Statistical Analysis of Network Attacks Based on Passive Traffic Monitoring Through Honeypot

Sanjeev Kumar[1], Amandeep kaur[2], Dr. G.N. Verma[3]

M.Tech Scholar, Computer Science and Engineering,Sri. Sukhamani Institute of Engg. & Technology-Punjab Technical University[1]
Assistant Professor, Computer Science and Engineering,Sri. Sukhamani Institute of Engg.& Technology- Punjab Technical University[2]
Principle, Sri. Sukhamani Institute of Engg. & Technology- Punjab Technical University[3]
Email:ror.sanjeev@gmail.com[1],amandeep76@gmail.com[2], drvermagh@gmail.com[3]

Abstract-In the domain of network security, many technologies and tools have been used throughout the years to create and test systems of network intrusion detection system and to test the strengthen of computer security system. One of these is the honeypot. A honeypot appears as a normal part of the network but in actual way it is an isolated environment which monitors the malicious activities in a network. The main benefit of the honeypot is to collect the attack data which are not logged and detected by the network intrusion detection system. Nowadays, security system is very important to any organization to protect their data or any information kept in their computer from the intruders to access. Unauthorized user is able to connect to the organization's computers and control it in some form to view or access the files. Many of us know how to use the computer but do not have enough information to secure the computer especially for the system administrators. "A Statistical Analysis of Network Attacks Based on Passive Traffic Monitoring through Honeypot", this is exactly implemented during the course of our research area to get the internal things about honeypots and to collect the attack data. A honeynet based attack data collection framework is designed and presented in this paper. The isolated environment in the form of honeypots is established as a data feed of malicious traffic to statistical analytical algorithms. The distribution and graphical representations of statistical algorithms are presented which help any normal user or system administrators to see the distributions of attack data. The network packet dumps recorded on a honeypot are treated as malicious traffic which is combined with the normal traffic to learn the machine to detect the malicious activities in a network as real time scenarios.

*Index Terms-**Malwares, Virtualized environment, Statistical Analysis, Network Security, Intrusion Detection System, Honeynet.***

1. INTRODUCTION

In terms of computing environment, various resources in the form of hardware and softwares are present which are known as assets of computing environments. The computer security involves providing the protection of such assets. It can be further expressed as a function of confidentiality(C), integrity (I) and availability (A) of information to authorized users.

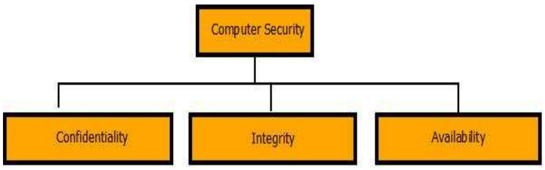

Figure 1: Computer Security Triad

Depending upon the security requirement of an organization, professionals attempt to maximize the value of security. In the following section, the above mentioned concepts of computer security are explained in more details.

- Confidentiality

Confidentiality deals with the secrecy or privacy of assets. It ensures that only authorized users are allowed to access computer assets. This 'access' incorporates any kind of access including reading, writing, printing or even the knowledge that a particular asset exists. In short, as quoted from [2], confidentiality means that’s only authorized people or systems can access protected data".

- Integrity

The concept of integrity makes sure that assets can only be modified by authorized users. Modification of an asset may include tasks like changing, deleting, writing, changing status and creating. According to Clark and Wilson [3], integrity is maintained when "No user of the system, even if authorised, may be permitted to modify data items in such a way that assets or accounting records of the company are corrupted." According to the Orange Book [4], integrity may also be defined as ensuring external consistency

- Availability

This property is concerned with the proper availability of information or services to authorized users whenever desired. It primarily aims to prevent any denial of service [5]. Apart from the above properties, there are other properties which may be considered a part of computer security. These include authentication, accountability, reliability, fault-tolerance and assurance [6].

2. BACKGROUND AND RELATED WORK

2.1. *Intrusion Detection System*

The intruder spreading in an organization can be detected by Intrusion Detection which is caused by people external to the organization or within the organization. The external violation is performed by the attackers who are outside the organization whereas the attacks can also be performed by insider employees of the organization [7-11].

