IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 5, July 2014.
www.ijiset.com
ISSN 2348 – 7968

Design and simulation of wireless network for Anomaly detection and prevention in network traffic with various approaches

Niraj Telrandhe*, Mangesh Wanjari**

M.tech Scholar*, Asst.Prof**

Wainganaga College of Engineering& Management, Nagpur*

Ramdevbaba Kamla Nehru Engineering College, Nagpur**

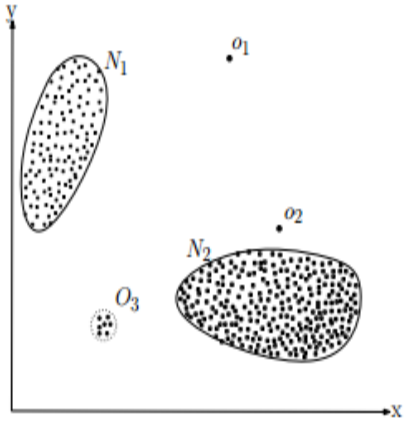
niraj.telrandhe@gmail.com*, mangeshwanjari@gmail.com

Abstract: *Anomaly detection is the process of identifying unusual behavior. It is widely used in data mining, for example, to identify fraud, customer behavioral change, and manufacturing flaws. We discuss how a probabilistic framework can elegantly support methods to automatically explain why observations are anomalous, assign a degree of anomalies, visualize the normal and abnormal observations and automatically name the clusters.*

This paper discusses a statistical approach to analysis the distribution of network traffic to recognize the normal network traffic behavior The Research proposals in anomaly detection typically follow a four-stage approach, in which the first three stages define the detection method, while the last stage is dedicated to validate the approach method to detect anomalies in network traffic, based on a non restricted α -stable first-order model and statistical hypothesis testing. Here we focus on detecting and preventing two anomaly types, namely floods and flash-crowd .Here we use NS2 simulator to calculate result.

I. What are anomalies?

Anomalies are patterns in data that do not conform to a well defined notion of normal behavior. Figure 1 illustrates anomalies in a simple 2-dimensional data set. The data has two normal regions, N1 and N2, since most observations lie in these two regions. Points that are sufficiently far away from the regions, e.g., point's o1 and o2, and points in region O3, are anomalies



simple example of anomalies in a 2-dimensional data set.

Anomalies might be induced in the data for a variety of reasons, such as malicious activity, e.g., credit card fraud, cyber-intrusion, terrorist activity or breakdown of a system, but all of the reasons have a common characteristic that they are interesting to the analyst.

Recognize network anomalies are serious for the timely mitigation of events, like attacks or failures that can affect the security, SLAs, and performance of a network. Anomalies can come from action with malicious intentions (e.g., scanning, DDoS, prefix hijacking), or from mis configurations and failures of network components (e.g., link failures, routing problems, outages in measurement

equipment), or even rightful events such as strangely large file transfers or flash crowds. Traffic analysis and anomaly detection are extensively used to understand and characterize network traffic behavior, as well as to identify abnormal operational conditions such as malicious attacks. However, techniques for traffic analysis and anomaly detection are typically carried out independently in different parts of the network, either in the edge or in the core networks alone. In fact, different traffic characteristics and anomalies can normally be better observed in a specific part of the network, although they affect the network as a whole Most works these days center on flow-level data. At least five minutes(net flow data) delay is predictable even for the online detection methods, so anomaly detection methods depend on flow-level data are usually use for the warning/alerting to the network manager and hard to be used for the next generation intrusion detection system design. Ideal IDS, besides warning, should identify the anomaly packet in real time and block it. Hence, exploring detection methods based on packet-level data is indispensable. Our work mainly focuses on anomaly detection for the packet-level data. A number of techniques have been proposed in order to identify anomalies by analyzing network traffic. They all seek to expose anomalies by detecting deviations from some underlying model of normal traffic. Usually, these kinds of models have to be learned from days or weeks of anomaly-free traffic traces, which is a practical problem since the training data is never guaranteed to be clean and training should be performed periodically.

Research proposals in anomaly detection typically follow a four-stage approach, in which the first three stages define the detection method, while the last stage is dedicated to validate the approach. So, in the first stage, traffic data are collected from the network (data collection). Second, data are analyzed to extract its most relevant features (data analysis). Third, traffic is classified as normal1 or abnormal (inference); and fourth, the whole approach is validated with various types of traffic anomalies.

