Mathematical Analysis of Sensor Fusion for Intrusion Detection Systems

Ciza Thomas and N. Balakrishnan
Indian Institute of Science, Bangalore

Outline of the talk

- Network Security
- Intrusion Detection System: An Introduction
- Present status of IDS
- Need of Sensor Fusion for IDS
- Mathematical Basis of Sensor Fusion
- Data-Dependent Decision Fusion
- Modified Dempster-Shafer Theory
- Experimental Evaluation
- Summary

Security

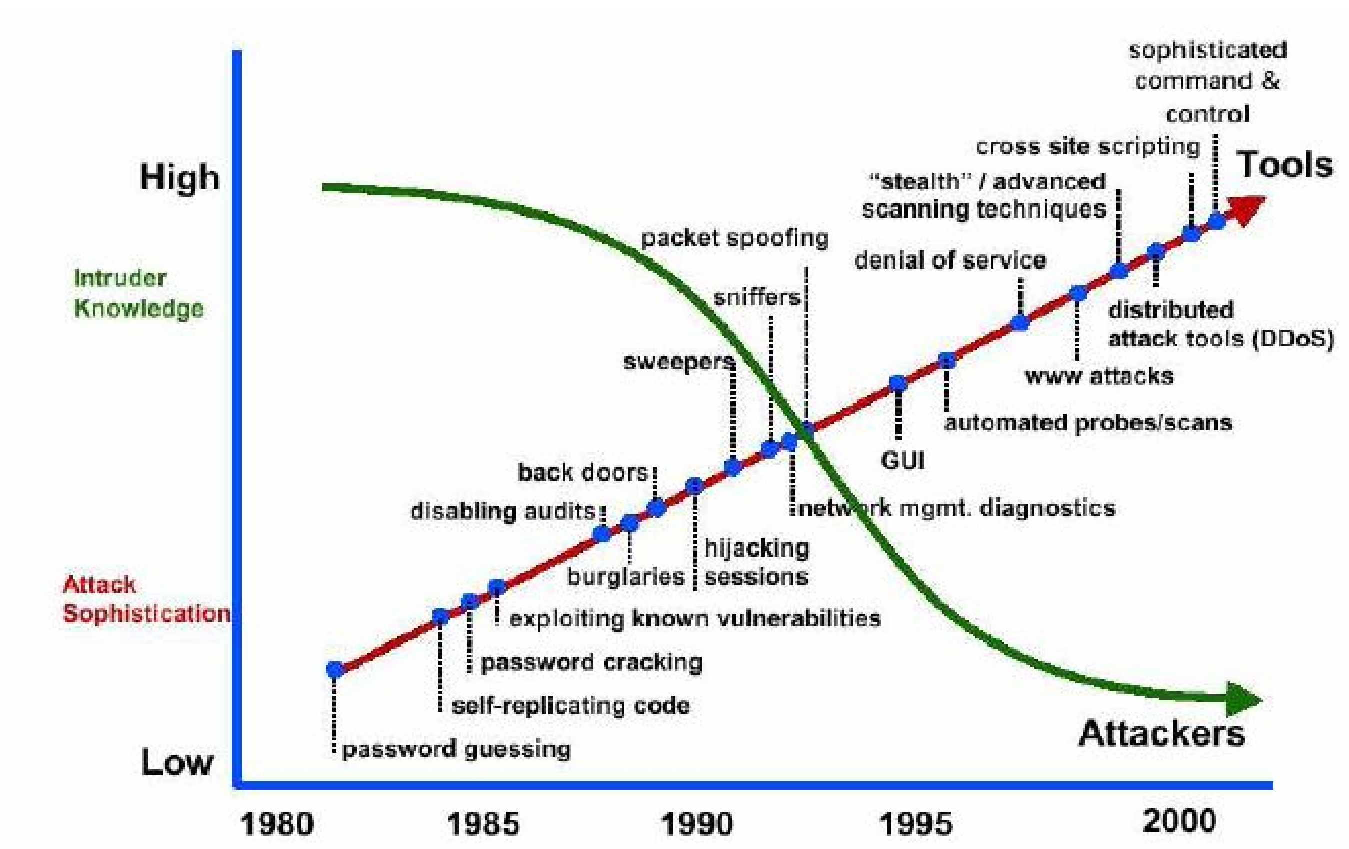
Bruce Schneier’s “Beyond Fear” defines security as preventing the adverse consequences from intentional and unwarranted actions of others.