Types of IDS:

- Signature-based detection (also known as Misuse detection)
- Specification based detection
- Anomaly based detection

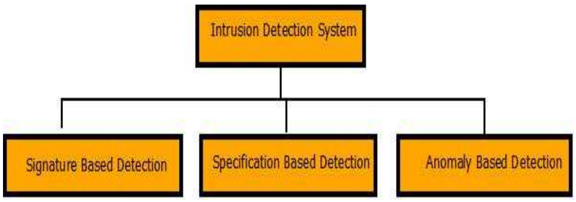
Intrusion Detection System
Signature Based Detection
Specification Based Detection
Anomaly Based Detection

Figure 2: IDS categorizations

2.2. *Introduction of Proactive Network Security*

Honeypots

A Network security resource whose value lies in it being scanned, attacked, compromised controlled and misused by an attacker to attain his malicious goals. Lance Spitzner defines Honeypots as "A Honeypot is an information system resource whose value lies in unauthorized or illicit use of that resource" [12]. Honeypots can be classified into two main categories. Firstly, they can be based upon their level of interaction with an attacker. This can be further categorized as [13-16]. As shown the honeypots are categorized as based on the level of interactions, type of attacked resources and kind of honeypots. In broad way, the honeypots are classified as:

- Server Honeypot- The honeypots on which vulnerable ports and services are configured and they are passively waiting on the internet to be attacked by the attackers
- Client Honeypots- The kind of honeypots which actively browse the internet (web links) and capture the potential class of attack data such drive by download attacks.

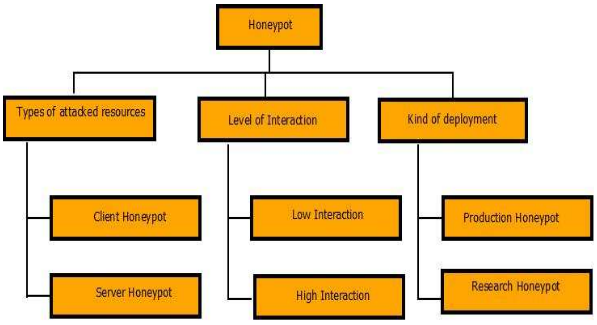

Figure 3: Honeypot’s Classifications

2.3. *Introduction of Statistical Analysis*

In the field of computer security, the incident is happening almost on daily basis. These incidents are in the form of small, from normal to complex one such as malware stealing the user’s credentials. One such incident is reported in [12], which is a botnet for distributing the denial of service attacks. The botnets are special class of attacks which is remotely controlled by the attackers.

Over the past years, due to exponential growth of computer attacks, there is a need to collect more and more data which need a high computational resource to analyse the huge attack data. Thereby to reduce the complexity of analysing the attack data, there is need to apply intelligent algorithms on the fly to get the distribution of attacks, trends determinations and further for future predictions for situational awareness. The role of statistical and mathematical domains and algorithms are getting increasing to analyse the huge data sets. Here in this paper, we try to implement the statistical analysis in the form of attack distributions by applying the statistical algorithm. By applying such algorithms, the users or system administrators or a system analyst may get a feel of which is most target trend or configuration targeted by the attacker.

3. DESIGN AND IMPLEMENTATION

Here we present the design methodology for implementation of honeypot as attack data collection mechanism and further processing and filtering of attack data to produce the statistical inference for classification and inferential output is presented. Furthermore, in depth analysis and implementation of honeypot as network attack data collections tools and techniques is presented in which we have incorporated the passive traffic monitoring to log the network data as dump files in the form of PCAP (packet captured) files. With the help of network traffic monitoring tools such TCPDUMP which give us the flexibility to log each and every packets coming towards the honeypots and originating from the honeypots are being monitored and logged for further analysis.

3.1. *Honeynet Implementation Scenarios*

- Facing the Internet: In this implementation, the honeypots are assigned the public IP addresses which are directly visible to the internet. The ports and services hosted on a honeypots configured in a virtual environment is directly accessible to the outside world.