1) Statistics Collection.
2) Statistics analysis (feature extraction).
3) Conclusion (classifying normal vs. anomalous traffic).
4) Justification.

Statistics Collection is typically carried out by polling one or more routers periodically, so that traffic data are collected and stored for posterior analysis in the second stage. Some authors sample data at the packet level, gathering information from headers, latencies, etc., while others prefer to use aggregated traffic as the source of information, often through the use of the Simple Network Management Protocol (SNMP). Sampling data at the packet level provides more information, but at the cost of a higher computational load and dedicated hardware must be employed. Aggregated traffic, on the other hand, gives less information from which to decide for the presence or absence of anomalies, but is a simpler approach and does not need any special hardware.

In the statistics analysis phase, several techniques can be applied to extract interesting features from current traffic. Some of them include information theory [1], [9] wavelets [6], statistics-based measurements [3], and statistical models. Of these techniques, the use of statistical models as a means to extract significant features for data analysis has been found to be very promising, since they allow for a robust analysis even with small sample sizes (provided that the model is adequate for real data). Moreover, with a traffic model, its set of parameters can be used as extracted traffic features, since any traffic sample is determined by the model parameters.

The fact that these models do not account for high variability may have a negative impact on capturing traffic properties and, as a consequence, on detecting anomalies.

ISSN 2348 – 7968

High variability manifests itself in the marginal (first-order) traffic distribution and states that traffic is inherently burst. This results in traffic distributions exhibiting heavy tails which cannot be properly modeled with, e.g., Gaussian functions. Long-range dependence, on the other hand, states that traffic is highly dependent over a wide range of time scales, i.e., its autocorrelation function exhibits a slowly decaying tail Several statistical distributions are capable of modeling the high variability property. One of such distributions is the α-stable family, which has been previously used to model network traffic (where the detection problem is not addressed). To the best of our knowledge, these distributions have never been applied to anomaly detection. Moreover, in addition to properly modeling highly variable data, α-stable distribution are the limiting distribution of the generalized central limit theorem, a fact that sets them as good candidates for aggregated network traffic. Regarding the time evolution model and long-range dependence, the first-order α-stable model is

II. RELATED WORK:

2.1 Volume Depends anomaly detectors

Volume depends loom are monitoring the number of bytes, packets or flows broadcast more time and aims at detecting irregular variances that represent abusive usages of network resources or resource failures. Several technique have been proposed to effectively recognize local and global traffic volume variances that place for respectively short and long lasting anomalies. For example, bar ford et al. [5] proposed a technique based on wavelet that inspects the traffic volume at different frequencies. Their loom makes use of the wavelet examination to dissect the traffic into three distinct signals instead of local, normal and global variances of the traffic. The rotten signals are analyzed by a detection procedure that finds the irregularities and information the period of time they occur. Since the three signals represent the traffic at dissimilar time scales this approach is able to report short and long lasting anomalies. Nevertheless, as the whole traffic is collective into a single signal analyze the detected anomalies is challenging and anomalous flows or IP addresses are left unknown.

Lakhina et al. [6] proposed a recognition method that perceive and diagnoses anomalies in large scale networks. First, their approach checks the traffic using a matrix in which each cell symbolizes the traffic volume of a link of the network at a certain time interval. Second, the main behavior of the traffic is removing from the matrix with the principal component analysis (PCA) and anomalies are detected in residual traffic. Finally, the origin and destination nodes of the network that are exaggerated by the anomalous traffic are recognized and reported.
traffic volume in matrices. The main idea fundamental their approach is to represent in a matrix the traffic between nodes of a large network and remove the normal traffic using a Kalman filter. The remaining traffic is analyzed with a statistical method that detects anomalous traffic and reports the pair of nodes exaggerated by the anomalous traffic.

These volume-based anomaly detectors successfully report volume anomalies while their false positive rate is low. Their plan, however, restrict them to report only a few classes of anomaly, thus, network operative need additional detectors to identify threats that are invisible in the interchange volume (e.g., network scan or port scan).

