Modified Evidence Theory for Performance Enhancement of Intrusion Detection Systems

Ciza Thomas
SERC, Indian Institute of Science, India
cizathomas@yahoo.com

N. Balakrishnan
balki@serc.iisc.ernet.in

Abstract—Sensor Fusion using heterogeneous Intrusion Detection Systems are employed to aggregate different views of the same event in order to improve the detection through detector reinforcement or complementarity. The fusion technique proposed in this paper is expected to combine the Intrusion Detection System outputs with subjective judgements. In this paper, a new evidence model which is an extension and improvement of the classical Dempster-Shafer theory is proposed. The feasibility of this method is demonstrated via an analysis case study with several simulated detectors using the replayed DARPA data set. The experimental results are validated and a discussion on why and how the new model is useful is provided. The result shows an improvement in the probability of detection along with a reduction in the false alarm rate with the proposed fusion algorithm.

Index Terms—Intrusion Detection Systems (IDS), Sensor Fusion, Dempster-Shafer (DS) method, belief, plausibility, ignorance, uncertainty, conjunctive operator, disjunctive operator, context-dependent operator, quasi-associativity, idempotence, dipolarity

I. INTRODUCTION

Intrusions in a computer network basically exploit weaknesses of the network transmission protocol and the vulnerabilities and bugs exhibited by system and application software. There does not exist a unique Intrusion Detection System (IDS) suitable to any application domain, and many a times the decision made by different IDSs are complementary. This complementarity is usually seen to be achieved by using different feature sets or different training sets on multiple sources. If they get appropriately combined, the resultant outperforms the best individual Intrusion Detection System.

There is a factor of uncertainty in the results arising from vagueness and imprecision in the outputs of most of the IDSs available in literature. Sensor Fusion is the process of merging information from heterogeneous classes of sensors for a more intuitive, accurate and reliable result. Sensor fusion, in the case of intrusion detection aims to reduce uncertain and inaccurate observations of detectors by combining both redundant as well as complementary data. One of the techniques of sensor fusion is the Dempster-Shafer (DS) evidence theory [1], [2], which can be used to characterize and model various forms of uncertainty.

Even though our previous research works in enhancing the performance of Intrusion Detection Systems using advances in sensor fusion [3], [4] have given encouraging results, we have realized that there was no possibility of detecting the novel attacks because of the difficulty of generalizing from any previously observed behavior. As a result, we are pursuing in using a Neural Network Learner that understands the reliability of each one of the IDSs corresponding to the data and accordingly provide a weight to every IDS decision and then make use of a context dependent selection of a fusion operator. Specifically the interest is in the capability of a neural network to learn the confidence to be assigned to every detector and the context dependent operator to optimally fuse the IDSs.

This paper is organized as follows: Section II briefly recalls the Dempster-Shafer theory of evidence with its weighted extension and also its disadvantages. Section III discusses the related work and section IV illustrates the disjunctive combination of evidence which helps in evidence aggregation. Section V explains the modified evidence approach and section VI includes more detailed observation on the performance of this approach. Section VII is a general discussion on the proposed approach for the particular application of sensor fusion in Intrusion Detection. Section VIII includes experimental evaluation and section IX, the impact of this work. In section X the conclusion of this paper is drawn.

II. DEMPSTER-SHAFER COMBINATION METHOD

Dempster-Shafer (DS) theory is a generalization of the classical probability theory with its additivity axiom excluded or modified, where the probability mass function (p) is a mapping which indicates how the probability mass is assigned to the elements. The Basic Probability Assignment (BPA) function (m) on the other hand is the set mapping, and the two can be related by a belief structure. The mass m is very near to the probabilistic mass p, except that it is shared not only by the single hypothesis but also to the union of the hypotheses.

In the case of Dempster-Shafer theory, Θ is the Frame of Discernment (FoD), which defines the working space for the desired application. FoD is expected to contain all propositions of which the information sources (IDSs) can provide evidence. The belief of how likely is the traffic in an anomalous state, is detected by various IDSs by means of a mass to the subsets of the FoD. In DS theory, rather than knowing exactly how the probability is distributed to each element E ∈ Θ, we just know by the BPA function m, that a certain quantity of a probability mass is somehow divided among the focal elements. Because of this less specific knowledge about the allocation of the probability mass, it is difficult to assign exactly the probability associated with the subsets of the FoD, but instead we assign two measures: the (1) belief and (2) plausibility, which correspond to
the lower and upper bounds on the probability,

\[ \text{Bel}(A) \leq p(A) \leq P_l(A) \]

i.e., \( \text{Bel}(A) \leq p(A) \leq P_l(A) \) where the belief function, \( \text{Bel}(A) \), measures the minimum uncertainty value about proposition \( A \), and the Plausibility, \( P_l(A) \), reflects the maximum uncertainty value about proposition \( A \).

