

Research on Large-scale Network Traffic Model

Zhongqi Xin

Xi'an Fanyi University

Keywords: network; ON/OFF model; TSG; self-similarity

Abstract: This paper analyzes the current network traffic model, and studies and discusses the related problems of network traffic simulation. It proposes a multi-level queue flow selection NSTS algorithm and out-of-order simulation PDRS algorithm based on network self-similarity. A generator TSG capable of generating a self-similar network stream in a single machine or distributed simulates the generation of network traffic with self-similarity by aggregating multiple ON/OFF sources that are subject to Pareto distribution. It can be seen from the actual simulation results that the generator can simulate the main network protocols, common network conditions and self-similarity of the network. The actual test shows that TSG meets the design requirements and meets the relevant research and testing work of UTM.

1. Introduction

With the rapid development of the Internet, the Internet has developed into a complex and massive nonlinear system, and network security issues have become more and more prominent. Unified Threat Management (UTM) is a new type of integrated network security management system first proposed by IDC in September 2004. IDC has integrated concepts such as anti-virus, firewall and intrusion detection into a new category called Unified Threat Management, which has attracted widespread attention in the industry and has become a trend in the future of security devices. UTM has become a hot Issues.

The next-generation UTM equipment adopts hardware acceleration technology, hardware acceleration card and software scheduling require a lot of debugging and testing work, and it faces great difficulties. In terms of testing: 1. The network flow generated by a single machine has far failed to meet the testing requirements; 2. The real backbone network environment is difficult to obtain (for reasons of security and management). In terms of equipment: millions or even millions of professional hardware test equipment far exceeds the affordability of the average user. Therefore, it is necessary to research and develop a high-performance, highly flexible and inexpensive network traffic generator to serve the debugging and testing of the next-generation UTM equipment. How to generate approximate network traffic offline in an experimental environment, and comprehensive testing of next-generation UTM devices has become an important direction of research. Also choose the appropriate simulation granularity to avoid excessively heavy analog calculations.

2. Research on network traffic model

Today, there are two general approaches, log-based traffic generators and model-based traffic generators. The former is to automatically extract and record feature parameters from real network traffic, and use these parameters to configure network traffic. Thereby achieving the purpose of generating network traffic, such as Harpoon. The latter is to establish a mathematical model for network traffic by studying the characteristics of the network. The generator sends data packets according to the mathematical model to generate network traffic conforming to network characteristics.

2.1 Network self-similarity

The traditional telecommunication network widely uses the Poisson flow model to describe the traffic characteristics of the computer network. Based on the measured analysis of network traffic,

Vern Paxson and Sally Floyd pointed out that the Poisson flow model is not suitable for describing modern network traffic characteristics. Will E. Leland et al. deeply studied the distribution characteristics of data streams in the network and found that network traffic has self-similarity. Subsequent series of studies have confirmed that VBR video, backbone network, wireless network and so on have statistical self-similarity. Since then, network researchers have analyzed and compared the mechanism of self-similar traffic, and also developed several self-similar flow generation methods, such as RMD algorithm, FFT transform, etc., but they are all Lack of clear physical meaning can not explain the cause of self-similarity.

Self-similarity means that a stochastic process has the same statistical characteristics at all-time scales, which is the whole and part of a complex system, and the similarity between the part and the fine structure or nature of that part, or from the whole The local (local) extracted in the middle can reflect the basic characteristics of the whole. It is closely related to fractals and chaos and exists in many phenomena in nature. For network traffic, it appears to be bursty on various time scales or at least over a large range.

Assume $X_t = Y(t) - Y(t-1)$

$X_n^{(m)}$ M-order aggregation process for discrete stochastic processes, If :

$$X_s^m = \frac{1}{m} \sum_{k=s-m+1}^{s+m} X_k, i=1,2,3,4,\dots,$$

$$\mu = E[X_s], \sigma^2 = E[(X_t - \mu)^2],$$

$$r(k) = E[(X_t - \mu)(X_{t+k} - \mu)],$$

$$r^{(m)}(k) = E[(X_s^{(m)} - \mu)(X_{s+k}^{(m)} - \mu)],$$

If, $X(m)$'s Autocorrelation function $r(m), r^{(m)}(k) = r(k)$, For all $m=1,2,3,\dots, k=1,2,3,\dots$, The random process is said to $X(m)$ is self-similar..

2.2 Heavy-tailed distribution

Boston University's Mark E. Crovella, Azer Bestavros conducted a large number of experiments on network traffic, speculating that the heavy-tailed distribution process is likely to cause network self-similarity.

