Alert Confidence Fusion in Intrusion Detection Systems with Extended Dempster-Shafer Theory

Dong Yu
Department of Computer Science
University of Idaho
+1 (425) 707-9282
dongyu@csds.uidaho.edu

Deborah Frincke
Pacific Northwest National Laboratory
University of Idaho

deborah.frincke@pnl.gov

ABSTRACT
Accurate identification of misuse is a key factor in determining appropriate ways to protect systems. Modern intrusion detection systems often use alerts from different sources such as hosts and sub-networks to determine whether and how to respond to an attack. However, alerts from different locations should not be treated equally. We propose improving and assessing alert accuracy by incorporating an algorithm based on the exponentially weighted Dempster-Shafer (D-S) Theory of Evidence. Our approach uses D-S theory to combine beliefs in certain hypotheses under conditions of uncertainty and ignorance, and allows quantitative measurement of the belief and plausibility in our detection results. Our initial evaluations on the DARPA IDS evaluation data set show that our alert fusion algorithm can improve alert quality over those from Hidden Colored Petri-Net (HCPN) based alert correlation components installed at the demilitarized zone (DMZ) and inside network sites. Due to alert confidence fusion in our example, the detection rate rises from 75% to 93.8%, without adversely affecting the false positive rate.

Categories and Subject Descriptors
C.2.0 [**Computer Systems Organization**]: *Computer-Communication Networks – Security and protection*

General Terms
Algorithms, Experimentation, Security.

Keywords
Alert Confidence Fusion, Intrusion Detection System, Dempster-Shafer Theory of Evidence, Hidden Colored Petri-Net, Alert Correlation.

1. INTRODUCTION
Accurate identification of misuse activity is a key factor in system protection, particularly when appropriate responses need to be determined. Distributed intrusion detection systems (IDS) in combination with other system defenses, such as firewalls, are often the primary means of misuse identification and response. Much of the early IDS research emphasized comprehensive identification (low false negatives) over accuracy (low false positives). However, when one considers implementing automated response in particular, it can be more important to gauge how much confidence to place in a raised alert than to be certain that all possible alerts have been received. For instance, alerts in which we have high confidence are more suitable for driving automated response systems. Our work presents, and performs preliminary assessment of, a technique for improving the quality of IDS alerts drawing on multiple sources of information, through extending the Dempster-Shafer (D-S) Theory of Evidence.

Informally, we refer to an intrusion detection system (IDS) as producing "high quality alerts" if the IDS both avoids false positives (identifying behavior as misuse when it is acceptable) and false negatives (missing misuse behavior).

Methods for improving the quality of IDS alerts have been a subject of considerable study [3][6][11] [12][14][15][18][20]. One technique for enhancing alert quality, particularly reducing false positives, is correlating information about activities to increase confidence that a particular alert set actually signifies a specific event. Correlation approaches may operate in either the time or space domain [6][11][14]. Our Hidden Colored Petri-Net (HCPN) [20] based framework is an example of such an approach. We showed in [20] that HCPN-based alert correlation can significantly reduce both the total number of alerts and the number of false alarms. However, we also determined that *alert confidence fusion* is needed to address the different reliability of alerts from alert analyzers, particularly when these are installed at different sites, which is a typical scenario. As has been indicated elsewhere [9], information from different sites is not equivalent in trustworthiness, and this factor should be incorporated into the data when one assesses confidence in an alert.

We propose a method for performing alert confidence fusion based on the weighted Dempster-Shafer (D-S) Theory of Evidence – our extension to the basic D-S Theory of Evidence. Informally, the basic D-S theory combines confidence scores directly with individual scores tied to the likelihood that observed information supports a given conclusion. In contrast, our extended D-S theory first determines weights based on the sources of our gathered observations and then combines individual confidence scores using these weights. The weightings can incorporate factors such as our level of trust in specific observers, and our belief in the capacity of specific observers to make particular observations. For example, alerts from remote sites are not considered to be as trustworthy as alerts from local sites [9]. If we weight our confidence in remote site observations differently from local site observations, we should therefore be able to produce more accurate results. Our initial observations using

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
43rd ACM Southeast Conference, March 18-20, 2005, Kennesaw, GA, USA. Copyright 2005 ACM 1-59593-059-0/05/0003…$5.00.

Defense Advanced Research Projects Agency (DARPA) IDS evaluation data indicate that this is a fruitful approach.