Also the additivity axiom of probability theory \( p(A)+p(\bar{A}) = 1 \) is modified as \( m(A) + m(\bar{A}) + m(\Theta) = 1 \), in the case of evidence theory, with uncertainty introduced by the term \( m(\Theta) \). \( m(A) \) is the mass assigned to \( A \), \( m(\bar{A}) \) is the mass assigned to all other propositions that are not \( A \) in FoD and \( m(\Theta) \) is the mass assigned to the union of all hypotheses when the detector is ignorant. This clearly explains the advantages of evidence theory in handling an uncertainty where the detector’s joint probability distribution is not required.

The equation \( \text{Bel}(A) + \text{Bel}(\bar{A}) = 1 \), which is equivalent to \( \text{Bel}(A) = P_l(A) \), holds for all subsets \( A \) of the FoD if and only if \( \text{Bel} \)'s focal points are all singletons. In this case, \( \text{Bel} \) is an additive probability distribution.

The problem is formalized as follows: Considering the DARPA data set, assume a traffic space \( \Theta = \text{DOS} \cup \text{Portsweep} \cup \text{R2L} \cup \text{U2R} \cup \text{Normal} \), of five mutually exclusive classes. Each IDS assigns to the traffic, the detection of any of the traffic sample \( x \in \Theta \), that denotes the traffic sample to come from the class which is an element of the FoD, \( \Theta \). With \( n \) IDSs used for the combination, the decision of each one of the IDSs is considered for the final decision of the of the fusion IDS.

A. Motivation for choosing Dempster-Shafer Combination Method

The paper presents a method to detect the traffic attacks with an increased degree of confidence by making use of a fusion system composed of different detectors. Each detector observes the same traffic on the network and detects the attack traffic with an uncertainty index. The frame of discernment consists of singletons that are exclusive (\( A_i \cap A_j = \emptyset, \forall i \neq j \)) and are exhaustive since the FoD consists of all the expected attacks which the individual IDS detects or else the detector fails to detect by recognizing it as a normal traffic. All the constituent IDSs that take part in fusion is assumed to have a global point of view about the system rather than separate detectors being introduced to give specialized opinion about a single hypothesis.

Dempster-Shafer method is a powerful tool that can deal with subjective hypothesis for evidence as well as statistical data combination: intuitively, it is much like a voting mechanism. It doesn’t have the requirement like Bayesian that the sensor set be predefined and sensor’s joint observation probability distribution be known beforehand.

The DS rule corresponds to conjunction operator: it builds the belief induced by accepting two pieces of evidence, i.e., by accepting their conjunction. Shafer developed the DS theory of evidence based on the model that all the hypotheses in the FoD are exclusive and the frame is exhaustive. The purpose is to combine/aggregate several independent and equi-reliable sources of evidence expressing their belief on the set.

The DS combination rule gives the combined mass of the two evidence \( m_1 \) and \( m_2 \) on any subset \( A \) of the FoD.

\[
m(A) = \frac{\sum_{X \cap Y = A} m_1(X)m_2(Y)}{1-\sum_{X \cap Y = \emptyset} m_1(X)m_2(Y)}
\]

(1)

The denominator of the equation (1) is \( 1 - k \), where \( k \) is the conflict between the two evidence. This denominator is for normalization, which spreads the resultant uncertainty of any evidence with a weight factor, over all focal elements and results in an intuitive decision. i.e., the effect of normalization consists of eliminating the conflicting pieces of information between the two sources to combine, consistently with the intersection operator.

Whether normalized or not, the DS method satisfies the two axioms of combination:

\[
0 \leq m(A) \leq 1 \text{ and } \sum_{A \subseteq \Theta} m(A) = 1 .
\]

The third axiom \( \sum m(\emptyset) = 0 \) is not satisfied by the unnormalized DS method. Also independence of evidence is yet another requirement for the DS combination method.

