International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 4, No.2 (Apr-2015)
http://dx.doi.org/10.12785/ijcds/040201

Examining Self-Similarity Network Traffic Intervals

Hengky Susanto and ByungGuk Kim

[1] *Department of the Computer Science , University of Massachusetts at Lowell, USA*

Received 21 Nov. 2014, Revised 19 Jan. 2015, Accepted 7 Mar. 2015, Published 1 Apr. 2015

Abstract: Understanding network traffic pattern and its impact on the Internet provides valuable insights in designing new network protocols, particularly in designing one for applications with a tendency to generate bursty traffic of data, such as Voice over IP (VoIP). To capture the behavior, network traffic can be illustrated on many scales using the notation of *self-similarity* because network traffic is statistically self-similar. In this paper, we propose a study on analyzing the length of a traffic interval by self-similarity based on the difference between arrival times of packets. We examine the dependency between fast and slow interval as well as a study on the data transition between both intervals.

Keywords: network traffic, self-similarity, Probability

1. Introduction

Computer networks can provide better quality of service (QoS) when upcoming network traffic patterns are known. Many of today's real time applications, such as teleconferencing, Video on Demand (VOD), Voice over IP (VOIP), and others similar rely heavily on the quality of the network connection. Such real-time applications will certainly benefit from knowing the traffic condition ahead of time. For example, a system can be better prepared to anticipate upcoming traffic by adjusting the playout mechanism in VOIP or by seeking a new alternative path to support the minimum required bandwidth.

Many studies have been done to measure and predict traffic patterns on the Internet that show the presence of fractal or self-similar properties. [1,2,3,4] Network traffic can be illustrated on many scales using the notation of self-similarity. Self-similarity means that the statistical patterns may appear similar at different time scales, which can vary by many orders of magnitude. In the other words, self-similarity is a fractal property of traffic patterns in which appearances are unchanged regardless of the scale at which they are viewed; this ranges from milliseconds to minutes or even hours. There are a number of models that are used to describe bursty data stream in the Internet such as the Pareto distribution model or Poisson distribution related-models (for example Poisson-batch, Markov-modulated Poisson, packet train models, Markovian Input model, or a fluid flow model).

Based on a statistical description of traffic illustrated by Pareto and Poisson (Exponential) distribution models, we can compute the probability of bursty data stream occurring at the next interval of time T or when a particular packet burst will end. The idea is to evaluate whether the next arriving packet comes in a burst by analyzing the probability of the transition from burst to non-burst and non-burst to burst. However, predicting the arrival of future packets does not tell us whether packets come in large or small bursts.

There are other researchers who take advantage of self-similar traffic patterns. Self-similar traffic is also used to predict future events on the Internet and to improve network performance. In [6] the author proposed a new algorithm for predicting audio packet playout delay for VOIP conferencing applications. And the proposed algorithm uses hidden Markov model to predict the playout delay. Similarly, in [5], a study was done using a Markovian distribution model to predict queuing behavior with self-similar input.

As stated previously, the objective of this study is to present a statistical method to predict when the next burst in network traffic occurs and when it ends based on the transition from burst to long silence, and vice versa. We also compare the outcome of self-similar input between Pareto and Poisson distributions. However, Poisson distribution may not represent the real Internet traffic

E-mail: hsusanto@cs.uml.edu, kim@cs.uml.edu

because the packet arrivals time are not exponentially distributed [7]. The models for network traffic essentially become uniform. Hence most experiments in this paper use Pareto distribution, and Poisson distribution is only used for as a comparison purpose. Through our simulation on Internet traffic pattern by using traffic models with a self-similar characteristic, we demonstrate that the probability of transition from between slow and fast intervals packet arrival is almost constant.

