# Survey and Research Directions on Intrusion Detection in UNIX Environment

**Zafar Sultan**[†]

School of Science and Technology, University of New England, Armidale, NSW, Australia

**Summary**

Although UNIX is considered a very stable and secure platform, the development of Intrusion Detection Systems is essential as current and future generations of hackers are continuously attempting to undermine its integrity. There are few intrusion detection systems in UNIX for detecting multiple threats in a distributed networking environment. Researchers have applied different statistical models that involve data fusion. The most common and popular approaches include Bayesian theory, Dempster Shafer Evidence Theory, Parametric and Non-Parametric techniques, and Markov Chain. With few exceptions, almost all these detection models cater only for single threat. Thus, there is a genuine need for research on multisensor data fusion model in intrusion detection systems that enhance its capability to detect multiple simultaneous threats, particularly in the UNIX environment. In this paper, I'll survey existing intrusion detection system s and detection models in the literature, followed by a discussion of my research directions on intrusion detection in UNIX environment.

***Key words:***
*Multiple Simultaneous Threat Detection; Intrusion Detection Systems; Bayesian Theory; Dempster Shafer, Multisensor Data Fusion; UNIX.*

## 1. Introduction

A large number of Intrusion Detection Systems have been developed for computer security but more development is required as attackers are very shrewd these days and have developed different approaches and programmes to penetrate into computer systems and have succeeded many times in breaking all security walls. Thus hackers in fact not only have stolen valued and critical business data but also forced computer industry and businesses to develop advance software to monitor and block their attacks. As a result the companies have to spend billion dollars to develop codes for this purpose [31]. For example Microsoft spent $1.2 billion to stop Sapphire/Slammer worm [35]. Integration of UNIX with Firewall protection and CISCO technology were considered very secure systems but hackers have also broken such security measures. The way the security field is progressing, it looks like this is a continuous battle between security professionals and hackers. Hackers are in reality people familiar with all types of computer systems like cyberspace, networks, operating systems and their thousands of applications. Hackers know the gaps of information technology systems, they exploit system weaknesses and misuse their expertise to perform illegal functions on business critical systems such as stealing of important information, business secrets, damaging data or systems etc. etc. The hardest problem in tracking these types of attack, their origin and quantity of damage depends upon attacker’s software and techniques. Hackers may attack from multiple sites and hide their identity by continuously changing their IP addresses. Sometime they do physical damage to the systems or their applications, but if they just steal important information, the security experts may be unaware of it for many months until they apply a new security update or hackers does any physical damage to any process or data of that particular business [38].

False positives and false negatives are additional issues in computer security and also in UNIX systems. False alarm results because alarms are set at low levels of security. The present monitoring and intrusion detection system analyze data taken from system processes, memory, CPU, disk utilization and log messages and track or batch or log files. Attacks are checked based on pattern matching with existing situations of the processes and systems attributes [12].

A vast majority of intrusion detection system have Multisensor data fusion techniques that detect only single threats at a time. False alarms are classified as misuse and anomaly detection. In Pattern matching algorithms the events that do not match with the known pattern are known anomalous. Others use mathematical and statistical or machine learning techniques for anomaly detection and data fusion [32].

Manuscript received December 5, 2009
Manuscript revised December 20, 2009

The aim of this paper is the survey of Intrusion Detection System in Multisensor data fusion, its various approaches and techniques. However, the main emphasis of my research is to detect multiple simultaneous attacks in UNIX environments. My research will help in building multiple simultaneous threat detection systems for computer security in general and for UNIX environments in particular.

