Distributed Network Forensics Framework: A Systematic Review

Gurpal Singh Chhabra
Thapar University, Patiala

Prashant Singh
Thapar University, Patiala

ABSTRACT
Network forensics is a branch of digital forensics, which applies to network security. It is used to relate monitoring and analysis of the computer network traffic, that helps us in collecting information and digital evidence, for the protection of network that can use as firewall and IDS. Firewalls and IDS can't always prevent and find out the unauthorized access within a network. This paper presents an extensive survey of several forensic frameworks. There is a demand of a system which not only detects the complex attack, but also it should be able to understand what had happened. Here it talks about the concept of the distributed network forensics. The concept of the Distributed network forensics is based on the distributed techniques, which are useful for providing an integrated platform for the automatic forensic evidence gathering and important data storage, valuable support and an attack attribution graph generation mechanism to depict hacking events.

Keywords: Network security, Distributed framework, Agent, Proxy

1. INTRODUCTION
A computer network is a telecommunication network which allows computer to exchange data. A lot of digital devices like PCs, notepads and terminators are connected with in network via the wired or wireless links. main motive is to make network secure. Firewall and IDS is used for securing a network, but the strength of these measures is restricted. This paper mainly emphasize on the deterrence and uncovering over the network intrusion. There is a large amount of crucial information transferred through the Internet like, financial information, electronic money and digital signature. In the current system, reaction of an event always comes out after the attack has happened, which always leads to loss of some important data. For the network security a approach is based on defensive mechanisms, which is used to thwart the network crimes. In this approach attacker usually emphasize to detect the network vulnerabilities and then make a cover to obstruct these vulnerabilities. There are some unique methods to obstruct malevolent communication from outside. These methods are not meant to obstruct the network criminals, but to recognize them, and to accumulate significant proof of these crimes. There is a provision to punish the cyber criminals for theirs illegitimate activities, thereby providing a restriction to online crime. In the current scenario, network forensic investigations held manually, which used to check different types of logs and with the help of these logs, recreate the actions that led to an attack. Manual methods of forensics are effective for small data. But in the case of large amounts of data, these manual techniques make the investigation infeasible. There is a need of infrastructures, which helps to collect, store and analysis data for forensic purpose. Most of the time logging mechanism and audit trails reserved for hosts. These mechanisms are applicable to the local host and tell us that what is happening on the local machine. These mechanisms do not show the relationship between local hosts and other network nodes. The current network forensic model supports the process of forensic procedure whenever any illegal action occurs. Due to the quick disappearance of digital evidence, this type of model cannot find the crucial evidence. To resolve all these problems, there is a need of a distributed system that is used to monitor massive networks. There is a need of a distributed system for gathering and accumulate the forensic data concurrently that is located at different locations. We collect and store data in this manner that helps in forensic analysis properly. Network forensics is a step ahead of computer forensics. Floor-Net is a good network forensic system, but there is a limitation, it does not support the cooperative attack attribution. It works as a network data group and inquiry system. This paper presents a framework of a distributed network forensics system. Primary objective of this system is to provide proper method for data collection and storage in order to analyze this information for the forensics purpose. This system should provide significantly evidence gathering, and immediate reaction to network criminals. Work objective of this framework is to provide three types of contribution for the network forensics: data storage method based on efficient distributed storage, pre-selection and specific data compression, distributed techniques supports the general framework for easy integration of known attribution methods and effective cooperation within the system, an attack attribution graph generation mechanism to illustrate hacking procedures.

2. LITERATURE SURVEY
Youngling Tang and Thomas E. Daniels (2005) improvised a simple framework for the purpose of distributed forensics. It provides a network forensics framework which depends on the distributed techniques. These are used to provide a unified platform for automatically collection of forensic evidence and the storage of data in an efficient manner, and also support the integration of well-known attribution methods easily. It does provide efficient support and an attack attribution graph generation mechanism, which helps to verify hacking actions.

Tang Hong, Zou Tao, Jin Qi, Zhang Jianbo (2011) suggests a distributed framework for network forensics, which is used to confine data as a digital proof, and store the digital information. This information comes through the network. Distributed data agents and forensics canter are used to design the architecture of the framework. Firstly, extract and compress the information from all required network transmissions specifying address of the suspected host, which involves in illegal transmission or wrong information in the network. This classification of information depends on the data proof.

