

From Traffic Matrix to Routing Matrix: PoP Level Traffic Characteristics for a Tier-1 ISP

Vijay Kumar Adhikari, Sourabh Jain, and Zhi-Li Zhang
Department of Computer Science & Engineering, University of Minnesota
Minneapolis, MN

viadhi@cs.umn.edu, sourj@cs.umn.edu, zhzhang@cs.umn.edu

ABSTRACT

Traffic matrices play a pivotal role in the management of an ISP’s network such as various levels of traffic engineering and capacity planning. However, it is unclear how the interaction between the internal traffic routing policies chosen by the ISPs and large-scale content providers, and the ongoing trend of “cloud-computing” affect the traffic matrices. In this paper, we use network data collected by a Tier-1 ISP to understand the characteristics of a PoP-level traffic matrix. We also shed light on the role of “routing matrix” in shaping the characteristics of the traffic matrix. Two of the most important observations in this study are: a) multi-exit prefixes and use of early-exit routing are the major reasons why PoP level traffic for the large ISPs do not follow the gravity model (i.e. proportional distribution based on size) as assumed by previous works, and b) routing plays a fundamental role in shaping the traffic matrix.

1. INTRODUCTION

With rapid growth of the Internet and the accompanying traffic, network traffic measurement plays an ever critical role in how network service providers and operators manage and plan network operations. For instance, the rise of data centers and emergence of cloud computing are making this measurement more complex, where content or service providers employ load-balancing (among multiple data centers) to dynamically adapt to user demands. Understanding the flow of traffic in such networks will help in improving the operations, management and security of today’s IP networks as well as emerging services.

Traffic matrix – which represents the flow of data from each ingress point to each egress point through a network – is an important piece of information needed to plan, manage and understand any network. Using (sampled) flow-level network data collected by a tier-1 ISP at its various PoP locations in the US and Europe, in this paper we study the key characteristics of the *PoP-level*¹ traffic matrix, with the goal to understand how various factors (e.g., where and how

¹PoP stands for “Point-of-Presence.” We note that traffic matrices can be defined in terms of various granularity: for example, besides PoPs, the ingress and egress points of the traffic can also be router interfaces, individual routers, IP prefixes or autonomous systems (ASes). Which granularity is appropriate often hinges on the network management tasks at hand. Although this study is primarily concerned with PoP-to-PoP traffic, in understanding the characteristics of the PoP-level traffic matrix we do take into account effects of traffic (and routing) at finer granularity such as IP prefixes and ASes.

ASes are inter-connected, routing policies they adopt) affect the traffic matrix dynamics, and the estimation and management thereof. Based on the actual traffic matrix derived from the ISP network data, we first demonstrate that the traffic matrix does not follow the so-called *gravity model* at the “global” scale (i.e., when viewed from the perspective of the entire ISP network): the gravity model and its variants have been a key premise used in several early studies on traffic matrix estimation from (SNMP-based) link-level byte counts [2, 9], see Section 2 for more discussion.

To explore the factors behind this phenomenon, we examine the traffic at finer AS and prefix levels (at which inter-domain routing is performed) and introduce the notion of *routing matrix*. The routing matrix is derived from BGP routing tables at routers of each ingress PoP, and records the egress PoP used for each prefix. We refer to prefixes having multiple egress PoPs as *multi-exit* prefixes, while prefixes having a single egress PoP are referred to as *single-exit* prefixes. Using the routing matrix, we find that multi-exit prefixes are prevalent (more than 60%) in the ISP due to a number of reasons (e.g., multiple peering or interconnection points between the ISP and some of its AS neighbors, typically other large ISPs or content providers, multi-homed customers, etc.). The prevalence of multi-exit prefixes and routing policies (e.g., early-exit routing) are main factors why the traffic matrix does not follow the gravity model. In fact, when considering only single-exit prefixes, the traffic matrix can still be well approximated by the gravity model. Additionally, for multi-exit prefixes we show that there is a strong *regional affinity* among PoPs: where given a prefix, PoPs within certain geographical regions are likely to pick the same egress PoP – typically a dominant PoP (in terms of both ingress and egress traffic) within the region. Our findings suggest several new directions for estimating and managing PoP-level traffic matrices of large ISPs. For instance, we can utilize the routing matrix (which can be obtained from BGP routing tables independent of traffic measurement) to partition the (global) traffic matrix into several regional ones, and then apply gravity model and other techniques (e.g., selective prefix-level flow sampling) to estimate the regional traffic matrices. By explicitly incorporating the routing matrix into the traffic matrix estimation, we can also account for the effects of network failures and routing changes, and explore “what-if” scenarios to investigate the impact of such failures and routing changes on the overall network performance. These new directions will be part of on-going work. The remainder of the paper is organized as follows. We provide background and related work in Sec-