## 2. Survey of Intrusion Detection System

2.0 Data Fusion Approaches in UNIX

Bayesian, Dempster Shafer, fuzzy rules, parametric / non parametric and Kalman Filter are widely used data fusion techniques [6]. Chapman-Kalmogorov prediction model has also been used as an integral model with Bayesian and Dempster Shafer, inferences regarding threats, location and other attributes are made from these models. These models fuse data from the Multisensor systems on the same or different networks. Fusion model behave exactly the same way like human brain process data and take actions or decisions [24]. UNIX system's Intrusion Detection System gets their data from different sensor created by systems commands and networks packets. Data may be sniffer packets; sys log files, SNMP traces, system messages and other similar activities of the network. Data fusion model after processing this information send its outputs in form of alarms to security people and system engineers and warn of any expected threat on a particular subnet. Though data fusion models work like cognitive approach but in fact they are not really intelligent enough to cope with different type of changes or attacks if their information does not already exist in the Intrusion Detection System database. The Langley attack lost million dollars and they could not find email bombs until their business server crashed.

The current Multiple Intrusion Detection Models are unable to auto track, identify, and block all suspected threats. Advance Intrusion Detection System is required to deliver enhanced reliability and precision in threat detection [1]. Thus additional development is required in the field of multiple sensor data fusion models of Intrusion Detection System in UNIX [27].

### 2.1 Other Data Fusion Approaches

A large amount of research work and literature is available on Multisensor data fusion of Intrusion Detection System in defense and other related fields. However, there is a little work in the field of UNIX, only few scientists worked on multiple simultaneous threat detection in UNIX. It is, therefore, a relatively new area to work on. Though a few years back UNIX was one of the secure environments from outside hackers but intruders now have broken many databases and applications in UNIX network whilst all critical business like credit cards, client profiles and financial transactions are online and need more security ever than before.

Majority of the workers used Bayesian, Dempster Shafer, parametric / non parametric and few others inference engines for Multisensor data analysis in their intrusion detection system.

Dong and Deborah (2005) [16] worked on DARPA intrusion detection system evaluation data set show in their experiments that improved threat detection rates from 75 to 94 % with their hybrid models. In another study, Dong and Deborah emphasized that hybrid model of Bayesian is the best technique to improve the intrusion detection precision for intrusion detection system.

Christos and Basil (2004) [34] worked on multiple data fusion model and concluded that the use of Multisensor data analysis increases threat detection accuracy. They used a Bayesian and Dempster Shafer detection engine.

Huadong Wu, Mel Siegel and Rainer (2002) [17] identified relationship between Bayesian and Dempster Shafer theory and compared with the probability method and concluded that combined mathematical inference models will be a promising area for Multisensor data analysis in intrusion detection system.

A. Habib, M Hefeeda (2003) [2] and Christos (2004) [34] worked on DoS in an intrusion detection system and found a increase in precision by using classical Bayesian methods for data analysis.

Diego Zamoni (2000) [14] used a pattern matching detection model to detect new attacks, however, he did not mention any particular fusion model in his experiment.

V. Chatzigiannakis et al. (2002) [38] found that their fusion model is more effective than single metric analysis. They used Principal Component Analysis for Multisensor data fusion for intrusion detection.

Vladimir G et al. (2002) [39] suggested that combining a decision model is better in thereat detection precision than a Meta model in Intrusion Detection System.

Kapil K S (2000) [25] worked on Intrusion Detection System architecture and found that rule set knowledge, expert systems state models and string match are useful parameters in the development of an advance threat detection model.

Hervaldo S. Carvalho, et al. (2003) [19] experimented in "General Data Fusion" and concluded that data fusion architecture can be applied to different domain and applications.

Hugh Durrant-Whyte and Mike Stevens (2005) [20] described mathematical model for their fusion model. They analysed data using Kalman Filter and theoretical methods derived from Bayesian theorem.

S Terry Brugger (2004) [33] worked on offline data fusion model, used data mining approach in her INTRUSION DETECTION SYSTEM. However, she did not produce any particular model during her experiment.

In the view of all above literature reviews, it is obvious that there is enough material on Multisensor data fusion models of Intrusion Detection System. However, very little was reported in the UNIX. And almost negligible work was found if we search material or study on multiple simultaneous threat detection in the field of UNIX.