Ren Wei proposed a Model of distributed Cooperative network forensics system. The term distributed cooperative forensics system means provide an effective integration between different security products. Due to the lack of integration between the different security products, leads to

the incomplete coverage of information, this can't give the complete view of the network misuse behaviour. Network forensics is a new approach which follows two aspects, incident investigation and emergence response, which helps to enhance the security of networks in many different aspects.

Wei Ren, Hai Jin(2005) proposed distributed agent based real time network intrusion forensics system architecture design, know that in the previous network forensics, system both the process forensics and investigation always occur after the attack take place, which leads to drop some valuable instantaneous evidence. There is a requirement of a system which helps to integrate the analysis of the log, with the help of audit system and network traffic provide effective monitoring of the traffic.

Tao Li, Sunjun Liu, Jianhua Zhang, Caiming Liu (2007) proposed a dynamic network based on immune agent. Current network forensics systems are not dynamic in nature. And these do not depend on real-time phenomena. The concept of dynamic network forensics model illustrates with the help of artificial immune theory and multi-agent theory, to solve the problem of storages referred to as DNF, is found here. This paper, describes architecture of the model, definitions and its components, and all these things depend on the immunity theory. With the help of this model capable to assure the authenticity, integrity and authorization of digital proof.

Emmanuel S. Pyle, R. C. Joshi,Rajdeep Niyogi (2010) improvised the network forensics framework. In this paper there is a description of various network forensics frameworks. A standard process model for network forensics improvised on several phases of the forensics process. What is the Definition, what the thing is that motivate for network forensics are undoubtedly mentioned here. And it highlights the major challenges occur during the process of forensic investigation.

Kulesh Shanmugasundaram, Nasir Menon, Anubhav Savant, and Herve Bronnimann proposed Fornet: A distributed Forensics Network. This paper introduces Fornet, which is based on the distributed technique such as a network logging mechanism. The logging mechanism is used to assist digital forensics over WAN.

3. BACKGROUND

Network forensics is being researched for a long time, but it still looks new and lots of concerns are ambiguous. Here the definition, categorization, and motivation for this forth coming field are discussed below.

3.1 Motivation of network forensics

There is the large number of cyber attacks or cyber crime happening now days, which affects many organizations in malicious way. So there is a need of network forensics to overcome these problems. This concludes that, this is the main reason behind that drive the network forensics to investigate these attacks. Firewalls and intrusion detection systems are the defensive approach. The main objective of this system is to investigate the cyber crime. There are many approaches exist for the investigation process. Firewall and IDS is used for securing a network, but the strength of these measures is restricted. This paper mainly emphasize on the deterrence and uncovering over the network intrusion.

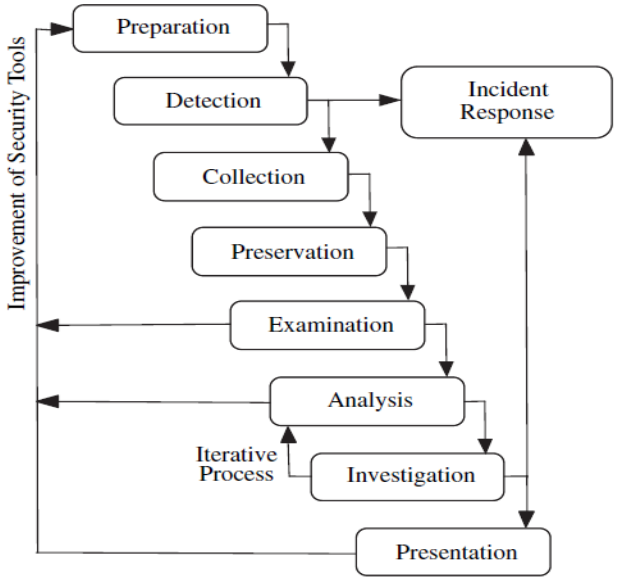

Fig 1: Generic Process model of network forensics

3.2 3.2 Generic process model

3.2.1 Detection

The term detection refers to detecting unauthorized events and anomalies.

3.2.2 Incident Response

It is another important phase of the process model. This phase is to inform about any unauthorized and illegal action that has happened through intrusion detection system.

3.2.3 Collection

It is the main important phase of the generic price model, because all the things depends upon the effective data collection. We collect data in such a way that it gives us to maximum information regarding illegal activities, and it does not affect the privacy of the victim.

