Intrusion Detection: Using Log and Activities of User on the Computer

Kurundkar G.D.[1]
Department of Computer Science
S.G.B.College,Purna(Jn.) (MS) India

Khamitkar S.D.[2]
School of Computational Sciences
S.R.T.M. University, Nanded(MS) India

Abstract: Intrusion detection consists of actions for detection of illegal activity of (intruders) system that recognize the intruders. Some important intrusion avoidance activities are writing and implementing good activity information security rule, planning and performing effectual information security programs, installing and testing technology-based information security system for counting intruders activities such as intrusion detection and prevention systems. In Information security intrusion detection systems (IDS) works like a burglar alarm in that it detects devastation and activates an alarm or system response such as sending message about intruder. Recently new technology for IDS systems is the intrusion prevention system (IPS), which can detect an intrusion and also prevent that intrusion from attacking the organization. There is a system called intrusion detection/prevention system (IDPS).Recently Snort is a very useful tool for Network based Intrusion detection. A Snort is tool which can use for collecting log and activities of intruder over the computer system. After collecting logs and activities of system user. By using data analysis Intruder is find out.

Keywords: *Intruder, prevention, activities, quartile, deviation, abnormal*

Introduction:

Intrusion Detection System collects information from a collection of system and network sources, and then examines the information for secret code of intrusion (attacks coming from outside the organization) and misuse (attacks from inside the organization.)

Intrusion detection systems help computer systems prepare for and deal with attacks. They collect information from a variety of vantage points within computer systems and networks, and analyze this information for security problems. Susceptibility assessment systems check systems and networks for system problems and configuration errors that represent security vulnerabilities. Both intrusion detection and vulnerability assessment technologies allow organizations to protect themselves from losses associated with network security problems.

Intrusion detection is a set of techniques and methods that are used to detect suspicious activity both at the network and host level. Intrusion detection systems (NIDSs) that passively monitor a network link are the ability of a skilled attacker to evade detection by exploiting ambiguities in the traffic stream as seen by the NIDS [1].

Intrusion detection systems fall into two basic categories: signature-based intrusion detection systems and anomaly detection systems. Intruders have signatures, like computer viruses, that can be detected using software. You try to find data packets that contain any known intrusion-related signatures or anomalies related to Internet protocols. Based upon a set of signatures and rules, the detection system is able to find and log suspicious activity and generate alerts. Anomaly-based intrusion detection usually depends on packet anomalies present in protocol header parts. In some cases these methods produce better results compared to signature-based IDS. Usually an intrusion detection system captures data from the network and applies its rules to that data or detects anomalies in it. Snort is primarily a rule-based IDS, however input plug-ins are present to detect anomalies in protocol headers.

For intrusion detection Snort is applicable, This snort tool is a freeware tool which collects all system logs which helps to find out Intruder in the current system or Network, Snort uses rules stored in text files that can be modified by a text editor. Rules are grouped in categories. Rules belonging to each category are stored in separate files. These files are then included in a main configuration file called snort .conf. Snort reads these rules at the start-up time and builds internal data structures or chains to apply these rules to captured data. Finding signatures and using them in rules is a problematic work, since the more rules you use, the more processing power is required to process captured data in real time. It is important to implement as many signatures as you can use as few rules as possible. Snort comes with a rich set of pre-defined rules to detect intrusion activity and you are free to add your own rules at will. You can also remove some of the built-in rules to avoid false alarms.

This document explains how intrusion detection and vulnerability assessment products fit into the overall construction of security products. It includes case histories outlining scenarios in which the products have been used by customer organizations. Finally, the concepts and definitions section provides information about product features, explaining why they represent effective countermeasures to hacking and misuse. Protecting important information systems and networks is a multifaceted operation, with many tradeoffs and considerations. The effectiveness of any security solution approach depends on selecting the right products with the right combination of features for the system environment one wishes to protect. In this document, we provide the information one needs in order to be a ability client in the areas of intrusion detection and vulnerability assessment.

