Data Fusion for Improved Situational Understanding

Ambareen Siraj
Department of Computer Science
Tennessee Tech University
asiraj@tntech.edu

ABSTRACT

Situational understanding is a difficult problem in cybersecurity where the challenge lies in effective integration of evidence reported by multiple sources or sensors monitoring cyberspace. While sensor diversity contributes to better coverage of attack space, it makes analysis of diverse data more difficult. Our work employs a possibilistic approach to intelligent fusion of data from multiple sources to better understand security situation of information systems. This paper summarizes our previous work in this area and presents goal of our ongoing work.

Categories and Subject Descriptors
K.6.m [Miscellaneous]: Security

General Terms
Security, Reliability

Keywords
Network security, Intrusion detection systems, Sensor fusion, possibilistic approach

1. INTRODUCTION

In 2005, the InfoSec Research Council (IRC) identified situational understanding and attack attribution as one of the eight “hard-questions” areas in cybersecurity [4]. The report describes the need for methodologies that would aid to achieve “common operational understanding from different vantage points” for improved situational awareness. This paper describes our effort in that direction where we employ a possibilistic approach to intelligent fusion of data from multiple sources to better understand security status of information systems.

Through the notion of fuzzy sets, possibilistic theory makes it feasible to deal with uncertainty in terms of vagueness and approximation [12]. Possibilistic approaches are different from probabilistic approaches as the former deals with vague but coherent knowledge and later deals with precise but varied knowledge [6]. While a probability distribution describes the probability that a given variable is of certain value, possibility distribution describes the possible value of the certain variable [11]. In other words, in our context, while probabilistic approaches can be used to answer the question how likely is it that the system is under attack, possibilistic approach can be used to answer the question to what extent is the system under attack (1 being completely possible and 0 being impossible)?

We address situational understanding of the security status by accommodating uncertainty in terms of impreciseness and vagueness of information. Besides lack of solutions in this area using possibilistic approaches and the fact that probabilistic approaches requires extensive information from experts, we believe security status of information systems is not a binary metric rather it must accommodate the uncertain attributes inherent in the dynamic cyberspace to some extent. This calls for a different approach than reasoning under certainty where the world is viewed as exact and certain.

2. APPROACH

In an intrusion detection environment, different types of sensor fusion techniques can be utilized for better situational understanding. These techniques primarily vary in terms of their mission objectives. Some prioritize alerts for alert reduction, some correlate alerts to identify common attack patterns, and some cluster alerts to identify multi-staged attacks. Each of these tasks has its own merits. In [9], we particularly address the problem of fusing results of alert clustering and alert correlation for the determination of systems’ overall security health using a possibilistic approach.

We refer to intelligent alert fusion as the process of interpretation, combination and analysis of alerts reported by sensors to determine and provide a quantitative assessment of systems’ security status, which is representative of the degree of concern in the system [7, 8]. As Debar and Wespi point out in [2] that alerts may not seem significant when they are isolated, but their importance may intensify when association(s) can be discovered among them. While structural association is useful for identifying common patterns in attacks, causal association is useful for identifying alerts that can be attributed to multi-staged attacks.

Therefore we employ a unified alert fusion model provides an overall condensed view of the network by assessing the health of the resources in the network [9]. The fusion model is resource centric, i.e., all analyses are centered upon the resources in the system. The input to the unified alert fusion model comes from the low-level intrusion detection systems (IDS)s that are treated as sensors. The model incorporates the following (Figure 1):

- Prioritize alerts;
- Identify alert associations by:
  - clustering similar alerts; and
  - correlating related alerts;
- Assess the overall security situation by fusing results.