The classical DS theory treats all sensors democratically, but this is not realistic, since some are more precise and accurate than others. Hence, the weighted Dempster-Shafer method can also be used. In simplified situations, this weight factor matches with the prior probability in the classical Bayesian inference method. The weighted and extended DS can be used to:

- Realize differential trust scheme on sensors
- Mitigate conflicts that cause counter-intuitive results using classical DS evidence combination rule.

B. Disadvantages of the Dempster-Shafer Combination

In the case of full contradiction between the bodies of evidence, \( k = 1 \), and such a case occurs when there exists \( A \subset \Theta \) such that \( \text{Bel}_1(A) = 1 \) and \( \text{Bel}_2(\bar{A}) = 1 \) as in Table (1).

Table I

<table>
<thead>
<tr>
<th>( \cap )</th>
<th>( A )</th>
<th>( B )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_1 )</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( m_2 )</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

The computation of the combined evidence is done for the DS method and its other alternative methods like Yager, Smet's TBM and the Murphy's averaging [5], [6], [7], [8].

- DS method: \( m(A) = 0, m(B) = 0 \)
- Yager method: \( m(A) = 0, m(B) = 0, m(\Theta) = 1 \)
- Smet's TBM: \( m(A) = 0, m(B) = 0, m(\emptyset) = 1 \)
- Murphy's averaging: \( m(A) = 0.5, m(B) = 0.5 \)

One more case of contradiction between the bodies of evidence is shown with a different example in Table 2.
The conflicts in evidence are seen to result in either non-intuitive results as with DS, conflicts get ported over to uncertainty as in Yager or to the null set as in Smet’s TBM, or averaged as in Murphy’s. But none of them seem to be intuitive or reasonable from the point of view of improving the belief.

It can be concluded that the conflict is not expected to draw to a clear conclusion in one step. Final decision should be made depending on the collection of additional evidence. Until more evidence is collected, an attempt to forcefully converge the conflicting evidence should not be made. Hence, without suppressing any of the available evidence as in the case of DS, it is better to aggregate evidence by the union operator.

Another major drawback with DS and its alternatives except for Murphy’s is that since it uses conjunctive combination, if any one or more detectors fail to give evidence on a particular class, the evidence from other detectors on that particular class will have no effect and the intersection becomes a null set. A detector might fail to give evidence in cases when it is not tuned for that particular class of attack due to the shortcomings of the technology used or due to some other reason. This disadvantage was overcome by Murphy’s averaging, but this result also looks counterintuitive since if one evidence fails, the belief of that hypothesis gets weakened. This is illustrated in Table 3 with one of the detectors in the fusion being totally unreliable resulting in counter-intuitive results with DS.

### TABLE II

<table>
<thead>
<tr>
<th>Evidence with conflict</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\cap$</td>
</tr>
<tr>
<td>$m_1$</td>
</tr>
<tr>
<td>$m_2$</td>
</tr>
</tbody>
</table>

### TABLE III

<table>
<thead>
<tr>
<th>Evidence with four detectors with one unreliable</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\cap$</td>
</tr>
<tr>
<td>PHAD($m_1$)</td>
</tr>
<tr>
<td>ALAD($m_2$)</td>
</tr>
<tr>
<td>Snort($m_3$)</td>
</tr>
</tbody>
</table>

DS method: $m(A) = 0$, $m(B) = 1$, $m(C) = 0$
Yager method: $m(A) = 0$, $m(B) = 0.01$, $m(C) = 0$, $m(\Theta) = 0.99$
Smet’s TBM: $m(A) = 0$, $m(B) = 0.01$, $m(C) = 0$, $m(\phi) = 0.99$
Murphy’s averaging: $m(A) = 0.45$, $m(B) = 0.1$, $m(C) = 0.45$

**The use of data fusion in the field of Denial of Service (DoS) anomaly detection is presented by Siatetlis and Maglaris [9]. The Dempster-Shafer theory of evidence is used as the mathematical foundation for the development of a novel DoS detection engine. The detection engine is evaluated using the real network traffic.**

The superiority of data fusion technology applied to intrusion detection systems is presented in the work of Wang et al. [10]. This method used information collected from the network and host agents and application of Dempster-Shafer theory of evidence. Another work incorporating the Dempster-Shafer theory of evidence is by Hu et al. [11]. The Dempster-Shafer theory of evidence in data fusion is observed to solve the problem of how to analyze the uncertainty in a quantitative way. In the evaluation, the ingoing and outgoing traffic ratio and service rate are selected as the detection metrics, and the prior knowledge in the Distributed Denial of Service (DDoS) domain is proposed to assign probability to evidence.

