*Impact Factor (SJIF): 3.632*

**International Journal of Advance Research in Engineering, Science & Technology**

*e-ISSN: 2393-9877, p-ISSN: 2394-2444*

***Volume 3, Issue 4, April-2016***

# AN APPROACH TO ASSIST NETWORK FORENSICS ANALYSTS

**Yogesh Pandey[1], Ravi K Sheth[2]**

[1]*M.Tech (Cyber Security), Raksha Shakti University, Ahmedabad*
[1]*yogeshakayo@outlook.com*
[2]*Assistant Professor (IT), Raksha Shakti University, Ahmedabad*
[2]*rks.it@rsu.ac.in*

***Abstract— Network forensics is a critical area of research because, in the digital age, information security is vital. Network Forensics can help us to determine whether a network is truly under attack or a user has carelessly installed an untested patch or some custom scripts. A lot of time and resources get wasted in determining whether a bug in a custom program or an untested open-source program caused the "attack." So, Network Forensics can reduce this time and resources and can minimize downtime of a network. One must appropriately differentiate malicious network traffic from normal network traffic based on the patterns in the data transfers. With confidential information such as social security numbers, credit card information, and government records stored on a network, the potential threat of identity theft, credit fraud, and national security breaches increases. In this paper we are presenting a review of how we can reduce the network traffic data carved for Forensics by implementing Piecewise Polynomials to save memory and also by implementing NetStore (a storage infrastructure) for storing our captured data and access those data in a fast manner with simplified query options so that a Forensics Analyst can have a good resource for investigating cyber-crime across the network.***

***Keywords— Cyber-crime, Data, Forensics Analyst, Information, Network, Network Forensics, Network Traffic Piecewise Polynomials, Security.***

## I. INTRODUCTION

In this era of Digital world, we are witnessing the exponential growth of Internet with the increase in the number of smartphones, tablets, and other PDAs. With the growing Internet, we are regularly noticing the increase in cyber-crimes these days. As the cases are increasing day by day, the cases are not solved within given time and also, there are a lot of technical problems faced by experts in catching the cyber criminals. Also, the cases are regularly forwarded to Digital Forensics Experts to get the traces of the attack so that attacker can be caught easily. But the Digital Forensics experts in the field of network attacks can be called as Network Forensics Analysts are using many techniques and tools to get the details of the network traffic and do the Forensics with the data they fetched. But the problem is that, do they have enough resources or efficient tools which can help them to do the Network Forensics in that manner which can give the results that much accurate as they want and also in the given time to solve the case and finally catching the criminal. The answers are not positive and this leads to the development or motivation of this review paper which focuses on the resources available to Forensics Analyst are useful for them or not and the problems faced by them in carrying out the network forensics.

To know the core of what are the problems and difficulties faced by Network Forensics Analysts to investigate cyber-crime related with the network attack, we first define what actually network forensics means. Then we will give the overview of the network traffic data storage options available with the Forensics Analyst and other problems related to the storage of the captured network logs. In the later sections, we will give the brief of the related work done in this area followed by problems with the existing system and resources to carry out network forensics. In the remaining sections, we will give the proposed approach followed by the conclusion.

### 1.1. Network Forensics

Network Forensics can be defined as the process of collecting and analyzing raw network traffic data and tracking network traffic systematically to find out how an attack was carried out or how an event occurred on a network. Network forensics refers to the scientific study of the network-based evidence, commonly applied to legal proceedings related to cyber-crime. Network forensics is a field of study and deep analysis independent of any specific cyber-crime case, and many of the scientific advances, techniques, and tools developed for the purposes of cyber-crime investigation can also be used for historical analysis, scientific exploration, and social study of network environments [9]. Network Forensics examiners must establish standard procedures for how to acquire data after an attack or intrusion incident. Typically network examiners want to find compromised machine, get them offline, and restore them as early as possible to

minimize downtime of the network. Apart from this, the application of network forensics is not limited to Forensics only it enables e-commerce and service organizations to verify transactions, including source, recipients, and data transmitted. This analysis can be used not only for troubleshooting but also for customer service. Also, it provides many other options such as comprehensive data collection, rich data analysis, network performance benchmarking, precise data recording etc. that IT engineers can use to give several benefits to the network of their organizations.