tion 2. Section 3 analyzes the PoP-to-PoP traffic matrix and Section 4 examines the routing matrix. We discuss future directions and conclude the paper in Section 5.

Dataset. The dataset consists of (sampled) netflow records and BGP routing tables collected at 24 PoPs for 6 days in Spring 2008. For anonymity and related reasons, we will use the term *ISP-X* to denote the ISP where data was collected and no actual PoP names are used (except for rough geographic locations). Most figures and plots only show *normalized* statistics instead of absolute statistics (e.g., byte counts, etc.)

2. RELATED WORK

Most existing studies on network traffic matrices focus primarily on how to estimate the (PoP- or router-level) traffic matrix based on link load measurements (e.g., SNMP link counts). The problem is reduced to solving an acutely under-constrained mathematical equation of the form $Y = AX$ for X where Y is the vector of measured link loads, A is a “routing matrix” (different than what is defined in this paper) indicating which ingress-egress pairs use which links in the network and X is the vector representing the traffic matrix. Medina *et al.* [3] provides a good survey of the earlier studies which attempt to solve this problem by imposing additional constraints by assuming certain (often unrealistic) models that the traffic matrix is supposed to follow. In the same paper, the authors propose the “choice model” (a variant of the gravity model) for modeling the probability an ingress PoP to send packets to other PoPs. Following up on [3], Roughan *et al.* [5] formalize the notion of *gravity model* (in its simplest form, it assumes that each ingress point sends traffic to other egress points in proportion to their sizes), and also incorporate other information (e.g., structure, configuration and AS relations) to further improve the performance of the model. A *generalized* gravity model is further proposed in [9] where more detailed “tomographic” information is used, and traffic from peers is excluded in the traffic matrix estimation. In [2], a comprehensive comparison of various techniques for estimating the traffic matrix based on link loads is conducted using real data, and the paper also points out the limitations of the (simple) gravity model in estimating traffic matrices. This paper observes that PoP-level traffic matrix does not follow gravity model but does not clearly explain why. The “Independent connection model” is proposed in [1] which provides yet another and more intuitive form of the gravity model based on “independent” connections (instead of independence of individual packets). In [6] the impact of routing changes on the network traffic matrix is studied, whereas in [4] Nucci *et al.* show how routing changes can be utilized for traffic matrix estimation.

3. POP-LEVEL TRAFFIC MATRIX AND GRAVITY MODEL

In this section we examine the characteristics of the PoP-to-PoP traffic. We first look at the distribution of PoP level ingress and egress traffic totals. Next, we investigate the existence of gravity-model based proportional distribution in the PoP-level traffic matrix.

In order to understand the basic traffic characteristics at the PoP level, we first look at the amount of egress and ingress traffic at each PoP. From the collected raw flow data we extract the total amount of traffic that enters and exits

the ISP network at each PoP. We plot the sizes of PoPs in terms of their ingress and egress traffic volumes in Fig. 1. In this figure X-axis shows the various PoPs of *ISP-X*, and Y-axis shows the normalized ingress/egress traffic.

We see that PoP sizes vary widely. We also see that egress traffic at any PoP is very similar to the amount of ingress traffic at the PoP. This shows that *larger PoPs, in general, attract larger amount of traffic.*

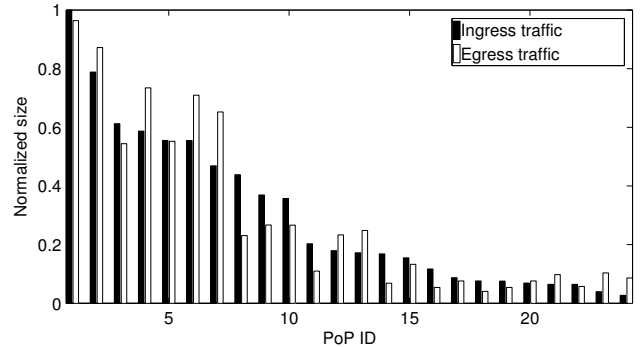


Figure 1: Ingress and egress traffic at different *ISP-X* PoPs.