2.2 Abnormality Exposure

Detecting abnormal traffic is a research topic that had recently established a lot of attention. We classify this topic into two domains; network intrusion detection and Internet traffic anomaly detection. The goal of intrusion detection is to protect a

network from remote threats, thus, the detection method is monitoring the traffic at the edge of the protected network where complete flows and packet payload are usually accessible. In contrast, Internet traffic anomaly detection aims at identifying anomalous traffic that is transiting in the core of the Internet where the monitored traffic is asymmetric due to routing policies, thus, flows are incomplete. For the last decade researchers have taken a strong interest in anomaly detection and proposed different detection methods that are basically monitoring traffic characteristics and discriminating outliers. We differentiate different categories of anomaly detection method; the methods monitoring the traffic volume and those monitoring the distribution of traffic features.

2.3 Traffic features Depend Abnormality Detectors:

In order to conquer the drawbacks of volume-based anomaly detectors re-searchers proposed to purify the traffic features that are inspect by the anomaly detectors. For example, as many anomalies cause abnormal operation of ports or addresses, inspecting the sharing of the traffic into the port and address spaces permits to identify anomalous traffic that is not reported by volume-based detectors (e.g., port scan). Nevertheless, due to the size of analyzed traffic examine detailed traffic features are costly and impose researchers to complicated effective traffic aggregation schemes. The main challenge in collective network traffic is the tradeoff between maintaining a concise representation of the traffic and preserving its interesting characteristics. We distinguish four groups of detection method in regard to their traffic aggregation scheme; namely, (1) Recognition methods aggregating the traffic in a single signal,(2) those collective the traffic in traffic matrices, (3) methods collective traffic in histograms, and (4) the other methods.

2.4 Packet Filtering for Flow-Based information:

In packet filtering, packet flows are sampled by capturing the IP headers of a select set of packets at different points in the network Information gathered from these IP headers is then used to provide detailed network performance information. For flow-based monitoring, a flow is identified by source destination addresses and source-destination port numbers. The packet filtering approach requires sophisticated network sampling techniques as well as specialized hardware at the network devices to do IP packet lookup. Data obtained from this method could be used to detect anomalous network flows. However, the hardware requirements required for this measurement method makes it difficult to use in practice.

2.5. Data from Routing Protocols:

Information about network proceedings can be gain through the use of routing peers. For example by using an open shortest path first (OSPF) peer, it is possible to get together all routing table updates that are sent by the routers. The data collected can be use to build the network topology and provides link status updates. If the routers run OSPF with traffic engineering (TE) extensions, it is possible to get link operation levels. Since routing updates occur at recurrent gap, any change in link utilization will be updated in near real time. However, since Routing updates must be kept small; only limited information pertaining to link statistics can be propagated through routing updates [9]

III. EXPERIMENTAL SETUP:

This Research work design and implemented in NS2. NS (version 2) is an object-oriented, discrete event driven network simulator developed at UC Berkelywritten in C++ and OTcl. NS is primarily useful for simulating local and wide area networks. Tcl is a general purpose scripting language. While it can do anything other languages could

possibly do, its integration with other languages has proven even more powerful.

In this section we present the experimental setup of our research work with complete result. As mentioned we use the NS2 to calculate the result. Basically we focus on to detecting and preventing flood and flash crowd anomaly in network. Here we consider the 10 nodes in network and sending the packet at regular interval of time and providing the proper threshold to calculate the anomaly in network. The generalized ratio test can be used to divide the anomalous network. And draw the result through graph.

3.1 Flash Crowd Anomaly:

A flash crowd occurs when there is a surge in demand for a service and is typically manifested by a large number of clients trying to access network resources.

Flash-crowd anomalies encompass traffic patterns which are caused by a net growth of (usually human) users trying to access a network resource. Typical flash-crowd anomalies are related to overwhelming web server usage patterns.

3.2 Flood anomaly:

Flood anomalies include attacks, or any other circumstances, which result in a net growth of instantaneous traffic. One can think of flood anomalies as having one or more relatively constant traffic sources added to otherwise normal traffic. DDoS attacks typically give rise to anomalies of this kind.

Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks are one of the most common malefic actions over the Internet. This type of attacks consumes the resources of a remote host or network that would otherwise be used to serve legitimate users. Nowadays a diversity of tools is available to accomplish DoS and DDoS

```
Agent/My Agent set my Val 0
Agent/My Agent set bottle_neck 10; #set a bottle
neck for the transmissions
Agent/My Agent set RTSEQ 4200; #set some value
for RTSEQ
Agent/My Agent set NUM_NODES $opt(num Of
Nodes)
Agent/My Agent set PACKET_THRESHOLD 10
#SET THE SOURCES AND DESTINATIONS
Agent/My Agent set source00 $sources (0)
Agent/My Agent set dest00 $dest(0)
Agent/My Agent set source01 $sources (1)
Agent/My Agent set dest01 $dest(1)
Agent/My Agent set source02 $sources (2)
Agent/My Agent set dest02 $dest(2)
Agent/My Agent set source03 $sources (3)
Agent/My Agent set dest03 $dest(3)
Agent/My Agent set source04 $sources (4)
Agent/My Agent set dest04 $dest(4)
Agent/My Agent set source05 $sources (5)
Agent/My Agent set dest05 $dest(5)
Agent/My Agent set source06 $sources (6)
Agent/My Agent set dest06 $dest(6)
Agent/My Agent set source07 $sources(7)
Agent/My Agent set dest07 $dest(7)
Agent/My Agent set source08 $sources (8)
Agent/My Agent set dest08 $dest(8)
Agent/My Agent set source09 $sources (9)
Agent/My Agent set dest09 $dest(9)
Agent/My Agent set source10 9
Agent/My Agent set dest10 2
Agent/My Agent set source11 9
Agent/My Agent set dest11 7
#Agent/My old Agent set my Val_ 10
```

Using Statistical Approach we statistically indicate source & destination. Here we declare source node 1,2,3---& destination source 2,3,4,5 ---.Source 1 will send

ISSN 2348 – 7968

packet to destination 2 only & so on. If any abnormal activity occurs just like sends packet to destination 7, then anomaly is detected, called Flash Anomaly. If any node receives large no packet and cross the threshold limit called flood anomaly. If any unwanted movement occurs the packet would not be send

IV. PROPOSED WORK AND METHODOLOGY

The methodical work that is followed to differentiate network traffic and to get anomaly information connected with the traffic examines. The method occupies the steps followed to produce anomaly result. The steps start with examining of the simulated data by using (NS2) and ends with a graph representing the abnormal traffic and normal traffic in a time interval.In research proposed method to detect and prevent the anomaly in network traffic, by using the statistical approach and α-stable model.

4.1. STATISTICAL ANOMALY DETECTION:

The potential to detect unknown attacks is the strength of statistical anomaly detection systems. Anomaly detection systems derive a model of the normal behavior of a network or system and detect divergence from this normal profile. This enables them to detect known and source 1

unknown malicious activities likewise. The normal profile has been derived based on different Information such as system calls on a single host, payload byte patterns in received traffic, or volume and entropy Information over the traffic in a whole network

4.2 .Statistical Anomaly Detection Algorithm:

STEP: 1 Node Initialization
I =1 to 10
Initialize Threshold =value
STEP: 2 Transfer Packets in Sequential Node
For I=1 to 10
Xmt (node [i], node (i+1))
STEP: 3 If (xmt (node (i), node (i+1)!)
Display "Anomaly Detected"
Then, If (Threshold==n)
(a) Count the Packet on each Node= Counter
Threshold ≥ Counter
(b) DDoS attack Detected i.e. Flood anomaly
detected
Else, Display "No anomaly found"
Packet Received (node (i), node (i+1)
(c) Display Counter on Node [i]
If (i==10)
Xmt (node [i-(i-1)], node [i])
Display "Flash Anomaly Detected",
go to call (b)

