International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 03 | Mar -2017 www.irjet.net p-ISSN: 2395-0072

IDENTIFYING MALICIOUS DATA IN SOCIAL MEDIA

M.Sai Sri Lakshmi Yellari[1], M.Manisha[2], J.Dhanesh[3] ,M.Srinivasa Rao[4] ,Dr.S.Suhasini[5]

[1]*Student, Dept. of Information Technology, Velagapudi Ramakrishna Siddhartha Engineering College, Andhra Pradesh, India*

[2]*Student, Dept. of Information Technology, Velagapudi Ramakrishna Siddhartha Engineering College, Andhra Pradesh, India*

[3]*Student, Dept. of Information Technology, Velagapudi Ramakrishna Siddhartha Engineering College, Andhra Pradesh, India*

[4]*Student, Dept. of Information Technology, Velagapudi Ramakrishna Siddhartha Engineering College, Andhra Pradesh, India*

[5]*Associate professor, Dept. of Information Technology, Velagapudi Ramakrishna Siddhartha Engineering College, Andhra Pradesh, India*

Abstract - *Network anomaly detection is a broad area of research. The use of entropy and distributions of traffic features has received a lot of attention in the research community. Disclosing malicious traffic for network security using entropy based approach and power law distribution is proposed. To calculate entropy feature considered is packet size. Malware, most commonly known as malicious data is prevalent, arising a number of critical threat issues. With the increasing volume of contents users share through social media, the user is going to share large amount of information. Using power law distribution malware is detected which the users share in social media by making a comparison with Shannon entropy technique*.

***Key Words*: Social media, Entropy, Malware, Power law, security, traffic.**

1. INTRODUCTION

A network consists of two or more computers that are linked in order to apportion resources (such as printers and CDs), exchange files, or allow electronic communications. The computers on a network may be linked through cables, telephone lines, radio waves, satellites, or infrared light beams. A network may be composed of any coalescence of LANs, or WANs. Network traffic can be defined in a number of ways. But in the simplest manner we can define it as the density of data present in any Network. Network data security should be a high priority when considering a network setup due to the growing threat of hackers endeavoring to infect as many computers possible. Due to the cumbersomely hefty utilization of the network now a day's sundry attacks are been occurring and malicious data is injected into the user's profile or document. Due to lack of security in the organization the data breaches. In this paper, we study the comparison between Entropy based anomaly detection mechanism and Power law distribution. Entropy based anomaly detection captures more fine grained traffic patterns as compared to normal volume based metrics. Many traffic features such as IP address, port number, flow size etc are considered as attributes is calculating entropy where as Power law (also called the scaling law) states that a relative change in one quantity results in a proportional relative change in another.

1.1 Malicious Data

Malicious data is data that, when introduced to a computer—usually by an operator unaware that he or she is doing so—will cause the computer to perform actions undesirable to the computer's owner. Malicious practices done by the local networks users that do not allow efficient sharing of the network resources. Common threats are: Unauthorized Access, Data Destruction, Administrative Access, System Crash/Hardware Failure, Virus. Malware is short for malicious software, denoting software that can be habituated to compromise computer functions, steal data, bypass access controls, or otherwise cause harm to the host computer. Malware is a broad term that refers to a variety of maleficent programs. This post will define several of the most mundane types of malware; adware, bots, bugs, rootkits, spyware, Trojan horses, viruses, and worms.

1.2 Problem Description

Online social networks are widely use these days for the purpose of communication. Users can share more type of information among friends. But there exist some social network users who misuse the features of these social networks and promote the spreading of malicious content. They do this by uploading the malicious files. These contents spread at a fast rate. There is no proper mechanism to detect these malicious files immediately and remove it effectively.

Convivial network sites like Face book, Twitter, and Google+ are experiencing incredible magnification in users. There are more than a million users as of now. Besides just engendering a profile and linking with friends, the gregarious networks are now building platforms to run their website. These platforms are built predicated based on the profile details. These social applications are anon becoming an example of online communication which makes utilization of the user's private information and activities in convivial links for sundry accommodations. The gregarious networks are popular denotes of communication among the cyber world users.