### IV. DISJUNCTIVE COMBINATION OF EVIDENCE

The union approach can be considered under situations where different intrusion detection systems are specialized on different types of attack detection and hence may not respond to a certain attack. Thus the Union approach focuses on the best-case behavior of each detector. The combination of detectors is done so as to utilize the strength of each detector.

All the hypotheses are singletons since the hypothesis includes each IDS in the binary state with the traffic detected as a particular attack or as normal. Hence it is more simple than the generalized case with the hypotheses taking any possible subset of the power set of FoD.

The equation to find the mass of singletons is given by:

$$m(A) = \frac{\sum_{X \cap \Theta} m(A) \cap X = \phi}{\sum_{X \cap \Theta} m(A) \cap X \neq \phi} (m_1(X) + m_2(Y))$$

Here the evidence $m_1$ is unreliable, and by giving equal weight factor to all evidence, the result turns out to be counter-intuitive. If proper weight factor is given to the evidence depending on their reliability, the Murphy’s method gives acceptable results as in the modified averaging method.

In all the above cases where the DS fails, the union operator is proposed to be used for aggregating the evidence. This is mainly because of two reasons. In case of conflict, on receiving further evidence it is possible to properly converge it and in case of zero evidence from any one or more of the detectors, the union operator works the same as the averaging operator, which is the best that can be thought of. Thus if the intersection of the evidence is not empty, the sources overlap and the combination rule can be intersection. Otherwise, at least one of the sources is necessarily wrong, and a more natural combination rule is the union, which assumes that all the sources are not wrong.
\[ X = \frac{\sum_{Y} A(m_1(X) + m_2(Y))}{\sum_{X=Y} A(m_1(X) + m_2(Y))} \]  

(2)

The numerator of the above equation relates to the disjunctive combination and the final mass is calculated by the normalization with respect to the entire power set \(2^\Theta\), which is closed under union, intersection and complement and hence is a sigma algebra. This normalization allows the disjunctive combination equation to satisfy all the axioms of the evidence theory.

Also, conflicts can be thought of as due to uncertainty whereby the IDS cannot take the decision correctly and the collective information will be ambiguous. Hence only if the reliability of the detectors are known, a conclusion need to be drawn by suppressing some evidence over the others. Else, it is better to aggregate all the evidence and combine conjunctively with other evidence agreeable to the aggregated evidence. The normalization is not done until the conjunctive combination is done for converging the results.

The properties of associativity and commutativity are satisfied with the disjunctive combination if normalization is done only at the final combination stage. In the case of intrusion detection with only singletons as the expected hypotheses, \(Bel(A\cup B) = Bel(A) + Bel(B)\), since DS simplifies to Bayesian. However, the advantage of the evidence combination method over the Bayesian combination method is that more evidence can be combined in a single step, without the knowledge of the associated probability distribution.

V. CONTEXT-DEPENDENT OPERATOR

The DS operator is most acceptable except for its two disadvantages that were highlighted in section II. Hence, a Context-Dependent operator is needed for the decision fusion which is supposed to utilize enough available information before making a final decision. This operator is expected to be:

- Conjunctive, if the sources have very low conflict and also when all masses are non-zero. The fusion should then behave as a severe operator where the common or redundant part gets chosen and reduces the mass of the less certain information.
- Disjunctive, if sources conflict and also when any one or more beliefs happen to be zero.
- Compromise or average, in case of partial conflict.

The context-dependent operator is expected to have an adaptive feature of combining information related to one class in one way, and information related to another class in another way.

The proposed hybrid operator works the same way as the DS operator except in the case of conflict and when any belief mass happens to be zero, and also when varying reliability need to be introduced on the different detectors. The combined operator has a modified mass \(l\), which can exceed one at intermittent stages due to para-consistency and is given by:

\[
l = \left( \frac{\sum_{i=1}^{n} w_i l_i}{\prod_{i=1}^{n} w_i l_i} \right)^{1-k} + \left( \sum_{k=0}^{n} w_i l_i \right)
\]

(3)

where \(w_i\) is the weight associated with each detector, \(k\) is the conflict between the combining detectors and \(l_i\) is the mass associated with each detector.