The above approaches can be summed up into following points;

o A large quantity of work was found about Multisensor data fusion models of Intrusion Detection System using Bayesian, Dempster Shafer inference models. Literature regarding development of Multisensor data fusion models based on Set Covers rule set is very weak and does not exist in the area of UNIX networks.

o Only a few of papers speak about theoretical knowledge Multisensor data fusion models in UNIX environments. This means theory of inference for data fusion in UNIX is at its infancy stage.

o Mathematical / Statistical approaches such as Bayesian, parametric / non parametric and Dempster have been the dominant data analysis tools for Multisensor data fusion models in Intrusion Detection System in general and especially for UNIX environment.

o Almost all of the Multisensor models with few exceptions were used to detect single threat detection.

o Testing, development and comparison of data fusion models in multiple threat detection are few in number.

## 3 Research Directions

In this research, I'll identify a multiple simultaneous threat detection model. This model will be a hybrid of Bayesian and Dempster Shafer theory of inferences with Set Cover theory. The new model will increase the precision in threat detection and reduce the volume of false alarms in UNIX environment. The use of the model will assist in decreasing the data security expenses, particularly web based businesses. Researchers will get also benefit for future Intrusion Detection System developments in UNIX.

The new multiple simultaneous threat detection model will be able to detect more than one threat simultaneously. Another advantage is that the results of this research can be applied in high speed networks like cyberspace. There are also some additional situational parameters that will be generated as a result of this work such as high level architecture of multiple threat detection model, identification of proper Multisensor environment based on hybrid model, and identification of middle tiers of the research.

### 3.0 Original Contributions

This study is different in many ways from other's work in Multisensor data fusion in Intrusion Detection System development. The idea of using Set Covers for data set rule in hybrid Bayesian and Dempster model is novel as no one has used this before. In order to detect multiple threat detection, researchers are making efforts to develop suitable data fusion model based on advance mathematical and statistical techniques. However, most of the models detect single threats, few models are advanced but the work in multiple threat detection is rare in UNIX environment [36].

UNIX is a leading operating system for critical business databases and applications in the computer business. Therefore, UNIX security is really an important issue in the industry. Companies spend million dollars every year to track them. This happens because of incorrect signature. Weakness' in security and false alarms cause excessive process cost for the businesses.

Development of additional security measures requires comprehensive security knowledge of the systems. New security aspects will be combined with the existing Multisensor data fusion model to achieve more precision and reliability in the new hybrid model [7].

### 3.1 Novelties of this Research

This research, in fact, is a step forward that will addresses the additional precision in multiple threat detection process as compared to the existing threat detection approaches in UNIX and it is different in many ways from other's work in Multisensor data fusion in IDS development.

1) I'll used hybrid model of Multisensor data fusion comprised of basic Bayesian, Dempster Shafer, and Extended Dempster Shafer theory of inference in multiple simultaneous threat detection of UNIX environment.

2) Set Cover as a middle tier data fusion tool in hybrid Bayesian and Dempster model is a novel approach as no one has used it before.

3) Generalized Evidential processing (GEP) presents a better evidential combination and separate propositions and the decisions. GEP will be implemented very first time in a distributed Multisensor network of an UNIX environment.

### 3.2 Approach and Methodology

In a large number of Multisensor data fusion model, Bayesian and Dempster Shafer have been used for data analysis. Most of the existing work was in single threat detection [17]. Only couple of researchers tried to focus on multiple threat detection without using Set Cover theory. Set Cover has been identified as a new area which can be used to prioritize and schedule rule set on certain criteria in the fusion process. On this topic there are only a few papers available in the UNIX environment, therefore, it is difficult to compare literature on Multisensor threat detection in UNIX [14].

