FACTOR ANALYSIS OF NETWORK FLOW THROUGHPUT MEASUREMENTS FOR INFERRING CONGESTION SHARING

Doğu Arifler[†] *and Brian L. Evans*[‡]

[†] Eastern Mediterranean University
Department of Computer Engineering
Gazimağusa, TRNC
email: dogu.arifler@emu.edu.tr

[‡] The University of Texas at Austin
Department of Electrical and Computer Engineering
Austin, TX 78712 USA
email: bevans@ece.utexas.edu

ABSTRACT

Internet traffic primarily consists of packets from Transmission Control Protocol (TCP) flows. Based on passive, flow level TCP network measurements, our previous work has focused on using the principal component method to perform factor analysis on flow class throughput correlation matrices in order to infer which classes of TCP flows are sharing bottlenecks in the network. In this paper, we present a first-order autoregressive model for congestion at a bottleneck to analyze the need for filtering out a subset of the collected flow measurements before analysis. We demonstrate the successful application of our statistical methods in inferring congestion sharing after filtering out small- and large-sized flow samples.

1. INTRODUCTION

Determining which network flows are sharing congested resources might be the first step in analyzing and eliminating the causes of poor network performance. For instance, information on congestion sharing might be used by content providers to replicate content at other locations to reduce the load on the congested portions of the network, and by service providers to diagnose problems and direct traffic sharing a bottleneck onto disjoint routes. However, deciding which network flows are sharing congested resources in the Internet is usually difficult without access to the complete routing information for the network. In general, network managers have information only about their network domain, and have little or no information about the properties of the other domains.

Previous research has focused on using packet level statistics such as packet loss [1, 2], packet delay [2], packet order [3], or entropy of inter-packet spacing [4] to *infer* congestion sharing. Our previous work on inferring congestion sharing has differed significantly from the previous approaches in that we have considered *flow level* (higher level traffic object) instead of *packet level* statistics. Our methods have relied on *passive*, flow level Transmission Control Protocol (TCP) measurements (flow records) collected at a network node (e.g. router, gateway, server). In particular, in [5], we have shown that the correlation structure of flow class throughputs obtained by flow level measurements for a number of TCP flow classes can often be captured by a fewer number of *latent factors*. The latent factors represent *congested resources* and can be used to infer which flow classes are sharing resources in the network. In [6], we have applied *factor analysis* to analyze voluminous amounts of flow level TCP network traffic measurements collected at a single measurement site to infer which classes of TCP flows are sharing congested resources.

In this paper, using a first-order autoregressive (AR(1)) model [7], we analyze the need for filtering out a subset of the collected TCP flow measurements before factor analysis. We demonstrate that filtering out small- and large-sized flows enables us to infer congestion sharing flow classes in the Internet. We present the results of congestion sharing inferences using two sets of real data.

2. MODEL

A commonly accepted definition of an Internet Protocol (IP) *flow* is a unidirectional sequence of packets, which are close to each other in time and share a common identifier such as a common source and destination IP address. State-of-the-art networking equipment that runs traffic monitoring tools (such as Cisco's NetFlow) is capable of generating *flow records*. A flow record is a measurement object that contains the source and destination IP addresses, TCP or User Datagram Protocol (UDP) port numbers, IP protocol type, type of service fields in IP headers, start (s_f) and end (e_f) times, and the number of packets and bytes in a flow f. We define an IP *flow class* as a collection, or aggregation, of flows that are emitted successively and in parallel, and that have a common attribute (e.g. all flows sharing common source and destination IP address prefixes). We define the perceived *throughput* of a flow as the amount of data it carries (its size in bits) divided by the duration of the flow (in seconds). The *throughput of a flow class* at a given time is an average over the flows in that class (class average) that are active at that time.

It has been shown that TCP flows that are operating in additive-increase multiplicative-decrease mode of congestion control share the capacity of a bottleneck link roughly fairly when the flows have similar round-trip times and packet loss rates [8]. As an example, Fig. 1 illustrates how two temporally overlapping TCP flows share available bottleneck capacity in an idealized model. However, in general, TCP flows take some time to discover the congestion state, or the available capacity of the network, and especially, very small flows (due to TCP's Slow Start) may not have an opportunity to "learn" the capacity available to them during their sojourn. As a consequence, the throughput perceived by short flows will not reflect the congestion state of the network during their sojourn.