The conditions and requirements of using this operator are the following:

- The proportionate detector weighting factor is used since the intrusion detection systems used for the combination are binary in nature. The axiom of combination; i.e.,

\[
\sum_{A \subseteq \Theta} m(A) = 1
\]

gets satisfied only when exponential weighting factors are made use of.
- The weights assigned to the detectors should add to one; i.e.,

\[
\sum_{i=1}^{n} w_i = 1
\]
- The value of \(k\) lies between 0 and 1 and is the parameter that controls the degree of compensation between the intersection and union parts. The value of the conflict factor (k) between any two detectors can be calculated as the Euclidean distance between the two detectors. \(k\) takes a value of zero if the detectors are in consensus and non-zero in case of conflict.

The method proposed with this operator gives the most intuitive result and works as follows:

- Disjunctive combination is done on diverse pairs of IDSs (averages out without suppressing any evidence). Pair-wise disjunctive combination is done on all the detectors which are necessarily not redundant (since it is intuitive to think of a stronger evidence in case of redundancy rather than averaging out which does not give any additional support even though both detectors support it).
- The results are then conjunctively combined if not totally contradicting after the pair-wise aggregation, since at this stage suppression of any evidence will not happen to a higher extend and also suppression of certain evidence helps in faster convergence.
- In the case of redundancy, if the disjunctive combination is used, it is required to work without normalization in all the intermittent combinations for the sake of making the strong evidence still stronger.
- In order to satisfy all axioms of evidence theory it is required to do normalization at the final stage so that all the masses of a particular evidence sum to one.
- It can be concluded that the proposed method combines detectors reasonably well under all conditions by disjunctively combining diverse or contradicting pairs and finally suppressing the weak hypotheses by pair-wise conjunctive combination.
- In the very specific case of binary evidence to singletons, it can be observed that there is no additional support happening with the addition of redundant evidence. Hence it is better to go for disjunctive combination in all the intermittent
combinations until the final step where the conjunctive combination results in a faster convergence.

VI. PERFORMANCE OF THE PROPOSED COMBINATION OPERATOR

The proposed operator (⊗) is:
- Same as the DS operator which is a consensus operator. This consensus operator cannot provide information from a set of measures among which one or more are zeros.
- Union operator in case of conflict and also when the mass/es are zeros. The union operator is a para-consistent combination operator and hence the combination mass can exceed one.

The operator satisfies the commutative, continuity, monotonicity (after normalization), quasi-associativity (if normalization is done only at the final stage of combination) and no idempotence (para-consistency of union operator) properties.

The reason for using the context dependent operator is that the union operator is totally acceptable in case of diversity or conflict because of uncertainty. But disjunction gives averaging with redundant observations. Even though reasonable to use an average value where the additional observation by one more detectors does not add to the increase in the belief, intuition makes us think that some method which increases the belief of the strong hypothesis is a requirement.

The proposed operator subsumes the celebrated DS operator except for the cases of conflict and zero evidence and hence all the axioms of the DS theory appear with the proposed operator also:

$$\sum_{A \subseteq \Theta} m(A) = 1, \ m(\phi) = 0 \text{ and } 0 \leq m(A) \leq 1.$$ 

The independence of the sources which combine is another assumption taken in the case of applicability of the combination.

When working with real-time attacks, zero-day attacks can be expected and most of the commercial IDSs which are signature-based will not be able to identify it. Then either misclassification or False Negative (which is again misclassification, since the FoD contains the hypothesis "normal" which is an expected output of the IDS) is expected. Hence the proposed combination operator can assume a closed-world assumption as in the case of DS method.

The same example which was used to illustrate the performance of the DS and other operators is taken again to illustrate the performance of the proposed operator.

Applying the proposed method with aggregation of information from the conflicting evidence and subsequently converging down
B. Advantages of the Proposed Operator