NEED FOR INTRUSION DETECTION

A computer system should provide confidentiality, integrity and assurance against denial of service. However, due to increased connectivity with the network like Internet and the vast range of financial possibilities that are opening up, more and more systems are subject to attack by intruders. If there are attacks on a system, we would like to detect them as soon as possible (preferably in real-time) and take appropriate action. Intrusion detection is very important to make data secure. This is essentially what an Intrusion Detection System (IDS) does. An IDS does not usually take preventive measures when an attack is detected; it is a reactive rather than pro-active agent. It plays the role of an informer.

When we are working on the Internet it becomes our responsibility make our network more secure by using Network monitoring tools and making security settings and there are several other reasons to use an Intrusion Detection System.

- To detect attacks that are not prevented by other security measures
- To detect and deal with attacks
- To perform as quality organize for security design and administration, especially of large and complex enterprises
- To provide useful information about intrusions that do take place, allowing improved finding, improvement, and correction of contributing factors.

Who Misusing the System

There are two terms to explain the intruder: **hacker** and **cracker**. A hacker is a basic term for a person who likes getting into things. The benign hacker is the person who likes to get into his/her own computer and understand how it works. The malicious hacker is the person who likes getting into other people's systems. The benign hackers wish that the media would stop bad-mouthing all hackers and use the term 'cracker' as an alternative.

Following are the types of intrusion detection systems:-

i. Intrusion Detection System or IDS is software, hardware or combination of both used to detect intruder activity. Snort is an open source IDS available to the general public. IDS software’s are available with different facilities. There are two general methodologies of detection: *misuse* and *anomaly* detection [5,6].

ii. Network IDS or NIDS: NIDS are intrusion detection systems that capture data packets traveling on the network media e.g. data transmission cables & wireless, an alert is generated or the packet is logged to a file or database. One major use of Snort is as a NIDS. A fundamental problem for network intrusion detection systems (NIDSs) that passively monitor a network link is the ability of a skilled attacker to *evade* detection by exploiting ambiguities in the traffic stream as seen by the NIDS [1]. NIDS can eliminate much of the ambiguity if it has access to a sufficiently rich database cataloging the particulars of all of the end-system protocol implementations and the network topology. [2] Network intrusion detection systems (NIDSs) that passively monitor a network link is the ability of a skilled attacker to *evade* detection by exploiting ambiguities in the traffic stream as seen by the NIDS [3].

iii. Host IDS or HIDS: Host-based intrusion detection systems or HIDS are installed as agents on a host. These intrusion detection systems can look into system and application log files to detect any intruder activity. Some of these systems are reactive, meaning that they inform you only when something has happened. Some HIDS are proactive; they can sniff the network traffic coming to a particular host on which the HIDS is installed and alert you in real time. HIDS(host intrusion detection systems) take a different approach to detecting attacks, they look at each host separately. HIDS must be deployed on each system that is to be protected. HIDS have similar methods for detecting and identifying attacks. [4]

iv. Signatures: A signature is used to detect one or multiple types of attacks. e.g. packet going to your web server may indicate an intruder activity. Signatures may be present in different parts of a data packet depending upon the nature of the attack. e.g. you can find signatures in the IP header, transport layer header (TCP or UDP header) and/or application layer header or payload. **v. Alerts:** Alerts are any sort of user notification of an intruder activity. When an IDS detects An intruder, it has to inform security administrator about this using alerts. Alerts may be in the form of pop-up windows, logging to a console, sending e-mail and so on. Alerts are also stored in log files or databases where they can be viewed later on by security experts. Snort can also send the same alert to multiple destinations.

vi. Logs: The log messages are usually saved in file. By default Snort saves these messages under snort\log directory. However, the location of log messages can be changed using the command line switch when starting Snort. Log messages can be saved either in text or binary format. The binary files can be viewed later on using Snort or tcpdump program. A new tool called Barnyard is also available now to analyze binary log files generated by Snort. Logging in binary format is faster because it saves some formatting overhead. In high-speed Snort implementations, logging in binary mode is necessary.