2. PROPOSED SYSTEM

Malware (for "malicious software") is any program or file that is harmful to a computer user. Thus, malware includes computer viruses, worms, Trojan horses, and also spyware, programming that gathers information about a computer user without permission. It is hard to detect and distinguish malicious packets and legitimate packets in the traffic. The behavior of internet traffic is very far from being regular. Malicious are abnormal traffic may look very similar to normal traffic.

2.1 Shannon Entropy

Security breaches on a network server can result in the disclosure of critical information or the loss of a capability that can affect the entire organization. Therefore, securing network servers should be a significant part of your network and information security strategy[1]. Many security problems can be avoided if servers and networks are appropriately configured. So, we approached entropy based approach to detect anomalous traffic with altered packet size help of packets that sent. Entropy based anomaly detection techniques captures more fine grained traffic patterns as compared to normal volume based metrics. To calculate entropy we have some parameters like Source and destination, IP address, port numbers, Packet size, Connection time and the total number of packets flowing.

Definition 1: Entropy

Entropy is a disorder or randomness of system [3].

Algorithm:

AIM: To detect the altered packets using Entropy based approach by making use of Shannon Entropy Algorithm.

1) Capture and add packets in the current queue L.
2) Compute the current queue length.
3) Select the desired features required for calculations i.e. IP address of source and destination, port number of source and destination, packet size, packet rate and connection time[2].
4) Calculate the entropy.

$$\mathbf{H(X)} = \sum \mathbf{P\ (xi)\ log\ (xi)}$$

Here, X for a fixed time window w is,

P(xi) = mi/m, Where mi is the frequency or number of times we observe X taking the value xi as m = mi.

H (X) = - (mi /m) log (mi/m).

H (X) = Entropy

If we want to calculate probability of any source (destination) address then,

mi = number of packets with xi as source (Destination) address.

M = total number of packets

P (xi) =Number of packets with xi as source/destination address/M.

Here total number of packets is the number of packets seen for a time window T.

Similarly we can calculate probability for each source (destination) port as P (xi)= Number of packets with xi as source (destination) address/M

Normalized entropy calculates the over all probability distribution in the captured flow for the time window T.

5) Normalized entropy = (**H/log (n)**) Where n is the number of distinct xi values in the given time window.
6) Determine the threshold on the basis of the maximum and minimum deviations calculated for a number of times.
7) If the result exceeds the threshold an attack is found, else no attack is found.

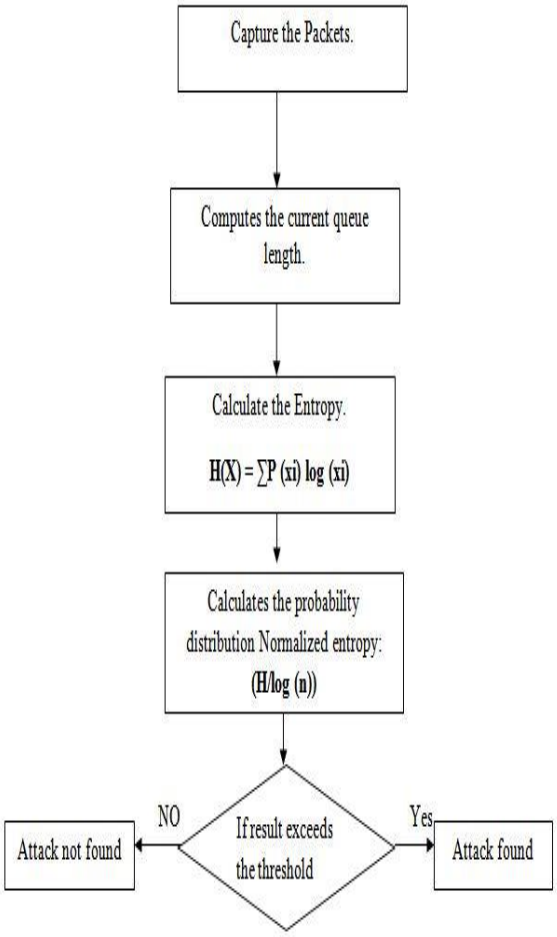


Fig -1: Flow diagram of Shannon Entropy

Features extracted for the detection of anomaly based attack are as follows:

- Entropy of source IP address and port number.
- Entropy of destination IP address and port number.
- Entropy of packet type. Occurrence rate of packet type.
- Number of packets per unit time.
- Entropy of packet size.