3.2.4 Preservation

After the collection of data stored on a backup device. The collected data resides in the form of the traces and logs. Preservation also refers to a hash of all the data. Copy of data has been analyzied instead of original data.

3.2.5 Examination

After the preservation of data move a step ahead, that is examining data. Data consists in the form of traces and logs. The data obtained from the various places integrated and fused into a single data unit on which analysis could be performed.

3.2.6 Analysis

It is also an important phase of the process model. In this phase attack pattern is analyzed from obtained data and matched this pattern to the attacker pattern with the help of existing statistical, soft computing and data mining approaches.

3.2.7 Investigation

The main motive of this phase is to find out the path between a victim network and the point of origination through intermediate system and communication pathways.

3.2.8 Presentation

It is the last and important phase of this process model. The presentation of the information and digital evidence must be unambiguous and in an understandable language for legal personnel.

3.3 Benefits of Distributed Framework

Cyber communication is distributed in nature, Out of which some network which are used for forensics evidence collection, lack effective attack attribution. Network forensics framework based on the distributed techniques which provide an integrated platform for automatic forensic evidence gathering and efficient storage of data, which support easy integration of known attribution methods, effective cooperation and an attack attribution graph generation mechanism to illustrate hacking procedures. Distributed framework for network forensics which attempt to confine and store the digital proof of the data disclose via network. The frameworks of distributed agent-based real time network intrusion forensics system, is kept in LAN environment. Current network forensics systems are not dynamic and not real-time. Distributed system framework is dynamic in nature. There's a need to collect these logs, fuse them and analyze on the central server.

4. TYPES OF NETWORK FORENSICS FRAMEWORK

Many forensics, network forensics improvised with various phases in the previous section.

4.1 A distributed system based framework

Internet and LANs are distributed in nature. The events related to the network attack are logged in as client at various location. For carrying the digital proof which obtained from the crime site there are some network systems and the current systems also lack efficient attack attribution.

Shanmugasundaram et al. (2003) suggests Fornet a distributed network logging mechanism. Which is used to add cyber forensics over internet. There are the two components, SynApps and a forensics server. SynApps is used to perform summarization and remember network events. And the second is forensics server, which act as a concentrated authority and handles a particular domain of control a set of SynApps in that domain.

4.2 Soft computing based framework

The soft computing approach comes into picture after the collection of data. It is used to examine the collected data. A step ahead, it classified the attacked data. In this approach fuzzy and neural network tools are used for the validation of attack occurrence.

Kim et al. (2004) develops a fuzzy logic based expert system for the analysis of computer crime and provide the facility to make digital evidence automatically. This analyzed information is used for forensic expert and for the reduction of cost and time.

4.3 Honeypot based frameworks

Honeypot framework is a type of trap, which attract the attackers. It helps us in understanding the process and methodology of the attacker. This system help to improve the defensive mechanism.

Honeytrapes (Yasinsac and Manzano, 2002) developed a tool that confused the attacker. It is a type of fraud to attract the attacker for the purpose of gathering information about illegal activities and learn their techniques to generate the defensive mechanism against these black hat activities. Honeypot or Honeynet is based on the concept of Deception. It attracts the attacker to enter a host through a weakness of a system. After the attacker entering into the honeypot, data is captured to identify and trace his action.

4.4 Attack graph based framework

Wang and Daniels (2008) suggest a graph based approach for the network forensics analysis. There is a facility of evidence graph model for automated reasoning and effective evidence presentation. This framework contains six phases: collection of evidence, evidence, pre-processing of evidence, based on attack knowledge, asset knowledge based, evidence graph manipulation, and attack reasoning.

4.5 Formal method based module

Ranchos et al. (2008) implements a system for Digital Forensic in Networking (Dig For-Net), this system helps us to examine security events and make clear the attacker steps, followed by the attackers. Dig For-Net uses the expert knowledge of intrusion response. This system integrate examined results performed by the IRT on a compromised system with the help of Incident Response Probabilistic Cognitive Maps (IRPCMs).Which offers a proper method framework, that helps to recognize potential attack patterns with the help of Investigation-based Temporal Logic of Actions (I-TLA).

4.6 Aggregation framework

Network forensic analysis consists of several phases. There are many security tools. These tools can be used for particular phase. The main aim of this framework is to handle the strength of these tools rather than constructing a new tool from scratch. Almulhem and Traore (2005) suggest a Network Forensics System (NFS) that account data either host level or network level. The system contains three main modules – marking, capture and logging.