### 1.2. Overview of Network Traffic Data Storage.

The most important thing with respect to Forensics examination of the network under attack is the collection of network traffic data. The tools used for collecting them are many need no mention here but the important thing which is to be mentioned is what they actually captured, the size of the data captured and storage system available for the entire capture. We have mentioned the impact of Internet on network forensics but one thing we need to revise is that the Internet uses individual packet to transfer data from one node to other and Internet is to be modeled as a discrete collection of individual data points. Discrete processes are difficult for analysis; therefore, continuous model is preferred as they can be used for different types of analysis for e.g. extrapolation and interpolation. The storage of the packets is area of concern as huge amount of packets needs to be stored and for that we need memory which we cannot afford to increase as it will only make the task of the examiners tedious and the analysis of the collected data will take lot of time and will definitely hinder the accuracy of the results. Therefore, the collected data are examined using graphs (for e.g. Evidence Graphs), statistical methods (with the help of least square approximation) to model network traffic data and later they help in classifying different network events which are mostly complex, in the form of patterns. These patterns are stored in the discrete manner as explained above and later they are represented in a discrete manner for analysis purpose with the use of methods such as interpolation and extrapolation. The problem also arises when these packets are needed to be stored for longer periods as they can be used later to compare new traffic data and the old one to classify any network events or malicious things happened across the network. The storage of captured packet is not feasible as it will only increase the memory storage needs and therefore they are deleted so that new data can be stored [7].

## II. RELATED WORK

In several papers we have studied what Network Forensics is all about and how much important is the research on this sub-branch of Digital Forensics is since the increase of usage of Internet due to easy access of smartphone and increased network data due to this leads to an almost impossible situation when any cyber-crime takes place to distinguish the network traffic data and find out the attackers footprint for investigation purpose. For the purpose of providing some support in the area of Digital Forensics to catch the cyber criminals we have studied a lot about how can we reduce the storage of network traffic data and aid the investigators to fast access of stored data for differentiating the network traffic data to find the evidence from that about the crime takes place and catch the criminal.

There are lot existing systems which are nowadays available but they are not as much sufficient and goal oriented as the system we trying to make. Some researcher applied dynamic modelling techniques to detect intrusions using anomaly detection. This particular form of modelling was only used for identifying intrusions and not for analyzing them or conducting a forensic investigation. Few used modelling techniques to try to predict network traffic and took a polynomial approach that utilized Newton's Forward Interpolation method to predict and model the behaviour of network traffic. This technique used interpolation polynomials of arbitrary order to approximate the dynamic behaviour of previous network traffic. Wang et al.'s technique is useful for modelling general network behaviour, but using the polynomial approach for intrusion analysis is another issue. Wang et al.'s technique proved that general network behaviour can be predicted and modelled using polynomials, but did not prove whether individual network events can be distinguished and categorized through the use of polynomials [3][6][8][11].

Below we will discuss the related work done towards the efficient developments and advancements in the field of network forensics. Also, we will discuss the related work regarding the storage issues and enhancements in the systems developed for network forensics analysis.

### 2.1 . Growing Hierarchical Self-Organising Map

Growing Hierarchical Self-Organising Map (GHSOM) tries to find out the problems of application of Self-Organising Maps for Network Forensics and proposing the new approach called GHSOM. The GHSOM tries to overcome the limitation by generating a hierarchical architecture that is automatically determined according to the input data and reflects the inherent hierarchical relationships among them. Moreover, the proposed GHSOM has been modified to correctly treat the qualitative features that are present in the traffic data in addition to the quantitative features. Experimental results show that this approach can be very useful for a better understanding of network traffic data, making it easier to search for evidence of attacks or anomalous behaviour in a network environment [2].

### 2.2 . Design and Implementation of VAST

The design and implementation of Visibility Across Space and Time (VAST), a distributed database to support interactive network forensics, and libcppa, its exceptionally scalable messaging core. The extended actor framework libcppa enables VAST to distribute lightweight tasks at minimum overhead. They have shown how VAST enables forensics analysts to deal with the huge amounts of data related with cyber-crime investigations. They compose it as a fully distributed system

of lightweight using native actors embodied in the libcppa programming environment. VAST unifies in a single framework retrospective data analysis and proactive measures to automatically apply previously developed queries to events that may occur in the future [5].

**2.3 . Existing Solutions for Network Forensics Analysis**

Since network forensics depends on the analysis done on the collected evidence and the solutions are very important to collect lots of evidence and they later can be used for analysis part to finds the trails of the attacks or any other suspicious activities. In this section we will discuss the existing solutions for analysis of the network forensics which comprises of tools, systems and other methods.