Clearly, making good decisions about how to weight observers and specific categories of observations is a key element in the success of alert confidence fusion. As one way of making such estimates, we demonstrate that the weights in our extended D-S theory can be estimated by using past data points. We applied this alert confidence fusion model to our HCPN-based alert correlation system and showed further improvement in resolving conflicts between alerts from different alert analyzers and in reducing false alerts.

The paper is organized as follows. In section 2, we introduce the basic Dempster-Shafer Theory of Evidence and its limitations when applied to the alert confidence fusion task in the intrusion detection systems. We extend the D-S theory to resolve the limitations. In section 3, we introduce how the extended D-S theory can be applied to our alert confidence fusion system. Specifically, we discuss the architecture and some practical considerations related to the alert confidence fusion task. We report our experiment results on DARPA IDS evaluation data set in section 4, and conclude in section 5.

2. The Dempster-Shafer Theory of Evidence

In this section, we first introduce the basic Dempster-Shafer Theory of Evidence formally. Then we indicate the limitations of the basic D-S theory when it is applied to the alert confidence fusion task in intrusion detection systems. We address these limitations by proposing an exponentially weighted D-S theory.

2.1 The Basic Dempster-Shafer Theory

The *basic Dempster-Shafer's (D-S) Theory of Evidence* was first formulated by Shafer in 1976 [17]. This theory can be considered as a generalized Bayesian theory. It can be interpreted from either a probabilistic or an axiomatic point of view [13]. The D-S theory has been applied in the fields of statistical inference, diagnostics, risk analysis, and decision analysis.

In a Dempster-Shafer reasoning system, the *frame of discernment* is the set of all possible mutually exclusive and complete facts (or events) $\Theta = \{\theta_i \mid 1 \le i \le N\}$. It consists of all hypotheses for which the information sources can provide evidence. A hypothesis *H* is a subset of *Θ*. Each hypothesis is assigned a value by an observer from the *mass probability function m*, which is defined as:

$$m : 2^{\Theta} \to [0,1] \tag{1}$$

and satisfies the following conditions:

$$m(\phi) = 0 \tag{2}$$

$$m(H) \ge 0, \quad \forall H \subseteq \Theta \tag{3}$$

$$\sum_{H \subseteq \Theta} m(H) = 1 \tag{4}$$

The probability that the evidence supports hypothesis *H* is within the interval defined by the belief and plausibility:

$$[Bel(H), Plaus(H)] \tag{5}$$

where

$$Bel(H) = \sum_{B \subseteq H} m(B) \tag{6}$$

$$Plaus(H) = 1 - \sum_{B \cap H = \phi} m(B) \tag{7}$$

The gap $|Plaus(H) - Bel(H)|$ indicates the ignorance about the hypothesis. Note that the D-S theory calculates the probability that the evidence *supports* a hypothesis rather than the probability of the hypothesis itself. In other words, the probability being computed is tied to confidence that a particular suite of evidence is being interpreted correctly, rather than that some particular hypothesis is correct.

A powerful feature of the D-S theory in a distributed intrusion detection system is its usefulness in combining evidence provided by different observers. Assume we have *K* observers O_i (e.g., sensors or analyzers in the IDS). Each observer provides its own perceived state (or evidence) to a central information processor whose task is to infer the true state of the system by combining the evidences from all observers. Also assume that the observer O_1 believes that hypothesis *H* is true with confidence $m_1(H)$ and O_2 believes that hypothesis *H* is true with confidence $m_2(H)$. The D-S theory provides a rule to combine evidences from independent observers O_1 and O_2 into a single and more informative hint:

$$m_{12}(H) = \frac{\sum_{B \cap C = H} m_1(B) m_2(C)}{\sum_{B \cap C \neq \phi} m_1(B) m_2(C)} \tag{8}$$

A useful property of this combining rule is that it does not rely upon *a priori* probability distributions on the possible system states (which is needed in a Bayesian approach [3]) and so it is useful even if we don't have *a priori* knowledge about the system.

The combining rule (8) can be generalized by iteration. For example, suppose we define $m_{12}(H)$ as the evidence provided by the combined observer O_{12} (O_1 and O_2); we can then treat this evidence as though it came from a single source. By extension, we can combine evidences from any number of observers to produce a single result. This property also means that we can incorporate new evidence and update our beliefs as we acquire new knowledge through observations. In other words, we can apply this not only to multiple observations taken "at once", but also to multiple observations taken from many places over a period of time.