5. PROPOSED DISTRIBUTION FRAME-WORK

This is the proposed distributed network Forensics Frameworks. It contains an agent, proxy and proxy servers. The internet and LANs are distributed in nature. So there is a high chance of network attacks events. All the existing systems only support the collection and storage of forensic evidence and they do not emphasize on the effective attack attribution mechanism. This distributed network forensics, system depends on the distributed techniques. Distributed system is not used only to collect and store the forensics data, but also used simultaneously gather and accumulate data at various locations and manage the critical information which is very useful to support forensics analysis.

5.1 Agents

Agents and proxies are the basic units of this network forensics system. Agents provide help to examine, design and realization of distributed system. Agents usually monitor the network. The framework of data agent contains four modules. The first module of data agent refers the data capturing module. In this module agent collects the data which are useful for all target hosts. The second module is concerned with data processing. It is used to process the relevant information, and take out the significant information text from the initial network data. A step ahead, third module is used to record the attribute of communication called as log model. It records communication between the intended hosts and the

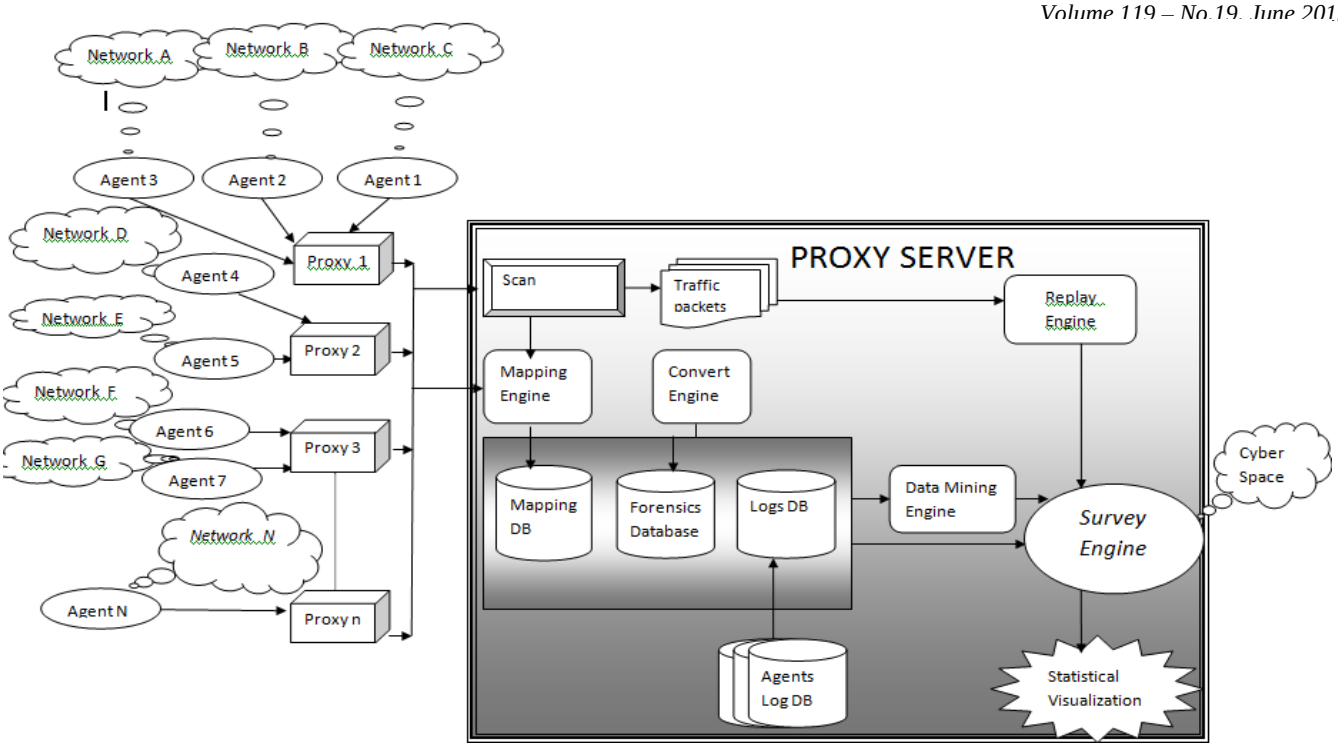

Fig 2: Network forensics distribution framework

other terminal. And the last module is called communication module that is used to present the evidence data. The agent is used to hide the complication of the network infrastructure and makes the heterogeneity of the networks. Agents are accountable for data gathering and easy analysis.