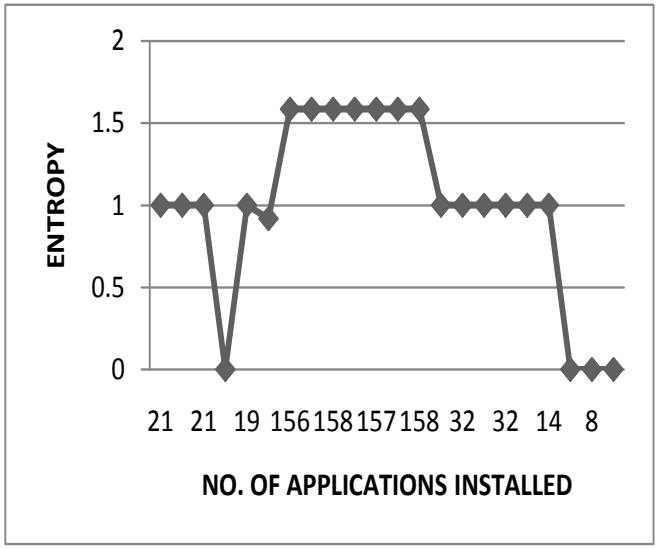


Chart -1: Analysis of Shannon Entropy

The above analysis is done by considering the github facebook dataset in which we concentrated on the no. of applications installed part to find the malicious data.

2.2 Power Law Distribution

MALWARE are malicious software programs deployed by cyber attackers to compromise computer systems by exploiting their security vulnerabilities. Motivated by extraordinary financial or political rewards, malware owners are exhausting their energy to compromise as many networked computers as they can in order to achieve their malicious goals. A compromised computer is called a bot, and all bots compromised by a malware form a botnet. Botnets have become the attack engine of cyber attackers, and they pose critical challenges to cyber defenders[4]. In order to fight against cyber criminals, it is important for defenders to understand malware behavior, such as propagation or membership recruitment patterns, the size of botnets, and distribution of bots.

The proposed two layer epidemic model and the findings are the first work in the field. Our contributions are summarized as follows.

- We propose a two layer malware propagation model to describe the development of a given malware at the Internet level. Compared with the existing single layer epidemic models, the proposed model represents malware propagation better in large-scale networks.
- We find the malware distribution in terms of networks varies from exponential to power law with a short exponential tail, and to power law distribution at its early, late, and final stage, respectively. These findings are first theoretically

proved based on the proposed model, and then confirmed by the experiments through the two large-scale real-world data sets.

- **Definition 1: Power Law**

The **power law** (also called the scaling law) states that a relative change in one quantity results in a proportional relative change in another.

Algorithm:
AIM: To detect malicious content by making use of Power Law Distribution.

1) Enter the Web Application.
2) Open User Registration form, enter Login credentials.
3) If Login credentials are invalid, it displays that the credentials entered are invalid.
4) If valid, the user can view the further steps.
5) The steps include Uploading Files, Search Users, Search File, My Search History, Malicious Document, Friends Request and Response Details, My Files, My Friend Files.
6) The Power Law Distribution is applied at My Friends Files option.
7) Calculate the Power Law Distribution using,
   a. $Y = KX^{\alpha}$
   b. Where Y and X are variables of interest (here X taken as Size of the file),
   c. α is law's exponent,
   d. K is constant.
8) The Power Law Distribution discloses the Good files and the Malicious files respectively.
9) It is done internally as ,if value of Y is equal to the value of the file uploaded by the user's friend then the file is said to be malicious.
10) User can log out of his account after his actions are performed.
11) The Admin who monitors overall users, blocks the account of malicious user.