Based on the findings on capacity sharing by TCP flows at a bottleneck, we consider a simple model of congestion level seen by a flow[1]: $\{B(i)\}$ is an AR(1) process that represents the instantaneous bandwidth (or capacity) available to each flow at the bottleneck. $\{B(i)\}$, whose mean is denoted by μ_B, is then defined by

$$B(i) - \mu_B = \alpha\,(B(i-1) - \mu_B) + Z(i),$$

where $\{Z(i)\} \sim N(0, \sigma_Z^2)$, $|\alpha| < 1$, and $Z(i)$ is uncorrelated

[1] We assume that the congestion process is roughly independent of the flow; i.e., the flow makes only a small contribution to the overall congestion.

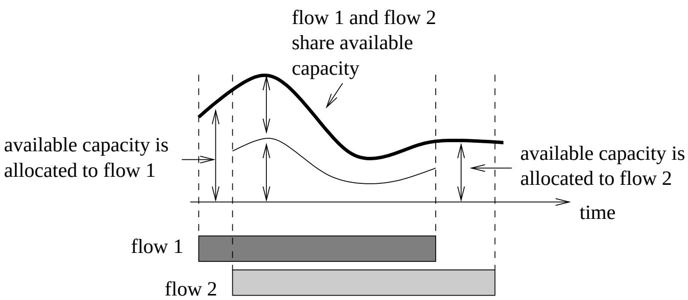

Figure 1: Sharing of instantaneous available resource capacity by two temporally overlapping flows. The length of the rectangle corresponds to the flow's duration. In this particular example, both flows have similar packet round trip times and packet loss rates, and the instantaneous bandwidth sharing is roughly fair during the time period over which they overlap.

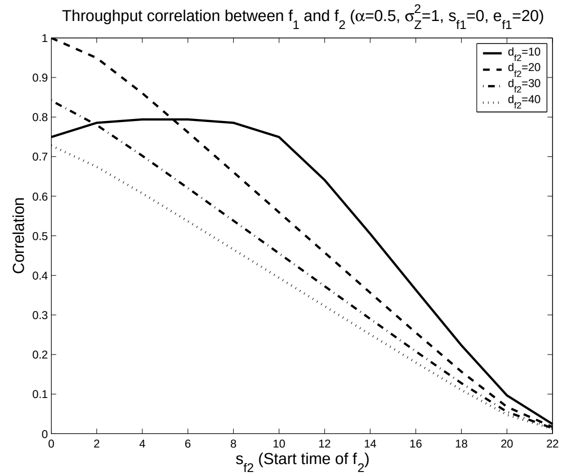

Figure 2: The effect of flow duration and temporal overlap on the correlation in (1) between throughputs of f_1 and f_2 that share a congested resource. The correlation values shown are for $\sigma_W^2 = 0$. Flow 1 starts at time 0 and ends at time 20.

with $B(j)$ for each $j < i$ [2]. For now, let us assume that start and end times (s_f and e_f) are discrete times. A given flow f carries an amount of data equal to

$$V_f = \sum_{i=s_f}^{e_f} B(i).$$

In the discrete-time AR(1) model, the duration of a flow f is given by $d_f = e_f - s_f + 1$. For simplicity, consider two flows f_1 and f_2 with given start and end times, and suppose that $s_{f_1} = 0$ and $s_{f_1} \leq s_{f_2}$ without loss of generality. The throughputs perceived by these flows are

$$Y_{f_1} = \frac{\sum_{i=0}^{e_{f_1}} B(i)}{d_{f_1}} + \frac{W_{f_1}}{d_{f_1}} \quad \text{and} \quad Y_{f_2} = \frac{\sum_{j=s_{f_2}}^{e_{f_2}} B(j)}{d_{f_2}} + \frac{W_{f_2}}{d_{f_2}},$$