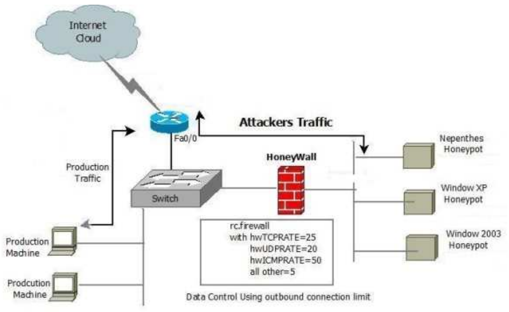

Figure 4: Honeypots facing the Internet

- Inside the Network: In this scenario, the honeypots are placed in a private network such organizational network. This implementation is very well suitable to capture the internal threats spreading in the network. The attacks which are propagating inside the network through some medium such USB stick, or internal users are captured by these kinds of implementations.

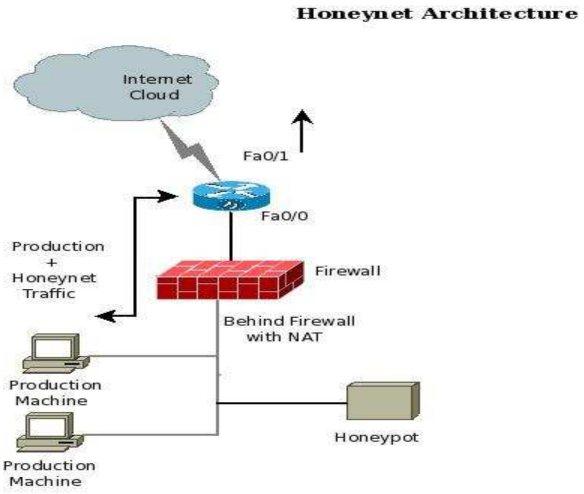

Figure 4: Honeypots inside the network

- In DMZ zone: This is one of the crucial implementation of honeypots deployment which are placed in parallel to other resources such as DNS, Web servers etc.

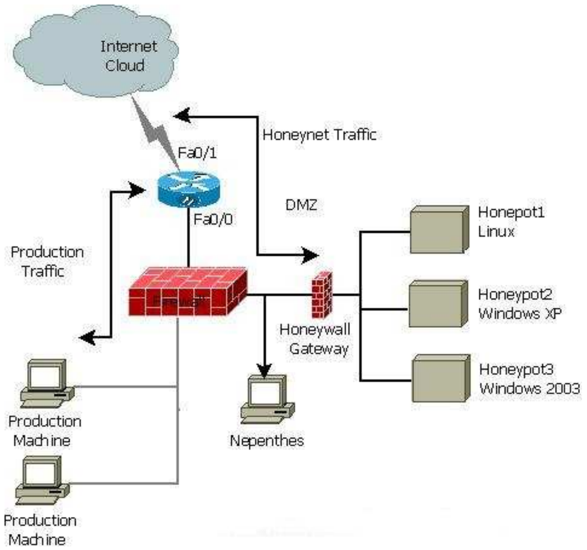

Figure 5: Honeypots in DMZ zone

4. ALGORITHMIC IMPLEMENTAIONS AND RESULTS

Here we present the implementation results in the form of box plots, histograms, and cumulative distribution of data sets with respect to class defined in our data set.

Feature Selection

Selection of features which have unique properties among different applications is no doubt the vital part of any classification effort ever made, as it includes minute details and deep investigation of data packets. A single data packet has large number of features which can be studied, but not all of them are useful for classification. Furthermore, it is important to know the aim one wants to achieve, for example classification between different applications requires different features, to be extracted. The main task here is to select the attributes which have different values for normal and malicious traffic. The number of studies presented in Brugger's work shows that the most common features used for particular classification as ours are:

< protocol, Timestamp, Dest Port, Src Address, TCP flags, Total source bytes, total destination bytes, durations.>