- The proposed operator can combine evidence from two detectors with different FoD. Then the combined FoD will be the union of the FoDs.
- The closure property is satisfied so as to stay within a given mathematical framework.
- This operator works under all conditions and states of the individual detector.
- This operator has been developed quite intuitively and hence the result is most intuitive. The conjunctive operator is acceptable when all sources happen to be reliable and similar, whereas the union operator corresponds to the data aggregation from weaker reliable sources. Thus conjunctive combination makes sense when the mass distribution significantly overlaps and if not, the combination will have at least one of the sources as wrong and it is better to choose disjunctive combination. Also, the intuition was that a certain diversity among detectors assures versatility whereas a certain redundancy assures reliability.
- The combination operation is simple and easy.
- The combination ensures that no information is unnecessarily suppressed, but at the same time convergence is assured.
- The non-idempotence property is counted as an advantage in sensor fusion, since the same observation from two sensors should improve the belief in that observation, rather than idempotence. Bel ⊕ Bel ≠ Bel; even though Bel ⊕ Bel will favor the same subsets as Bel but with, as it were, twice the weight of evidence. This is because if each source supports a hypothesis for independent reasons, it is natural to conclude that the hypothesis is strongly supported. Also, adopting idempotence is a matter of context; acceptable when sensors are homogeneous.
- Other properties like commutative, continuity, and distributive properties are satisfied.
- Associativity is not absolutely required, since the combination algorithm is not associative (since it considers an ordering of sources). A weaker property such as quasi-associativity is often sufficient, if the normalization is delayed till the end of the combination.
- This operator subsumes the most celebrated DS method of evidence and hence all axioms of DS theory of evidence is incorporated as it is.
- This operator has the property of bipolarity, which means the more one proposition is supported by all of the evidence, the more it can obtain belief masses after combination. This property is seen to be satisfied by the DS operator also.
- This operator is relatively tolerant of inaccurate, incomplete and inconsistent evidence.
- This operator aggregates the conflicting evidence and then as the next stage comprehend the aggregated pairwise results for an improved sensitivity and for a false alarm suppression.

C. Disadvantages of the Proposed Operator

- Choice of the operator function depends on the context and hence has to be carefully chosen.
- Sources need to be independent for the combination. The Dempster’s idea on the combination of independent sources of information can be stated as follows: Suppose there are n pieces of evidence which are given in the form of n probability spaces \( (\Theta_i, \chi_i, m_i) \) where \( \chi_i \) is a subset of \( P(\Theta) \), the power set of the FoD. Each of which has a mapping relation with the same space S through a multivalued mapping. These n sources are independent and the explanation by Dempster is as follows: “opinion of different sensors based on overlapping experiences could not be regarded as independent sources.” Dempster assumes statistical independence of sources as “different measurements by different observations on different equipment would often be regarded as independent ...”
- For a parallel combination of any model, the basic requirement is that the combination should be associative.

VII. Discussion

1) The selection of the detectors is done by choosing the detectors with minimum correlation. Detectors are redundant when the correlation coefficient is one, similar when the correlation coefficient is greater than 0.5, diverse when the correlation coefficient is less than or equal to 0.5 and totally contradict when the correlation coefficient is zero.

2) It is quite intuitive to think that the fusion method should work with a minimum number of detectors and get the advantages of fusion techniques what ever be the fusion technique used. Every best detector when merged should improve the confidence of the existing evidence and thereby converge faster i.e., the fusion method makes strong evidence stronger so that confusion (uncertainty) is avoided.

3) There are inherent advantages in using best detectors in the fusion scheme. This makes sense intuitively also in the case of evidence theory of fusion, i.e., if one detector gives its evidence and the second also gives a similar evidence, the belief is reinforced stronger, or the contradiction in evidence reduces the belief.

4) The DS method of combination implicitly has a closed world assumption, i.e., the set of possible hypotheses is perfectly known. \( \Theta \) is assumed to represent a set of states that are mutually exclusive and exhaustive. In the intrusion detection application, when dealing with real time traffic, \( \Theta \) need not necessarily be exhaustive, since the traffic may contain many novel attacks not included in the \( \Theta \). However, in such cases, intrusion detection systems may also be unable to detect it and hence it appears as “normal” which is also a hypothesis included in the FoD. Hence an additional label denoting “none of the above” gets included in the hypothesis “normal” by the evidence only or else this novel attack gets misclassified as some other attack only.

5) Applying DS assumes that the sources are independent, \( m(A) = m_1(A) \star m_2(A) \). Even though the IDSs are trained on the same training set, the two are independent of each other and they in turn depend on different features of the training data set for detection. When detectors depend on the same features in the data, they may give a consensus in their decision, which is different from the dependence between the detectors. When fusing by means of mathematical decision rules, it is necessary to have independent detectors, because this will simplify construction of the rule and enhance its efficiency.