where $W_{f_1}, W_{f_2} \sim N(0, \sigma_W^2)$ are included to model the "noisy" throughputs perceived by short TCP flows, and are independent of each other, $\{B(i)\}$, and $\{Z(i)\}$. In this context, a "noisy" throughput means that the throughput perceived by a short flow is not a typical one for the class to which the flow belongs due to its inability to discover the congestion state of the network. For flows with long sojourn times, the "noise" terms become negligible. The autocorrelation function of $\{B(i)\}$ is denoted by $\gamma(h)$, and is equal to $\sigma_Z^2 \alpha^h / (1 - \alpha^2)$ for $h \geq 0$. The correlation between Y_{f_1} and Y_{f_2} is

$$\mathrm{Corr}(Y_{f_1}, Y_{f_2}) = \frac{1}{d_{f_1} d_{f_2} \sigma_{Y_{f_1}} \sigma_{Y_{f_2}}} \sum_{i=0}^{e_{f_1}} \sum_{j=s_{f_2}}^{e_{f_2}} \gamma(|j-i|), \quad (1)$$

where $\sigma_{Y_{f_1}}$ and $\sigma_{Y_{f_2}}$ are the standard deviations of throughputs of f_1 and f_2, respectively. The standard deviation of the throughput of f with $s_f = 0$ is given by

$$\sigma_{Y_f} = \sqrt{\mathrm{Cov}(Y_f, Y_f)} = \sqrt{\frac{1}{d_f^2} \left(\sum_{i=0}^{e_f} \sum_{j=0}^{e_f} \gamma(|j-i|) + \sigma_W^2 \right)}. \quad (2)$$

[2]For similar approaches to modeling available bandwidth at a bottleneck, refer to [9]. Note that in our AR(1) model, available bandwidth can become negative. However, we introduce this simple model only to provide an insight into the nature of flow throughput correlations without attempting an exact traffic model in any way.

Note that perceived throughputs of long flows have smaller standard deviation than those of short flows since d_f^2 dominates in (2) [10]. These results agree with the observations reported for the perceived throughputs of small and large flows in [11].

In order to illustrate the behavior of (1) with different flow durations and different amounts of temporal overlap between the two flows, we set $\alpha = 0.5$, $\sigma_Z^2 = 1$, and $\sigma_W^2 = 0$ (no noise), and in Figs. 2 and 3, we exhibit the correlation as a function of s_{f_2} for different d_{f_2} values when $e_{f_1} = 20$ and $e_{f_1} = 30$. From the AR(1) model, we can draw the conclusion that the correlation between perceived throughputs of congestion sharing elastic flows is largely determined by the amount of temporal overlap between flows relative to the (product of) durations and standard deviations of flows. The throughput correlation is high when the two flows temporally overlap, and then decreases with increasing s_{f_2} (i.e. as the amount of overlap decreases). Furthermore, the correlation between overlapping flows decreases as the duration of the first flow is increased (see Fig. 3).

As a consequence, for a set of flow records, we expect throughput samples associated with long flows that have large amounts of temporal overlap to result in high throughput correlations. However, note that the occurrences of such samples are rare since the Internet is currently dominated by short flows [12, 13]. Furthermore, throughput samples associated with long flows overlapping with short flows will give a lower value for throughput correlation. On the other hand, throughput samples associated with short flows are noisy, and will not exhibit high throughput correlation. Therefore, leaving out long and very short flows may be desirable when estimating throughput correlations that are due to congestion sharing. Since flows with long durations will typically be large (in size)[3], we study the effect of different size thresholds to filter out large flows, and similarly, consider the impact of different size thresholds for omitting small flows. Unlike the duration of a flow, the size of a flow is invariant regardless of the capacity of network links. Hence, flow size is the proper flow attribute to consider for filtering out flows.

[3]Based on the processor sharing approximation of TCP bandwidth sharing at a bottleneck, we assume that the flows of different sizes experience the same slowdown.