Six systems are used to setup our research experiment.

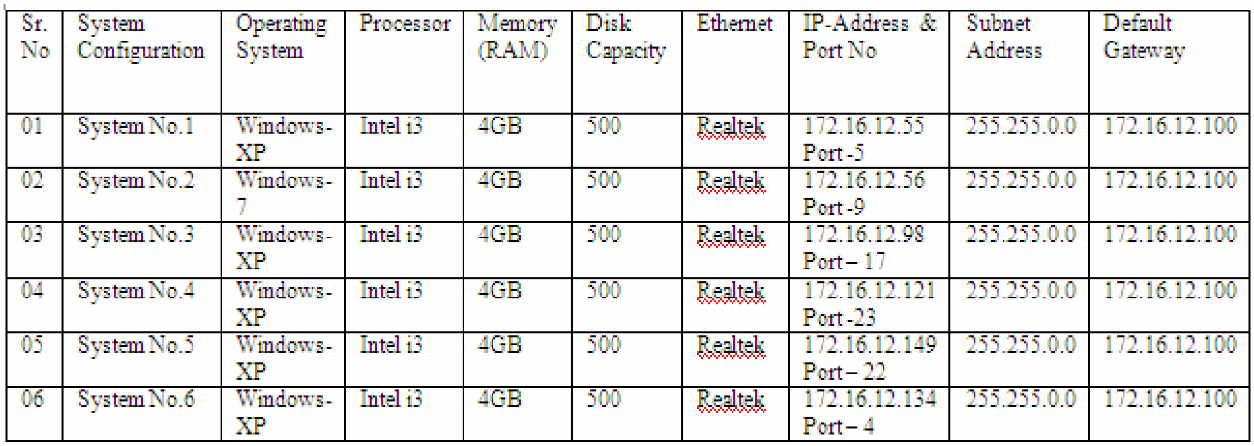
Intruders can be classified into two categories.

Outsiders: Intruders from outside your network, and who may attack you external presence (forward spam through e-mail servers, etc.). They may also attempt to go around the firewall to attack machines on the internal network. Outside intruders may come from the Internet, dial-up lines, physical break-ins, or from partner (vendor, customer, reseller, etc.) network that is linked to your corporate network.

Insiders: Intruders that legally use your internal network. These include users who misuse privileges (who masquerade as higher privileged users (such as using someone else's terminal). A frequently quoted statistic is that 80% of security breaches are committing by insiders.

Design of Experimentation:

For our Research Experiment we have setup small Computer Network and following equipments we have chosen managed switch with local ip address 172.16.12.100
Make: DIGISOL DG-GS4528S is a managed stackable Gigabit Ethernet the general configuration of this switch is



Sr. No	System Configuration	Operating System	Processor	Memory (RAM)	Disk Capacity	Ethernet	IP-Address & Port No	Subnet Address	Default Gateway
01	System No.1	Windows-XP	Intel i3	4GB	500	Realtek	172.16.12.55 Port -5	255.255.0.0	172.16.12.100
02	System No.2	Windows-7	Intel i3	4GB	500	Realtek	172.16.12.56 Port -9	255.255.0.0	172.16.12.100
03	System No.3	Windows-XP	Intel i3	4GB	500	Realtek	172.16.12.98 Port – 17	255.255.0.0	172.16.12.100
04	System No.4	Windows-XP	Intel i3	4GB	500	Realtek	172.16.12.121 Port -23	255.255.0.0	172.16.12.100
05	System No.5	Windows-XP	Intel i3	4GB	500	Realtek	172.16.12.149 Port – 22	255.255.0.0	172.16.12.100
06	System No.6	Windows-XP	Intel i3	4GB	500	Realtek	172.16.12.134 Port – 4	255.255.0.0	172.16.12.100

System Hardware Configuration:

Initially the behaviors (historical based) of the users which are using these machines are observed on following merits for one month.