The intentional and unwarranted actions are performed by an intruder or an attacker.

Securing the information Systems is highly complex, and decisions involve trade-offs!

Trade-off between a high rate of false alarms or a significant number of missed alarms, which are the two possibilities of system failure.

Attack Sophistication vs Intruder Knowledge



- Botnets to storm botnets
- Sophisticated array of malware
- blended, cross-vector and targeted attacks

Source: Julia Allen, CMU

Protection Techniques

✓Physical protection for hardware
✓Passwords, access tokens, biometrics, etc. for authentication
✓ Access control lists for authorization
✓ Cryptography for secrecy
✓ Backups and redundancy for availability
✓ Trusted operating systems for authenticity
✓ Firewalls for network protection
… and so on

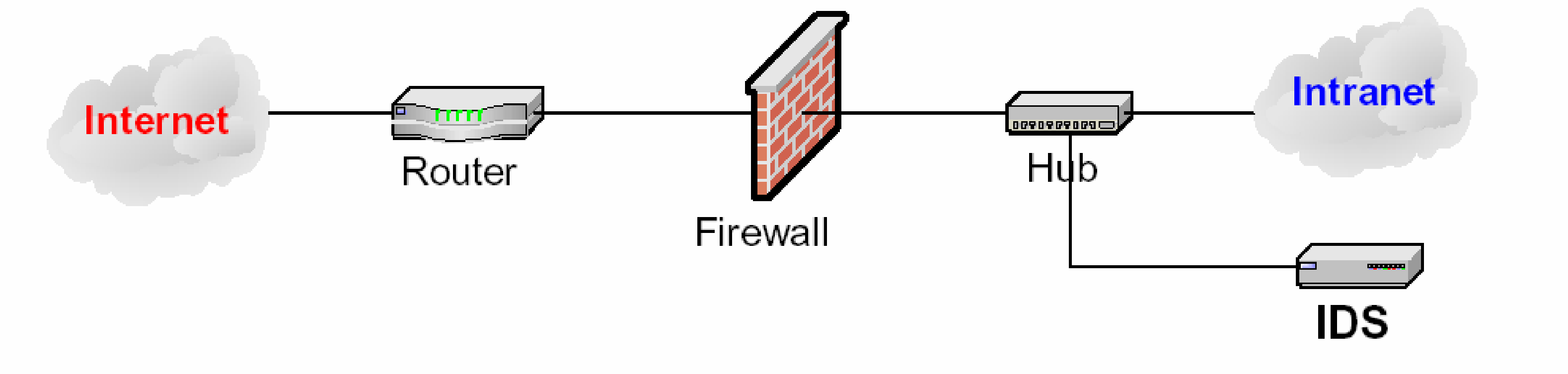
"Absolute" security cannot be guaranteed

Intrusion Detection System (IDS)

- Intrusion is a set of actions that attempts to compromise the integrity, confidentiality or availability of any resource on a computing platform. Intrusion is unauthorized access to, and/or activity in, an information system

- IDS is a hardware or a software tool that attempts to detect an intruder hacking into a system or a genuine user exploiting system resources. IDS is the process of identifying that an intrusion has been attempted, is occurring, or has occurred

Typical Security Scenario in any network



Present Status of IDS

The state-of-the-art in IDS has not reached a stage of acceptable perfection and does not adequately address the threat of computer-based attacks.

Many of the research works in the field of IDS try a feasible approach to an improved detection rate.

With the increasing traffic and increasing complexity of attacks, there is high demands for a very high detection rate (usability) and an extremely low false alarm rate (acceptability) .

Most of the IDSs available in literature show distinct preference for detecting a certain class of attack with improved accuracy while performing moderately for the other classes of attacks.

How to Improve IDS

In view of the enormous computing power available with the present day processors, combining multiple IDSs to obtain best-of-breed solutions has been attempted earlier.