Statistical /mathematical models such as Parametric / Non Parametric, Bayesian, Dempster are the most commonly used theory of estimation in data fusion in UNIX environment. These experimental models were used in detection of DoS, email bombs and buffer overflow attacks with many limitations.

This research is about identification of multiple simultaneous threat detection models. Data fusion will be done using hybrid model of Bayesian and Dempster. Set Cover will be used to identify data groups and scheduling [3]. The hybrid model will provide a increase in precision of threats detected and additional theoretical and technical knowledge about multiple threat detection for computer security, especially for UNIX [15].

The success of simulating good "Multiple Simultaneous Threat Detection model" depends upon the quality of test environment that will be set up to achieve this result. To set up an environment, I'll develop a software programme in Perl and shell script to identify type of attacks and monitoring and evaluation of the results. The main emphasis would be to track multiple simultaneous threats based on Multisensor data fusion model [6].

Multisensor data of the test environment including Java, database, and web servers will be collected, analysed with Bayesian and Dempster inference models.

As it has already been discussed before that the multiple simultaneous threat detection is a future prospect for Intrusion Detection System development in the UNIX environment. In order to make it a useful work, I'll identify the following parameters in my research;-

Identify and implement multiple simultaneous threat detection models targeting future Intrusion Detection systems

identify high-level model/architecture that can address Multiple Simultaneous attacks in UNIX

Identify proper Multisensor data environment to use in the fusion model

Identify and Implement and run testing environment for the data fusion algorithms in multiple simultaneous threat detection systems

Identify if my new research on multiple simultaneous threat detection model works well or not? And provide all possible reasons in any case

I'll provide an excellent comparison of models based on different mathematical inferences

When I complete all possible tests of my research model. The results of these tests will provide guidelines that which multiple simultaneous threat detection model work well; I will give a detail overview that how it works. I do hope if it work well in my tests then it will also work well in other environments like Wintel etc. as long as they have similar network layers and protocols. If these models do not work, I will explain about the problems and issues with my model. The expected problems in preventing my research models could be multiple threat detection architecture, Multisensor data fusion model, network or application layer data or whatever. I'll also give theoretical description why my research model works or do not work.

## 4. Conclusions

Survey of Intrusion Detection System indicates that Bayesian, Dempster Shafer, fuzzy rules, parametric / non parametric and Kalman Filter are widely used data fusion techniques. However, the existing Intrusion Detection System's data fusion models are unable to identify simultaneous multiple threats. Thus in order to get better results, there is a requirement for advance development in this area particularly, multiple threats' detection in UNIX environment. .

According to research directions, the proposed models and its optimization will be only achieved through a combination of mathematical and statistical models. Theory of combination in data set rules also known as Set Covering. Set Covering problems are rules of combination and scheduling. Bayesian and Dempster Shafer and Set Coving are all developed models that can assist in determining a more precise model for multiple simultaneous threat detection in distributed UNIX environment. Latest approach to apply these models will be referenced in this research. I've covered this in literature review. Further details will also be given toward end of this research.

This research will transform the Intrusion Detection problem into a Set Covering Problem by applying advanced and appropriate statistical and mathematical modelling and transformations. However, optimization is not guaranteed before I get final results. Efforts will be made to centralize the data fusion process using leaning methods. Lean is the worlds most successful process optimization standard and is becoming very popular in IT industry.

Finally in order to achieve high level precision in my multiple simultaneous threat detection models I'll use mathematical and statistical techniques known as Bayesian modelling, Dempster Shafer optimisation and extended Dempster Shaffer Theory.

## 5. Acknowledgments

I would like to thank Dr. Paul Kwan (my principal supervisor) for his suggestion on the initial research direction and his valuable comments on how to structure this paper.

I also thank to Thomas Kerin, IBM, Capacity Planner, Melbourne, Victoria, Australia for his suggestions.