- how much time they spend on the system
- Which web sites are they frequently visited.
- Which search engine is frequently used by the users?
- What are their requirements while searching on net
    - Entertainment
    - Education
    - Adult
    - Official

- Business
- Research Related
- Social

Depending on the observed data the rules for SNORT are manually designed for the implementation of the proposed practical work.

We have used one dedicated system for installation of SNORT and which is working on same network. The system where snort is installed monitors all the incoming and outgoing data of the systems which are connected on the same network.

The log was recorded on snort installed system. The experiment is carried out for the web sites which are in the rule file of Snort. We have created our own rules for some websites and these are used with snort for recording logs. The log is collected in text (.CSV) format.

We have used different ports on each machine for packet capturing , e.g. 5,22 etc.

The system wise analysis of the outcome of log which is recorded using SNORT is compared with the observed reading which are observed before implementation of the SNORT statistically. The observation and analysis of the result is totally depending on the behavior of the used on the visited web sites by the user after and before implementation of SNORT.

Procedure for Data Analysis:

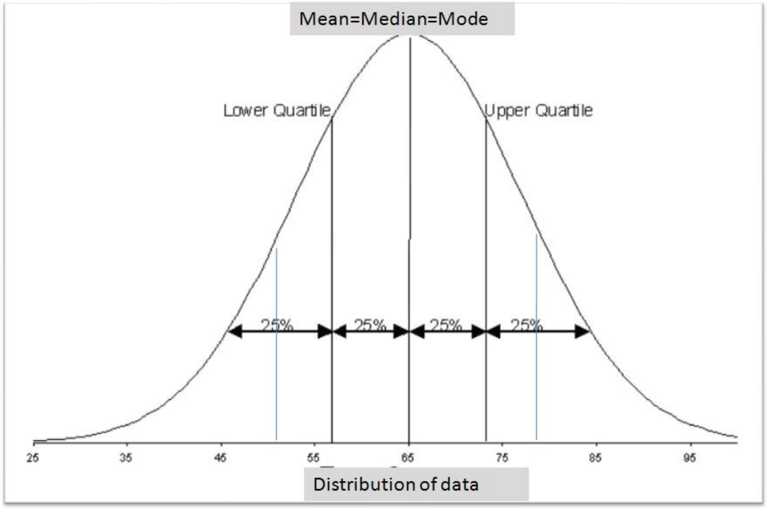
Data analysis is an integral part of the experimental studies to understand, explore and make inferences from given data set. Such data sets are assumed normally distributed because most of the measurable things, variables in the nature are sprayed in normal distribution. Regression toward the mean is an idea that states values will tend to cluster around the mean with few values toward the trailing ends or "tails" of the distribution. As a result, most things we measure will tend to have a normal shape.

The normal distribution is one that is unimodal and symmetric. The normal distribution consist two kinds of tendencies amongst which one is central tendency and another is variation. Mean, median and mode are the measures of central tendency and for a standard normal probability curve the values of mean, mode and median are one and same.

A deviation from the mean is the difference between a score and the mean. So, when we say the sum of the deviations about the mean must always equal zero is just a way of saying that there are just as many differences between values above the mean and the center as there are differences between values below the mean and the center.