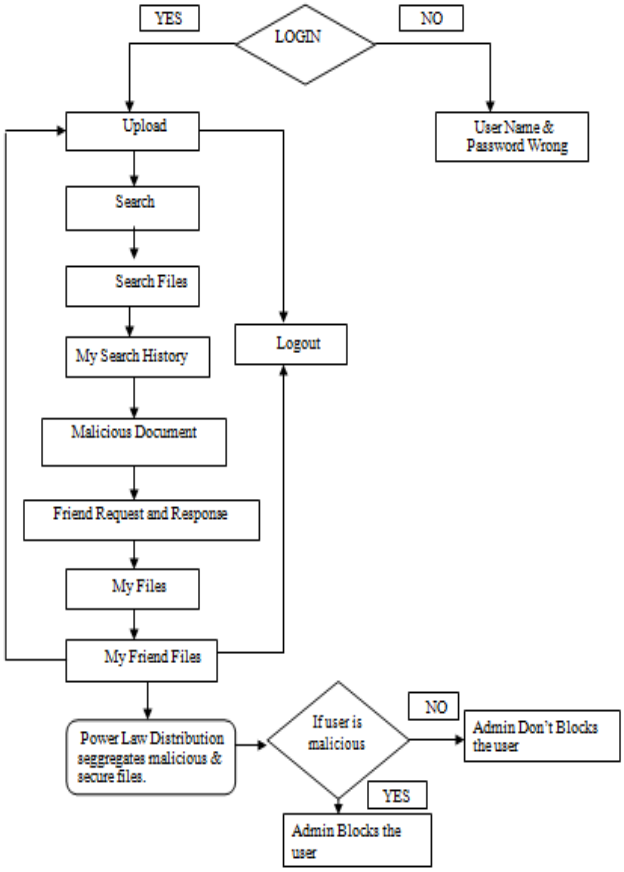


Fig -2: Flow diagram of Power Law Distribution.

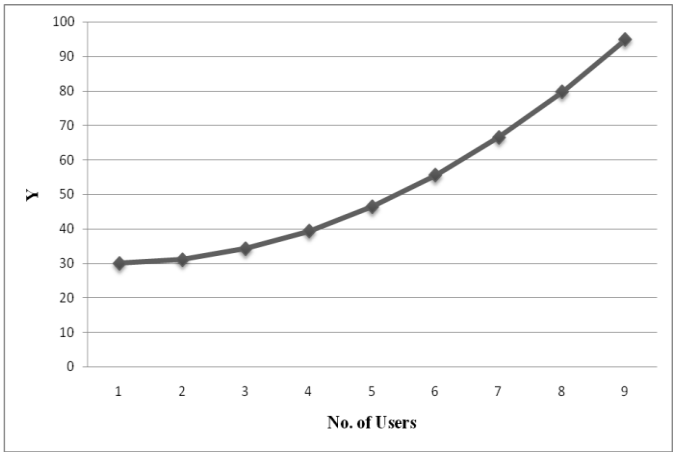
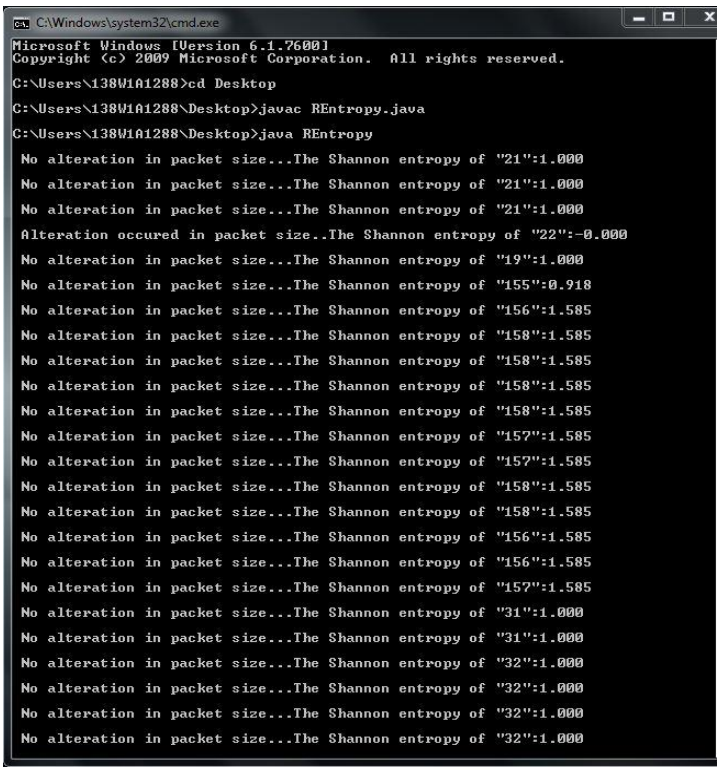
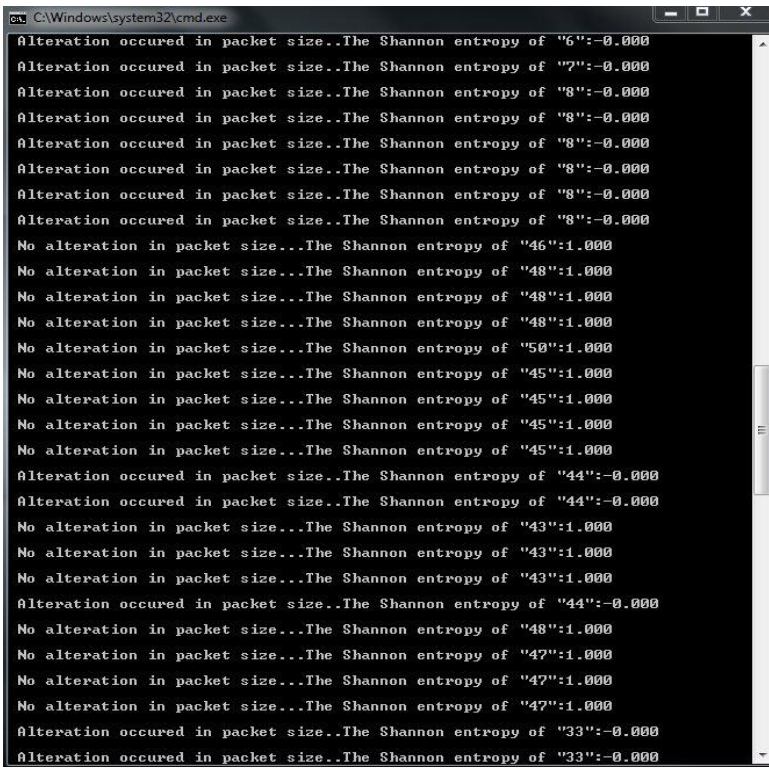