Having observed that larger PoPs, in general, attract larger amount of traffic, we next want to see if the way the traffic is distributed is consistent across all the PoPs. That is, we want to see if the largest PoP, for instance, is the most popular destination for all the ingress PoPs. For this, we extract the PoP-to-PoP traffic matrix T_{pop} where each row represents the PoP at which the traffic enters the ISP and each column represents the PoP at which the traffic exits the ISP network and each entry in the matrix represents the total number of bytes sent from the ingress PoP to the egress PoP. We also extract the proportions in which ingress traffic at each PoP is distributed to all the egress PoPs by normalizing the traffic matrix T_{pop} by dividing the total traffic going from i th ingress PoP to an egress PoP j by the total ingress traffic at i . We refer to this row-wise normalized form of traffic matrix T_{pop} as \bar{T}_{pop} .

Fig. 2 graphically shows traffic matrix T_{pop} and Fig. 3 shows normalized traffic matrix \bar{T}_{pop} . For anonymity reasons, numbers in Fig. 2 are normalized by dividing each entry by the maximum number in the whole matrix. We can see in these figures that each ingress PoP seems to have different proportions in which the traffic is distributed to various egress PoPs. They do not seem to distribute the total ingress traffic in some fixed proportions to all the egress PoPs. Even the most popular egress PoP is different for different ingress PoPs. Furthermore, we see that normalized traffic matrix is dominated by the diagonal entries, which shows the presence of significant amount of traffic that exits the network from the same PoP where it enters the ISP network. We refer to such traffic as `local traffic`.

The visual inspection of the traffic matrix shown in Fig. 2 and 3 suggests that there are no consistent proportions in which traffic at each ingress PoP is distributed to all egress PoPs. Next, we mathematically verify our observations. We use singular values [8] based analysis of the traffic matrices. The basic idea here is that if all the rows of the matrix follow the same proportions, then it is an approximately rank-1 matrix and should have only one significant singular value.

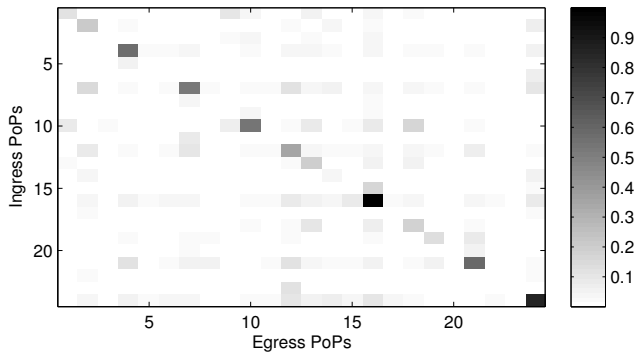


Figure 2: PoP-to-PoP traffic matrix.

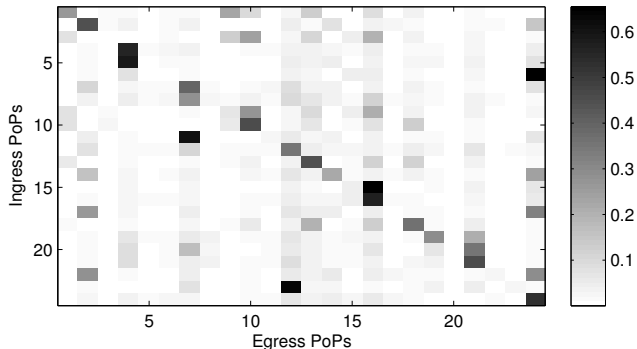


Figure 3: PoP-to-PoP traffic proportions.

Fig. 4 shows the normalized magnitudes of the singular values of T_{pop} . We can clearly see that the matrix has several significant singular values, which implies that T_{pop} is not even close to being a rank-1 matrix. This clearly suggests that there is no fixed global proportion at which the traffic is being distributed.