Especially, mean and standard deviations are claimed to be most reliable measures because they are the product of mathematical operations and the impact of each value given in the data set. Therefore, these are the best descriptive measures but they are not competent to avoid the effect of extreme values and thus may fail to represent the group. For such as situation median and quartile deviation are most useful to describe the nature of data set. Quartiles are points in a distribution which divide that distribution into quarters. These quartiles are shown in Figure 1.1

Figure 1.1: Normal distribution of data



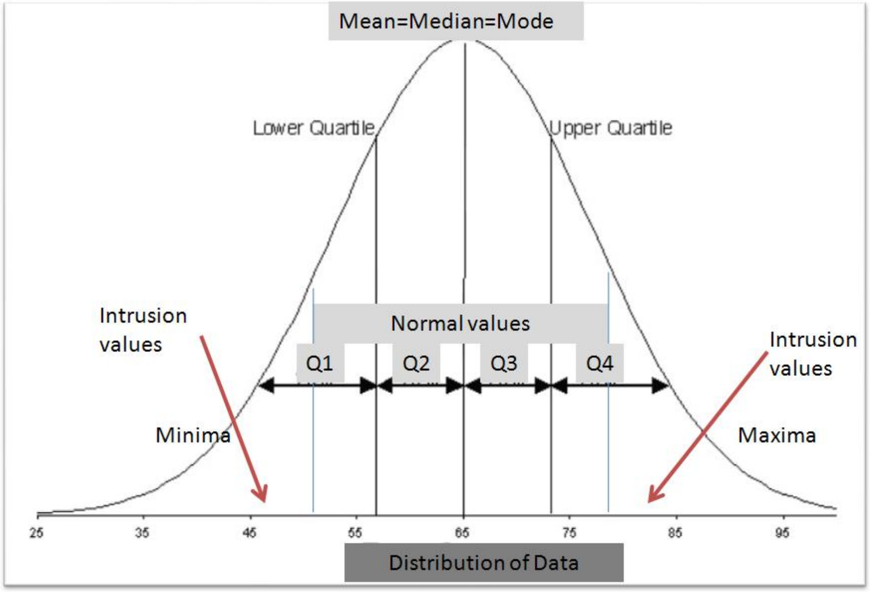
The above figure 1.1 shows that a normally distributed data set is divided in symmetrically in four sections. Each section is called as a quartile which covers the 25 % of values in given data. Thus, the first quartile (Q1) is the 25th percentile. The second quartile (Q2) is the 50th percentile or the median. The 75th percentile is the third quartile (Q3).

The upper limit and lower limit of the distribution are calculated by means of coefficient of first and third quartile. The first quartile is the median of the numbers located **below** the median; the third quartile is the median of the numbers **above** the median. The lower limit is the ratio of first quartile with sum of the data which is converted in to percentage by multiplying by 100. The upper limit is the ratio of third quartile with sum of the data which is converted in percentage by multiplying by 100. The boundary percentage of the data set (to identify the error tolerance capacity of the data set) is the average of upper limit percentage and lower limit percentage. The maxima and the minima of the distribution are calculated by adding and subtracting the boundary percentage value in to mean value respectively.

The values which fall in the within the maxima and minima indicate to be "normal" and the values beyond this boundary are identified as the "intrusions".

The critical regions for identification of the behavioral pattern (either "Normal" or "Intruder") in a normal probability distribution of the data extracted from the log maintained during the experimentation period is shown in the figure given below.

Figure 1.2 – Critical region for normal and intrusion values.