## 6. References

[1] A. Bendjebbour, Y. Delignon, et al., "Multisensor Image Segmentation Using Dempster-Shafer Fusion in Markov Fields Context", IEEE Transaction on GeoScience and Remote Sensing, Volume 39 Issue 8, August 2001; pp. 1-10

[2] A. Habib, M. Hefeeda, and B. Bhargava. Detecting service violations and DoS attacks. In NDSS Conference Proceedings. Internet Society, 2003; pp. 439-446

[3] Aickelin U (2002): 'An Indirect Genetic Algorithm for Set Covering Problems', Journal of the Operational Research Society, 53(10), pp. 1118-1126.

[4] Alexei Makarenko, Hugh F. Durrant-Whyte: Decentralized Bayesian algorithms for active sensor networks. Information Fusion 7(4): 2006; pp. 418-433

[5] Ambareen Siraj et. al. Intrusion Sensor Data Fusion in an Intelligent Intrusion Detection System Architecture. Proc. of a conference 2004. Hawaii: International Conference on System Sciences; pp. 1-10

[6] Ben Grocholsky, Alexei Makarenko, Hugh F. Durrant-Whyte: Information-theoretic coordinated control of multiple sensor platforms. ICRA 2003: pp. 1521-1526

[7] Braun, J. (2000) Dempster-Shafer theory and Bayesian reasoning in multisensor data fusion, Sensor Fusion: Architectures, Algorithms and Applications IV; Proceedings of SPIE 4051, pp. 255–266

[8] Burroughs J , Wilson F and Cybenko V (2002 ) Analysis of Distributed Intrusion Detection Systems Using Bayesian Methods ;pp. 329-334

[9] COMPUTER EMERGENCY RESPONSE TEAM (CERT). CERT advisory CA-2001-35 recent activity against secure shell daemons, Dec. 2001. accessed on.. http://www.cert.org/advisories/CA-2001-35.html. Accessed 24 July 2002

[10] Cooper, M., Miller, M. (1998) Information gain in object recognition via sensor fusion, Proceedings of the International Conference on Multisource-Multisensor Information Fusion (Fusion '98), pp. 143–148

[11] Cuppens F et al (2002), 'Correlation in an Intrusion Process', Internet Security Communication Workshop (SECI'02); 2-7

[12] D. Hall. Mathematical Techniques in Multisensor Data Fusion. Artech House, Norwood, Massachusetts, 1992;pp. 99-105

[13] DANYLIW, R., AND HOUSEHOLDER, A. CERT advisory CA-2001-19 "Code Red" worm exploiting buffer overflow in IIS indexing service DLL, August 2001. accessed on.. http://www.cert.org/advisories/CA-2001-19.html Accessed 13 August 2002.

[14] Diego Zamboni. "Doing intrusion detection using embedded sensors" CERIAS Technical report 2000-21, CERIAS, Purdue University, West Lafayette, IN, Oct, 2000;pp. 1-9

[15] Don Koks and Subhash Challa, 2005, An Introduction to Bayesian and Dempster-Shefer Data Fusion; pp. 1-52

[16] Dong and Deborah (ACM 2005) Alert Confidence Fusion in Intrusion Detection Systems with Extended Dempster-Shafer Theory; pp. 142-147

[17] 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;1-6

[18] Hall, D., Garga, A. (1999) Pitfalls in data fusion (and how to avoid them), Proceedings of the Second International Conference on Information Fusion (Fusion '99), pp. 429–436

[19] Hervaldo S. Carvalho, Wendi B. Heinzelman, Amy L. Murphy, Claudionor J. N. Coelho 2003 "General Data Fusion Architecture" University Of Rochester; pp. 1-8

[20] Hugh F. Durrant-Whyte: Data fusion in sensor networks. IPSN 2005, pp. 545-565

[21] J. Burroughs, L. F. Wilson and George V. Analysis of Distributed Intrusion Detection Systems Using Bayesian Methods, presented at IPCCC 2002, April 2002; pp. 142-147