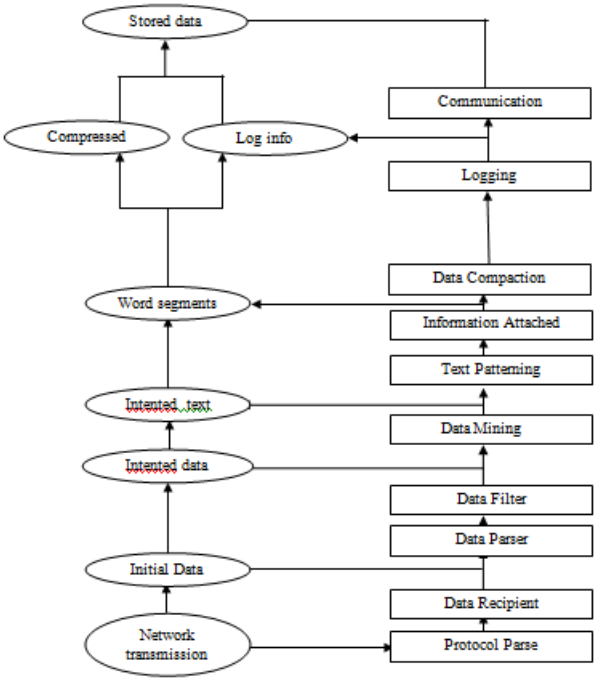

Fig 3: Workflow of Agent

The original data are carried via the network on internet. Firstly, data agent comes into the picture, and begins to analyze the protocol of network such as ICMP, and TCP/IP protocol transmission. Because it contains the user's text content information. The other useless data will be dumped. The initial data consists of multimedia or document which is not readable. So it should be analyzed again to be differ. The data agent is used to filter the irrelevant data. The target data categorized in many categories like pdf/doc/txt. The term data mining is used to mine the relevant information from the intended data. Extracted information will be in the readable form which consists of all information about the target data. After the data mining phase the target text has to be split into the word segment.

After the data mining phase the target text has to be split into the word segment. In this phase, append some extra information such as IP/Mac address. For the compression of word segment used a bloom - filter algorithm. Log information files are used to store the information regarding the data such as the time and the address of the hosts. The last step of the proxy server is to precede the compressed information and the log information which is resides in the database for forensic purpose in the future.

5.2 Proxy

Proxies are connected to each other and also connected to the agents. Proxy acts as a server which receives the data from the agents. Communication among the proxies may be difficult, due to the complexity of communication among the proxies. For this purpose use COBRA which is a communication mechanism don't use direct socket connection for the communication among the proxies. All proxies can work together as a single node with the help of COBRA. When the forensics request comes from the agents to a particular proxy, then remote service is activated on related proxies. It is based on the distributed technique. So the environment of the system will be scalable and open. This system is applicable for both large networks and small networks due to the scalability feature. In case of large network more proxies and for small network less proxy. This system provides an environment; with the help of this environmental system that uses different data analysis techniques to ensure the attack imposition. Which depends upon the different situation and gather the imposition result as the forensic proof?

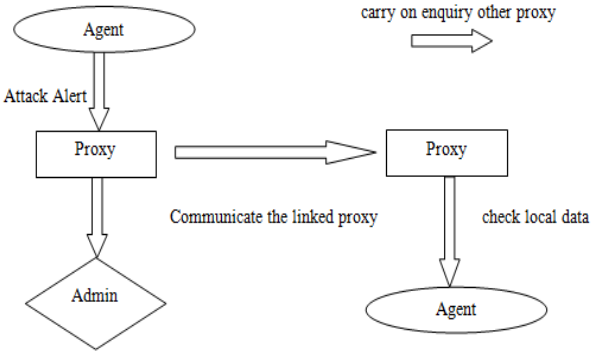

Fig 4: Process of Proxy

In this figure proxies and agents are connected together and a proxy is connected to the other proxies via the alert information. When an attack occurs, then the agent sends a message to the related proxy and proxy sends attack alert message to the admin. When admin gets this alert message, then application of Corrective action is done.

