extension of cost-sensitive response selection to a distributed environment, and explore further automation to support the quantifying related values.

## REFERENCES

- [1] D. Schnackenberg, H. Holliday, R. Smith *et al.*, "Cooperative intrusion traceback and response architecture," in *Proceedings of the IEEE DARPA DISCEX I*, 2001.
- [2] A. Somayaji and S. Forrest, "Automated response using system-call delay," in *Proceedings of the USENIX Security*, 2000.
- [3] W. Lee, W. Fan, M. Miller, S. J. Stolfo, and E. Zadok, "Toward cost-sensitive modeling for intrusion detection and response," *J. Comput. Secur.*, vol. 10, no. 1-2, pp. 5-22, 2002.
- [4] I. Balepin, S. Maltsev, J. Rowe, and K. Levitt, "Using specification-based intrusion detection for automated response," in *Proceedings of RAID*. Springer, 2003, pp. 136-154.
- [5] Y.-S. Wu, B. Foo, Y.-C. Mao, S. Bagchi, and E. H. Spafford, "Automated adaptive intrusion containment in systems of interacting services," *Comput. Netw.*, vol. 51, no. 5, pp. 1334-1360, 2007.
- [6] N. Stakhanova, S. Basu, and J. Wong, "A cost-sensitive model for preemptive intrusion response systems," in *Proceedings of the IEEE AINA*, 2007, pp. 428-435.
- [7] S.-H. Wang, C. H. Tseng, K. N. Levitt, and M. Bishop, "Cost-sensitive intrusion responses for mobile ad hoc networks," in *Proceedings of RAID*, 2007, pp. 127-145.
- [8] N. Stakhanova, S. Basu, and J. Wong, "A taxonomy of intrusion response systems," in *International Journal of Information and Computer Security*, vol. 1, no. 1/2, 2007, pp. 169-184.
- [9] B. Foo, M. W. Glause, G. M. Howard, Y.-S. Wu, S. Bagchi, and E. H. Spafford, *Intrusion Response Systems: A Survey*. Morgan Kaufmann Publishers, 2008, ch. 13, pp. 377-412.
- [10] T. Toth and C. Kruegel, "Evaluating the impact of automated intrusion response mechanisms," in *Proceedings of the ACSAC*. Washington, DC, USA: IEEE Computer Society, 2002, p. 301.
- [11] M. Jahnke, C. Thul, and P. Martini, "Graph based metrics for intrusion response measures in computer networks," in *Proceedings of the IEEE LCN*, 2007, pp. 1035-1042.
- [12] A. Arora, D. Hall, C. A. Pinto, D. Ramsey, and R. Telang, "Measuring the risk-based value of it security solutions," *IT Professional*, vol. 6, no. 6, pp. 35-42, 2004.
- [13] C. Strasburg, N. Stakhanova, S. Basu, and J. S. Wong, "Intrusion response cost assessment methodology," in *Proceedings of the ASIACCS*, 2009.
- [14] A. A. Cárdenas, J. S. Baras, and K. Seamon, "A framework for the evaluation of intrusion detection systems," in *Proceedings of the IEEE S&P*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 63-77.
- [15] G. Gu, P. Fogla, D. Dagon, W. Lee, and B. Skorić, "Measuring intrusion detection capability: an information-theoretic approach," in *Proceedings of ASIACCS*, 2006, pp. 90-101.