Multi-sensor fusion meets these requirements by a refinement of the combined responses of different IDSs.

Sensor fusion can be defined as the process of collecting information from multiple and possibly heterogeneous sources and combining them to obtain a more descriptive, intuitive and meaningful result.

Why sensor Fusion in IDS?

To meet the requirements of a better than the best detection by a refinement of the combined response of different IDSs with largely varying accuracy.

To get a result more reliable, complete, meaningful and certain than any of the individual IDSs.

A better analysis of existing data gathered by various individual IDSs can detect many attacks that currently go undetected.

Mathematical Basis for Sensor fusion

- The mathematical basis for sensor fusion provides enough support for the acceptability of sensor fusion in performance enhancement of IDSs.

- Formulate the problem of fusion of multiple heterogeneous IDSs and examine whether the improvement in performance could be achieved through sensor fusion.

- To study the performance of the theoretically best fusion approach using mathematical analysis.

Mathematical Basis for Sensor fusion

- This results in the ability to automatically exploit the strengths and weaknesses of each IDSs resulting in an improved estimate of the intrusion detection better than with a single sensor alone.
- The theoretical study undertaken justifies why and how the sensor fusion algorithms work, when the decisions from multiple detectors are combined.
- This describes the central concept underlying the work and a theme that ties together all the arguments in this work.

Mathematical Basis for Sensor fusion

- The theoretical model is initially undertaken without any knowledge of the available detectors or the monitoring data.
- The mathematical analysis of decision fusion develops a rational basis which is free from the various techniques, sensors or data used.
- The empirical evaluation to augment the mathematical analysis is also illustrated using the DARPA data set. The empirical results is used to validate the analytical findings.

Mathematical Basis for Sensor fusion

- Variance of the IDSs determines how good their average quality is when each IDS acts individually.

- Covariance among detectors measures the dependence of the detectors. The more the dependence, the lesser the gain benefited out of fusion.

- When the participating sensors are independent, for each access x_j, n responses are available and are used independently of each other. The average of variance is given as:

$$(\sigma_{av}^j)^2 = \frac{1}{n}\sum_{i=1}^{n}(\sigma_i^j)^2$$

Mathematical Basis for Sensor fusion

- With dependent sensors, all n responses are used together and are combined using the mean operator; the variance can be calculated as follows:

$$(\sigma^j_{fusion})^2 \leq (\sigma^j_{av})^2$$

- Fusion of the scores reduces variance, which in turn results in reduction of error. To measure explicitly the factor of reduction in variance,

$$\frac{1}{n}(\sigma^j_{av})^2 \leq (\sigma^j_{fusion})^2 \leq (\sigma^j_{av})^2$$

Mathematical Basis for Sensor fusion

Factor of reduction in variance, $\mathrm{v}_r = \frac{(\sigma_{av}^j)^2}{(\sigma_{fusion}^k)^2}; 1 \leq v_r \leq n$

- Reduction in variance is more when more detectors are used.
- The reduction in variance of the individual classes results in lesser overlap between the class distributions. Thus the chances of error reduces, which in turn results in a better detection.
- This forms a trivial argument for why fusion using multiple detectors works for intrusion detection application. Experimental results provide strong evidence to support this claim.

Solution Approaches

- Even if fusion is expected to reduce the variance and improve the detection, there is uncertainty associated with every IDS
- In that case, how to do a meaningful combination is the task at hand
- There is an arsenal of different theories of uncertainty and methods based on these theories for making decisions under uncertainty.
- There is no consensus as to which method is most suitable for problems with epistemic uncertainty, when information is scarce and imprecise.

Uncertainty Sources

Uncertainty handling

In order to handle uncertainty and to result in a better combination result, we need reasoning models, the commonly used ones being Probability theory, Fuzzy set theory and the Evidence theory.

The random usage of a model for combination usually leads to inappropriate matching.

Understanding ignorance

The various forms of ignorance can be encountered simultaneously and it is necessary that we are able to integrate them.

The basic assumption in probabilistic analysis is the randomness of uncertainty for which the numerical data should be free from imprecise subjective opinion and additionally the uncertainty should not be caused by deficient knowledge.