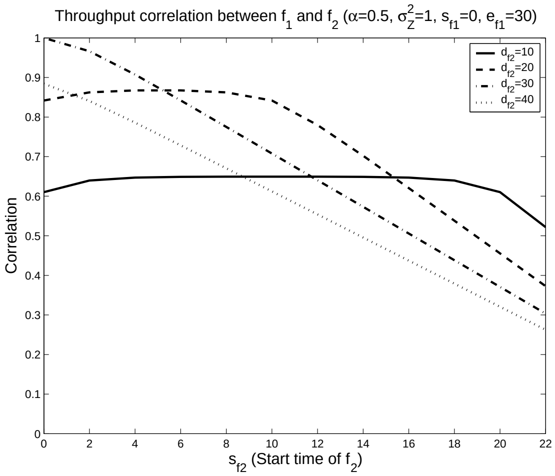

Figure 3: The effect of flow duration and temporal overlap on the correlation in (1) between throughputs of f_1 and f_2 that share a congested resource. The correlation values shown are for $\sigma_W^2 = 0$. Flow 1 starts at time 0 and ends at time 30.

3. ANALYSIS OF REAL TCP DATA

We apply factor analysis to TCP flow class throughput correlation matrices that are constructed using actual TCP flow records collected by networking equipment. With actual TCP flow measurements, a validation of the inferences of flow classes sharing congestion is extremely hard, if not impossible, since routing information about all the domains that flows visit and the congestion status of the servers that provide the incoming traffic are not available. However, bootstrap confidence intervals [14] can be used to demonstrate the statistical accuracy of the inferences. One may refer to [10] for the 95% bootstrap confidence intervals of the parameters that are used in making inferences in this section.

3.1 Description of Datasets

We use NetFlow records collected at the border router of The University of Texas at Austin (UT Austin) on November 6, 2002, between 12:58 PM and 2:07 PM CST, and on January 21, 2004, between 12:58 PM and 1:26 PM CST. The records that are collected in 2002 are referred to as Dataset2002, and those that are collected in 2004 are referred to as DataSet2004. Dataset2002 consists of 5,173,385 TCP flow records out of a total of 5,866,602 flow records. Dataset2004 consists of 4,440,697 TCP flow records out of a total of 6,556,674 flow records. The records contain both the incoming and outgoing traffic from UT Austin. The IP addresses belonging to UT Austin were made anonymous to protect privacy.

We assume that the packets from a given TCP flow follow the same route[4]. Such assumptions, although idealized, are not completely unrealistic for our one-hour long (or less) flow measurements.

3.2 Methodology

In NetFlow records, the start time of a flow is the time of arrival of the first packet in the flow, and the end time is the time of arrival of the last packet in the flow. Since the time between the first and the last packet is zero, flow throughput (size divided by duration) is not defined for flows consisting of one packet. Hence, one-packet flows will be omitted. Based on the premises of Section 2, we select to filter out flow records whose sizes are less than 8 kB or greater than 64 kB. The choices for these thresholds are based on some practical, empirical considerations: For example, Estan and Varghese [16] define "small" flows as those that send less than 0.1% of the link capacity during a given measurement interval, say 1 second. For instance, for a (bottleneck) OC-1 (optical carrier level 1) link of 51.84 Mbps, a small flow will be one that transports less than 7 kB. When choosing the upper threshold value for filtering out flows, we took into account the measurement studies that find that 50 kB Web objects (carried by TCP) are becoming common in the Internet [17]. Therefore, we can consider a flow whose size is larger than 64 kB as "large". In addition, in the Internet, packets belonging to flows that consist of only a few packets can sometimes arrive back to back (or with a very small inter-packet spacing). In this case, it is unreasonable to assume that such large flow throughputs are typical for that flow class. Hence, we will also omit all flows whose durations are shorter than one second[5].

We choose to analyze incoming traffic (flow records with source IP addresses) associated with AOL and HotMail, since one can reasonably expect that traffic belonging to these content providers potentially experience congestion at their source due to high demand for their content. We define two flow classes for traffic from each provider: AOL1 and AOL2 (class 1 and class 2) from AOL, and HotMail1 and HotMail2 (class 3 and class 4) from Microsoft Corporation. Assignment of flows into AOL1 or AOL2 (and similarly for HotMail1 and HotMail2) is performed by randomly splitting all flows from AOL (and HotMail) into two sets.