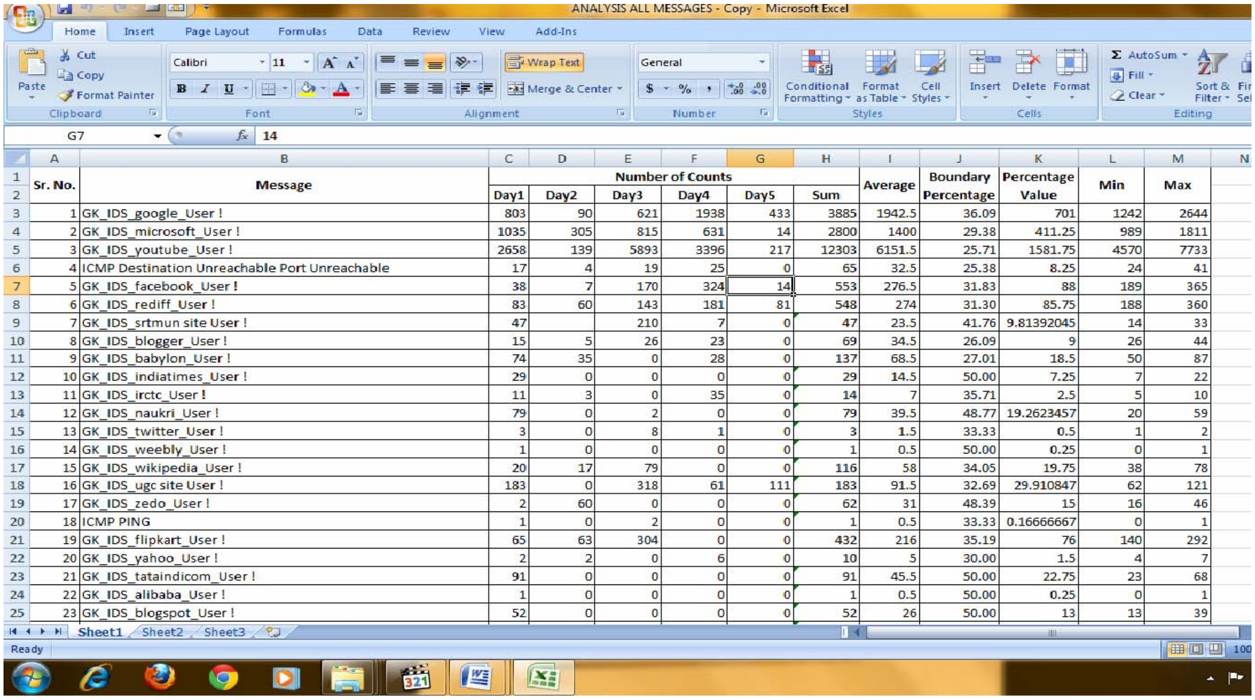


The region of intrusion values lies on the both/ opposite sides of the distribution. The intrusion value lay in the in the positive side which is beyond the median value of forth quadrant. The boundary of maxima for normal behavior also lay at this side. However, the intrusion value lay in the in the negative side which is before the median value of first quadrant. The boundary of minima for normal behavior also lay at this side.

On the basis of theoretical frame of data analysis discussed above researcher followed the procedure of data analysis in the steps given below**.**

Conversion of Data in to Spread sheet:

The day wise data regarding the users, IP addresses, Websites visited etc is converted from .CSV format into excel sheet using SNORT IDS tool and predefined rules. The spread sheet is shown in the following figure.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Sr. No.	Message	Number of Counts						Average	Boundary	Percentage	Min	Max
2			Day1	Day2	Day3	Day4	Day5	Sum		Percentage	Value		
3	1	GK_IDS_google_User !	803	90	621	1938	433	3885	1942.5	36.09	701	1242	2644
4	2	GK_IDS_microsoft_User !	1035	305	815	631	14	2800	1400	29.38	411.25	989	1811
5	3	GK_IDS_youtube_User !	2658	139	5893	3396	217	12303	6151.5	25.71	1581.75	4570	7733
6	4	ICMP Destination Unreachable Port Unreachable	17	4	19	25	0	65	32.5	25.38	8.25	24	41
7	5	GK_IDS_facebook_User !	38	7	170	324	14	553	276.5	31.83	88	189	365
8	6	GK_IDS_rediff_User !	83	60	143	181	81	548	274	31.30	85.75	188	360
9	7	GK_IDS_srtmun site User !	47		210	7	0	47	23.5	41.76	9.81392045	14	33
10	8	GK_IDS_blogger_User !	15	5	26	23	0	69	34.5	26.09	9	26	44
11	9	GK_IDS_babylon_User !	74	35	0	28	0	137	68.5	27.01	18.5	50	87
12	10	GK_IDS_indiatimes_User !	29	0	0	0	0	29	14.5	50.00	7.25	7	22
13	11	GK_IDS_irctc_User !	11	3	0	35	0	14	7	35.71	2.5	5	10
14	12	GK_IDS_naukri_User !	79	0	2	0	0	79	39.5	48.77	19.2623457	20	59
15	13	GK_IDS_twitter_User !	3	0	8	1	0	3	1.5	33.33	0.5	1	2
16	14	GK_IDS_weebly_User !	1	0	0	0	0	1	0.5	50.00	0.25	0	1
17	15	GK_IDS_wikipedia_User !	20	17	79	0	0	116	58	34.05	19.75	38	78
18	16	GK_IDS_ugc site User !	183	0	318	61	111	183	91.5	32.69	29.910847	62	121
19	17	GK_IDS_zedo_User !	2	60	0	0	0	62	31	48.39	15	16	46
20	18	ICMP PING	1	0	2	0	0	1	0.5	33.33	0.16666667	0	1
21	19	GK_IDS_flipkart_User !	65	63	304	0	0	432	216	35.19	76	140	292
22	20	GK_IDS_yahoo_User !	2	2	0	6	0	10	5	30.00	1.5	4	7
23	21	GK_IDS_tataindicom_User !	91	0	0	0	0	91	45.5	50.00	22.75	23	68
24	22	GK_IDS_alibaba_User !	1	0	0	0	0	1	0.5	50.00	0.25	0	1
25	23	GK_IDS_blogspot_User !	52	0	0	0	0	52	26	50.00	13	13	39

Figure 1.3 shows statistics analysis for log collection in

Statistical and logical Operations:

The statistical operations were applied to calculate the mean, SD, Boundary percentage etc. of the given data sets. Then, the logical operations were carried to identify “Normal” and “Intruder” by applying different association rules.

Compilation of Results:

The results obtained for each day were compiled after completion of experimentation period and they were put together for further analysis which is presented as following.

Data Analysis and Results:

The data of normal users and intruders obtained in the experimentation period is shown in the table given below.

	Duration of the data	**No. of normal users (Average)**	**No. of intruders (Average)**	**No. of intruders (Percentage)**
1	**8 weeks**	**20**	**2**	**10**
2	**6 weeks**	**18**	**3**	**16.7**
3	**4 weeks**	**19**	**2**	**10.5**
4	**2 weeks**	**21**	**2**	**9.5**
5	**1 week**	**17**	**6**	**35.3**
6	**3 days**	**22**	**1**	**4.5**
7	**2 days**	**21**	**0**	**0**
8	**1 day**	**18**	**0**	**0**
9	**Average (the difference of 0.5 is neglected)**	**20**	**2**	**10**

The above table shows that the average of users is 20 amongst which average of intruders is 02 (10%) . The maximum frequency of intruders is found in the data set of one week. It is seen that when data set is small (one or two day) , the intruders were with zero frequency and when the data set becomes large (4 to 8 weeks) the intruders lay in range of 02-03 (10 to 17%) but for a duration of one week the frequency of intruders found was 06 (35%).

For the purpose of comprehension of the usefulness of data set for identification of intruders the duration of the data and frequencies of intruders given in the table no. 1.1 are shown in the graph given below.