Why not the familiar Bayesian?

Probability models are too restrictive to model quantified beliefs as they appear in diagnostic contexts (deals only with Aleatory Uncertainty and not with Epistemic Uncertainty)

•A piece of evidence could support a hypothesis A without necessarily refuting it.

i.e., the truth of the hypothesis A with an evidence need not be equivalent to the falsehood of its absence.

•The need for the priori knowledge of the probability distribution is yet another disadvantage of the Probability model.

How do we combine?

The relevance to this work:

- When several sensors provide information, how do we recognize the nature of ignorance involved and select the appropriate model?

- How do we deduce them into more compact forms?

- How do we combine them?

- How do we take into consideration the redundancies, the correlations and the contradictions?

ALL THESE PROBLEMS MUST BE STUDIED AND THE IMPLEMENTATION OF POTENTIAL SOLUTIONS TESTED.

Sensor Fusion using Evidence Theory

Sensor Fusion primarily combines the information from heterogeneous sensors to take a better decision than from a single sensor, by reducing imprecision and uncertainty and increasing completeness.

The fusion can make use of the theory of evidence, where the events to which degrees of belief are assigned are related to the decision problem at hand being the presence of different types of attack traffic.

The degree of belief is the basic probability assignment (bpa) in Evidence Theory.

Basics of Evidence Theory

The FoD consists of all possible outputs of the IDS:

Say, **Θ = { Probe, DoS, R2L, U2R, Normal}**

It is composed of mutually exhaustive and exclusive hypotheses and its power set is closed under union and intersection.

The portion of total belief that is assigned exactly to a proposition through **basic belief assignment function** (*m*) defined as:

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

Basics of Evidence Theory

It satisfies the conditions:, $0 \le m(A) \le 1$, $m(\Phi)=0$, $\sum_{A \subseteq \Theta} m(A) = 1$

Belief function: $Bel(A) = \Sigma_{B \subseteq A}\, m(B)$

Plausibility function: $Pl(A) = \Sigma_{A \cap B \neq \Phi}\, m(B)$

Combination Rule of Evidence

In the case of imperfect data, fusion is an promising solution to obtain more relevant information. Evidence theory offers appropriate aggregation tools.

Two bpas m_1 and m_2 can be combined to yield a new bpa m by a combination rule. The DS combination of evidence is the most celebrated in this area and is given by:

$$m_c(A_i) = \frac{\sum_{B_j \cap C_k = A_i} m_1(B_j) m_2(C_k)}{1 - K}$$

$$K = \sum_{B_j \cap C_k = \varnothing} m_1(B_j) m_2(C_k)$$

Dempster’s Rule of Combining

BBA structures

Combined BBA structure

m_1 m_2 m

Algebraic properties : commutative and associative

Dempster’s rule disregards every contradiction

Dempster-Shafer Combination

- Dempster-Shafer (DS) theory is required to model the situation in which a classification algorithm cannot classify a target or cannot exhaustively list all of the classes to which it could belong.
- This is most acceptable in the case of unknown attacks or novel attacks or the case of zero *a priori* knowledge of data distribution.
- DS theory does not attempt to formalize the emergence of novelties, but it is a suitable framework for reconstructing the formation of beliefs when novelties appear.

Dempster-Shafer Combination

- Fusion should result in a thought on the decisions and not merely a response that aggregates the decisions.
- The aim of using the DS theory of fusion is that with any set of decisions from heterogeneous detectors, sensor fusion can be modeled as utility maximization.
- The DS rule corresponds to conjunction operator since it builds the belief induced by accepting pieces of evidence, i.e., by accepting their conjunction.
- All the hypotheses in the FoD are exclusive and the frame is exhaustive.

Dempster-Shafer Combination Method

	A	*B*	*C*
m_1	0.8	0.1	0.1
m_2	0.4	0.2	0.4
m_3	0.3	0.3	0.4
m_4	0.4	0.3	0.3

DS performs with

m(A)=0.85

m(B)=0.04

m(C)=0.11

Implementation

Illustrations with heterogeneous sensor outputs reveal the best truth regarding the propositions of interest in terms of practical utility.