After filtering out small and large flows, we use temporal throughput samples of the remaining flow records on a discretized time axis to compute pairwise correlations between throughputs of p flow classes (see Fig. 4). Since throughputs of congestion sharing flows exhibit positive correlation when they temporally overlap (see Section 2)), we use the samples from (discrete) times when both flow classes are active to compute pairwise correlations. A $p \times p$ correlation matrix $\mathbf{R}$ is then constructed using the pairwise correlations. Using the *orthogonal factor model* [18], one can express $\mathbf{R}$ as

$$\mathbf{R} = \mathbf{\Lambda}^* \mathbf{\Lambda}^{*T} + \mathbf{\Psi}$$

where $\mathbf{\Lambda}^*$ denotes a $p \times m$ (rotated) *loading matrix*, and $\mathbf{\Psi}$ is a diagonal matrix. The number of latent factors m is determined by the number of eigenvalues of $\mathbf{R}$ that are above 1 [10]. The elements of the loading matrix $\mathbf{\Lambda}^*$, Λ^*_{ij}, capture the degree of correlation exhibited between a given factor (corresponding column in $\mathbf{\Lambda}^*$) and variable (corresponding row in $\mathbf{\Lambda}^*$). The elements of the diagonal matrix $\mathbf{\Psi}$, ψ_i, are used to compute the explanatory power of factors: $1 - \sum_{i=1}^{p} \psi_i / p$. Explanatory power gives an indication about the sufficiency of the number of common factors that are used in the model. Estimates $\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{\Psi}}$ for $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ can be determined by using the *principal component method*. The flow classes that have the largest (the most significant) loading (in magnitude) with a common factor are identified as classes that are likely to share a congested resource in the network [10].

3.3 Results

Using 95% (bootstrap) confidence intervals, we find that $\mathbf{R}$ has two eigenvalues that are above 1 (i.e. $m = 2$) in both datasets: Four classes share two different network infrastruc-

[4]This assumption is supported by the empirical measurements in [15].

[5]Flows can also be categorized according to their duration. Brownlee and Claffy [17] term flows whose durations are less than 2 seconds as "dragonflies".

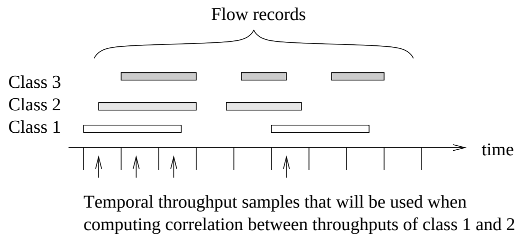

Figure 4: An example arrangement of flow records belonging to three classes on a discretized time axis. The length of the rectangles correspond to durations of flows.

tures. We then estimate factor loadings of four class throughputs based on $m = 2$ latent factors. For DataSet2002,

$$\hat{\mathbf{\Lambda}}^* = \begin{pmatrix} \boxed{0.7933} & 0.0711 \\ \boxed{0.7289} & -0.1315 \\ -0.0842 & \boxed{0.9088} \\ 0.0501 & \boxed{0.9240} \end{pmatrix}$$

and $\hat{\mathbf{\Psi}} = \mathrm{diag}\,(0.3656, 0.4514, 0.1669, 0.1437)$. For Dataset2004,

$$\hat{\mathbf{\Lambda}}^* = \begin{pmatrix} \boxed{0.8378} & -0.0451 \\ \boxed{0.8411} & 0.0044 \\ 0.0200 & \boxed{\text{-}0.7415} \\ 0.0260 & \boxed{\text{-}0.7351} \end{pmatrix}$$

and $\hat{\mathbf{\Psi}} = \mathrm{diag}\,(0.2961, 0.2926, 0.4497, 0.4589)$.

The explanatory power of the two factors is 72% in the case of Dataset2002 and 63% in the case of Dataset2004. By inspecting the significant loadings (which are boxed) on the loading matrices, we can conclude that classes 1 and 2 (flows belonging to AOL) share factor 1, and classes 3 and 4 (flows belonging to HotMail) share factor 2. In this case, factor 1 would be interpreted as the networking infrastructure belonging to AOL, and factor 2 would be the networking infrastructure belonging to Microsoft Corporation.