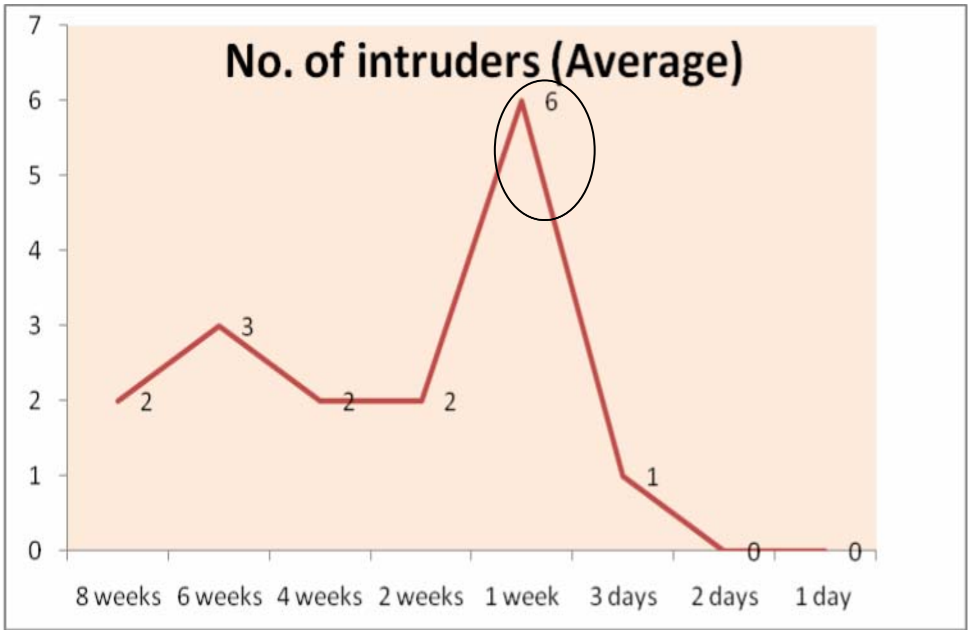


Figure 1.4: Average intruder

The graph show s the probability of finding maximum no. of intruders lays in the data set of one week than other small or larger durations' set.

Conclusion and Discussion:

The data set with duration of the 1 week gives the more discriminating results for identification of normal users and intruders. This, phenomenon can be explained in the frame of normal distribution and behavioral sciences.

We need a reference set of data to claim either normal or intruder. In a data set of one or two day every behavior is normal because they have no any reference set to categorize. Thus, the intruders can not be identified.

When a data set becomes large (of few weeks) then many behaviors are gathered for reference but simultaneously size of category of normal also increases which expands the upper limit (Maxima shown in the fig. no.1.2) and reduces the lower limit (Minima shown in the fig. no.1.2). Thus, the critical regions for identification of intruders become smaller and it makes difficult to identify intruders.

It is observed that to find out the behavioral change of user from Normal to Intruder and vice versa can be asserted by using data set of last week as reference set rather than the data set of more than one week.

References: References

[1] T. H. Ptacek and T. N. Newsham, "Insertion, Evasion and Denial of Service: Eluding Network Intrusion Detection", Secure Networks, Inc., Jan. 1998. http://www.aciri.org/vern/Ptacek-Newsham- Evasion- 98.ps

[2] Network Intrusion Detection: Evasion, Traffic Normalization, and End-to-End ProtocolSemantics by Mark Handley and Vern Paxson *AT&T Center for Internet Research at ICSI (ACIRI) International Computer Science Institute Berkeley, CA 94704 USA*

[3] T. H. Ptacek and T. N. Newsham, "Insertion, Evasion and Denial of Service: Eluding Network Intrusion Detection", Secure Networks, Inc., Jan. 1998. http://www.aciri.org/vern/Ptacek-Newsham- Evasion-98.ps

[4] An Agent-Based Intrusion Detection System by Theodor Richardson, Goran Trajkovski Department of Information Technology South University.

[5] Kumar, S., Spafford E. H.: software Architecture to Support Misuse Intrusion Detection. In Proceedings of the 18th National Information Security Conference. (1995) 194-204

[6] Forrest, S., Hofmeyr, S. A., Somayaji, A., Logstaff, T. A.: A Sense of Self for Unix process,Proceedings of 1996 IEEE Symposium on Computer Security and Privacy. (1996) 120-128