Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 12, December 2013, pg.298 – 302

SURVEY ARTICLE

Survey of the High Volume Network Traffic Data Monitoring

Birunda Devi.M[1], Vairachilai.S[2]

PG Student[1], Assistant Professor[2]

Department of Computer Science and Engineering,
NPR College of Engineering and Technology,
TamilNadu, India.
E-mail: birunda.devi@gmail.com

Abstract- Network Traffic monitoring describes the use of a system that constantly monitors a computer network for slow or failing components and that notifies the network administrator in case of outages. Network traffic measurement is the process of measuring the amount and type of traffic on a particular network. This is especially important with regard to effective bandwidth management. Network traffic is monitoring and analysis in order to maintain the network system stability and availability such as to fix network problems on time or to avoid network failure, to ensure the network security strength, and to make good decisions for network planning. The monitoring, detecting and analysis of this traffic may be against the goal of maintaining privacy of individuals whose network communications are being monitored. The network traffic monitoring can identify performance of the static and dynamic IP address.

Index terms: Traffic data; Netflow; tcpdump; Monitoring

I. INTRODUCTION

Network monitoring for a corporate network is a critical IT function that can save money in network performance, employee productivity and infrastructure cost overruns. A network monitoring system monitors an internal network for problems. It can find and help resolve snail-paced webpage downloads, lost-in-space e-mail, questionable user activity and file delivery caused by overloaded, crashed servers, dicey network connections or other devices.

Network monitoring can be done by on a high level into maintaining the network's current status, ensuring availability and improving performance. An NMS also can help you build a database of critical information that can use to plan for future growth. Network monitoring systems come with automatic discovery, which is the ability to continuously record devices as they're added, removed or undergo configuration changes.

The network monitoring can be based different types of networks like wired, wireless, LAN, VPN, WAN. Network monitoring can be achieved using various or a combination of plug-and-play hardware and

software appliance solutions. Virtually any kind of network can be monitored. It doesn't matter whether it's wireless or wired, a corporate LAN, VPN or service provider WAN. Monitor devices on different operating systems with a multitude of functions, ranging from Blackberry and cell phones, to servers, routers and switches. These systems can help you identify specific activities and performance metrics, producing results that enable a business to address various and sundry needs, including meeting compliance requirements, stomping out internal security threats and providing more operational visibility.

Performance-sensitive functions include voice over IP (VoIP), Internet Protocol TV (IPTV) and video on demand (VOD). Monitoring enables managers to allocate resources to maintain system integrity. Network monitoring can provide the Measure latency, or the delayed transfer of data. The monitoring, detecting and analysis of this traffic may be against the goal of maintaining privacy of individuals whose network communications are being monitored.

II. LITERATURE REVIEW

2.1 Monitoring Infrastructure for Converged Networks and Services

Network convergence is enabling service providers to deploy a wide range of services based IP networks. Traditional network management systems are not suitable to managing converged networks. This enables service providers to monitor and manage their networks to provide the necessary quality, availability, and security. In this monitoring infrastructure supports three primary functions: service assurance, traffic profiling, and fault detection and diagnosis.

Traditionally, IP networks have been managed by measuring aggregate parameters that are link utilization and packet losses, over interfaces of routers or other network elements. This proposed monitoring infrastructure supports three primary functions: service assurance, traffic pro-filing, and fault detection and diagnosis.

These requirements drive the need for a robust, scalable, and easy-to-use network management platform that enables service providers to monitor and manage their networks to provide the necessary quality, availability, and security. This paper describes the Monitoring mechanisms that give service providers critical information on the performance of their networks at a per-user, per-service granularity in real time. This allows the service providers to ensure that their networks adequately satisfy the requirements of the various services.
The proposed monitoring infrastructure has the following components:

Measurement source:

This provides the element for the monitoring applications from the necessary measurements. The monitoring applications use this information to deduce the performance of the network and to identify faults.

Topology inventory:

The data collected using the measurement sources can be used to identify problems in the network.

Measurement layer:

This layer provides an inter-face that can be used by monitoring applications to gain access to the data collected by the various measurement sources.

Monitoring applications:

Finally, the various monitoring applications use the data from the measurement layer to offer service providers insight into the performance of their networks for each service.

The proposed monitoring infrastructure consisting of various measurement methods and this provides a powerful platform to develop and deploy a wide variety of monitoring applications. Measurements through monitoring mechanisms are mapped by voice quality metrics to the actual quality of the end-user experience, referred to as mean opinion score (MOS).