2. SELF-SIMILARITY AND HEAVY TAIL

In a self-similar time property [1,8], given a stationary time series X, we have

$$X(t) = (X_t : t = 1,2,3,\dots).$$

with parameter H (0.5). Define m aggregated series

$$X^m = (X_k^m : k = 1,2,3,\dots)$$

by adding all the series of X over non overlapping block of size $m \geq 0$, such that

$$X_k^m = \frac{1}{m[X_{kt-m+1} + X_{kt-m+2} + \cdots + X_{kt}]}.$$

Then X is self-similar when X has the same autocorrelation function $r(k)$, which is defined as follows.

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

as the series $X^{(m)}$ for all m. [7] In the Poisson distribution series models, the network traffic become uniform when aggregated by a factor of 1000.

The distribution that is used in this paper has the property of being *heavy-tailed*. A distribution is a heavy tailed if

$$P[X \geq x] - x^{\alpha} \text{ as } x \to \infty, 0 < \alpha < 2.$$

Where X is a random variable and α is a shape parameter. The distribution has infinite variance when α is less than 2. The simplest heavy-tailed distribution is Pareto distribution. The distribution is hyperbolic over its entire range. We write the density function as

$$p(x) = \alpha k^{\alpha}\, x^{-\alpha-1}, \alpha, k > 0, \;\; x \geq k$$

and its cumulative distribution function is

$$F(x) = P[X \leq x] = 1 - \left(\frac{k}{x}\right)^{\alpha}.$$

The parameter k represents the smallest value of the random variable. When α ≤ 2, then the distribution has infinite variance and if α ≤ 1 then distribution has infinite mean. Thus, as α value decreases, the probability density is present in the tail of the distribution. The closer α parameter is to 1, the shorter the generated burst; closer α parameter to 2, then the larger the generated burst in the simulation.

3. TRANSITION BETWEEN INTERVALS

Initially packets arrive at different time. They may arrive in a burst or be delivered after a long silence between bursts. For example, n number of packets may arrive in x milliseconds or there is x milliseconds of silence between the arrival of two packets. This study shows that when packets come in bursts, it is very likely that the next packet to arrive will in also be as part of the burst. Also, the next packet that arrives after some number of packets that arrive with long silence in between is more likely come in a burst.

Packets are clustered according to the time difference of their arrival time and a constant Γ (**Tau**) where Γ is range from 1 to 100 milliseconds. If the time difference is less than or equal to Γ then the cluster is called *fast interval* and when the time difference is greater than Γ then the cluster is revered as *slow interval* as it is shown in figure 1. For instance, packet one arrives at time 0, packet two arrives at time 1, packet three arrives at time 5, and Γ is 2. If the arrival time difference between packet one and packet two is exactly 1 and it is less than the Γ value, than packet one and two are clustered in the fast intervals category. On the other hand, if the arrival time difference between packet two and three is larger then Γ, hence packet three is considered to be a slow interval. Moreover, think of Γ as a constant timeout value or a restriction that packet has to arrive at a certain time after the last arriving packet. For example, Γ = 4 and the last packet arrives at time 3, then the next packet hast to arrive before or at time 7.

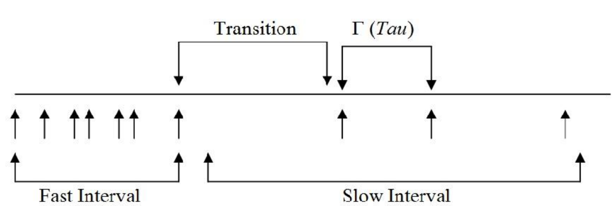


Figure. 1. *A Graph of arrival packets at time t_n, n = {1, 2, 3,.....n}.*

The starting point of computing whether the next arrival packet is bursty or non-bursty is to gather data about the packets' arrival time. Network traffic input is obtained through Pareto Distribution with the restriction 1 < α < 2 in order to obtain the long-range dependence. Γ is used to control the size of fast interval and slow interval. To compute the total number of packets in the interval (whether fast or slow), calculate the probability of the possible type of the next incoming packet whether it belongs to fast of slow interval. Given that $t_{arrival} > \Gamma$, the probability of $P[X]_{slow}$ and $P[X]_{fast}$ are defined as followed.

$$P[X]_{slow} = \frac{1}{1 + \sum_t TP_{fast\ interval}}$$

and

$$P[X]_{slow-packet} = \frac{1}{1 + \sum_t TP_{slow\ interval}}.$$

Then, the probability of average packet with slow and fast arrival is defined as follows.