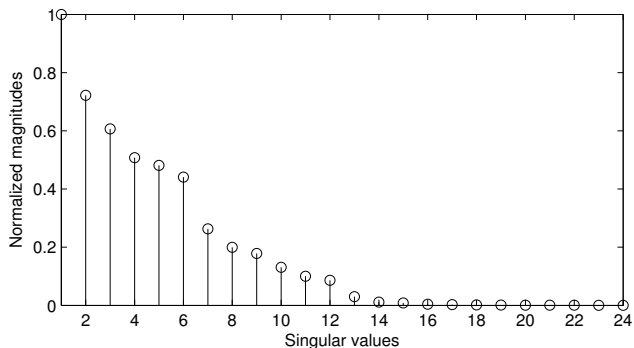


Figure 4: Distribution of singular values for PoP-to-PoP traffic matrix T_{pop} .

Ignoring local traffic: Several works related to PoP-to-PoP traffic estimation, such as [2, 3], either implicitly or explicitly ignore the local traffic. Since the local traffic is entering from a PoP and exiting from the same PoP, it might be useful to ignore them in some analysis such as when measuring the load in the ISP backbone because such traffic nev-

er enters the backbone. However, the gravity model is not well-defined at PoP-to-PoP level if we ignore the local traffic or set it to 0. Since the diagonal entries correspond to this local traffic, ignoring local traffic is equivalent to setting the diagonal entries to 0. Obviously, such matrix can not follow gravity model because there is no rank-1 matrix with 0s in the diagonal. This means that we do not have a meaningful way to check whether the rows of a PoP-to-PoP traffic matrix follow the same proportions or not if we ignore the local traffic.

Although ignoring local traffic at PoP-level traffic matrices make gravity-model ill-defined, we compared the ratios involving non-diagonal entries to see if there is a fixed proportion. We found that even when we compare pairs that did not involve diagonal entries the proportions at which one ingress PoP divides the traffic to other egress PoPs is, in general, very different than how another PoP divides. In summary, the rows of the matrices did not follow similar proportionality even when we ignore the local traffic. Therefore, when viewed from the perspective of the entire ISP network, the gravity model does not hold true.

4. ROUTING MATRIX

In Section 3, we have made a number of observations regarding the PoP-level traffic matrix. To understand the latent causes behind these observations, we need to understand the most important factor that drives the traffic from the ingress PoPs to the egress PoPs: *routing*.

In general, routing is done on the basis of IP prefixes that are announced by “neighbor ISPs” to each other through BGP sessions between their border routers. *ISP-X* peers with other large ISPs at multiple locations and those ISPs may announce similar set of prefixes at multiple locations. Additionally, many of its large customers including large content-providers and other multi-homed customers might announce their prefixes from more than one location. Due to this, traffic destined to a given IP address can potentially exit from the *ISP-X* network at many different PoPs. In this section, we investigate how *ISP-X* routes the traffic in the presence of several potential exit points. We use this information to explain the characteristics for the PoP-level traffic as observed in Section 3.

As mentioned above, some of the prefixes may be announced at multiple different locations by the neighboring ISPs. We use the term **multi-exit prefix** to refer to a prefix which is announced at more than one PoP locations, and therefore traffic destined to these prefixes may exit *ISP-X* network from multiple locations. On the other hand, some prefixes may be announced at only one PoP location, and therefore, traffic destined to these prefixes has only one exit PoP. In the following, we refer to such prefix as **single-exit prefix**.

To examine the traffic at the prefix level (at which inter-domain routing is performed), we introduce the notion of *routing matrix*. The routing matrix is derived from BGP routing tables at routers of each ingress PoP, and records the egress PoP used for each prefix. In this routing matrix $R_{pop \rightarrow p}$, an entry $R_{pop \rightarrow p}(i, p)$ represents the egress PoP for destination BGP prefix p at PoP i . We examine different characteristics observed in this routing matrix in the following subsections.

