Poster: Prioritizing Intrusion Analysis Using Dempster-Shafer Theory

students: Loai Zomlot, Sathya Chandran Sundaramurthy, *faculty*: Xinming Ou
Dept. of Computing and Information Sciences - Kansas State University
Manhattan, KS, USA - Email: lzomlot@ksu.edu

S. Raj Rajagopalan
HP Labs - Princeton, NJ, USA

***Abstract*—Intrusion analysis, *i.e.* the process of combing through IDS alerts and audit logs to identify true successful and attempted attacks, remains a difficult problem in practical network security defense. The major root cause of this problem is the large rate of false positives in the sensors used by IDS systems to detect malicious activities. This work presents an approach to handling such uncertainty through the Dempster-Shafer (DS) theory that uses a generalization of probabilities called beliefs to characterize confidence in evidence in support of a given hypothesis. We address a number of practical but fundamental issues in applying DS to intrusion analysis, including how to model sensors' trustworthiness, where to obtain such parameters, and how to address the lack of independence among alerts. We present an efficient algorithm for computing a belief score for a given hypothesis, *e.g.* a specific machine is compromised. The belief strength can be used to prioritize further analysis by a human analyst of the hypotheses and the associated evidence. We have implemented our approach for the open-source IDS system Snort and evaluated its effectiveness on a number of data sets as well as a production network. The verification of belief scores showed that it can be effective in taming the high false positive rate problem in intrusion analysis.**

I. Introduction

Intrusion analysis is the process of examining real-time events such as IDS alerts and audit logs to identify and confirm successful attacks and attack attempts into computer systems. The IDS sensors that we have to rely on for this purpose often suffer from a large false positive rate. It then becomes the responsibility of a human monitoring the IDS system to distinguish the true alarms from the enormous number of false ones. How to deal with the prevalence of false positives is the primary challenge in making IDS sensors useful, as pointed out by Axelsson [1] more than 10 years ago. Due to the lack of effective techniques to handle the false-positive problem, it has become a common practice to altogether disable IDS signatures that tend to trigger large amount of false positive. Turning off IDS signatures is like turning a blind eye to attack possibilities, which we believe is a dilemma due to the lack of effective techniques to *prioritize* investigating intrusions from the large amount of IDS alerts and audit logs.

There have been past attempts [9, 10] at prioritizing IDS alerts based on their trustworthiness – Bayesian analysis [5] has been the standard and there have been some approaches using alternative theories such as Dempster-Shafer theory [7]. However, a number of *fundamental issues* in applying these mathematical theories to intrusion analysis remain to be addressed. For Bayesian analysis, it seems difficult to establish adequate priors or determine the probability parameters in a robust manner. For Dempster-Shafer theory, it is not clear how to model sensor quality, where to obtain such parameters, and how to handle non-independent sources of evidence.

Our investigation reveals that Dempster-Shafer theory has its unique advantages in handling uncertainty in intrusion analysis, namely, the lack of need for specifying prior probabilities of all events and the ability to combine beliefs from multiple sources of evidence [2, 3, 9]. In this work we present an extended Dempster-Shafer model that addresses the fundamental issues in applying DS in intrusion analysis. We have implemented our method on top of an existing IDS alert correlation tool, so that one can calculate a numeric confidence score for each derived hypothesis and prioritize the results based on the scores.

II. Background on Dempster-Shafer Theory

A common example to illustrate the difference between probability theory and Dempster-Shafer theory is that if we toss a coin with an unknown bias, probability will still assign 50% for Head and 50% for Tail by the principle of indifference. Dempster-Shafer theory, on the other hand, handles this by assigning 0% belief to $\{Head\}$ and $\{Tail\}$ and assigning 100% belief to the *set* $\{Head, Tail\}$, meaning "either Head or Tail". More generally, the DS approach allows for three kinds of answers: *Yes, No, or Don't know*, the last option of allowing ignorance makes a big difference in evidential reasoning [4]. In DS theory, a set of disjoint hypotheses of interest, *e.g.*, $\{attack, no\text{-}attack\}$, is called a *frame of discernment*. The *basic probability assignment* (*bpa* function), distributes the belief over the *power set* of the frame of discernment and is defined as:

$$m_\theta : 2^\theta \rightarrow [0, 1] \tag{1}$$

Definition 1. *Let θ be a frame of discernment and m_θ a bpa function. The belief function is defined as*

$$For\ x \subseteq \theta\ \ Bel(x) = \sum_{y \subseteq x} m_\theta(y) \tag{2}$$

The belief function shows how much confidence we have in that one of the hypotheses contained in x holds (without specifying which). Dempster-Shafer has a combination method, the goal of which is to combine evidence for a hypothesis from multiple *independent* sources and calculate an overall belief for the hypothesis [6]. In general we have the following rule of combination known as the Dempster Rule.