**2.3.1. Diffusion and Spectral Methods using Evidence Graphs**

In this, a graph theoretic approach was taken into account with diffusion and spectral methods. These methods show potential in extraction of useful information of the attack scenario and very efficiently can characterize the attack case using their different techniques to extract the structure characteristics of the evidence graph. The evidence graph was made with the use of collected evidence from the attack locations and these graphs are then converted or processed to make different kinds of evidence graph models which are further used for the diffusion and spectral methods for the network forensics analysis. This approach has proposed a lot of potential in transforming the current ad-hoc forensic investigation process in the context of network forensics analysis into a systematic framework with well-grounded mathematical methods. This method has two important advantages in terms of latest trends in the analysis of network forensics; one is its ability to provide mathematically well- grounded approach to refine models by reverse solving the analysis results and the other is that it can utilize well established computational methods and software packages for network forensic analysis on a large scale [12].

**2.3.2. A Portable Network Forensic Evidence Collector**

This device is purposely designed with digital forensics and network forensics evidence collection in mind and is built using inexpensive embedded hardware and existing open source software. The device can be easily deployed and preconfigured in the forensics lab, and for its deployment, there is no need for exclusively trained professionals in digital forensics. The device offers three modes of operation for different live network evidence collection scenarios such as a network and single network nodes and the modes are investigator mode, server mode, and user mode and it operates at the link layer allowing the device to be apparently inserted under control between a single network node and the rest of a network. This includes the packet capturing in promiscuous mode which helps to enhance and assist evidence collection from remote network sources which can be DNS servers, FTP servers, peer-to-peer activity, websites, and etc. This device provides a lot of benefits for the forensics investigator such as easy to carry and configure, reduces time and resources, allow capturing of packets in any protocol or any network environment and does it work quietly without disturbing the surroundings[1].

**2.3.3. Network Forensics Analysis Using Artificial Intelligent Techniques**

In this work, two techniques of Artificial Intelligence were used, Artificial Neural Networks (ANN) and Support Vector Machines (SVM). They show that SVMs are more efficient than ANNs in the field of Network Forensics Analysis in three aspects: 1.Faster order of Magnitude for training time and running time; 2.Scalability; and 3. Accuracy. SVMs provide a generic mechanism to fit the surface of the hyper plane to the data through the use of a kernel function. The user may provide a function such as linear, polynomial and etc. to the SVMs during the training process, which selects support vectors based on these functions. The number of free parameters used in the SVMs depends on the margin that separates the data points, but not on the number of input features. Hence, SVMs does not a reduction in the number of features used for training process carved from the network traffic data so that over fitting problem stay at bay which is an obvious advantage in applications such as intrusion detection. Another advantage of SVMs over the ANNs is the low expected probability of generalization errors in the processing of network traffic data which is very useful in the context of network forensics analysis [10].

**2.4 . Source Address Validation Support**

The basic idea of this method is that it uses a probing technique to authenticate whether or not a sender node sent a packet with its real source address. This method also establishes where the real sender node of the packet is located by identifying the port number of the Layer 2 Switch (L2SW). In addition, this proposed system stores certain information about network users or cyber-criminal in order to deal with any cyber-crimes that may take place in future. This method provides substantial information as evidence about network users or cyber-criminal, which always indicates the true location of their source nodes. This helps Forensics examiner to identify and trace back attacks to their origin source nodes for most cases of network attack, even if much time has passed since the cyber-crime. Hence, by using this method, a cyber-criminal can no longer conceal its original node location. However, this method can be of no use when the attacker uses the IP spoofing attack to spoof the original packets and then the authentication of the source address in not possible to determine without any guarantee [4].

## III. PROPOSED APPROACH

**3.1 Overview for the Implementation of Piecewise Polynomials to approximate Network Traffic Data.**

We will implement Piecewise Polynomials for approximation of the network traffic data that we will capture using different packet sniffers available. After that we will try to implement the storage mechanism called NetStore practically to store the captured data and will see whether the data access and query is fast or not. To put it simply we will use the both idea together to get the job done. We will implement the Piecewise Polynomial in closed network and open network. To make closed network we will use two PC with Virtual Machine running on both. On first PC we will run one virtual machine and on another we will run three Virtual Machines. And for open network there is Internet and for this we will capture the packets of at least three websites. After that we will use Piecewise Polynomials to approximate the whole data that we will capture.