4.1 Prevalence of Multi-exit Prefixes and Early-exit Routing

We first look at the characteristics of different prefixes with respect to how many exit PoPs they have. In our dataset we see around 180k unique BGP prefixes learned by *ISP-X* and around 100k prefixes that are seen at more than one PoP location. Fig. 5 shows the distribution of these prefixes with respect to number of PoPs they are announced at. As seen in this figure, about 40% prefixes are only seen at one location. About 60% of the prefixes, on the other hand, are seen at more than one location.

Next we look at how many prefixes are reachable at any given PoP. Fig. 6 shows the fraction of prefixes seen at different PoPs. We see that the number of prefixes present at different locations varies significantly. However, the top 6 largest PoPs have approximately 40% of the prefixes reachable locally.

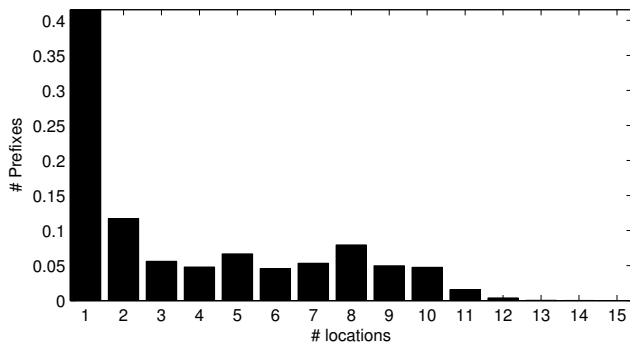


Figure 5: Exit-PoPs for prefixes.

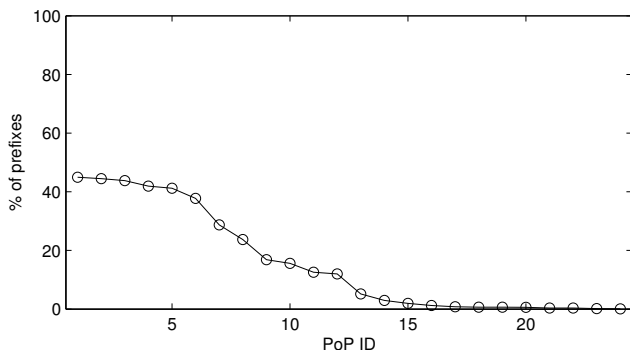


Figure 6: Prefix availability.

We have seen that a large number of prefixes are reachable through more than one exit-PoP. Therefore, ingress PoPs can route traffic destined to those prefixes through any of the egress PoPs where they are advertised. Next we want to see how the egress PoPs are chosen by the ingress PoPs for those multi-exit prefixes. Intuitively, *ISP-X* should try to reduce the load on its backbone network by routing the traffic through the “closest” egress PoP. This is usually referred to as early-exit (or hot potato) routing [7]. To see the extent of this early-exit routing in the data, we first obtained the geographical distances between all the PoP pairs and computed the percentage of time a PoP sends traffic to the nearest exit PoP from the list of possible exit PoPs. Fig. 7 shows that

for almost all PoPs, nearest-exit PoP is chosen for almost all the traffic. This result leads to two related observations. First, we see that early-exit routing is being used heavily and second, geographic distance closely captures the underlying “cost” of transporting data between PoPs. The cases where an exit-PoP other than the nearest one was chosen mostly involved cases such as Düsseldorf preferring Frankfurt over Amsterdam where the relative distances did not differ much.

4.2 Proportionality for “Single-Exit” Prefixes

Since a large portion of prefixes can be reached through multiple exit PoPs, the preferences chosen by each individual PoP to route its ingress traffic play an important role in shaping the overall PoP-to-PoP traffic matrix. Therefore, in order to separate the effects of the choices made by the ISP in selecting the egress PoPs, we look at the distribution of traffic involving only single-exit prefixes.

For each of the single exit prefixes, all the ingress traffic destined to it exits from one fixed PoP. Therefore, there is no effect of routing decisions made by different PoPs. We try to see if the PoP-level proportionality holds for the traffic destined to these single-exit prefixes. For this, we construct a PoP-to-PoP traffic matrix (T_{pop}^1) by only considering traffic destined to these single exit prefixes.