2.2 The Extended Dempster-Shafer Theory

The Dempster-Shafer combining rule (8) implies that we trust observers O_1 and O_2 equally. This assumption normally does not hold in a distributed-sensor intrusion detection system that spans domains. There are several reasons for this. First, information provided by remote sensors and analyzers is considered less trustworthy than that provided by local sensors and analyzers as noted earlier [9]. Second, even identical sensors (and analyzers) installed at different locations may have different detection capabilities since the raw events captured by these sensors are different. Third, different kinds of sensors and analyzers which detect the same type of attack may do so with a different level of accuracy. For example, a sensor may detect misuse activities such as Distributed Denial of Service (DDoS) attacks with 90% accuracy but detect escalation of privilege, such as Local to Root (L2R) attacks with only 50% accuracy. Another, perhaps host-

based, sensor may not detect DDoS at all, but be quite accurate in detecting escalation of privilege. We would therefore expect to place greater confidence in DDoS alerts generated by the first sensor but have less confidence in its reported L2R alerts. In contrast, we would trust more on the L2R alerts generated by the second sensor.

To address the reality that we cannot trust all observers equally, and that a given observer may have different effectiveness in identifying individual misuse types, we extend the D-S theory to incorporate a weighted view of evidence and propose the following modified combining rule:

$$m_{12}(H) = \frac{\sum_{B \cap C = H} \left[m_1(B)\right]^{w_1} \left[m_2(C)\right]^{w_2}}{\sum_{B \cap C \neq \phi} \left[m_1(B)\right]^{w_1} \left[m_2(C)\right]^{w_2}} \quad (9)$$

where w_i is the weight for the observer O_i. When $w_1 = w_2 = 1$, (9) is reduced to the basic D-S combining rule.

The weights can be estimated in several ways. For instance, they can be estimated based on the overall estimation correctness rate in the past and may be further adjusted based on other available information about the source (e.g., local vs. remote) of the information. In our system, we estimate the weights based on the Maximum Entropy principle [6][15] and the Minimum Mean Square Error (MSEE) criteria. We will describe our estimation algorithms in section 3.3.

Wu et al [19] have also proposed a weighted D-S combining rule (for other purposes). Our approach is different from theirs. In their approach, confidence scores are weighted proportionally; while in our approach they are weighted exponentially. With the usage of proportional weights in [19], Eq. (4) no longer holds for the combined evidence.

3. Alert Confidence Fusion

We have just described the exponentially weighted D-S theory to perform alert confidence fusion. We now show how it can be used to support alert confidence fusion in distributed intrusion detection systems.

3.1 Architectural Consideration

Our alert confidence fusion approach can be applied at different levels. For example, we can directly fuse alerts from sensors. However, this simple architecture does not take into account the relationships between different alerts.

A better architecture, albeit more complex, is the one depicted in Fig. 1. Here, alerts are first analyzed using alert correlation components. This will reduce false positives. The outputs of alert correlation components are then fused to further improve the alert quality. There are three advantages of using this architecture.

First, in a sophisticated attack, attackers may perform a series of actions with previous steps preparing for the later ones [6][14][20]. An alert correlation component which uses this prerequisite-consequence relationship can dramatically reduce the number of false alerts [20].

Second, a major challenge for applying alert confidence fusion is to determine the mass probability function over the observed alarms. We will show in section 3.2 that our alert correlation component can provide the probabilities naturally.

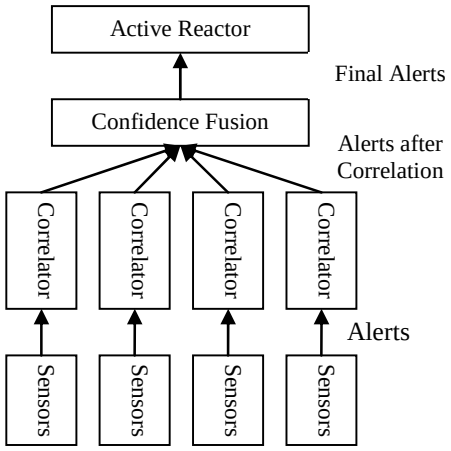

Fig. 1: Architecture used in our alert confidence fusion system. Alerts from different set of sensors are first correlated to reduce the number of total alerts and to improve the quality of the alerts. Alerts after correlation are then fed into the confidence fusion component.