Experiments were conducted using the DARPA data sets as well as real-time traffic and various learning methods, and compared the experimental results to theoretical predictions.

The results show an improvement in the probability of detection and reduction in the false alarm rate for the fusion algorithm.

It supports our claim that synergistic interaction between sensor fusion and intrusion detection facilitates the sensor fusion for detection improvement.

Test Setup

- Consists of three Pentium machines with Linux Operating System
  - One performing as an attacking machine
  - Second running the network sniffer to be provided to the intrusion detection systems. The network sniffer machine has the Snort IDS installed on it and also the PHAD IDS also running on it
  - Third machine installed with ALAD IDS
  - Inclusion of a collection of heterogeneous IDSs distributed across a single subnet observing the same domain

Data Set

- MIT-DARPA dataset (IDEVAL 1999) was used to train and test the performance of Intrusion Detection Systems

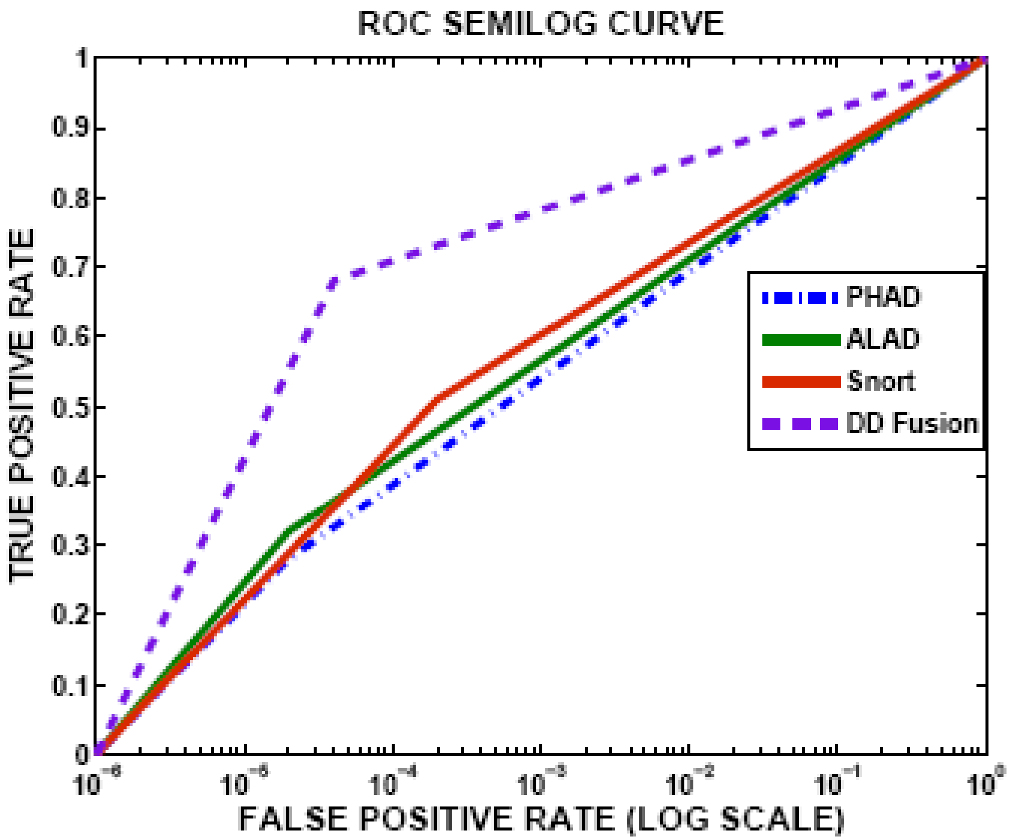
    - The data for the weeks one and three were used for the training of the anomaly detectors PHAD and ALAD and the weeks four and five were used as the test data
    - The DARPA 1999 test data consisted of 190 instances of 56 attacks which included 37 Probes, 63 DoS attacks, 53 R2L attacks, 37 U2R/Data attacks

- The real time normal traffic was embedded with the http attacks generated by the tool libwhisker.