4.3 FLOWCHART FOR ANOMALY DETECTION:

ISSN 2348 – 7968

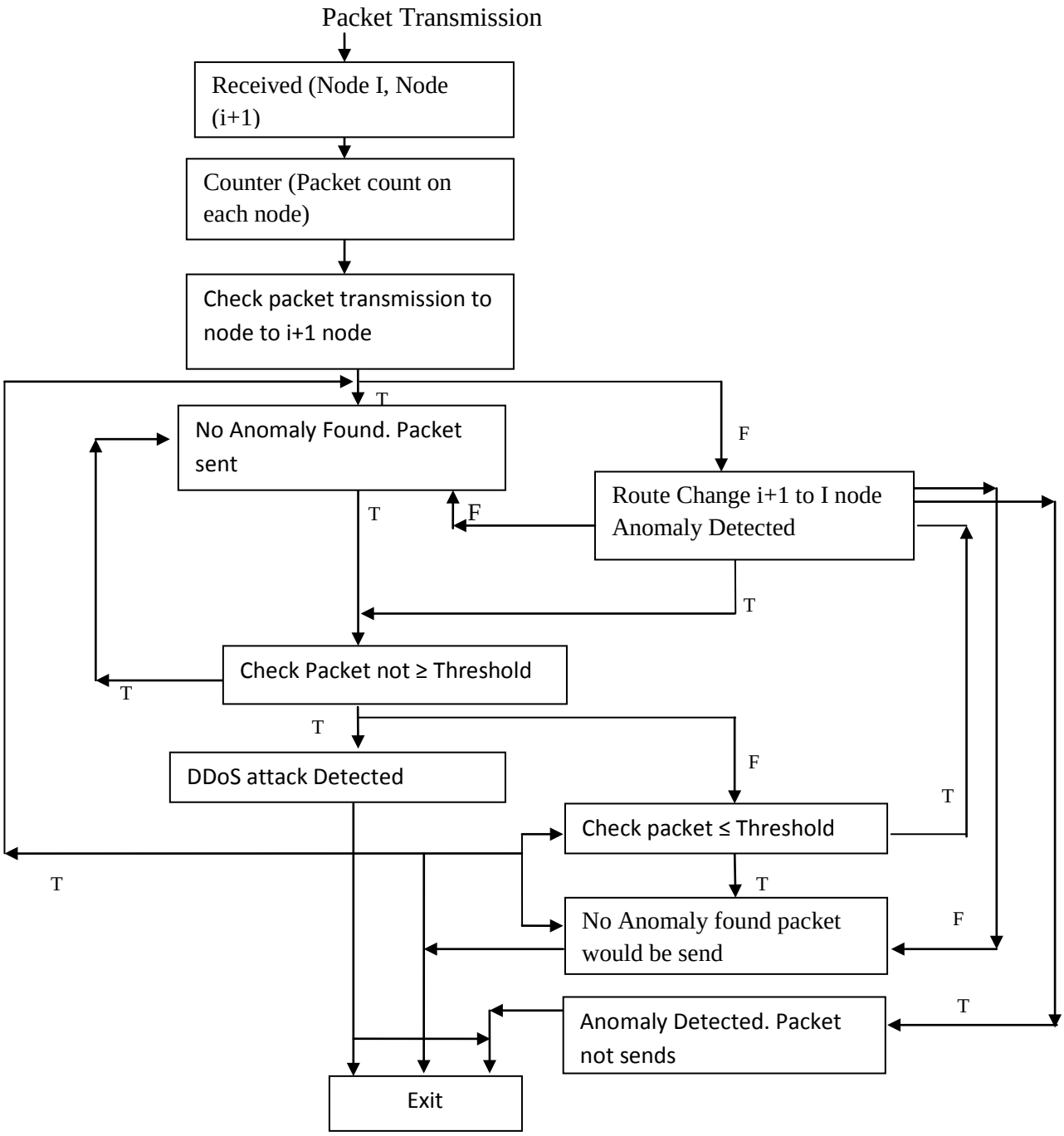


Figure.4.1 Statistical Approach for Anomaly finding in Network

V. DESIGN OF NETWORK:

In this section, we present the experimental result of our research work. As mentioned we use the NS2 to calculate the result. Here we focus on to detecting and preventing flood and flash crowd anomaly in wireless network. Here we consider the 10 nodes in network and sending the packet at regular interval of time and finding out the behavior of network and providing the proper threshold to calculate the flood anomaly in network. The generalized ratio test can be used to divide the anomalous network.

IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 5, July 2014.
www.ijiset.com
ISSN 2348 – 7968