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, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CSIIRW ’09, April 13-15, Oak Ridge, Tennessee, USA
Copyright © 2009 ACM 978-1-60558-518-5 ... $5.00
There are three major steps in the dynamic fusion process [9] - shown in Figure 2.0. The first step is the fuzzy inference process for the possibility distribution generation of the inputs where the crisp inputs denoting IAS and CAS are transformed into possibility distributions. The possibility distributions map the given domain of IAS and CAS into an internal of [0,1]. The inputs are fuzzified and aggregated and at the end of the fuzzy inference step, possibility distributions of Degree of Concerns (DOCs) are obtained. The actual fusion takes place in the next step where the fuzzy DOCs are combined into a new possibility distribution for overall degree of concern (ODOC). Therefore,

$$\Pi_{ODC} = \Pi_{IA} \otimes \Pi_{CAS}$$

where $\Pi_{IA}$ represents the possibility distribution of the degree of concern conveyed by IAS (DOC-IAS), $\Pi_{CAS}$ represents the possibility distribution of the degree of concern conveyed by CAS (DOC-CAS), $\Pi_{ODC}$ represents the global possibility distribution of ODOC and $\otimes$ represent a possibilistic information combination operator. In possibility theory, there are many choices for the $\otimes$ operator with different behavior properties [1]. Among them, we used context dependent operators so that fusion is not only dependent on the inputs but also on some external context [9]. As such context, agreement or consensus between the inputs is used so that the fusion process is sensitive to agreement/disagreement between sources. When two inputs totally agree, there is no conflict between them. If they don’t, there can be weak or strong conflict between them depending on the degree of agreement. When the two inputs totally disagree, there is total conflict between them. Use of conflict in information fusion has been proposed by Dubois and Prade in [3] and used by Rokos in [6] for Satellite Image classification. Using conflict (or degree of agreement) allows the fusion process to behave in one way when the inputs agree and in another way when they don’t. Readers are courteously referred to [9] for a full account of this technique.

For experimentation, MIT Lincoln’s Lab’s DARPA (LLD) 2000 Intrusion Detection Evaluation (IDEVAL) Scenario Specific dataset [5] was used. To generate the multi-sensor alert report, Snort and RealSecure intrusion detection systems were configured with different security policy to monitor the same attack traffic simultaneously and report results in a common repository. Following are two examples of possibilistic fusion when sources are in agreement and when they are not.

In the case of LLDOS 2.0.2 Inside Zone dataset, for one of the actual victim host mill: 172.016.115.020, incident association was reported as 80.86% (IAS) and cluster association was reported as 79.0% (CAS). Combining these resulted in an overall degree of concern of 83.26%, relating to a SEVERE level of concern for the host. Figure 3 shows the possibility distribution of the inputs and the output of the dynamic fusion process in this case.
Both of the possibility distributions of degree of concern related by IAS (DOC-IAS) and degree of concern related by CAS (DOC-CAS) consist of High fuzzy sets and show weak conflict between them. As expected, the combined output fuzzy set shows additive behavior until the level of consensus (0.634) is passed and conflict occurs. After that output fuzzy set follows compromised behavior. The arrow points to the defuzzified output value at the lower end of the Medium region.

In the case of LLDOS 1.0 Inside Zone dataset, for one of the host crow: 172.016.113.148, there was no report of incident association and cluster association was reported as 80.86% (CAS). This is a host which was not compromised but for which there were evidence of some form of anomalous activities present in the sensor report that resulted in suspicious clusters generated. The fusion process resulted in an overall degree of concern of 45.78%, relating to a CAUTIOUS level of concern for the host. Figure 4 shows the possibility distribution of the inputs and the output of the dynamic fusion process in this case.

The results of situation assessment experiment showed that when there was strong agreement between the results reported by abstract alert correlation and multi-level alert clustering, the combined overall concern increased. In contrast, whenever there were disagreements between the results of incident and cluster associations, the combined overall concern decreased in accordance with the extent of the conflict between the two and undertook a compromised assessment.

While the work described in the previous section focus on decision level fusion of sensor techniques analyzing alert correlation and alert clustering, in [10] we describe a sensor corroboration technique with data level fusion that we use for anomaly sensors. For anomaly sensors that are not able to indicate the specific nature of the attacks, the overall degree of concern for each resource in the network primarily depends on the extent of the anomaly reported for the particular resource. As anomaly sensors often suffer from false positives where legitimate deviation from normalcy is also mistakenly flagged as an intrusion, it makes sense to somehow substantiate the reports from anomaly sensors with some additional evidence in order to better understand the security situation.