The so-called heavy-tailed distribution means that the majority of the distinct elements in the set appear very small, but a small number of elements appear frequently (frequency) so much that they are in the majority of the set. The heavy-tailed distribution is defined as: Let X be a random variable with a cumulative distribution function $F(x) = P(X \leq x)$, and a residual distribution function $\bar{F}(x) = 1 - F(x) = P(X > x) \sim cx^{-a}, 0 < a < 2, c$ is constant.

Then the function is said to obey the heavy-tailed distribution. From the distribution function, the heavy-tailed distribution is characterized by the fact that $0 < a < 2$, the heavy-tailed distribution has an infinite variance at that time; at that time $a \leq 1$, the heavy-tailed distribution also has an infinite mean, so when a decreasing, most of the probability mass is concentrated at the tail of the distribution curve. When the variance is infinite, the random variable has high variability, which is the Noah effect.

The tail reduction of the heavy-tailed distribution model is not reduced in a linear manner, but is reduced in a manner similar to the hyperbolic function curve. Therefore, the asymptote of the general heavy-tailed distribution function curve is a one-sided hyperbolic function curve. There are many expression models for heavy-tailed distribution. The commonly used distribution models include Pareto model, Lognormal model and Weibull model.

The heavy-tailed distribution can well explain the formation mechanism of computer network traffic self-similarity. The measurement of file size in I/O mode on the Web also proves that they obey the heavy-tailed distribution. The CPU time consumed by Unix processes is also subject to heavy-tailed distribution. If the web browser is described as an ON/OFF source, the data it generates is exactly in accordance with the Pareto distribution. Therefore, it is inferred that the

phenomenon of heavy tail distribution may be the main cause of network self-similarity.

The Pareto distribution is a typical heavy-tailed distribution, the distribution function is:

$$F(x) = 1 - (k/x)^a, 0 < a. x > k.$$

k as the cutoff parameter, it determines the minimum value that the random variable can take. a is the shape parameter, it determines the mean and variance of the random variables.

For the time period obeying the Pareto distribution, the inverse function is used to:

$$(k/x)^a = 1 - F(x),$$

$$x = k / (1 - F(x))^{1/a},$$

$$x = k / U^{1/a}, \text{ the random variable } U \text{ is uniformly distributed within the range of } (0, 1].$$

3. Conclusion

Network traffic simulation is an important means to study the network. It has the significance of enhancing information security and meeting the testing requirements of new equipment. This paper starts from the self-similar mathematical means of network traffic, and self-similarity and heavy-tail distribution of network flows. The mathematical characteristics have been studied in depth. The correlation between self-similarity and heavy-tailed distribution is analyzed.

References

- [1] Joel Sommers, Paul Barford. SelfConfiguring Network Traffic Generation. Proceedings of the 4th ACM SIGCOMM conference on Internet measurement. 2004.
- [2] Kashi Venkatesh Vishwanath and Amin Vahdat. Realistic and Responsive Network Traffic Generation. Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications 2006.
- [3] Abdelnaser Adas. Traffic Models in Broadband Networks, IEEE Communications Magazine. Jul. 1997.
- [4] Victor S. Frost and Benjamin Melamed. Traffic Modeling for Telecommunications Networks, IEEE Communications. Mar. 1994.
- [5] Vern Paxson, Sally Floyd. wide area traffic:the failure of Poisson modeling. IEEE/ACM Transactions on Networking,June. 1995.
- [6] Will E. Leland, Murad S. Taqqu, Walter Willinger, Daniel V. Wilson. On the Self -Similar Nature of Ethernet Traffic (Extended Version). IEEE/ACM Transactions on Networking, 1994.
- [7] Mark W.Garrett, Walter Willinger Analysis,modeling and generation of self-similar VBR video traffic. ACM SIGCOMM Communication Review October 1994. 24(4): 269.
- [8] Yao L, Agapie M, Ganbar J, Doroslovacki M Long range dependence in Internet backbone traffic. IEEE International Conference on Communications. 2003.1611-1615.
- [9] Oliveira C, Jaime Bae Kim, Suda T Long-range dependence in IEEE 802.11b wireless LAN traffic: an empirical study. Proceedings of the 18 th Annual IEEE Computer Communications Workshop. 2003.
- [10] Abry Patrice, Baraniuk Richard G. Multiscale Nature of Network Traffic. IEEE Signal Processing Magazine. 2002, 19 (3) : 28-46
- [11] Horn G, Kvalbein A. An empirical comparison of generators for self similar simulated traffic. 2004,64(2):162-190.
- [12] Vem Paxson. Fast. approximate synthesis of fractional Gaussian noise for generating self-similar network traffic. ACM SIGCOMM Computer Communication Review. 1997.27(5):5-18.
- [13] Mark E. Crovella, Azer Bestavros. Self-similarity in World Wide Web traffic: evidence and possible causes. ACM SIGMETRICS Performance Evaluation Review. 1996.