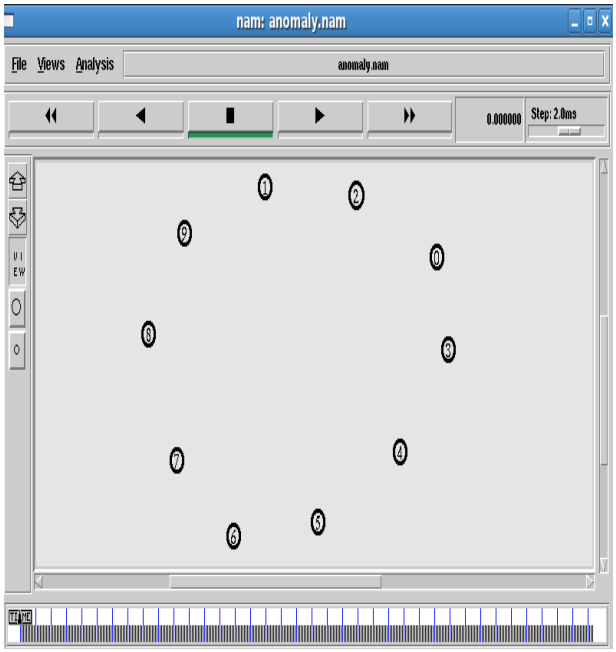


Figure 5.1 Nam output showing nodes in wireless networks

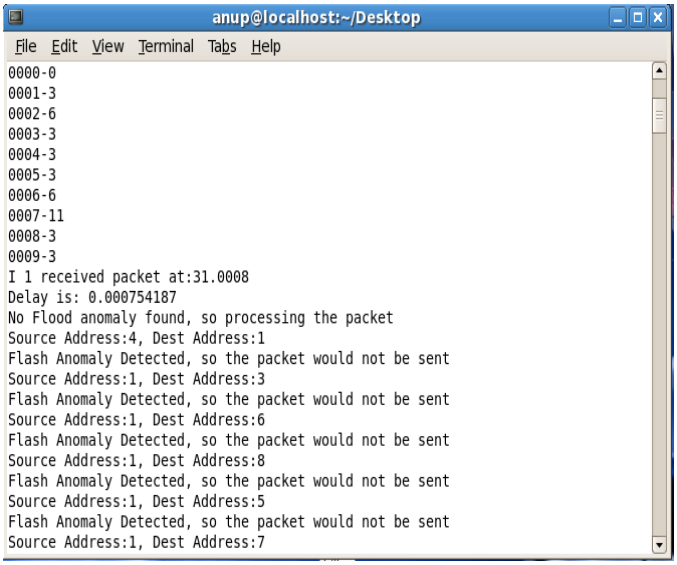


Figure.6.1 Node I1 Received packet at 31.008 time interval. Flash anomaly Result.

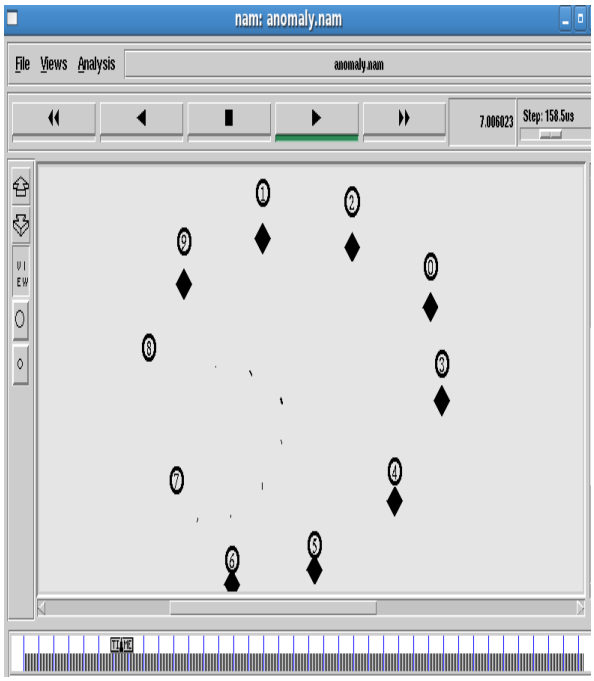


Figure 5.2 Packet Transmission of Node.

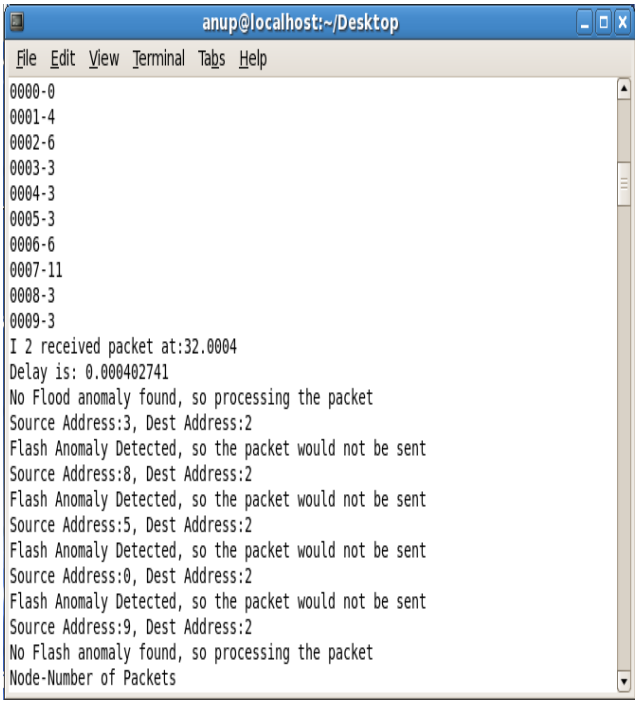


Figure.6.2 Packet Received by Node I2 with 32.004 time interval. Flash anomaly result.