Chart -2: Analysis of Power Law Distribution.

The above analysis depicts as the number of users go on increasing the amount of malicious data also increases resulting in poor performance.

3. RESULTS


C:\Windows\system32\cmd.exe
Microsoft Windows [Version 6.1.7600]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.
C:\Users\138W1A1288>cd Desktop
C:\Users\138W1A1288\Desktop>javac REntropy.java
C:\Users\138W1A1288\Desktop>java REntropy
No alteration in packet size...The Shannon entropy of "21":1.000
No alteration in packet size...The Shannon entropy of "21":1.000
No alteration in packet size...The Shannon entropy of "21":1.000
Alteration occured in packet size..The Shannon entropy of "22":-0.000
No alteration in packet size...The Shannon entropy of "19":1.000
No alteration in packet size...The Shannon entropy of "155":0.918
No alteration in packet size...The Shannon entropy of "156":1.585
No alteration in packet size...The Shannon entropy of "158":1.585
No alteration in packet size...The Shannon entropy of "158":1.585
No alteration in packet size...The Shannon entropy of "158":1.585
No alteration in packet size...The Shannon entropy of "158":1.585
No alteration in packet size...The Shannon entropy of "157":1.585
No alteration in packet size...The Shannon entropy of "157":1.585
No alteration in packet size...The Shannon entropy of "158":1.585
No alteration in packet size...The Shannon entropy of "158":1.585
No alteration in packet size...The Shannon entropy of "156":1.585
No alteration in packet size...The Shannon entropy of "156":1.585
No alteration in packet size...The Shannon entropy of "157":1.585
No alteration in packet size...The Shannon entropy of "31":1.000
No alteration in packet size...The Shannon entropy of "31":1.000
No alteration in packet size...The Shannon entropy of "32":1.000
No alteration in packet size...The Shannon entropy of "32":1.000
No alteration in packet size...The Shannon entropy of "32":1.000
No alteration in packet size...The Shannon entropy of "32":1.000