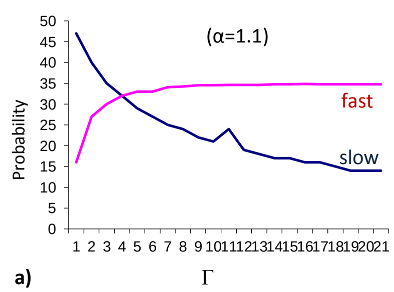
$$P_{\text{avrg slow packet}} = \sum_i P[X]_{i,slow-packet}\,,$$

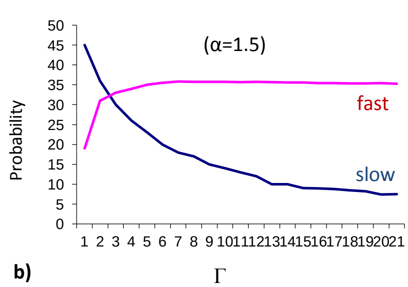
$$P_{\text{avrg fast packet}} = \sum_i P[X]_{i,fastpacket}\,,$$

where $i = \{1,2,3, \ldots, n\}$ and $P[X]$ is the probability that the next packet arrival time is greater or less than Γ and TP is the total number of packets that are in the interval. Next, we summarize the probability of the transition of the entire transaction up to time t by computing the average of the probability.

4. ANALYST AND SIMULATION

Our study shows that the average probabilities of the transition from slow interval to fast intervals form a hyperbolic curve. Similarly, the probability of the transition from fast intervals to slow intervals form a parabolic curve regardless the value of α parameter. As they are shown in the figure 2 (a, b, and c), where Γ = {1, 6, 11, 16, 21, 26, 31, 36, 41, 46, 51, 56, 61, 66, 71, 76, 81, 86, 91, 96, 101}.





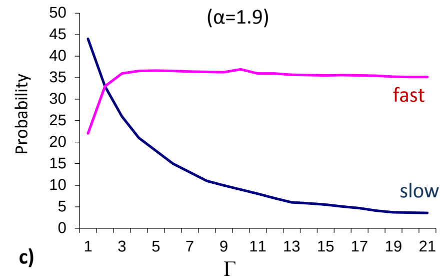


Figure. 2. Probabilities of fast and slow interval.

According to the graphs above, there is approximately a 35 percent chance that the next arrival packet will transition from slow to fast interval (packet arrives after Γ is expired or the transition from fast to slow interval) for most Γ values. On the other hand, the probability of a packet arrives before Γ (the transition from slow to fast interval) strictly depends on the value of Γ.

With smaller value of Γ, it is more likely that the probability of the transition from fast interval to slow interval will fall between 30 to 40 percent because of the small number of packet in fast interval. However, as Γ value grows, the probability becomes smaller regardless the value of α. On the other hand, the probability of the transition from slow interval to fast interval will be closer to 40 percent when the value of Γ is higher.

In a number of the experimental scenarios regardless of the value of $\Gamma > 5$ the number of packets in fast intervals is always larger than the number of packets in slow intervals. It shows that the Pareto distribution generates more packets with small intervals between packets' arrivals time. For example, two experiment scenarios with two different Γ values, two different distinct average of interval between packets' arrival time, and different α values, but the probability of the transition from slow to fast interval is almost constant, about 30 to 35 percent. In contrast, the probability of the transition from fast to slow interval gradually drops to less than 10 percent because the higher value of Γ. That means every increase of Γ value that there will be more packets in fast intervals and less number of slow intervals. The reason of the stable value of the probability for the transition from slow to fast intervals is that Pareto generates a cluster of two or three packets with long interval in between, and the clusters are well distributed in the distribution. In addition to that, the silence can be very lengthy hence there is always slow interval in the distribution even though the Γ value is significantly greater than the average interval time between arrival packets.