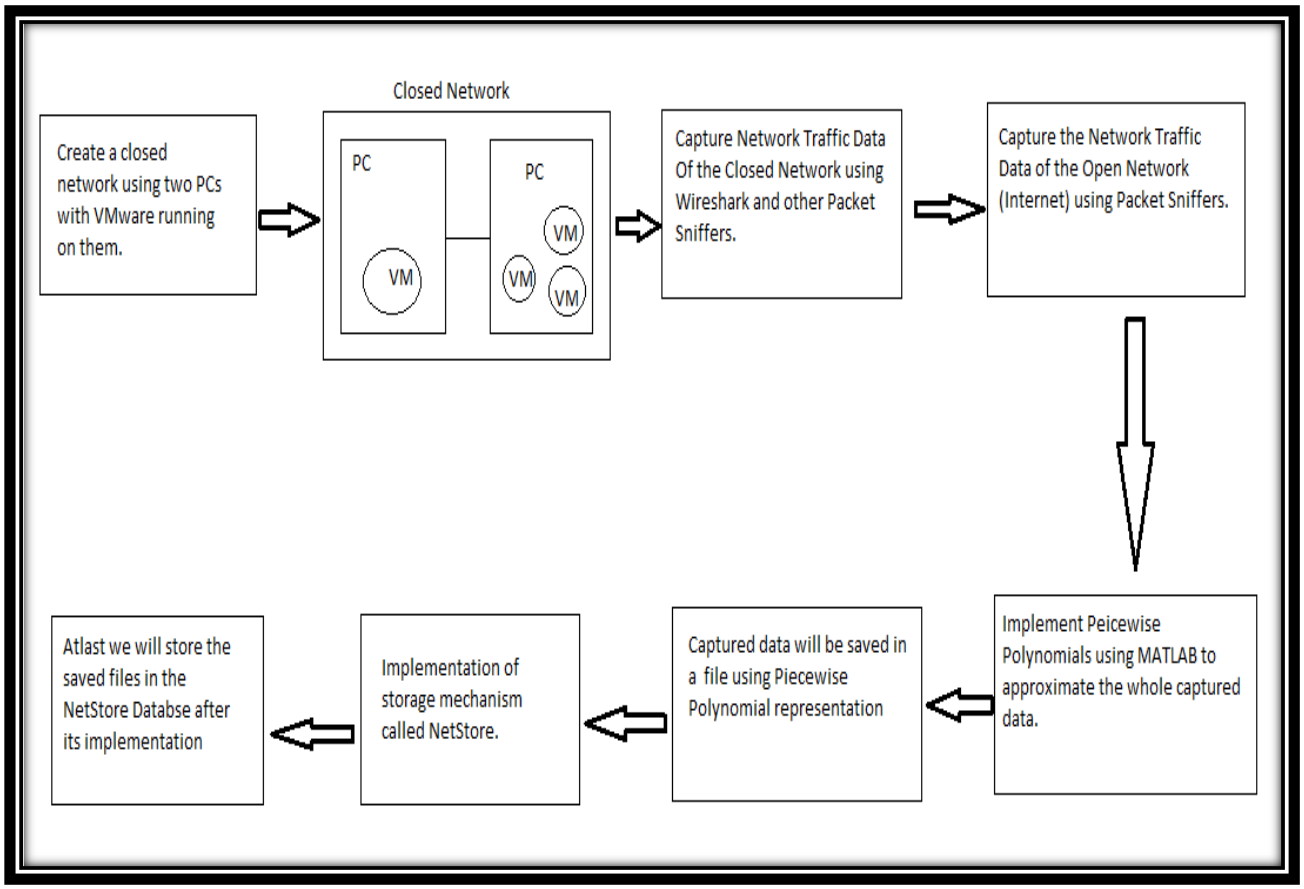


***FIGURE 1. Proposed Approach.***

The implementation of Piecewise Polynomials will be done using Matlab. The packet capture will be done of three websites for open network and then the files will be saved as .pcap file and then these files will be converted into .csv (comma separated values) file. Then these files will be further processed in Matlab using built-in functions where the input data is fed into these functions to compute polynomial and coefficients of polynomial. Input and output data of the functions will be represented as vectors in Matlab. The reason for use of Piecewise Polynomials is that a single polynomial cannot be used to model more than one network event, because it will not be able to represent the individual different network events that it is composed of. A piecewise polynomial model will address this issue by modeling each network event as an individual polynomial. If the order of the network events (segments) were changed, the individual polynomials would just occur at different time intervals, but each segment will remain the same. In other words, in a piecewise polynomial approximation each segment is represented by a distinct polynomial. The basic concept is that while the network will not behave the same all the time, it will behave the same in certain pieces. If network traffic can be quantified using piecewise polynomials, investigators can apply signature and anomaly detection techniques to identify and investigate events from a forensics perspective. Piecewise polynomial approximations will be effective because they will approximate the behavior pattern of a network with enough resolution to differentiate network traffic[].Memory storage is also of primary concern when modeling network data. The Internet packet capture shows that the discrete representation of data utilized 72Kb of memory storage, while the polynomial representation utilized 12Kb of memory storage. This result shows that polynomial processes utilize roughly six times less memory storage compared to

discrete processes. This size difference indicates that storing network traffic as polynomials instead of a collection of individual points significantly saves memory and this is important in network forensics because network events can now be archived for a longer amount of time than before as this is a major problem for the forensics investigator which we have mentioned earlier also. This extra storage allows for more extensive and detailed investigations which will provide better results to solve the cyber-crimes [7].