$$m_{1,2}(h) = \frac{1}{1-K} \cdot \sum_{h_1 \cap h_2 = h} m_1(h_1) \cdot m_2(h_2) \tag{3}$$

$$K = \sum_{h_1 \cap h_2 = \{\}} m_1(h_1) \cdot m_2(h_2) \tag{4}$$

III. Approaches

A. Using "unknown" to capture sensor quality

Dempster-Shafer theory allows specifying a weight on "unknown" rather than specifying precise probabilities for every possible event in the space. We use this ability to represent lack of knowledge to capture the intuitive notion of IDS sensor quality (which usually turns out to be imprecisely described), without suffering the non-intuitive effects of aggregation that have been observed by researchers [9].

The nature of *unknown* matches naturally with how humans interpret IDS alerts. When an alert is fired, we will have some degree (say 10%) of belief that an attack is going on. But we do not have 90% belief that an attack is *not* going on. Positively asserting that an attack is not going on after seeing an alert is counter-intuitive. Adopting the simple *true* and *false* case to capture the information provided by an alert would require us to know the prior probability of attack, which is hard if not impossible to obtain. By using DS, we can assign 0.1 belief to "attack"($\{true\}$), 0 belief to "no-attack" ($\{false\}$), and the 0.9 goes to "Don't know" ($\{true, false\}$). Another consequence of this model of sensor quality is that there will be no conflict among alerts. When we do not trust an alert, we just say "Don't know" whether the hypothesis is true, rather than assert that the hypothesis is false. This will not contradict the fact that we may trust another alert which derives the same hypothesis being true.

B. Accounting for lack of independence among alerts

A long-standing assumption in DS theory is that multiple pieces of evidence are independent, which is a property that is hard to confirm in

practice. This is especially a problem in IDS alerts since many alerts are triggered by the same or similar signatures. In combining these alerts to derive the overall belief on the attack status, it is important that such non-independence be appropriately accounted for so that the result is not skewed by over-counting. To the best of our knowledge, our method is the first in applying sound non-independent DS belief combination in IDS alerts.

We adopt an idea proposed by Shafer [8] which interprets combined bpa's as joint probabilities. Based on this, we develop a set of customized combination formulas to correctly account for the dependence in evidence when combining beliefs in the alert correlation graph. For non-independent evidence, multiplication of bpa's from two sources is no longer valid [8]. Instead of $m_1(h_1) \cdot m_2(h_2)$, we use $\psi[h1, h2]$ to denote the joint bpa of the two sources. We obtain the following new formula for combining possibly non-independent evidence.

$$m_{1,2}(h) = \sum_{h_1 \cap h_2 = h} \psi[h1, h2] \tag{5}$$

In our system, the only possible h_i's are $\{true\}$ (referred to as t hereafter) and $\{true, false\}$ (referred to as θ hereafter). The following equations calculate $\psi[h_1, h_2]$, where r_i is an *overlapping factor* that can be estimated from the sources that support two inference paths

$$\psi[t, t] = r_1 \cdot m_1(t) + (1 - r_1) \cdot m_1(t) \cdot m_2(t) \tag{6}$$
$$\psi[t, \theta] = (1 - r_1) \cdot m_1(t) \cdot m_2(\theta) \tag{7}$$
$$\psi[\theta, t] = (1 - r_2) \cdot m_1(\theta) \cdot m_2(t) \tag{8}$$
$$\psi[\theta, \theta] = r_1 \cdot m_2(\theta) + (1 - r_1) \cdot m_1(\theta) \cdot m_2(\theta) \tag{9}$$

C. Efficient calculation

A direct application of DS formulas can result in exponential (in the number of hypotheses – in our case, IP addresses) blow-up of belief combinations. We adopt a "*translate-then-combine*" approach so that beliefs are propagated in a correlation graph and only combined at join points in the graph. This produces an efficient algorithm with worst-case running time quadratic in the number of IP addresses in the input alerts.