Third, we have shown [20] that although our original alert correlation module could improve the quality of alerts, it is necessary to have an approach to combining alerts from alert analyzers at different locations. Our exponentially weighted D-S theory fulfills this requirement.

3.2 The Alert Correlation Component

We used the Hidden Colored Petri-Net (HCPN) [20] based alert correlation component in our system. HCPN is our extension to Colored Petri-Net. In HCPN, each token element is associated with a probability (or confidence), and observations are explicitly separated from transitions. In HCPN, colors represent agents; places represent resources; observations represent alerts; transitions represent actions; input arcs represent prerequisites of the action; and output arcs represent consequences of the action as shown in Fig. 2. The model is called "hidden" because the transitions are not directly observable. Instead, each transition has a probability to omit an observation. The true state transitions can be inferred through the observations only.

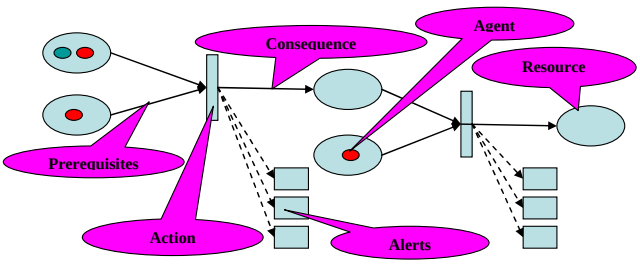

Fig. 2: Hidden Colored Petri-Net. In HCPN, colors represent agents; places represent resources; observations represent alerts; transitions represent actions; input arcs represent prerequisites of the action; and output arcs represent consequences of the action.

HCPN is suitable for correlating alerts where attackers' true actions are unknown and we may infer their actions only based on partial observations as embodied in alerts. We can indicate the situation in Fig. 2. Here, the HCPN model output gives the

probability that a resource has been undertaken by an attacker. This probability can be directly used in the alert confidence fusion component as the mass probability function.

3.3 The Alert Confidence Fusion Technique

The alert confidence fusion technique used in our system is based on our extended Dempster-Shafer Theory of Evidence as described earlier. Since we are interested in whether an alert A_k is true (a true positive) or false (a false positive), the *frame of discernment* is $\Theta_k = \{A_k, \neg A_k\}$. The possible hypotheses are:

$$2^{\Theta_k} = \{\phi, \{A_k\}, \{\neg A_k\}, \{A_k, \neg A_k\}\} \tag{10}$$

It is clear that:

$$m_i(\{A_k, \neg A_k\}) = 0 \tag{11}$$

$$m_i(\{\neg A_k\}) = 1 - m_i(\{A_k\}) \tag{12}$$

The combining rule then becomes:

$$m_{12}(\{A_k\}) = \frac{P(\{A_k\})}{P(\{A_k\}) + P(\{\neg A_k\})} \tag{13}$$

$$m_{12}(\{\neg A_k\}) = \frac{P(\{\neg A_k\})}{P(\{A_k\}) + P(\{\neg A_k\})} \tag{14}$$

where

$$P(X) = \left[m_1(X)\right]^{w_1^k} \left[m_2(X)\right]^{w_2^k} \tag{15}$$

We use the probability from the *i*th HCPN alert correlator as the mass probability function $m_i(A_k)$.

Now we illustrate the limitation of the basic Dempster-Shafer theory with a concrete example. Assume we have two observers. Observer O_1 reports $m_1(\{A_k\}) = 0.02$ and observer O_2 two reports $m_2(\{A_k\}) = 0.7$. According to the basic D-S theory (i.e., $w_1^k = w_2^k = 1$ in (15)),

$$m_{12}(\{A_k\}) = \frac{0.02 \bullet 0.7}{0.02 \bullet 0.7 + 0.98 \bullet 0.3} = 0.05 \tag{16}$$

If observer O_1 is very bad at detecting attack A_k, and observer O_2 detects A_k with high accuracy, we can see that the result in (16) does not give us the correct view of the system state, an indication that the basic D-S combining rule is not suitable for combining the results of the two observers under this condition. Actually, we should weight the report from O_2 more than that from O_1. Assume $w_1^k = 0.2$ and $w_2^k = 1$, then,

$$m_{12}(\{A_k\}) = \frac{0.02^{0.2} \bullet 0.7^1}{0.02^{0.2} \bullet 0.7^1 + 0.98^{0.2} \bullet 0.3^1} = 0.52 \tag{17}$$

which is a better combined view of the system.