In addition to Pareto experiment, we also experimented with Poisson distributions. The outcome was that the probability of transition from fast to slow interval and from slow to fast intervals are almost the same, as long as

$$\frac{TD}{2} < \Gamma < TD,$$

where TD is the average difference of the packet arrival time.

5. CONCLUSION

In this study, we have demonstrated that Pareto model has the tendency to generate more packets with a small interval of packet arrival time. Additionally, we also show that higher Γ value leads to a higher number of packets in fast interval than slow interval. This study provides us with a better understanding on what should be the acceptable packet rate in order to provide a better prediction or a define more precise Γ. Currently, the Γ value, which is used for experiment, is variable and is likely to exceed the actual Γ for some value of Γ. The experiment shows that the probability of transition from slow to fast interval is almost constant because Pareto distribution often generates a cluster of two or three packet with long silence in between. The question is whether this circumstance also occurs at the actual the Internet traffic, which would be featured in our future work.

Furthermore, in our future study, we would also consider the case where there are more than one Γ value to separate packets that arrive on time, in burst, or late (due to the network traffic), and analyze the probability of the next packet arrival in any of those categories. Furthermore, we would like to include self-similarity study with other "real world" problem such as peer to peer, online-gaming problem, multicasting, multimedia network, etc. Also, we would like to analyze that our studies is applicable to the "real world" networking problem.

Understanding traffic patterns and being able to predict the future of the traffic will patterns help systems provide better quality of service or make better decision selecting connection path.

REFERENCES

[1] Mark E. Crovella and Azer Bestavros., "*Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes"*. IEEE/ACM Transactions on Networking, Vol. 5, No 6, December 1997.

[2] Sahinoglu Z. and Tekinay S., "*On Multimedia Networks: Self-Similar Traffic and Network Performance"*. IEEE Communication Magazine, January 1999.

[3] Beran J., Sherman R., and Neame T., "*Fractal Traffic: Measurements, Modeling and Performance Evaluation"*, Proceeding of IEEE INFOCOM 1995.

[4] Leland W., Taqqu M., Willinger W., and Wilson D., "*On the self –Similar Nature of Ethernet Traffic"*, IEEE/ACM transactions on Networking , Vol. 2, No1, February 1994.

[5] Kasahara S., "*Internet Traffic Modeling: Markovian Approach to Self-Similar Traffic and Predict of Loss Probability for finite queues"*, IEICE Trans. Communication. Vol. E84-B. August 2001.

[6] Yensen T., Lariviere J., Lambadaris I., and Gaubran A. R., HMM "*Delay Prediction technique for VOIP"*. IEEE Transaction on Multimedia. Vol. 5 no.3, September 2003.

[7] Paxson V. and Floyd S., "*Wide-area Traffic: The failure of Poisson Modeing*", SIGCOMM, 1994.