2.2 An Information-Theoretic Approach to Network Monitoring and Measurement

In this paper include the information of packet measurement traces, monitoring collection ranges of packets and trace the join information. The information of packet measurements traces are made at either the flow level trace (NetFlow traces) or the packet/byte count level trace (SNMP traces). It propose and validate a flow-level model and it use to determine the information content of a packet trace collected at a single or multiple points in a network, and of sets of packet traces collected at separate points. This is used for to identify and quantify the potential benefit of network trace compression based on the network flow model.

Regarding traces collected at a single monitoring point and derive an information theoretic bound for the information content in those traces, or equivalently for the potential benefit of lossless compression on those traces. NetFlow or packet header capture and reporting can require a significant amount of bandwidth and storage. The techniques have the potential to significantly reduce those bandwidth and storage requirements.

The benefit gained from compressing a packet header trace gathered from one network monitor or traces collected at a set of network monitors.

Customer traffic largely goes to one of the two back-bone links. It is an ideal first step to analyse information redundancy without being overwhelmed by complicated routing. The full packet traces to deduce an SNMP-like utilization trace (referred to as utilization trace) and an unsampled raw Net-flow trace. NetFlow processes all the packets in all the flows. In a real network environment, both the number of flows and the number of packets generated by those flow are large. NetFlow can employ flow sampling, where only a fraction of flows are monitored, and packet sampling, where only a fraction of packets are counted, to reduce its cost in computation and memory, etc.

2.3 Data Pre-Processing Evaluation for Web Log Mining: Reconstruction of Activities of a Web Visitor

Presumptions of each data analysis are data themselves focus of selected analysis highly depend on the quality of analysed data. The analysis data are creating data matrices and web map based on these data will serve for searching for behaviour patterns of users. The data are tried to assess the impact based on the quantity and quality of the users' behaviour patterns. The observed variables into the log file are the necessary data like IP address, URL address, data and the time access etc.

Good quality data are a prerequisite for a well-realized data analysis. The purpose of using this as for the following

Survey sampling –to find out particular items in the questionnaire and a visitor of the knows objects of site,

Web log mining – to analyse the log file of the web server, this contains information on accesses to the pages of the visitor.

The aim of the is to find out using an experiment to what measure it is necessary to execute data preparation for web log mining and determine inevitable steps for gaining valid data from the log file. More accurately, the aim is to assess the impact of reconstruction of the activities of a web visitor on the quantity and quality of the extracted rules which represent the web users' behaviour patterns. Log file of the web server is a source of anonymous data about the user. These anonymous data represent also the problem of unique identification of the web visitor.

There exist several accesses from a single IP address with various versions of the browser or operating system, there is not only one user also assume that if such an access can be accessed as the one by other user. Session can be defined as a series of steps, which lead to the fulfilment of a certain task or as a series of steps, which lead to the reaching of a certain goal. The simplest method is to consider a session to be a series of clicks for a certain period of time. A real value for session can be derived based on empirical data.

The experiments for data preparation can be consisted of the following steps: data cleaning, identification of sessions and reconstruction of activities of a web visitor. The data cleaning is filtration of whole data preparation process. Identification of sessions of the user allows us to eliminate NAT and proxy devices, as well as identify various users alternating behind one computer. The analyses of the log file of the web server are performed efficiently. This provides the quantity and quality of the data matrix representation format for the users' web log files.

2.4 A Packet-Level Characterization of Network Traffic

A packet-level traffic characterization aiming at finding spatial and temporal invariance of TCP based applications, such as HTTP and SMTP. In traditional SMTP traffic characterization method is simply generalize and packet level are distinct behaviours of different traffic can be captured.

A methodology and software architecture to build packet level statistical characterization of network traffic based on large trace. There is a deeply investigated on HTTP traffic modelling and characterization at packet level. Maximum Transmission Unit (MTU) of the network interfaces on the hosts, which limits the size of packets that can be sent and received .Network traffic can be viewed at different abstraction levels are session, conversation, connection/flow, packet, and byte.

In this paper we focus our attention on the packet level a packet-level approach is not taken into consideration because of its dependency from network conditions and end-to-end flow control and originally expressed in about applications characterized by bulk transfers. Characterizing traffic generated by sources and related to specific application-level protocols, also with the purpose to conduct realistic network traffic simulations.

The decomposition of aggregate traffic into conversations and on packet-level traffic characterization of traffic inside conversations, that is, models based mainly on two variables: packet size (PS) and inter-packet

time (IPT). There are important advantages in this approach are do not need to make any assumptions regarding the kind of applications generating traffic, the same methodology is easily extensible to study other application-level protocols and mixes, packet-level models are simple and easily applicable to traffic simulations that can be used to study network-related issues (measuring delay, jitter, packet loss etc.)or to test network equipment.