To see how closely the rows of this matrix (T_{pop}^1) follows the fixed proportional distribution, we again analyze the distribution of its singular values. Fig. 8 shows the distribution of singular values for the matrix T_{pop}^1 . Clearly the first singular value is significantly larger than other singular values, and captures more than 45% of the total variance in the data. This shows that T_{pop}^1 is much closer to being a rank-1 matrix than the original traffic matrix (T_{pop}). Therefore, we can infer that the presence of multiple exit points for prefixes adds to significant distortion from gravity model.

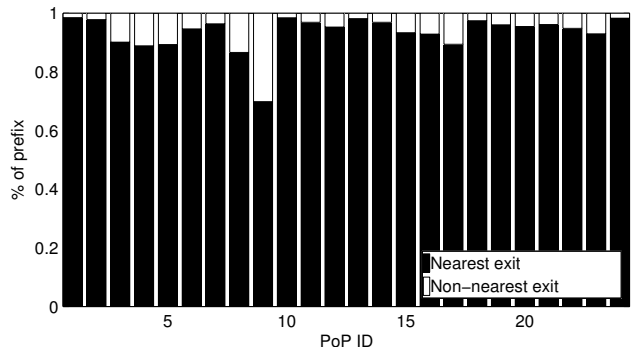


Figure 7: Prevalence of early-exit routing.

4.3 Classification of PoPs based on Regional Affinity

We have seen that in the presence of multiple egress PoPs for a given destination prefix, ingress PoPs generally prefer geographically closest egress PoPs. Therefore, geographically closer PoPs are more likely to have similar routing entries for different prefixes in their routing tables.

To evaluate how similar PoPs are in terms of the exit-PoP choices they make for all the prefixes, we cluster them using hierarchical clustering. For this we construct feature vector

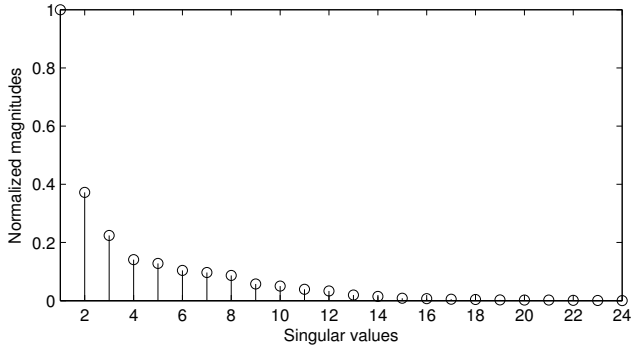


Figure 8: Singular values for T_{pop}^1 .

for each ingress PoP by using the egress PoPs for different destination prefixes, which is same as the rows of routing matrix $R_{pop \rightarrow p}$. For comparing feature vectors for any two PoPs we are interested in seeing whether they choose the same egress PoP to route the traffic destined to a given destination prefix or not. Therefore, we use *Hamming Distance* as a distance measure for the hierarchical clustering.

Fig. 9 shows the dendrogram for the hierarchical clustering of the PoPs. In this figure, X-axis shows various PoPs, and Y-axis shows the inter-cluster distance² between the clusters of the PoPs. The figure shows that, in general, each PoP is different from others, except for few pairs of PoPs which are very close to each other geographically. Similarly, clusters of PoPs which have similar routing choices are also close to each other geographically.

Interplay between Routing & Traffic Matrices. In Section 3 we observed that the gravity model or any other proportionality-based model does not hold true for the PoP-level traffic matrix. On the other hand we see strong geographical affinity among PoPs in terms of their routing choices. It leads us to an interesting research question: *Does this geographical affinity due to similar routing preferences also translate into similarities in the traffic distributions for the PoPs?* For instance, Fig. 9 shows that (NY, NJ) have very similar routing preferences, and so do (FL2, FL3). Similarly, (NY, NJ) also have very similar traffic distribution, and so do (FL2, FL3) (See Fig. 11). However, (NY, NJ) and (FL2, FL3) differ significantly with each other both in terms of routing preferences and traffic distribution. In order to understand the similarities among PoPs based on traffic distribution and routing preferences, we cluster the PoPs using the traffic matrix. For this clustering we use traffic distributions for all the PoPs as the feature vector, and correlation between them as the similarity measure. The traffic distribution for each PoP is given by the rows of the matrix \bar{T}_{pop} , and the similarity of the distribution implies that these PoPs divide the traffic to the egress PoPs at the same ratio. We observed that there are groups of PoPs that have similar distribution in general, however PoPs from different groups have very different traffic distribution proportions from each other. We also observe that, in most cases, the cluster of PoPs that have similar distributions are ge-

²We use the average pair-wise distance between two clusters as the inter-cluster distance between them. It is defined as the average of all the pair-wise distances, for the pairs of elements by choosing one element from each cluster.