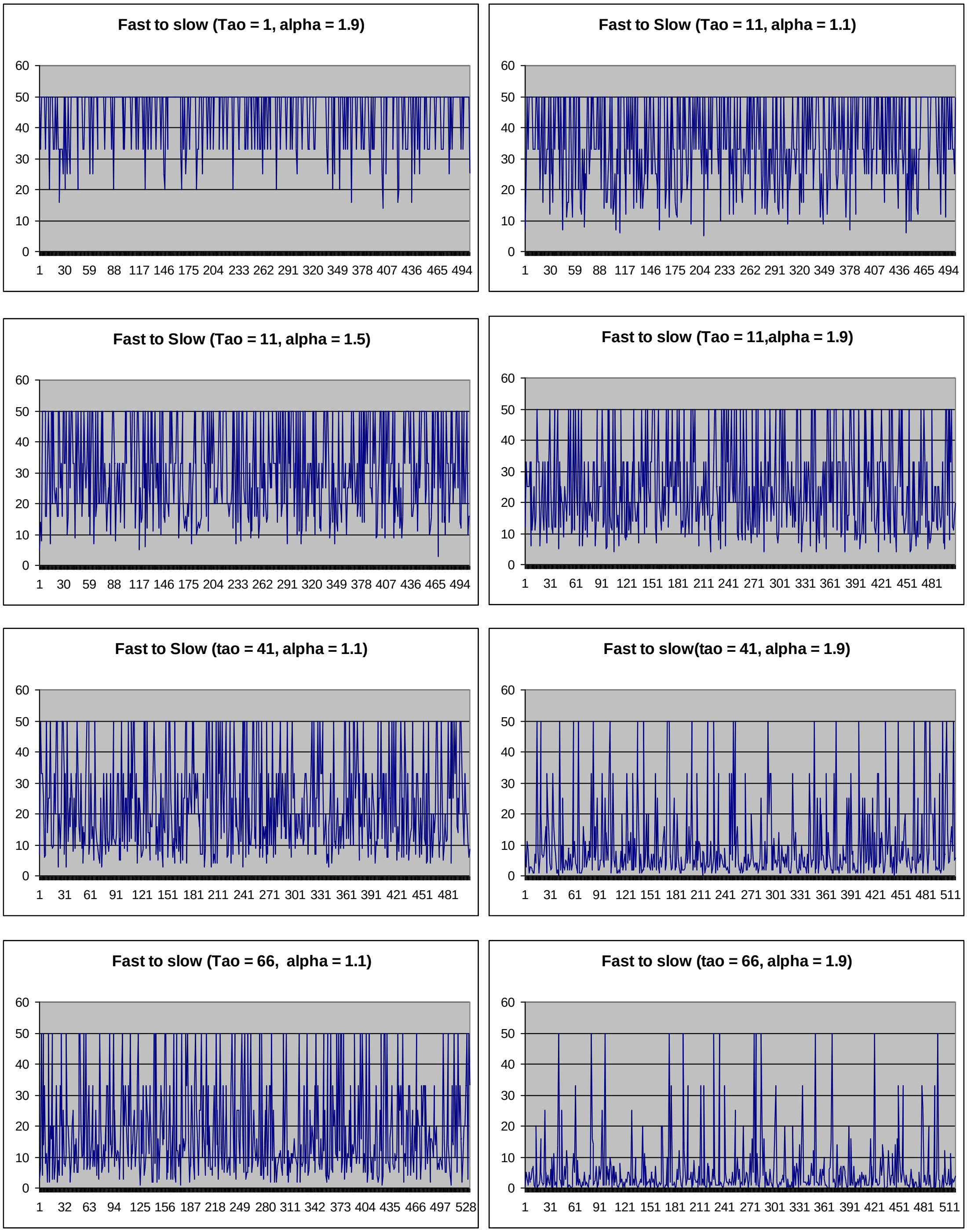
[8] Ulanovs P. and Petersons E., "*Modeling methods of self-similar traffic for network performance evaluation"*, Scientific Proceedings of RTO, Series 7 , Telecommunication and Electronics 2002.

[9] Cidon I., Khamisy A., and Sidi M., "*Analysis of packet Loss Processes in High-Speed Networks"*. IEEE transaction on information theory, VOL. 39, NO. 1, January 1993.

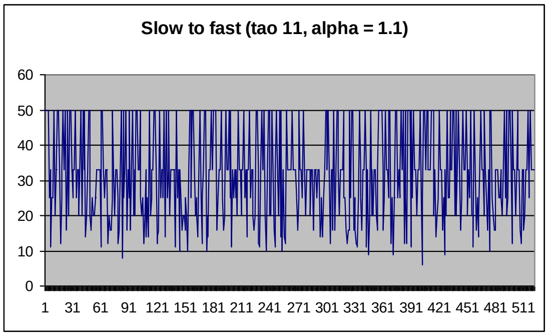
Hengky Susanto received his BS degree in computer science from University of Massachusetts Amherst in 1999, MS degree in computer science from University of Massachusetts Lowell in 2004. He did his post MS in network system administration at Tufts University. He received his PhD degree in computer science from University of Massachusetts Lowell. His main interests include wired and wireless network, network multimedia, protocol design, and network economic. He has also received several awards including the best USENIX LISA paper award 2005, EMC Achievement award 2000, and StorageNetwork Shine Award 2002.