ARFF file:
The following is the sample of the prepared ARFF data sets which include the automatic selection and extraction of features from network dumps. There are two main parts of an ARRF file Header part and Data part. Header part consist of all the features, last feature represents the class which determines whether the value is true or false for particular instance.

```
@relation Test-data1

@attribute frame.len numeric
@attribute tcp.dstport numeric
@attribute e numeric
@attribute tcp.srcport numeric
@attribute tcp.seq numeric
@attribute tcp.hdr_len numeric
@attribute tcp.window_size numeric
@attribute ip.proto {0x06,0x11,0x01}
@attribute Class {malicious,Normal}

@data
54,3389,?,6000,0,20,16384,0x06,malicious
54,3389,?,6000,0,20,16384,0x06,malicious
54,3389,?,6000,0,20,16384,0x06,malicious
54,3389,?,6000,0,20,16384,0x06,malicious
60,6000,?,3389,1,20,0,0x06,malicious
60,6000,?,3389,1,20,0,0x06,malicious
54,3389,?,6000,0,20,16384,0x06,malicious
54,3389,?,6000,0,20,16384,0x06,malicious
60,6000,?,3389,1,20,0,0x06,malicious
60,6000,?,3389,1,20,0,0x06,malicious
78,135,?,4445,0,44,65535,0x06,malicious
78,135,?,4445,0,44,65535,0x06,malicious
78,4445,?,135,0,44,65535,0x06,malicious
78,4445,?,135,0,44,65535,0x06,malicious
78,135,?,4445,0,44,65535,0x06,malicious
78,135,?,4445,0,44,65535,0x06,malicious
```

For the experimentations, we have used the R tool which is one of the popular tools for machine learning and statistical analysis. R is a free software environment for statistical computing and graphics [17].

In a similar way, the distribution of frame.len and IP protocol is depicted in figure 7. The IP proctol distribution is shown in green color in a total data set. Initially the attack distribution of attack data with respect to IP protocol is high and then it starts decreasing as whole data set is becoming less.

Figure 6 depict the distribution of frame.len and tcp.dstport by class in whole data set.

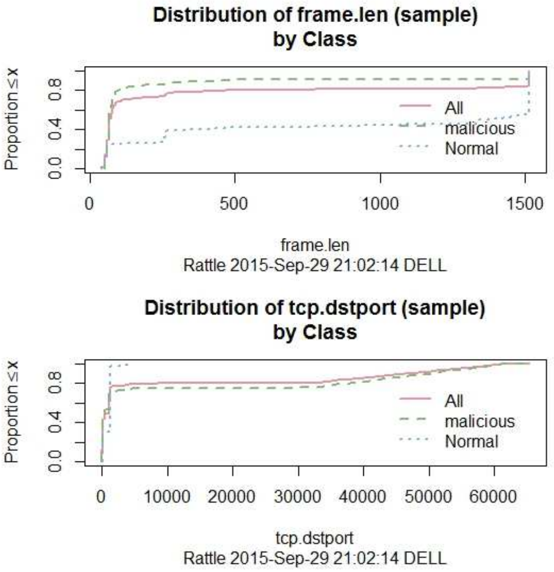

Figure 6: Distribution of frame.len and tcp.dstport by class.

We can see the distribution of the trends with respect to frame length and TCP destination port, the frequency is high initially on the starting ports, once the port number is high; the frequency of attacks is becoming less.

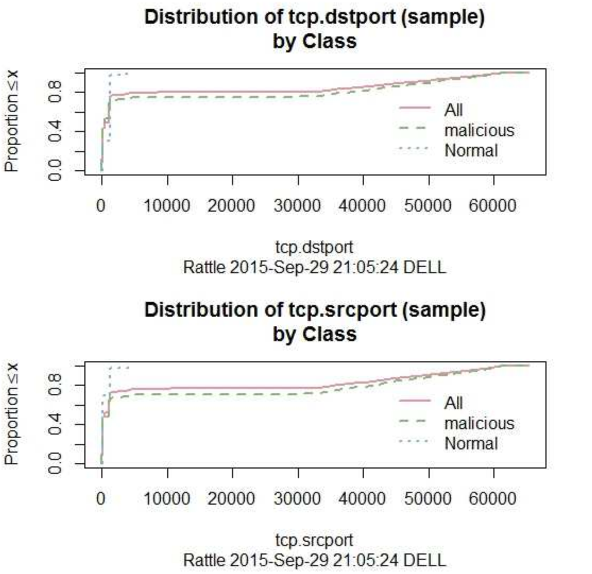

Figure 7: Distribution of tcp.dstport and tcp.srcport by class in whole data set

The distribution of tcp destination port and tcp source port is depicted in the figure 7. Initially the port ranges 0-10000 are targeted in exponential manner, then after that it become straight as shown. It can be concluded that standard ports are highly targeted initially.