6) It is important to note that mutual exclusiveness is not assumed when the union operator is used, \( m(A) = m_1(A) + m_2(A) \)
m_2(A)). The event that an IDS alert with \{DoS\} and the event that another IDS alert with \{DoS\} are not mutually exclusive. The elements in the FoD are mutually exclusive where as the two sets which are sample space of IDS1 and IDS2 are not mutually exclusive. Since they are not mutually exclusive, the result of union operator will be para-consistent. The union of two events IDS1 alerting DoS and IDS2 alerting DoS, denoted as \( IDS_1 \cup IDS_2 \) is the event containing all elements belonging to \( IDS_1 \) alerting as DoS or \( IDS_2 \) alerting as DoS or to both alerting as DoS. The idea is the aggregation of the support to the events so that the belief improves. But at the same time, disjoint training sets are used to train both the anomaly-based IDSs, PHAD and ALAD by training them on week one and week three of the DARPA’99 data set respectively.

7) Shafer describes the requirement of the different sensors to be independent and non-interacting, which is just that all their interaction should be in terms of the issues discerned by the FoD. That clearly says that the FoD should “discern the interaction of the evidence” (with the singletons A, B and C, the FoD consists of the eight possibilities which gives the total interaction within the FoD).

8) We concede that for highly conflicting cases, the proposed method is the same as an averaging operator. However, we argue that the proposed method still has considerable applicability due to weight factors and data-dependency which results in highly intuitive results.

### VIII. Experimental Evaluation

#### A. Data Set

MIT-DARPA dataset (IDEVAL 1999) [15] is being used to test the performance of intrusion detection systems. Supporting facts for the usefulness of this data set is given in the work of Thomas et al. [16]. The data for weeks one and three are used for training of intrusion detection systems and weeks four and five are used as test data. The data set is replayed for the test purpose by means of tcpreplay.

#### B. Test Setup

The test set up for the experimental evaluation consists of three Pentium machines running the Linux operating system. Different classes of Intrusion Detection Systems like signature-based, anomaly based, flow-based and packet-based are all simulated in this internal network. In many systems for good protection, a combination of shallow and deep sensors are necessary. The data from various simulated intrusion detection systems distributed across a single subnet and observing the same domain gets analyzed for their performance and are simultaneously fused for a trust worthy decision. A connection refers to a sequence of data packets related to a particular service, and the network intrusion detector detects connections related to malicious activities.

The intrusion detection systems used in this preliminary test set up consists of SNORT, PHAD, and ALAD. All the intrusion detection systems that form part of the fusion IDS were separately evaluated with the same data set as shown in tables VII, VIII, and IX. The combined fusion IDS is also evaluated and shown in table X. The experiments carried out clearly showed that no individual IDS was able to provide the minimum values of all performance measures, whereas the combined IDS outperformed all the individual Intrusion detection systems totally.

<table>
<thead>
<tr>
<th>Attack type</th>
<th>Total attacks</th>
<th>Attacks detected</th>
<th>% detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe</td>
<td>37</td>
<td>26</td>
<td>70%</td>
</tr>
<tr>
<td>DoS</td>
<td>63</td>
<td>27</td>
<td>43%</td>
</tr>
<tr>
<td>R2L</td>
<td>53</td>
<td>6</td>
<td>11%</td>
</tr>
<tr>
<td>U2R/ Data</td>
<td>37</td>
<td>4</td>
<td>11%</td>
</tr>
<tr>
<td>Total</td>
<td>190</td>
<td>63</td>
<td>33%</td>
</tr>
</tbody>
</table>

#### TABLE VII

**Attacks of each type detected by PHAD at 100 false alarms**

<table>
<thead>
<tr>
<th>Attack type</th>
<th>Total attacks</th>
<th>Attacks detected</th>
<th>% detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe</td>
<td>37</td>
<td>9</td>
<td>24%</td>
</tr>
<tr>
<td>DoS</td>
<td>63</td>
<td>23</td>
<td>37%</td>
</tr>
<tr>
<td>R2L</td>
<td>53</td>
<td>31</td>
<td>59%</td>
</tr>
<tr>
<td>U2R/ Data</td>
<td>37</td>
<td>15</td>
<td>27%</td>
</tr>
<tr>
<td>Total</td>
<td>190</td>
<td>78</td>
<td>41%</td>
</tr>
</tbody>
</table>

#### TABLE VIII

**Attacks of each type detected by ALAD at 100 false alarms**
ments conducted in this study have used only three detectors. It is possible that the use of more detectors will necessarily lead to higher performance improvement of the fusion IDS. Also, the context dependent operator can provide a generalizable solution for a wide range of applications. Thus, it supports the claim that synergistic interaction between sensor fusion and Intrusion detection facilitates the sensor fusion for detection improvement.