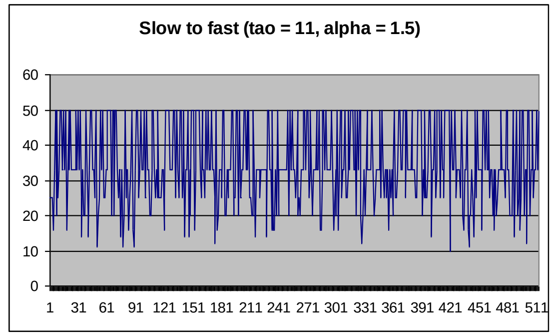
ByungGuk kim received his BS degree in electrical engineering from Seoul National University, South Korea, in 1975, MS degree in computer science from University of Massachusetts Amherst in 1978, PhD degree in computer science from University of Massachusetts Amherst in 1980. He is an Associate professor and undergraduate coordinator in the Department of Computer Science, University of Massachusetts Lowell. His research interests include network performance, wireless network, network multimedia, network traffic congestion and management, and Bioinformatics.

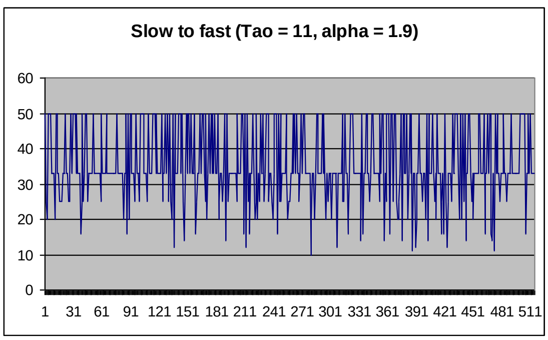
Graphic Transition Chart from Fast Interval to Slow Interval.

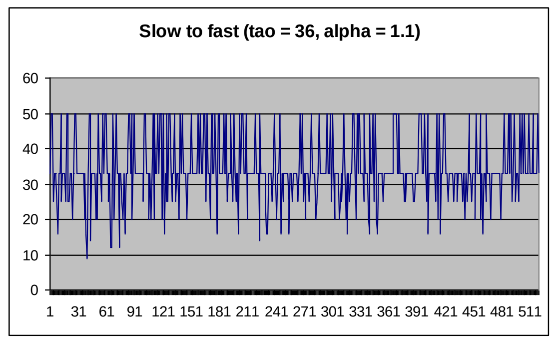


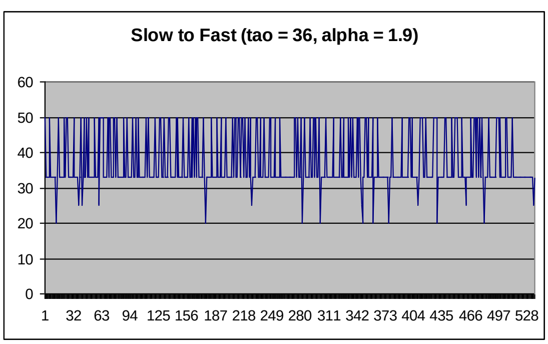
Graphic transition chart from slow interval to fast interval.