2.5 Traffic Data Repository at the WIDE Project

Traffic traces are collected by tcpdump and, after removing privacy information; the traces are made open to the public. WIDE project to collect a set of free tools build a traffic data repository containing detailed information of our backbone traffic. It not only on freely-redistributable software but also on freely-redistributable traffic data sets. The goals of the traffic repository are to promote traffic analysis research as well as to promote development of tools.

The monitoring tools in the UNIX based have the advantage of that users can use other software tools available on UNIX for manipulating and analysing packet traces. tcpdump is probably the most popular packet capturing tool in the UNIX community. tcpdump is based on a powerful filtering mechanism, the BSD packet filter (BPF).The packet capturing and filtering facilities of tcpdump are implemented in a separate library, pcap.

The motivation of the traffic information tends to be confined to a small set of members, and it is difficult to share detailed information without a framework to support sharing. In achieve a good use of this type of traffic trace repository have to solve two problems. One is to create a safety measure for handling traces that include privacy information. The other is automation of the trace acquisition process.

This paper has used the several tools to automatically maintain the traffic repository there are tcpdump, tcpdpriv, tcpdstat. tcpdump, puts the network interface into promiscuous mode to capture every packet going across the wire. tcpdump in the user space can read multiple frames in a single read from the store buffer in the kernel in an efficient manner. tcpdpriv to remove user data and scramble addresses.

Tcpdpriv removes the payload of TCP and UDP, and the entire IP payload for other protocols. tcpdpriv implements several address scrambling methods; the sequential numbering method and its variants, and a hash method with preserving address prefix. tcpdstat reads a tcpdump file using the pcap library and prints the statistics of a trace. tcpdstat is intended to provide a rough idea of the trace content. The output can be easily converted to a HTTP format.

III.CONCLUSION

In this paper describe the literature survey for expressing the different methods have used for the monitoring the network traffic data. And this monitoring can be performed for the improvement of the network usage performance and different techniques used in the network based on their situations. This Literature survey express the techniques have been used for monitoring the high volume data and there is some monitoring tools can also be used for the monitoring of the data. In the existing system can be expressed the different techniques for monitoring traffic data. Due to the proposed system can be the development of the efficient monitoring techniques for the improvement of the sensitivity and scalability for the traffic data.

REFERENCES

[1]. S. Agrawal, C. N. Kanthi, K. V. M. Naidu, J. Ramamirtham, R. Rastogi, S. Satkin, and A. Srinivasan, "Monitoring infrastructure for converged networks and services," Bell Labs Technical J., vol. 12, no. 2, pp. 63–77, 2007.

[2]. Y. Liu, D. Towsley, T. Ye, and J. Bolot, "An information-theoretic approach to network monitoring and measurement," in 2005 IMC.

[3]. Dainotti, A. Pescap`e, and G. Ventre, "A packet-level characterization of network traffic," in Proc. 2006 CAMAD, pp. 38–45.

[4]. M. Munk, J. Kapusta, and P. Svec, "Data preprocessing evaluation for web log mining: reconstruction of activities of a web visitor," Procedia Computer Science, vol. 1, no. 1, pp. 2273–2280, 2010.

[5]. K. Cho, K. Mitsuya, and A. Kato, “Traffic data repository at the WIDE project,” in Proc. 2000 USENIX Annual.

[6]. M. Peuhkuri, “A method to compress and anonymize packet traces,” in Proc. 2001 ACM Internet Measurement Workshop, pp. 257–261.

[7]. B. Claise, G. Sadasivan, V. Valluri, and M. Djernaes, “Cisco systems netflow services export version 9,” RFC 3954, Oct. 2004, Tech. Rep.

[8]. Pescap´e, “Entropy-based reduction of traffic data,” IEEE Commun. Lett., vol. 11, no. 2, pp. 191–193, Feb. 2007.

[9]. B. Pfaff, “Performance analysis of BSTs in system software,” in ACM SIGMETRICS Performance Evaluation Review, vol. 32, no. 1, pp. 410–411, 2004.

[10]. M. AdelsonVelskii and E. Landis, “An Algorithm for the Organization of Information”. Defense Technical Information Center, 1963.

[11]. G. Aceto, A. Botta, A. Pescap´e, and C. Westphal, “An efficient storage technique for network monitoring data,” in 2011 IEEE International Workshop on Measurements & Networking.