C:\Windows\system32\cmd.exe
Alteration occured in packet size..The Shannon entropy of "6":-0.000
Alteration occured in packet size..The Shannon entropy of "7":-0.000
Alteration occured in packet size..The Shannon entropy of "8":-0.000
Alteration occured in packet size..The Shannon entropy of "8":-0.000
Alteration occured in packet size..The Shannon entropy of "8":-0.000
Alteration occured in packet size..The Shannon entropy of "8":-0.000
Alteration occured in packet size..The Shannon entropy of "8":-0.000
Alteration occured in packet size..The Shannon entropy of "8":-0.000
No alteration in packet size...The Shannon entropy of "46":1.000
No alteration in packet size...The Shannon entropy of "48":1.000
No alteration in packet size...The Shannon entropy of "48":1.000
No alteration in packet size...The Shannon entropy of "48":1.000
No alteration in packet size...The Shannon entropy of "50":1.000
No alteration in packet size...The Shannon entropy of "45":1.000
No alteration in packet size...The Shannon entropy of "45":1.000
No alteration in packet size...The Shannon entropy of "45":1.000
No alteration in packet size...The Shannon entropy of "45":1.000
Alteration occured in packet size..The Shannon entropy of "44":-0.000
Alteration occured in packet size..The Shannon entropy of "44":-0.000
No alteration in packet size...The Shannon entropy of "43":1.000
No alteration in packet size...The Shannon entropy of "43":1.000
No alteration in packet size...The Shannon entropy of "43":1.000
Alteration occured in packet size..The Shannon entropy of "44":-0.000
No alteration in packet size...The Shannon entropy of "48":1.000
No alteration in packet size...The Shannon entropy of "47":1.000
No alteration in packet size...The Shannon entropy of "47":1.000
No alteration in packet size...The Shannon entropy of "47":1.000
Alteration occured in packet size..The Shannon entropy of "33":-0.000
Alteration occured in packet size..The Shannon entropy of "33":-0.000


C:\Windows\system32\cmd.exe
No alteration in packet size...The Shannon entropy of "32":1.000
No alteration in packet size...The Shannon entropy of "32":1.000
No alteration in packet size...The Shannon entropy of "45":1.000
No alteration in packet size...The Shannon entropy of "14":1.000
No alteration in packet size...The Shannon entropy of "60":1.000
No alteration in packet size...The Shannon entropy of "60":1.000
No alteration in packet size...The Shannon entropy of "139":1.585
No alteration in packet size...The Shannon entropy of "10":1.000
No alteration in packet size...The Shannon entropy of "110":0.918
No alteration in packet size...The Shannon entropy of "20":1.000
No alteration in packet size...The Shannon entropy of "20":1.000
No alteration in packet size...The Shannon entropy of "20":1.000
No alteration in packet size...The Shannon entropy of "204":1.585
No alteration in packet size...The Shannon entropy of "49":1.000
No alteration in packet size...The Shannon entropy of "49":1.000
No alteration in packet size...The Shannon entropy of "49":1.000
No alteration in packet size...The Shannon entropy of "48":1.000
No alteration in packet size...The Shannon entropy of "17":1.000
No alteration in packet size...The Shannon entropy of "19":1.000
No alteration in packet size...The Shannon entropy of "19":1.000
No alteration in packet size...The Shannon entropy of "43":1.000
No alteration in packet size...The Shannon entropy of "43":1.000
No alteration in packet size...The Shannon entropy of "43":1.000
No alteration in packet size...The Shannon entropy of "43":1.000
No alteration in packet size...The Shannon entropy of "40":1.000
Alteration occured in packet size..The Shannon entropy of "2":-0.000
Alteration occured in packet size..The Shannon entropy of "5":-0.000
Alteration occured in packet size..The Shannon entropy of "5":-0.000
Alteration occured in packet size..The Shannon entropy of "6":-0.000