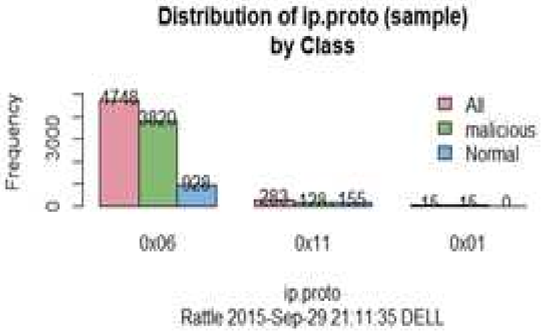

Figure 8 Distribution of ip.proto by whole class

CONCLUSIONS

In this research implementation, firstly we presented the significance of proactive network security in the form of honeypots compare to defensive network security such as intrusion detection system, firewalls, router etc. We discussed that there is a need to put an environment in the form of honeynet to capture, collect and further perform analysis on the collected attack data. Then we presented our research implementations in the form of creation of honeynet test bed which is deployable in different deployment scenarios such as public internet, DMZ and private network. With the help of deployed environment, we captured the malicious network data which are logged in the form of network PCAP data. To monitor the network traffic, we implemented the passive traffic monitoring through network tap in the form of network Ethernet card of honeypot machine. Every network communication is recorded from honeypots which act a data feed to our analysis engine. Then for analysis purposes, we used and implemented the two techniques 1) signature based detection and correlation engine 2) statistical analysis engine. Through signature based analysis, we are able to gather the intelligent information by processing the network traffic collected which we have used for creation of features set for statistical models. Wherever we found the indicative suspicious traffic with high severity, we considered that as a set of feature set.

Acknowledgments

I would like to acknowledge Ms. AmandeepKaur, Assistant Professor, *Sri. Sukhamani Institute of Engg.& Technology-Punjab Technical University*for her support and guidance during paper formulation and research implementation. I also would like to thank Dr. G.N. Verma, Principle, *Sri. Sukhamani Institute of Engg.& Technology-Punjab Technical University for his continuous guidance and mentoring the research on timely basis.*

REFERENCES

[1] Sandeep V. Sabnani, Computer Security: A Machine Learning Approach, Technical Report, RHUL-MA-2008-09, 07 January 2008

[2] Charles P. Pfleeger and Shari Lawrence Pfleeger. Security in Computing. Pearson Education, Inc, 2003.

[3] David D. Clark and David R. Wilson. A comparison of commercial and military computer security policies. IEEE Security and Privacy, 00:184, 1987.

[4] NIST. Trusted computer system evaluation criteria (orange book). http://csrc.nist.gov/publications/history/dod85.pdf , 1985.

[5] Dieter Gollmann. Computer Security. John Wiley & Sons, 1999.

[6] Jason Crampton. Notes on Computer Security, 2006.

[7] William Stallings. Network Security Essentials: Applications and Standards (3rd Edition). Prentice Hall, 2006.

[8] StephenWolthusen. Lecture 11 - Intrusion Detection and Prevention, notes in Network Security, 2006.

[9] Yihua Liao. Machine Learning in Intrusion Detection. PhD thesis, University of California (Davis), Department of Computer Science, 2005.

[10] LudovicM'e and C'edric Michel. Intrusion detection: A bibliography. Technical Report

SSIR-2001-01, Sup'elec, Rennes, France, September 2001.

[11] Rebecca G. Bace. Intrusion Detection. Sams, December 1999.

[12] Experiences with a Generation III Virtual Honeynet, Abbasi, F.H. ,Harris, R.J.; ATNAC, 2009

[13] JS Bhatia et.al, Honeynet Based Botnet Detection Using Command Signatures, WiMoA 2011/ICCSEA, CCIS, 2011.© Springer-Verlag Berlin Heidelberg 2011.

[14] Baecher, P., Koetter, M., Holz, T., Dornseif, M., Freiling, F.: The Nepenthes Platform: An Efficient Approach to Collect Malware. In: Zamboni, D., Krügel, C. (eds.) RAID 2006.LNCS, vol. 4219, pp. 165-184. ACM Press, New York (2006).

[15] Karasaridis, A., Rexroad, B., Hoeflin, D.: Wide-scale botnet detection and characterization. In-proceedings of the workshop on hot topics in understanding botnets (April 2007)

[16] Sanjeev Kumar et. al, Distributed Honeynet System using Gen III Virtual Honeynet, Proceedings of conference ICFN-2011

[17] https://www.r-project.org/