D. Linking to practical IDS tools

We have implemented our approach on the open-source IDS system Snort, and evaluated it continuously on our departmental network. Also we tested our prototype on *Lincoln Lab DARPA intrusion detection evaluation(98,99)* data sets. The objective of our evaluation is to examine whether the belief values calculated from our DS algorithm can help a security analyst to prioritize further investigation. To that end, we assign to an IDS alert a belief value which is the highest belief of the hypothesis it supports. Moreover, to show that it is indeed the application of customized Dempster-Shafer theory helps in the prioritization, we compare the performance of our DS algorithm against alternative methods. These methods are using sensor quality metrics only, the maximum sensor-quality metric in a correlation graph as the belief value for all alerts in the graph, and the belief values calculated from the standard DS rule of combination, instead of from our customized DS.

We used the truth files included in the datasets to determine which alerts are true alerts and which are false alerts, and compare this against the classification provided by the belief values. Our evaluation suggests that the scores computed from our algorithm provide an effective ranking for the correlated alerts based on the correlations' trustworthiness.

1) ROC curve Analysis: The ROC curve for one of the datasets is shown in figure 1. From the curves it is clear that our customized DS algorithm outperforms the other three alternative methods.

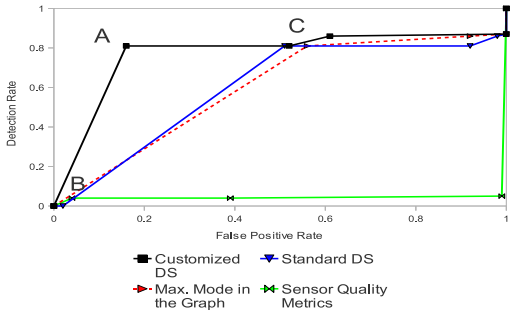

Figure 1. Lincoln Lab 1999 ROC Curves

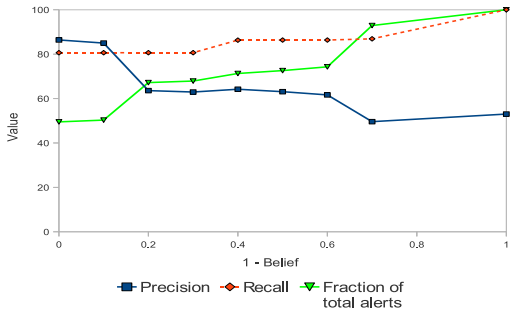

Figure 2. Prioritizing Effect (Lincoln Lab 1999)

2) Prioritization Effect: Figure 2 shows how the precision and recall change when the threshold decreases from 1 to 0 (note the X axis is 1-Belief). When one starts with alerts with high beliefs, the precision is high meaning more of the effort is devoted to useful tasks.

3) Sensitivity Analysis: We also did experiments to test how the variation in the sensor quality metrics, which are input to our algorithm, affect our algorithm's performance. We compare the results from multiple cases along with the default case in the ROC curves for both datasets. The results showed that our system is not sensitive for such changes.

IV. Acknowledgment

This material is based upon work supported by U.S. National Science Foundation under grant no. 1038366 and 1018703, AFOSR under Award No. FA9550-09-1-0138, and HP Labs Innovation Research Program. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, AFOSR, or Hewlett-Packard Development Company, L.P.

References

[1] Stefan Axelsson. The base-rate fallacy and the difficulty of intrusion detection. *ACM Trans. Inf. Syst. Secur.*, 3(3):186–205, 2000.
[2] Qi Chen and Uwe Aickelin. Anomaly detection using the Dempster-Shafer method. In *International Conference on Data Mining (DMIN2006)*, 2006.
[3] Thomas M. Chen and Varadharajan Venkataramanan. Dempster-Shafer theory for intrusion detection in ad hoc networks. *IEEE Internet Computing*, 2005.
[4] Joseph Y. Halpern. *Reasoning about uncertainty*. The MIT Press, 2005.
[5] Finn V. Jensen and Thomas D. Nielsen. *Bayesian Networks and Decision Graphs*. Springer Verlag, 2007.
[6] K. Sentz and S. Ferson. Combination of evidence in Dempster-Shafer theory. Technical report, Sandia National Laboratories, Albuquerque, New Mexico., 2002.
[7] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
[8] Glenn Shafer. Probability judgment in artificial intelligence and expert systems. *Statistical Science*, 2(1), 1987.
[9] Dong Yu and Deborah Frincke. Alert confidence fusion in intrusion detection systems with extended Dempster-Shafer theory. In *43rd ACM Southeast Conference*, Kennesaw, GA, USA, 2005.
[10] Yan Zhai, Peng Ning, Purush Iyer, and Douglas S. Reeves. Reasoning about complementary intrusion evidence. In *Proceedings of 20th Annual Computer Security Applications Conference (ACSAC)*, pages 39–48, December 2004.