In this regard, we used sensor corroboration which makes use of primary and secondary sensors for complementary evidence support [10]. Anomaly sensors are employed as primary sensors to monitor systems’ security status. This type of primary sensor reports an anomaly as event-based evidence. A set of secondary sensors are used to monitor different aspects of the system’s (i.e., protected resource’s) state. This type of secondary sensor reports system state attributes as state-based evidence, which can serve as complementary intrusion evidence to the event-based evidence or anomaly reported. Under normal conditions, the primary sensor monitors activities across the system environment and reports alerts or event-based evidence of possible intrusions. The fusion system calls upon the secondary sensor only when the criticality of the situation dictates and/or there is need to substantiate the intrusion evidence (i.e., in this case, when any anomaly is reported by the primary sensor). The secondary sensor is used to provide evidence of alteration of system state attributes and can typically reside on an individual system, where it monitors and stores system attribute data locally in order to save communication overhead. Data is sent to the fusion system only on an “on-demand” basis. Data from secondary sensor is used to corroborate or challenge the primary sensor’s reports and to evaluate the overall security situation. This type of fusion scheme should be particularly useful in resource constrained high performance cluster environments with a potential to save on communication overhead and resource utilization.

For anomaly situation assessment with primary and secondary sensors [10], the overall degree of concern for each resource in the network jointly depends on the extent of any event anomaly reported for the resource in a given time period and the alteration of any state attribute in the resource environment in that time period. To conduct data fusion at this level for situation assessment, the dynamic fusion approach used event anomaly and system state alteration report as deduced by the primary and the secondary sensors as measures of concern for the hosts in related security situations and employed agreement between the reports to guide the fusion process. After experimenting with synthetic data [10], we found that beyond the level of consensus when the two results disagreed, compromised behavior took place with bias
towards the results of the primary sensor, and within the level of consensus, when the results agreed, disjunctive behavior took place and when they did not, level of consensus was followed. Thus the fused result for final situation assessment was indicative of the collective extent of concern generated from both primary and secondary sensors with bias towards the primary sensor’s report.

In our ongoing work, we differentiate between sources/sensors that are more reliable as compared to the others. The observations reported by the sensors can be uncertain in terms of their validity. Past performance history of sensors can be used to discern reliability of sensors. In this respect, we adapt weighted, quantified and prioritized aggregation rules by Dubois and Prade for information fusion from parallel sources [3]. We assign weights to sources to depict their reliability (weight of 1 being most reliable) and allow the weight to restrict the information taken into consideration from a certain sensor to the extent of the weight. Once counted/rejected, the possibility distribution can be aggregated using choices of fuzzy conjunction/disjunction operators [1].

Also another fusion option is to give priority to the group of majority sensors that are more in agreement with each than the others. This is possible by pooling the possibility distribution for sources/sensors and determining a subset that majorly agree [3] with each other. In that way, we ensure more responsiveness to sensors with less conflict in findings and less responsiveness to sensor(s) with more conflict in findings. Aggregation rules can be adapted to select fusion operators that change behavior according to majority consensus.

In our previous work, the dynamic fusion approach combines information from symmetric sources. However in the real world, sources can be asymmetric. Therefore, currently we are also investigating development of fusion techniques for multiple asymmetric sources to address situational understanding for information systems. Candidates for such heterogenous sensors may be different types of IDSs (for example, combination of both host-based and network-based IDSs), vulnerability scanners (which scan systems for security loopholes), honey pots (which draw intruders to collect information about their intentions), and performance monitoring systems (which monitor physical system’s characteristics).

3. CONCLUSION

With the open nature of the Internet, network intrusion has become a serious problem in recent times and has proliferated the demand for effective monitoring information systems. As there is no “perfect” or “one for all” monitoring source/sensor, it is only natural to combine information from a suite of heterogeneous sources to increase trustworthiness in systems such that an inability and/or weakness of one is compensated by capability and/or strength of another. However, managing such information is extremely difficult and challenging. Therefore new and improved fusion techniques are all the more essential to maximize trustworthiness in information systems. Possibilistic approaches are promising because while extensive apriori knowledge about system variables is not needed, it allows to model uncertainty inherent in the real world.

REFERENCES