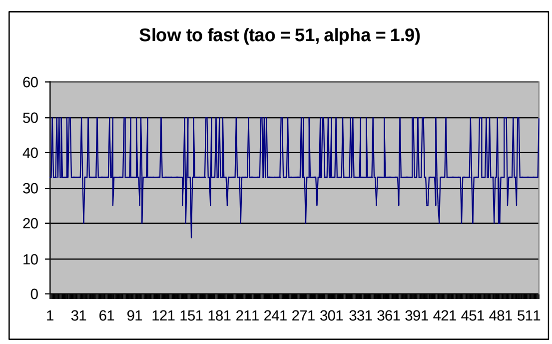


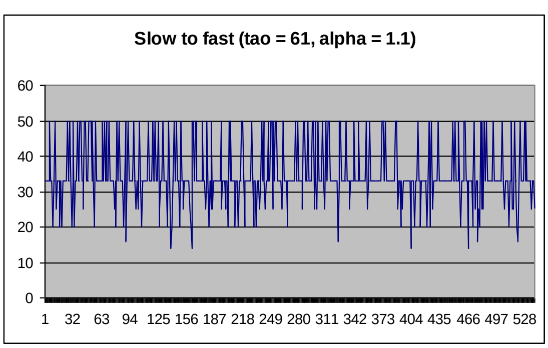


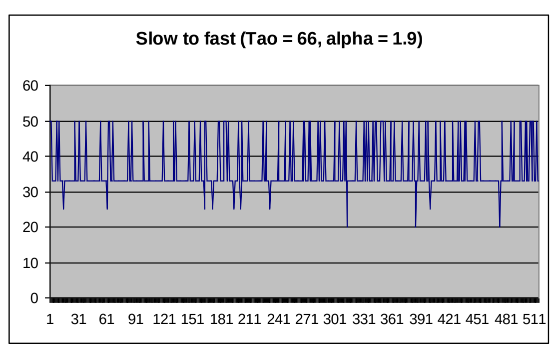




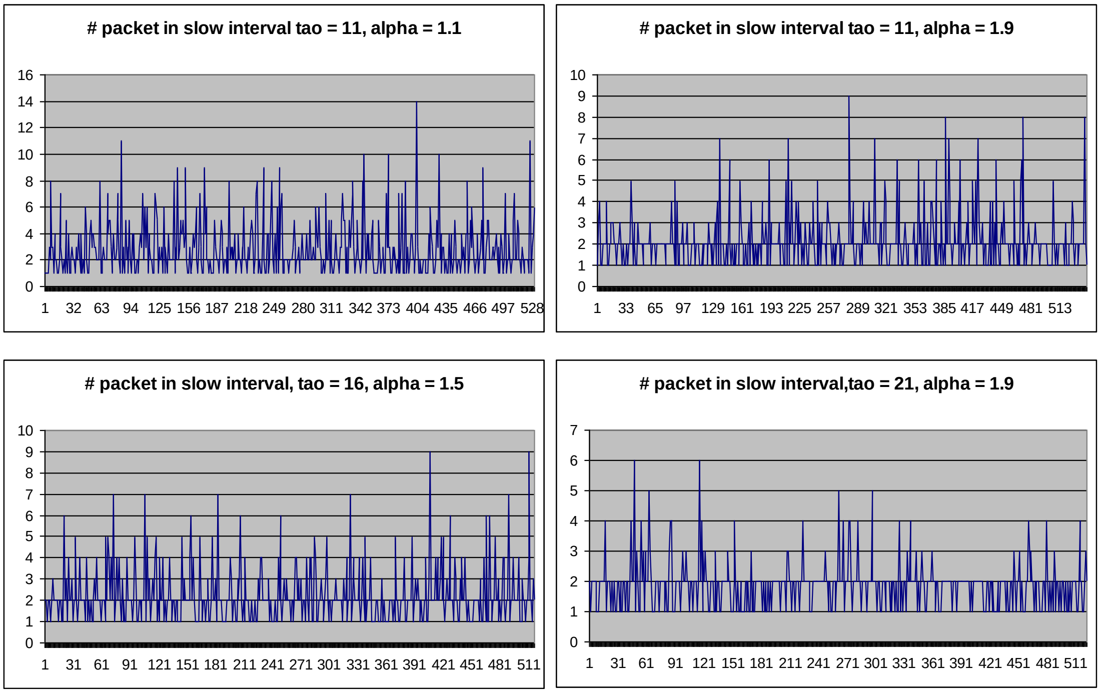








Graphic Number of Packet in slow interval.



Graphic Number of Packet in fast interval.