3.4 Discussion of Results

The potential power of this inference technique may be illustrated by considering the results in Section 3.3. For example, suppose that the users belonging to classes AOL1 and AOL2 at UT Austin were experiencing poor performance (long download times). Treating the external network as a "black box" (i.e., no knowledge about the utilization factors of access links or routing information of outside network), network managers could infer that poor performance was not due to the access links connecting UT Austin to the Internet, because the flow classes did not have one common factor that would indicate a bottleneck shared by all classes. The network managers could then hypothesize that the cause for poor performance was either at the content provider's server or a corresponding bottleneck link visited by pairs of flow classes (1 & 2 and 3 & 4) in the Internet.

REFERENCES

[1] K. Harfoush, A. Bestavros, and J. Byers, "Robust identification of shared losses using end-to-end unicast probes," in *Proc. IEEE Int. Conf. on Network Protocols*, Nov. 2000, pp. 22–36.

[2] D. Rubenstein, J. Kurose, and D. Towsley, "Detecting shared congestion of flows via end-to-end measurement," *IEEE/ACM Trans. on Networking*, vol. 10, no. 3, pp. 381–395, June 2002.

[3] M. Rabbat, R. Nowak, and M. Coates, "Network tomography and the identification of shared infrastructure," in *Proc. IEEE Asilomar Conf. on Signals, Systems and Computers*, Nov. 2002, pp. 34–38.

[4] D. Katabi, I. Bazzi, and X. Yang, "A passive approach for detecting shared bottlenecks," in *Proc. IEEE Int. Conf. on Computer Communications and Networks*, 2001, pp. 174–181.

[5] D. Arifler, G. de Veciana, and B. L. Evans, "Network tomography based on flow level measurements," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, vol. 2, May 2004, pp. 437–440.

[6] ——, "Inferring path sharing based on flow level TCP measurements," in *Proc. IEEE Conf. on Communications*, vol. 4, June 2004, pp. 2054–2059.

[7] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 2nd ed. Springer-Verlag, 2002.

[8] D. Chiu and R. Jain, "Analysis of the increase and decrease algorithms for congestion avoidance in computer networks," *Computer Networks and ISDN Systems*, vol. 17, no. 1, pp. 1–14, 1989.

[9] E. Altman, F. Baccelli, and J. Bolot, "Discrete-time analysis of adaptive rate control mechanisms," *High Speed Networks and Their Performance*, pp. 121–140, 1994.

[10] D. Arifler, "Network tomography based on flow level measurements," Ph.D. dissertation, The University of Texas at Austin, May 2004.

[11] S. B. Fred, T. Bonald, A. Proutiere, G. Régnié, and J. W. Roberts, "Statistical bandwidth sharing: a study of congestion at flow level," in *Proc. ACM Conf. on Appl., Tech., Arch., and Protocols for Computer Communications*, Aug. 2001, pp. 111–122.

[12] M. E. Crovella and A. Bestavros, "Self-similarity in World Wide Web traffic: evidence and possible causes," *IEEE/ACM Trans. on Networking*, vol. 5, no. 6, pp. 835–846, Dec. 1997.

[13] A. B. Downey, "The structural causes of file size distributions," in *Proc. IEEE Symp. on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, Aug. 2001, pp. 361–370.

[14] D. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Chapman & Hall, Inc., 1993.

[15] V. Paxson, "End-to-end routing behavior in the Internet," *IEEE/ACM Trans. on Networking*, vol. 5, no. 5, pp. 601–615, Oct. 1997.

[16] C. Estan and G. Varghese, "New directions in traffic measurement and accounting," in *Proc. ACM SIGCOMM Workshop on Internet Measurement*, Nov. 2001, pp. 75–80.

[17] N. Brownlee and K. Claffy, "Understanding Internet traffic streams: Dragonflies and tortoises," *IEEE Communications Magazine*, pp. 110–117, Oct. 2002.

[18] A. C. Rencher, *Multivariate Statistical Inference and Applications*. Wiley, 1998.