VI. RESULTS ON EACH NODE:

ISSN 2348 – 7968

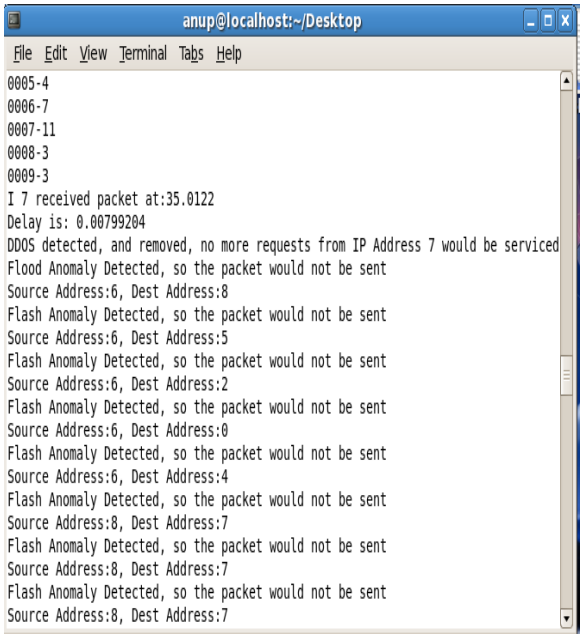


Figure 6.3 DDOS attack Detected on Node I7. Flood anomaly Result.

VII. CONCLUSION:

This Paper Presents the idea about the anomaly Detection in network Traffic, and also discusses statistical approach for anomaly Detection in Network Traffic. Ns2 is used for Design of Network and calculating the simulating Result.

References:

[1]. Federico Simmross, Juan Ignacio, Pablo Casaseca-de-la-Higuera, Ioannis A. Dimitriadis‖ Anomaly Detection in Network Traffic Based on Statistical Inference and α-Stable Modeling‖ IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 8, NO. 4, JULY/AUGUST 2011

[2] M. Thottan and C. Ji, ―Anomaly Detection in IP Networks,‖ IEEE Trans. Signal Processing, vol. 51, no. 8, pp. 2191-2204, Aug. 2003.

[3] C. Manikopoulos and S. Papavassiliou, ―Network Intrusion and Fault Detection: A Statistical Anomaly Approach,‖ IEEE Comm. Magazine, vol. 40, no. 10, pp. 76-82, Oct. 2002.

[4] Y. Gu, A. McCallum, and D. Towsley, ―Detecting Anomalies in Network Traffic Using Maximum Entropy Estimation,‖ Proc. Internet Measurement Conf., Oct. 2005.

[5] P. Barford, J. Kline, D. Plonka, and A. Ron, ―A Signal Analysis of Network Traffic Anomalies,‖ Proc. Second ACM SIGCOMM Workshop Internet Measurement, pp. 71-82, Nov. 2002

[6] A. Lakhina, M. Crovella, and C. Diot. Mining anomalies using traffic feature distributions. SIGCOMM '05, pages 217{228, 2005. (Cited on pages 12, 13, 25, 32, 40, 48, 57, 91 and 96.

[7] A. Lakhina, M. Crovella, and C. Diot. Mining anomalies using traffic feature distributions. SIGCOMM '05, pages 217{228, 2005. (Cited on pages 12, 13, 25, 32, 40, 48, 57, 91 and 96.)

[8] [48] X. Li, F. Bian, M. Crovella, C. Diot, R. Govindan, G. Iannaccone, and A.Lakhina. Detection and identi_cation of network anomalies using sketch subspaces. IMC '06, pages 147{152, 2006. (Cited on pages 3, 5, 12, 13, 22 and 57.)

[9] S. S. Kim and A. L. N. Reddy. A study of analyzing network traffic as images in real-time. INFOCOM '05, pages 2056{2067, 2005. (Cited on pages 14 and 32.)

[10] L. I. Kuncheva. Combining Pattern Classi_ers: Methods and Algorithms. Wiley-Interscience, 2004. (Cited on pages 15 and 61.)