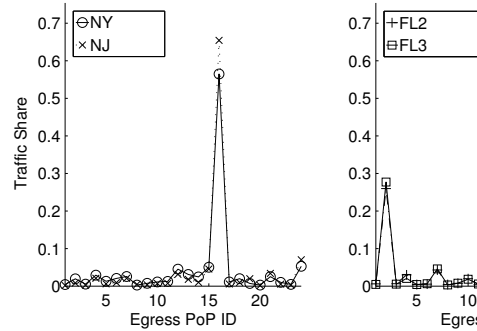


Figure 11: Examples of traffic distribution from four ingress PoPs

ographically close to each other. Fig. 10 presents the dendrogram showing clusters of PoPs that have similar traffic distribution proportions. The X-axis in this figure are the PoPs grouped together based upon their traffic distribution similarity and the Y-axis shows the intra-cluster distance.

The clusters seen in Fig. 10 are also consistent with the clusters seen on the basis of routing matrix as shown in Fig. 9. It shows a strong correlation between traffic and routing matrices, and therefore, shows that routing matrix plays an important role in shaping the traffic matrix.

In summary, we see that most of the observations that we made about the PoP-level traffic matrix can be elucidated by the routing matrix. We also observed that because of the prevalence of multi-exit prefixes large ISPs can and do employ early-exit routing extensively.

5. ONGOING WORK & CONCLUSION

In this paper we use flow-level network data collected by a tier-1 ISP at its various PoP locations to study the key characteristics of the *PoP-level* traffic matrix with the aim of understanding how various factors such as PoP location, prefix-level routing policies etc. affect the PoP-level traffic matrix observed by the ISP. Our paper makes two key contributions. First, we show that PoP-level traffic matrix does not follow the gravity model, as used by several earlier studies. Second, we explore the factors behind this phenomenon by examining the traffic at prefix level and introduce the notion of *routing matrix*. Moreover, our result shows that the prevalence of multi-exit prefixes and early-exit routing used by ISPs is one of the key factors why gravity-model does not hold for the PoP level traffic matrices. It is further corroborated by the fact that gravity model holds for the PoP level traffic matrix when only single-exit prefixes are considered. We also show that there is a strong regional affinity among PoPs in terms of how traffic destined to multi-exit prefixes is routed (i.e. egress PoPs chosen by these PoPs) as well as for the proportions in which traffic is distributed at PoP level.

Finally, our findings suggest several new directions for estimating and managing PoP-level traffic matrices of large ISPs. For instance, it might be possible to partition the global PoP level routing matrix into several regional ones using the prefix level “routing matrix”, and apply gravity model and other techniques (e.g., selective prefix-level flow sampling) to estimate the regional traffic matrices. In addition, by taking into account the routing matrix for traffic estimation, we can also model the effects of routing changes