The key to the successful use of our approach is the proper calibration of the weights. In our prototype system, we have used both the Maximum Entropy (MaxEnt) [6][15] based approach and the Minimum Mean Square Error (MMSE) based approach to estimating the weights.

To estimate the weights with the MaxEnt principle, we rewrite (13) and (14) to be:

$$m_{12}(H \mid t_i) = \frac{1}{N_k} \exp\left(w_1^k f_1(H, t_i) + w_2^k f_2(H, t_i)\right) \tag{18}$$

where t_i indicates the time,

$$f_i(H, t_i) = \log\left(m_i(H \mid t_i)\right) \tag{19}$$

and N_k is the normalization factor so that (4) holds:

$$N_k = \sum_{H \in \{\{A_k\}, \{\neg A_k\}\}} \exp\left(w_1^k f_1(H, t_i) + w_2^k f_2(H, t_i)\right) \tag{20}$$

(18) can be interpreted as combining features $\log\left(m_i(H \mid t_i)\right)$ with the standard exponential model. The *Generalized Iterative Scaling* (GIS) algorithm [8] and *Improved Iterative Scaling* (IIS) algorithm [8] thus can be used to estimate the weights. These algorithms have been shown to be able to minimize the Kullback-Leibler (K-L) distance between $m_{12}(H)$ and the empirical distribution $\tilde{p}(H)$ of a set of training samples $s_i\,(i = 1, \cdots, N)$:

$$D(\tilde{p} \,\|\, m_{12}) = \sum_H \tilde{p}(H) \log \frac{\tilde{p}(H)}{m_{12}(H)} \tag{21}$$

where $\tilde{p}(H)$ is defined as:

$$\tilde{p}(H) = \frac{c(H)}{N} \tag{22}$$

and $c(H)$ is the number of times hypothesis H is true in the training samples.

An alternative way to estimating weights is using the criteria of Minimum Mean Square Error (MMSE). The Mean Square Error (MSE) is defined as:

$$MSE = \frac{\sum_{j=1}^{N} \left(m_{c,j}(\{A_k\}) - r_j\right)^2}{N} \tag{23}$$

where N is the total number of samples in the training set; $m_{c,j}$ is the jth combined observation score; and r_j is the real (or true) value of the jth observation:

$$r_j = \begin{cases} 1 & \text{if } A_k \text{ is true} \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

We used the standard gradient descending algorithm to search for the best weights:

$$w_i^{t+1} = w_i^t - \lambda \frac{\partial MSE}{\partial w_i} \tag{25}$$

where λ is the step size.

The MMSE based approach provides us the same result as the MaxEnt based approach in our alert analysis experiments but outperforms the MaxEnt based approach in a more broad set of experiments.

Note that for each alert type A_k, there is a pair of weights w_i^k. In other words, if observer O_1 is good at detecting alert A_k but is not good at detecting alert A_j, w_1^k is large but w_1^j is small.

There are two practical considerations in our implementation:

First, when $m_i(H)$ is close to 0, $\log(m_i(H))$ might be very small. This may result in computational underflow. To prevent underflow from happening, all the probabilities in the system have a floor value of 0.01. If the confidence of a hypothesis is less than 0.01, we set it to be 0.01.

Second, an HCPN alert correlator reports to the alert confidence fusion component only if the perceived state of a resource has been changed, so alert correlators send the state change information asynchronously. To deal with this problem, we store every correlator's last reported evidence. When a new report regarding alert A_k comes in, we update the corresponding evidence and re-estimate the combined confidence for the alert A_k. We thus gauge confidence based on evidence "in hand", and increase or reduce confidence when new information arrives.

4. Experimental Results

We have performed off-line experiments using two DARPA 2000 DDoS intrusion detection evaluation data sets. These are not intended to be conclusive examinations of the efficacy of our technique – but rather provide some sense of how well our approach works.

In both data sets, an attacker probes and obtains access to the internal systems, installs a DDoS daemon, and launches a DDoS attack against an off-site server. Each dataset includes the network traffic data collected from both the demilitarized zone (DMZ) and the inside part of the evaluation network. Alerts are generated by RealSecure Network Sensor 6.0 with the maximum coverage policy to force the network sensor to save all the reported alerts. We use RealSecure network sensors because attack signatures are well documented.