[22] J. Kulik, W. Heinzelman, and H. Balakrishnan, "Negotiation- Based Protocols for Dissem inating Information in Wireless Sensor Networks," Wireless Networks, Vol. 8, 2002, pp. 169-185

[23] J. Llinas, Hall, D.L. "An Introduction to Multisensor Data Fusion" V6, Issue 6, 1998, pp. 537-540

[24] J. R. Boston, "A Signal Detection System Based on Dempster-Shafer Theory and Comparison to Fuzzy Detection", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Volume 30, Issue 1, February 2000;pp. 1-51

[25] Kapil Kumar S, 2000 Intrusion Detection and Analysis, University of British Columbia

[26] Krieg, M.L. (2002) A Bayesian belief network approach to multi-sensor kinematic and attribute tracking, Proceedings of IDC2002;pp. 1-50

[27] Lawrence A. Klein, "Sensor and Data Fusion Concepts and Applications" (second edition), SPIE Optical Engineering Press, 1999, ISBN 0-8194-3231-8; pp. 1-252

[28] Lee W, Fan W, et al (2002), 'Toward Cost-Sensitive Modelling for Intrusion Detection and Response', Journal 3. of Computer Security, Vol. 10, Numbers 1, Volume 2, 2002; pp. 5-22

[29] Lenny Zeltser " Intrusion Detection Analysis: A case Study, accessed on.. http://www.zeltser.com/intrusion-detection-analysis/

[30] Ma, Bing 2001 "Parametric and Non Parametric Approaches for Multisensor Data Fusion" PhD thesis, University Of Michigan;1-212

[31] Ning P, Xu D, Healey C and Amant R (2004), 'Building Attack Scenarios through Integration of Complementary Alert Correlation Methods', 11th Annual Network and Distributed System Security Symposium, pp. 97-111

[32] Rehman R, (2003) 'Intrusion Detection System with SNORT', accessed on.. http://www.snort.org/ ; pp. 1-288

[33] S Terry Brugger, 2004 Data Mining for Network Intrusion Detection – PP.8/55 accessed on.. www.bruggerink.com/~zow/papers/dmnid_qualpres.pdf

[34] Siaterlis C and Maglaris B (2004), 'Towards Multisensor Data Fusion for DoS detection', Proceedings of the 2004 ACM symposium on Applied Computing; pp.1-8

[35] SPAFFORD, E. H. The Internet worm incident. Tech. Rep. Purdue Technical Report CSD-TR-933, Department of Computer Science, Purdue University, West Lafayette, IN 47907-2004, 1991. LEM OS, R. Counting the cost of slammer, Jan. 2003; pp. 1-19

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

[37] Tim Bass and Dave Gruber. a glimpse into the future of id. Usenix. 18 Aug 2005. accessed on.. http://www.usenix.org/publications/login/1999-9/features/future.html.

[38] V. Chatzigiannakis, A. Lenis, C. Siaterlis, M. Grammatikou, D. Kalogeras, S. Papavassiliou & V. Maglaris 2002, Distributed Network Monitoring and anomaly Detection as a Grid Application ;pp. 1-13

[39] V.Gorodetski, O.Karsaev, I.Kotenko, and A.Khabalov. "Software Development Kit for Multi-agent Systems Design and Implementation". In B.Dunin-Keplicz and E.Nawareski (Eds.) "From Theory to Practice in Multi-agent Systems". Lecture Notes In Artificial Intelligence, vol. 2296, 121-130, Springer Verlag, 2002;pp. 121-130

## 7. Author's Profile

**Name**: Zafar Sultan
**Title**: Unix and Storage Engineer
**Education:**- Master in Computer Science and PhD student in IT at University of Armidale, NSW, Australia
**Skills**:- Over 18 years skills in UNIX and Storage environments with companies like IBM, Telstra, Compaq Computers, Vodafone etc. etc.
**Publications**: 8 international and 13 national publications
**Citizenship**: Australian