5.3 Proxy Server

Proxy server act as a server which receives data from the agent. And agent acts as clients which are responsible for the data collection and analyze the data in the initial situation. There are the methods for collection of both types of data active and passive. Proxy server responds the queries which arrive from the data agent.

6. CONCLUSION

This paper present a distributed framework for the network forensics. There are many benefits of this framework. The first advantage of this system, it can capture the data from the various locations over the internet. This data is dynamic in nature. It can capture unlimited data except the failure of distribution of data agent. In this paper, mentioned here various types of existing network forensics frameworks. It can collect the digital evidence from the distributed host, which are placed in various remote locations. All data monitored, host and investigation will be stored simultaneously as the evidence.

7. REFERENCES

[1] Y. Tang and E. Thomas, Daniels, 2005. A Simple Framework for Distributed Forensics In: Proceedings of the 25th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW'05) IEEE.

[2] T. Hong; Z. Tao; J. Qi; Z. Jianbo, 2011. A Distributed Framework for Forensics Based on the Content of Network Transmission, Instrumentation, Measurement, Computer, Communication and Control, 2011 First International Conference on , vol., no., pp.852,855, 21-23.

[3] R. Wei, 2004. On A Reference Model of Distributed Cooperative Network Forensics System The sixth International Conference on Information Integration and Web-based Applications Services, 27-29, Jakarta, Indonesia.

[4] W. Ren1,2 H. Jin 2, 2005. Distributed Agent-based Real Time Network Intrusion Forensics System Architecture Design. In: Proceedings of the 19th International Conference on Advanced Information Networking and Applications (AINA) IEEE.

[5] D. Wang, T. Li, S. Liu, J. Zhang, C. Liu, 2007. Dynamical Network Forensics Based on Immune Agent Third International Conference on Natural Computation (ICNC) IEEE.

[6] E. S.Pilli, R.C.Joshi, R. Niyogi, 2010. Network forensics framework: Survey and research challenges, Digital Investigation.

[7] K. Shanmugasundaram, N. Menon, A. Savant, and H. Bronnimann, 2003. ForNet: A Distributed Forensics Network. MMM-ACNS, LNCL 2776, pp. 1-16.

[8] Alex C. Snoeren, 2002. Single-Packet IP Traceback in IEEE/ACM Transactions on Networking (ToN), 2 Volumes 10, Number 6, December, Pages 721-734.

[9] Yin Zhang, Detecting Stepping Stones Http:// www.icir.org /vern/papers/stepping/.

[10] F. Gonzalez, J. Gomez, M. Kaniganti and D. Dasgupta, 2003. An Evolutionary Approach to Generate Fuzzy Anomaly Signatures, In Proceedings of the Fourth Annual IEEE Information Assurance Workshop, 251-259. West point, NY.

[11] M. G Noblett and M. Pollitt and L. A Presley, 2000. Recovering and Examining Computer Forensic Evidence Forensic Science Communications,28-44.

[12] Culley, 2003 Computer forensics past, present and future Information Security Technical Report, Vol.8, Vol.8 (No.2) :pp.32-36.

[13] G. Mohay. 2005. Technical challenges and directions for digital forensics[C] First International Workshop Systematic Approaches to Digital Forensic Engineering, 155-167.

[14] V. Mee, T. Tryfonas, I. Sutherland, 2006, The Windows Registry as a forensic artefact: Illustrating evidence collection for Internet usage, Digital Investigation, Volume 3, Issue 3,166-173, ISSN 1742-2876.

[15] V. Corey, C. Peterman, S. Shearin, M. Greenberg, and J. Van Bokkelen, 2002. Network forensics analysis. IEEE Internet Computing, 6(6):60-66..

[16] M. Reith, C. Carr, G. Gunsch, 2002. An Examination of Digital Forensic Models International Journal of Digital Evidence, Fall ,Volume 1, Issue 3.

[17] N. K. Jerne, 1974. Towards a Network Theory of the Immune System Annual Immunology,125(3): 373-389.

[18] S. Axelsson, 1998. Research in intrusion-detection systems: A survey, Technical Report No 98-17.

[19] S. Axelsson, 1999. The base-rate fallacy and its implications for the difficulity of intrusion detection. In Proceedings of the ACM Conference on Computer and Communication Security.

[20] B. Babcock, S. Babu, M. Datar, R. Motwani, and J. Widom, 2002. Models and issues in data stream systems. In Symposium on Principles of Database Systems, Madison, Wisconsin, USA,. ACM SIGMOD.