We first trained the HCPN-based alert correlators as in [20], then trained the confidence fusion weights based on the outputs from the alert correlators. The test was performed on the first dataset. We used two alert correlators in the system, one for the DMZ traffic and one for the inside network traffic. The correlation results from these correlators were combined using our alert confidence fusion algorithm.

Table 1 shows the detection and false positive rates for RealSecure Network Sensor 6.0. The top and bottom part of Table 2 show the results of our HCPN-based system with and without alert confidence fusion, respectively. We separated them into two tables because we want to compare the performance of the system with and without our alert confidence fusion approach.

Before further discussion, we note that the results listed in the top half of Table 2 differ from those in [20] due to the way we count the alerts. In [20], we emphasize the performance of individual HCPN-based alert analyzers and the number of observable alerts is limited to individual sites. In Table 2, we report the number of observable alerts from all the sources (e.g., DMZ and inside network).

In our experiments, the detection rate (DR) is defined as:

$$DR = \frac{\text{Number of True Attacks Reported}}{\text{Number of Total Observable Attacks}} \quad (26)$$

The False Positive Rate (FPR) is defined as:

$$FPR = 1 - \frac{\text{Number of True Alerts Reported}}{\text{Number of Alerts Reported}} \quad (27)$$

Table 1

Detection and False Positive Rates for RealSecure Network Sensor 6.0

Setting	NOA	NA	NTAD	DR	NRA	FPR
DMZ	89	891	51	57.30%	57	93.60%
Inside	60	922	37	61.67%	44	95.23%

NOA=Number of Observable Attacks,
NA=Number of Alerts,
NTAD=Number of True Attacks Detected,
DR=Detection Rate,
NRA=Number of Real Alerts,
FPR=False Positive Rate

Table 2

Detection and False Positive Rates of Our HCPN-based System with and without Alert Confidence Fusion

Setting	NOA	NA	NTAD	DR	NRA	FPR
DMZ	16	15	12	75.0%	12	20%
Inside	16	12	12	75.0%	12	0%
Basic D-S	16	9	15	56.3%	9	0%
Ext. D-S	16	15	15	93.8%	15	0%

Acronyms carry the same meaning as that in Table 1.

Table 3

Alerts Received by the Alert Confidence Fusion Component

		Total	From Both	From DMZ	From Inside
Input	All	18	9	6	3
	True	15	9	3	3
	False	3	0	3	0
Output	Basic D-S	9	9	0	0
	Ext. D-S	15	9	3	3

From Table 1 and the top part of Table 2, we see that the number of alerts and false positive rates are dramatically reduced by using the HCPN-based alert analysis component. The bottom part of Table 2 shows that our alert confidence fusion algorithm (Extended D-S) further increases the detection rate while keeping the false positive rate down by combining information from DMZ and inside networks. Without using alert confidence fusion, the detection rate for both analyzers is around 75%, and the false positive rate at the DMZ site is 20%. Using extended D-S algorithm based alert confidence fusion, we increase the detection rate to 93.8% while reducing the false positive rate to 0% under the experiment setting. Note, however, if only the basic D-S combination algorithm is used, the detection rate *decreases*. The extended D-S algorithm provides more than 30% absolute detection rate against the basic D-S algorithm in our experiments.

Table 3 shows details of the alerts processed by the alert confidence fusion component. The alert confidence fusion component receives 18 alerts from lower level alert analyzers; 9 are reported by both the correlators installed at the DMZ and inside network sites; 6 alerts are from the DMZ alert analyzer

only; 3 alerts are from the inside network analyzer only. The exponentially weighted D-S algorithm correctly combines the alerts with the estimated weights.

5. Conclusion

We expanded the HCPN-based alert correlation and understanding system by incorporating our novel alert confidence fusion component. The alert confidence fusion algorithm used in the system is derived from the exponentially weighted D-S theory – our extension to the basic D-S theory by weighing hypothesis confidence scores from different sources. Our experiments on the DARPA intrusion data set show that our alert confidence fusion model can potentially resolve contradictory information reported by different analyzers, and further improve the detection rate and reduce the false positive rate. This is particularly important in situations where the alerts will be used to drive either automated responses, or where they will be used as the primary basis for a system administrator's decisions on how to defend a system from a perceived attack.

