Paper Review
"An Algebraic Representation of Wireles Workload"

Stefan Karpinski, Elizabeth M. Belding, and Devin C. Almeroth

March 05, 2008

1 Summary

1.1 Topic and Solution

As the authors of the paper well pointed out, the realistic traffic generation plays crucial role in the credibility of conclusions drawn from simulation studies or experiments. However, the networking community has not yet been equipped with handy gadgets for generating realistic workload traffics. This paper proposed the General Matrix Model(GMM) in the hope that it will improve methodology for expressing traffic patterns and also for generating realistic experimental workload based on that model. A flow of traffic is generally regarded as a sequence of IP packets between hosts which shares the next level protocol of IP header, source and destination nodes, and source and destination port numbers. Naturally, the behavior of each flow is characterized by the following seven properties in this paper, each of which is represented as a vector in GMM: flow type(i.e., the next level protocol of IP packet), source node, destination node, start time, flow size, packet size and inter-packet interval. Then, the traffic pattern of the entire network is the aggregate behavior of all flows and is represented as a matrix, the rows of which are the behavior vectors of the flows. The authors mentioned that the biggest challenge in mapping flow behaviors into a vector space is to map them in a way that the summation of vectors correspond naturally and meaningfully to the aggregate behavior of flows. And it seems to me that this is the reason why the matrix becomes large and sparse. A way to produce an actual workload from a given instance of GMM is presented in the section III-F.

1.2 Strength and Novelty

They used the method called *paired differential simulation* to evaluate how realistic GMM is. Basically, it tests to see if a set of simlations performed using given synthetic models, called a test group, accurately produces the performance metrics exhibited by a set of simulations using real network traffic, called a control group. Note that the comparison is between the performance metric of two groups not between any of the aforementioned seven properies of flow behaviors of the two. The performance metrics used are average end-to-end delay, packet delivery ratio, received throughput, AODV control overhead, packet drops, and 802.11 control overhead.

In order to assess how realistic the traffic models are, they used the measures of error. In particular, they used the log-ratio to measure the differences between characteristics of the performance metrics measured from actual traced traffic and synthetic traffic based on models. The results are summarized in figure 3. GMM performed better than other traffic models in comparison and reflected closely the performance of the traced data it is based on.

The paper provided pursuasively why algebraic representation is preferred. First, the transition between the flow behavior to the aggregate behavior is as easy as the summation of vectors. Secondly, many common concepts and assumptions–e.g.,uniformity, marginality, and regularity–from traffic modeling can be succintly expressed using a few linear transformations. Lastly, Common properties of the network including the number of flows, the expected average packet size, or the interpacket interval, are shown to be easily computed using GMM. This paper also successfully showed

that the current standard practice in traffic modeling is overly simplistic and hence the conclusion drawn from simulations or experiments using those models fail to accurately reflect the important performance metrics of real networks. As an essential initial step to provide network researchers with a powerful tool to express and generate the realistic workload in a flexible way, GMM serves the purpose well by providing the first-order approximation of network behavior. The result of the mixed regularized model is enlightening so that the potential pitfall of mixing up models arbitrarily thinking that it will make model more realistic can be avoided.

2 Comment

The paper could be improved by addressing following concerns:

- (Page 6) It is clearly shown in section V that how common modeling concepts can be expressed using matrix transformation. However, it is not very clear how this transformation allow us to evaluate the effect of each model's simplistics on the realism of generated workload as the authors claimed.

- (Page 8)Due to the vast size of the matrix and its sparseness, the requirement of processing and computational power might be quite demanding in order to smoothly express and generate traffic based on the GMM. There is no detail about how the matrix representation from the trace data is derived. Therefore, it is not clear how GMM can efficiently express the characteristics of total 2.1 million flows traced for 24 hours.

- (Page 9) The authors used the six performance metrics to prove how accurately GMM capture the characteristics of the real network. However, what if some other performance metricse.g., delay jitter, MAC-layer bandwidth share, etc.are concerned? Can we say these metrics also be accurately refelected with high confidence?

3 Questions

The following questions occurred to me as I reviewed the paper:

1. What is a flow size exactly? Based on my understanding, it is the sum of all the bytes flown between a pair of host. If so, flow size, packet size, and inter-packet interval are closely related to each other. Is there any inconsistency problem when considering them separately?

2. Does 24-hour trace data used in the paper really good enough?

3. What theoretical and computational tools of modern linear algebra can be adopted and used for traffic generation specifically? One thing they mentioned as their current work is using matrix factorization and clustering technique to extract hidden structure from a large and diverse body of real traffic. But this haven't verified yet and isn't it too soon to claim this?

4. Wouldn't it be better if the authors come up with the criteria to determine whether the seven properties characterize the traced traffic adequately instead of proving the adequacy of GMM model by showing the six performance metrics closely resembled those of traced traffic? For example, by comparing distribution of packet size of GMM model and that of traced data.

5. Paxson and Floyd observed TCP traffic which protocol itself has closed-loop behavior. However, can we generalized this to the entire network as the authors mentioned?