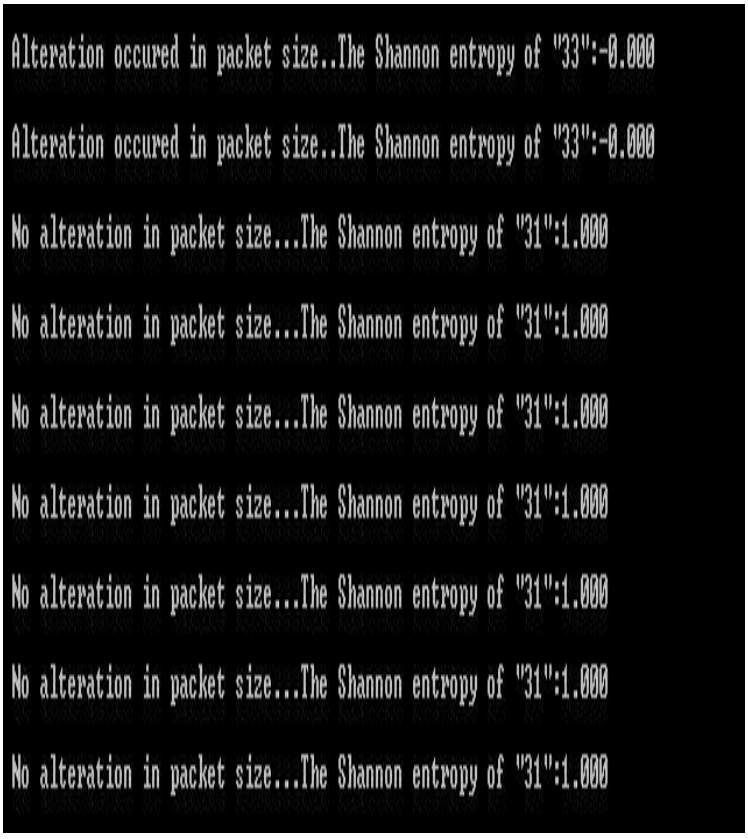

Alteration occured in packet size..The Shannon entropy of "33":-0.000
Alteration occured in packet size..The Shannon entropy of "33":-0.000
No alteration in packet size...The Shannon entropy of "31":1.000
No alteration in packet size...The Shannon entropy of "31":1.000
No alteration in packet size...The Shannon entropy of "31":1.000
No alteration in packet size...The Shannon entropy of "31":1.000
No alteration in packet size...The Shannon entropy of "31":1.000
No alteration in packet size...The Shannon entropy of "31":1.000
No alteration in packet size...The Shannon entropy of "31":1.000

Fig -3: Admin Login.

Fig -4: User Login.

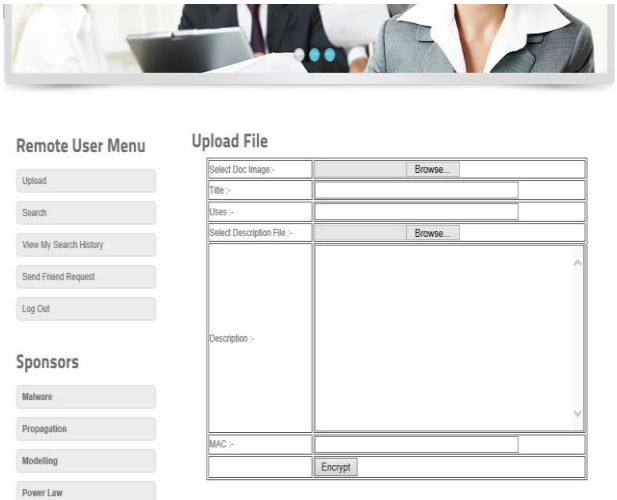
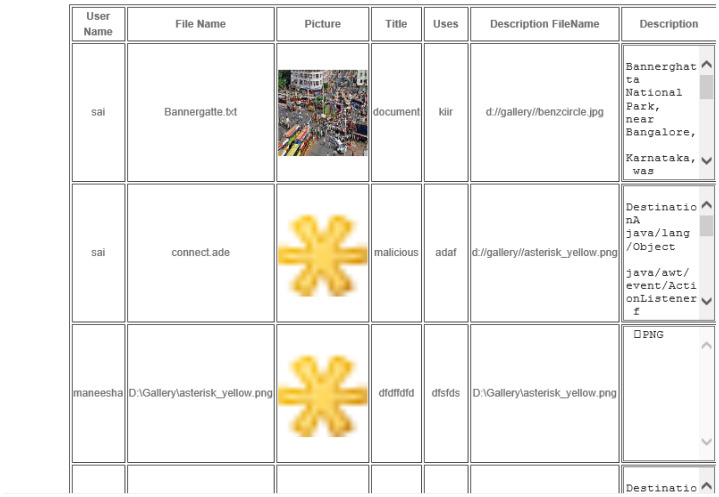


Fig -5: Upload Files.