After the implementation of Piecewise Polynomials for approximation of the network traffic data of the three websites we will capture, the files which will be represented as polynomials will be saved in the NetStore which is storage infrastructure for network forensics and monitoring but we will use it as storage infrastructure [6]. NetStore is a column-based Relational Database Management System which is very efficient for fetching the stored data in fast manner that enables the Forensics Investigators to carry out the analysis of evidence collected in fast manner and it is already proposed earlier by researchers. We will be using this concept to make a storage infrastructure using few concept of this NetStore and for small scale data as we are taking account the data of three websites. The detail explanation of NetStore and its implementation will be our future work. The successful implementation of both the concepts will be used in future to make useful tool for Network Forensics. According to our research and present resources available for network forensics, this tool will be a huge advantage and a major breakthrough in the field of this sub-branch of digital forensics called Network Forensics.

## IV. CONCLUSION AND FUTURE WORK

Network Forensics has the problem of storage of network data in a precise manner so that we can have less use of memory and also the solution for the Investigators to fetch the relevant data in a fast manner. Piecewise Polynomials will be used for approximation to represent the network traffic data so that it is easy for the Investigators to differentiate captured data to find out footprints of attackers. The main purpose of this project is to do something for the benefit of the Digital Forensic World. NetStore, a storage infrastructure to monitor network traffic data will be implemented as an efficient storage mechanism to store the data which we will capture in our implementation of Piecewise Polynomial for the approximation of the captured network traffic data. Both the idea will be clubbed together to get maximum advantage keeping an eye on the need for reduced memory and faster access to data and query execution to aid the Network Forensics Analysts in their work of finding the abnormal network events and malicious things in the network carried out by the hackers and cyber criminals. Our project will try to help this area of a segment of network forensics where the network related events go unnoticed and there are very few tools and equipment available with the Investigators.

## REFERENCES

[1] Bruce J. Nikkel, "A portable network forensic evidence collector", Digital Investigation 3, Elesevier, pp.127-135, 2006.

[2] E.J. Palomo, J. North, D. Elizondo , R.M. Luque, T. Watson, "Application of growing hierarchical SOM for visualisation of network forensics traffic data", Neural Networks 32, Elsevier, pp.275-284, 2012.

[3] Ilow J, "Forecasting Network traffic using FARIMA models with heavy tailed innovations", Proceedingsof the Acoustics, Speech, and Signals Processing IEEE, vol.6, pp.3814 – 3817, 2000.

[4] Khamphao Sisaat, Daisuke Miyamoto, "Source Address Validation Support for Network Forensics", Proceedings of the 1st Joint Workshop on Information Security, pp.1-8, 2006.

[5] Matthias Vallentin, Dominik Charousset, "Native Actors: How to Scale Network Forensics",Proceedings of the 2014 ACM conference on SIGCOMM, pp.141-142, 2014.

[6] Paul Giura and Nasir Memon, "NetStore: An Efficient Storage Infrastructure for Network Forensics and Monitoring.", 13th International Symposium, RAID, pp. 277-296 , September 2010.

[7] Sean Marcus Sanders, "Network Forensics Analysis Using Piecewise Polynomials", Spring 2010: The Tower, pp.25-36, 2010.

[8] Shah K, Jonckheere E, Bohacek S, "Dynamic Modeling of Internet Traffic for Intrusion Detection", EURASIP Journal on Advances in Signal Processing, Hindawi Publishing Corporation, May 2006.

[9] Sherri Davidoff, Jonathan Ham, "Network Forensics Tracking Hackers through Cyberspace", E-book, Pearson Education , May 2012.

[10] Srinivas Mukkamal, Andrew H. Sung, "Identifying Significant Features for Network Forensic Analysis Using Artificial Intelligent Techniques", International Journal of Digital Evidence, Volume 1, Issue 4, pp.1-17, 2003.

[11] Wang J, "A Novel Associative Memory System Based Modeling and Prediction of TCP Network Traffic", Advances in Neural Networks, Springer Berlin, pp.519-527, July2007.

[12] Wei Wang, Thomas E. Daniels, "Diffusion and Graph Spectral Methods for Network Forensic Analysis", New Security Paradigms Workshop, pp.99-106, 2006.