As its main advantage, our approach has the ability to quantify relative confidence in different alerts. We can use the modeling power of D-S theory in expressing beliefs in some hypotheses (alert diagnosis), the ability to describe uncertainty and ignorance in the system, and the quantitative measurement of belief and plausibility in our detection results. Our ongoing efforts will examine this technique in greater depth.

6. ACKNOWLEDGMENTS

We thank our colleagues in the department of computer science at University of Idaho for valuable discussions.

7. REFERENCES

[1] J. Allen, A. Christie, W. Fithen, J. McHugh, J. Pickel, and E. Stoner, *State of the Practice of Intrusion Detection Technologies*, *Technical Report CMU/SEI-99-TR-028*, 1999

[2] J.P Anderson, *Computer Security Threat Monitoring and Surveillance*, Technical report, James P Anderson Co., Fort Washington, Pennsylvania, April 1980.

[3] J. Burroughs, L. F. Wilson and George V. *Analysis of Distributed Intrusion Detection Systems Using Bayesian Methods*, presented at IPCCC 2002, April 2002.

[4] T. Bass, *Intrusion detection systems and multisensor data fusion*, Communications of the ACM, v.43 n.4, p.99-105, April 2000

[5] V. Berk, R. Gray, and G. Bakos, *Using sensor networks and data fusion for early detection of active worms*, Proc. of 2003 SPIE Aerosense Conference, Orlando, FL, April, 2003.

[6] A. Berger, S. Della Pietra, & V. Della Pietra, *A Maximum Entropy Approach to Natural Language Processing*, Computational Linguistics, 22(1):39–71, 1996.

[7] F. Cuppens, A. Miège, *Alert Correlation in a Cooperative Intrusion Detection Framework*, 2002 IEEE Symposium on Security and Privacy, May 12 - 15, 2002

[8] J. N. Darroch, & D. Ratcliff, *Generalized iterative scaling for log-linear models*, The Annals of Mathematical Statistics, 43(5), 1470–1480, 1972.

[9] D. Frincke, *Balancing Cooperation and Risk in Intrusion Detection*, ACM Transactions on Information and System Security (TISSEC) 3(1): 1-29 (2000).

[10] D. L. Hall, *Mathematical Techniques in Multisensor Data Fusion*, Artech House, Inc., Norwood, MA, 1992

[11] M.-Y. Huang, and T. M. Wicks, *A Large-scale Distributed Intrusion Detection Framework Based on Attack Strategy Analysis*, Web proceedings of the First International Workshop on Recent Advances in Intrusion Detection (RAID'98), 1998.

[12] G. Jiang and G. Cybenko, *Temporal and Spatial Distributed Event Correlation for Network Security*, 2004 American Control Conference, Boston, June 30 - July 3.

[13] J. Kohlas and P. Monney. *Theory of evidence - a survey of its mathematical foundations, applications and computational analysis*. ZOR- Mathematical Methods of Operations Research, 39:35--68, 1994.

[14] P. Ning, Y. Cui, D. S. Reeves, *Analyzing Intensive Intrusion Alerts Via Correlation*, in Proceedings of the 5th International Symposium on Recent Advances in Intrusion Detection (RAID 2002), LNCS 2516, pages 74--94, October 2002

[15] R. Rosenfeld, *A Maximum Entropy Approach to Adaptive Statistical Language Modeling*, Computer, Speech, and Language, 10, 1996.

[16] C. Siaterlis, and B. Maglaris, *Towards multisensor data fusion for DoS detection*, Proceedings of the 2004 ACM symposium on Applied computing, 2004.

[17] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.

[18] H. Svensson and A. Jøsang, *Correlation of Intrusion Alarms with Subjective Logic*. In the proceedings of the sixth Nordic Workshop on Secure IT systems, Copenhagen, Denmark, 1-2 November, 2001

[19] H. Wu, M. Siegel, R. Stiefelhagen, and J. Yang. *Sensor fusion using Dempster-Shafer theory*. In Proceedings of IEEE Instrumentation and Measurement Technology Conference, Anchorage, AK, USA, 2002.

[20] D. Yu, D. Frincke, *A Novel Framework for Alert Correlation and Understanding*, Springer's LNCS series, vol 3089. International Conference on Applied Cryptography and Network Security (ACNS) 2004.

[21] S. A. Yemini, S. Kliger, E. Mozes, Y. Yemini and D. Ohsie, *High speed and robust event correlation*, IEEE Communications Magazine, May, 1996.