User Name	File Name	Picture	Title	Uses	Description FileName	Description
sai	Bannergatte.txt		document	kiir	d://gallery//benzcircle.jpg	Bannerghatta National Park, near Bangalore, Karnataka, was
sai	connect.ade		malicious	adaf	d://gallery//asterisk_yellow.png	DestinationA java/lang/Object java/awt/event/ActionListener f
maneesha	D:\Gallery\asterisk_yellow.png		dfdffdfd	dfsfds	D:\Gallery\asterisk_yellow.png	PNG
						Destinatio

Fig -6: Admin views all the files.

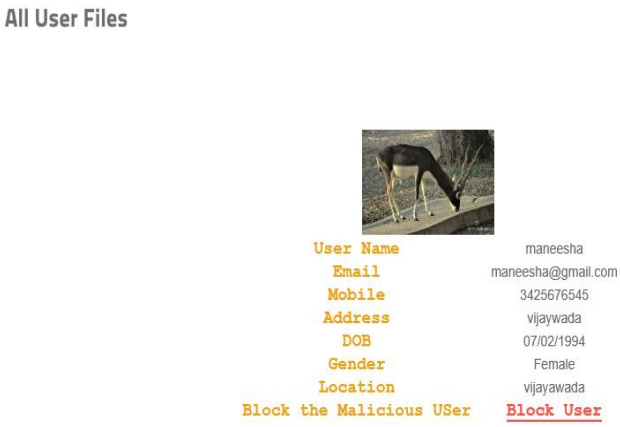


Fig -7: Admin Blocks.

4. CONCLUSIONS

The prevailing definition of network anomaly reports an occurrence that diverges from the normal behavior. However there are no known models available for normal network behavior. The major strength of the new scheme is that it can detect attack in the face book data set on the constraint of no. of applications installed. We are not claiming that our methods are superior to all other methods. It is well known that finding malicious traffic in a network or in a communication system has a wide scope for research. Using Entropy based technique we aim to detect the altered packet in a network or communication system. Experimental sample data set which we have taken is relatively small and hence this data-set won't cover all the attacks in the world. So, we are moving Power Law Distribution in which different from previous modeling methods, we propose a two layer epidemic model: the upper layer focuses on networks of

large-scale networks, for example, domains of the Internet; the lower layer focuses on the hosts of a given network. In regards to future work, we will first further investigate the dynamics of the late stage. More details of the findings are expected to be further studied, such as the length of the exponential tail of a power law distribution at the late stage. Second, defenders may care more about their own network, e.g., the distribution of a given malware at their ISP domains, where the conditions for the two layer model may not hold. We need to seek appropriate models to address this problem. Finally, we are interested in studying the distribution of multiple malware on large-scale networks as we only focus on one malware in this paper. We believe it is not a simple linear relationship in the multiple malware case compared to the single malware one.

REFERENCES

[1] S.Poornimavathi,Mr.K.Anandapadmanabhan,"Entropy-based analysis of multiple traffic anomalies Detection in network security", International Journal of Multidisciplinary Research and Development,IJMRD,ISSN:2349-5979,Vol 3,Issue 3,March,2016.

[2]Kamal Shah and Tanvi Kapdi, "Disclosing Malicious Traffic for Network Security", International Journal of Advances in Engineering and Technology, IJAET, ISSN:22311963, Vol 7, Issue 6, January,2015.

[3] C. G. Chakrabarti and I. Chakrabarty," Boltzmann Entropy : Probability And Information", Romanian Journal of Physics, P. 525-528,Volume 52, Numbers 5-6, , Bucharest, 2007.

[4] Shui Yu, Guofei Gu, Ahmed Barnawi, Song Guo and Ivan Stojmenovic, "Malware Propagation in Large-Scale Networks", IEEE Transactions On Knowledge And Data Engineering, Volume 27, January 2015.