Experimental Evaluation

Attack type	Total attacks	Attacks detected by PHAD	Attacks detected by ALAD	Attacks detected by Snort	Attacks detected by Fusion IDS
Probe	37	22	6	10	28
DoS	63	24	19	30	40
U2R/ Data	53	6	25	26	29
R2L	37	2	10	30	32
Total	190	54	60	96	129
Total Detection %	-	28%	32%	51%	68%
False Positive rate	-	0.00002	0.00002	0.002	0.000032

Experimental Evaluation



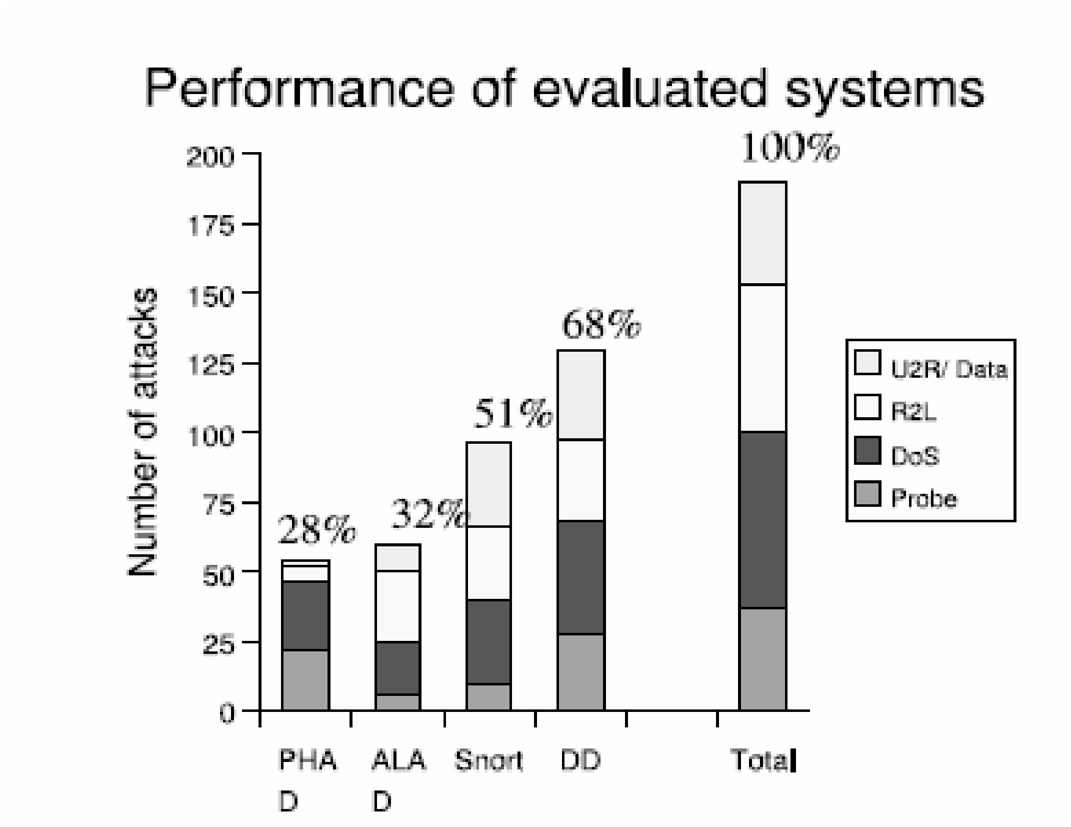
ROC curves showing the enhanced performance with fusion

Experimental Evaluation

Detection /Fusion	P	R	Acc.	Det. Perf.	AUC	F-score
PHAD	0.35	0.28	0.99	0.28	0.64	0.31
ALAD	0.38	0.32	0.99	0.32	0.66	0.35
Snort	0.09	0.51	0.99	0.51	0.75	0.15
Fusion IDS	0.39	0.68	0.99	0.68	0.84	0.50

Performance comparison of individual IDSs and DD fusion method

Experimental Evaluation



Recap

- Theoretically proved the acceptance of sensor fusion
- Validation of the data-dependent decision fusion using modified evidence theory with the DARPA dataset as well as the real time traffic has been undertaken
- The experimental results confirms the theoretical analysis

Comments?

Questions ?

Thank you