ACKNOWLEDGEMENT

We sincerely acknowledge the anonymous reviewers for their comments and corrections of the paper. Their suggestions have indeed enhanced the value of the paper.

REFERENCES


IX. EXPECTED IMPACT OF THE WORK

Different IDS have different detection rates and false alarm rates and these may be complementary, competitive or cooperative. What sensor fusion is all about is how to combine multiple sensor outputs to reveal the best truth regarding the objects of interest in terms of practical utility.

The fusion technique used in this paper is expected to combine IDS outputs with subjective judgements, the concept being nicely suited for intrusion detection, where the concern usually involves human subjects’ activity and intention. So the solution is to freely use subjective detectors, i.e., the detectors outputs can not only depend on observation of statistical process, but also depend on rational human reasoning. Since there are multiple detectors, there is the need to coordinate them and combine their results. The combination operator proposed in this paper for sensor fusion is mainly used because it is difficult to represent the information supplied by the detectors by means of single probability distributions, due to imprecision and/or lack of statistical evidence. The proposed hybrid operator which functions as conjunctive and disjunctive operator depending on the context is particularly suitable when the sources are heterogeneous. The proposed adaptive operator for heterogeneous sensors functions between conjunctive and disjunctive combination modes, depending on the consistency between the sources. The feasibility of this idea is demonstrated via an analysis case study with three simulated detectors using the replayed DARPA data set and artificially generated detector outputs distributed over a LAN network.

The technique gives a performance better than any of the individual intrusion detection systems which were fused and even though validated for a particular application, it should be a generalizable solution beyond any specific application case.

X. CONCLUSION

In summary, the context dependent operator was demonstrated to be feasible for sensor fusion. The research in this work has improved over the existing DS alternatives in that it can better handle uncertainty and ambiguity in sensed context. The experiments conducted in this study have used only three detectors. It

<table>
<thead>
<tr>
<th>Attack type</th>
<th>Total attacks</th>
<th>Attacks detected</th>
<th>% detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe</td>
<td>37</td>
<td>15</td>
<td>41%</td>
</tr>
<tr>
<td>DoS</td>
<td>63</td>
<td>35</td>
<td>56%</td>
</tr>
<tr>
<td>R2L</td>
<td>53</td>
<td>30</td>
<td>57%</td>
</tr>
<tr>
<td>U2R/ Data</td>
<td>37</td>
<td>34</td>
<td>92%</td>
</tr>
<tr>
<td>Total</td>
<td>190</td>
<td>115</td>
<td>61%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attack type</th>
<th>Total attacks</th>
<th>Attacks detected</th>
<th>% detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe</td>
<td>37</td>
<td>33</td>
<td>89%</td>
</tr>
<tr>
<td>DoS</td>
<td>63</td>
<td>45</td>
<td>71%</td>
</tr>
<tr>
<td>R2L</td>
<td>53</td>
<td>37</td>
<td>70%</td>
</tr>
<tr>
<td>U2R/ Data</td>
<td>37</td>
<td>35</td>
<td>95%</td>
</tr>
<tr>
<td>Total</td>
<td>190</td>
<td>129</td>
<td>79%</td>
</tr>
</tbody>
</table>

TABLE IX

ATTACKS OF EACH TYPE DETECTED BY SNIORT AT 1000 FALSE ALARMS

<table>
<thead>
<tr>
<th>Attack type</th>
<th>Total attacks</th>
<th>Attacks detected</th>
<th>% detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe</td>
<td>37</td>
<td>15</td>
<td>41%</td>
</tr>
<tr>
<td>DoS</td>
<td>63</td>
<td>35</td>
<td>56%</td>
</tr>
<tr>
<td>R2L</td>
<td>53</td>
<td>30</td>
<td>57%</td>
</tr>
<tr>
<td>U2R/ Data</td>
<td>37</td>
<td>34</td>
<td>92%</td>
</tr>
<tr>
<td>Total</td>
<td>190</td>
<td>115</td>
<td>61%</td>
</tr>
</tbody>
</table>

TABLE X

ATTACKS OF EACH TYPE DETECTED BY CONTEXT-DEPENDENT FUSION AT 160 FALSE ALARMS