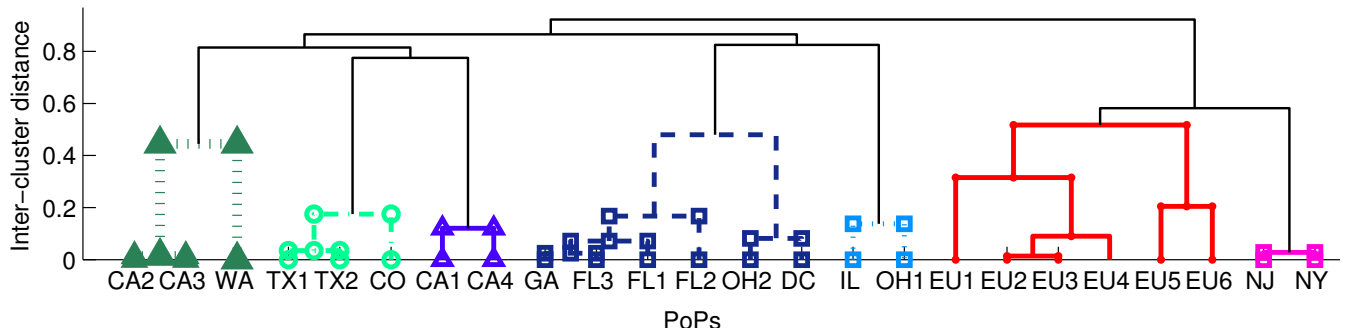


Figure 9: Dendrogram representing the clusters of PoPs using routing matrix.

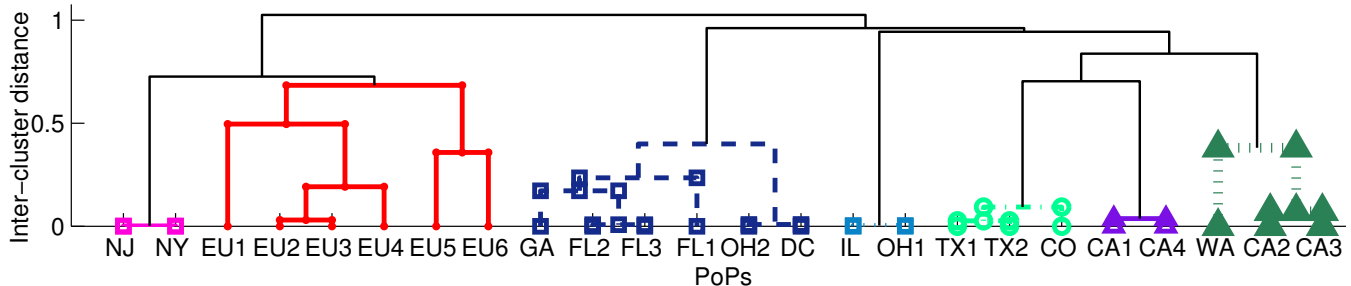


Figure 10: Dendrogram representing the clusters of PoPs using traffic matrix.

due to network failures or policy changes. It would not only provide a better model for network management, but also help us in exploring “what-if” scenarios to understand the impact of failures and policy changes on the overall network performance. These new directions are part of our current on-going work.

6. ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers and the discussant for our paper, Giuliano Casale, for their helpful feedback on this paper.

We gratefully acknowledge the support of our sponsors. The work is supported in part by the NSF grants CNS-0626812, CNS-0721510 and CNS-0905037. The opinions expressed in this paper are solely those of the authors and do not necessarily reflect the opinions of any of the funding agencies or the University of Minnesota.

7. REFERENCES

- [1] V. Erramill, M. Crovella, and N. Taft. An independent-connection model for traffic matrices. In *IMC*, 2006.
- [2] A. Gunnar, M. Johansson, and T. Telkamp. Traffic matrix estimation on a large IP backbone: a comparison on real data. In *IMC*, 2004.
- [3] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot. Traffic matrix estimation: Existing techniques and new directions. *ACM SIGCOMM Computer Communication Review*, 2002.
- [4] A. Nucci, R. Cruz, N. Taft, and C. Diot. Design of IGP link weight changes for estimation of traffic matrices. In *IEEE Infocom*, 2004.
- [5] M. Roughan, A. Greenberg, C. Kalmanek, M. Rumsewicz, J. Yates, and Y. Zhang. Experience in measuring backbone traffic variability: Models, metrics, measurements and meaning. In *IMC*, 2002.
- [6] R. Teixeira, N. Duffield, J. Rexford, and M. Roughan. Traffic matrix reloaded: Impact of routing changes. *Passive and Active Network Measurement*, 2005.
- [7] R. Teixeira, A. Shaikh, T. Griffin, and J. Rexford. Dynamics of hot-potato routing in IP networks. *SIGMETRICS*, 2004.
- [8] T. Will. Introduction to the Singular Value Decomposition. *La Crosse, WI*, 2003.
- [9] Y. Zhang, M. Roughan, N. Duffield, and A. Greenberg. Fast accurate computation of large-scale IP traffic matrices from link loads. *SIGMETRICS*, 2003.