What SNMP data can tell us about Edge-to-Edge network performance

Extented Abstract

Demetris Antoniades[1], Kejia Hu[2], Alex Sim[2], and Constantine Dovrolis[1]

[1] College of Computing
Georgia Institute of Technology
[danton,constantine]@gatech.edu
[2] Computational Research Division
Lawrence Berkeley National Laboratory
[kjhu,asim]@lbl.gov

With the high speeds of today's networks, monitoring information is most of the time either summarized or sampled. This policy is even more profound in network backbones, where aggregation of data from several sources and in very high speeds is often observed. SNMP link utilization counts, provide an inexpensive resource of information for network administrators and managers. In this paper, we provide evidence that using SNMP link counts, one can infer Edge-to-Edge information about network transfers taking place in the network. Our method first identifies significant variations in the link counts, and tags them as possible flow starting or ending points. Iteratively following these variations to the neighboring routers, we are then able to identify the path the specific flow traversed through the monitored network. Applications, like throughput prediction, traffic matrix estimation, anomaly detection, can use the resulting information, similarly or additionally to Netflow data.

The Simple Network Management Protocol (SNMP) is widely used to provide aggregated link usage data from network components. These data, even not enhanced with a great amount of detail, provide a valuable source for network administrators, aiding decisions about network routing, provisioning and configuration. SNMP data are simple to collect and maintain, thus also providing a low disk space for historical network usage log.

On the other end, Netflow data provides detailed information for end-to-end performance. Using Netflow, one can have accurate information about a host pair communication, the amount of data transferred back and forth. The enhanced information given by Netflow comes with the additional cost for its collection and many privacy concerns regarding the wide availability of the data. To reduce the cost of collecting Netflow data, aggressive sampling (i.e. 1:1000 packets) is often employed, even for relatively low-speed networks [4]. Sampling significantly affects the accuracy of Netflow data and may limit their application of use [2]. Netflow records of end-to-end information also include the IP addresses and port numbers used by the participating parties. Such content raises significant user privacy concerns [3, 5]. As a result, there is a limitation of Netflow data availability while private content is either censored or completely removed.

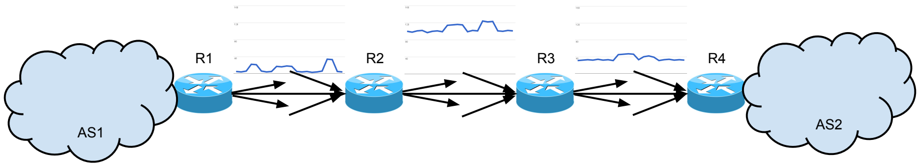

Fig. 1. Simple illustration: network events can be observed from SNMP link utilization data. These events can, also, be tracked down along the network path their traversing.

In this paper we provide evidence that by using SNMP link counts Edge-to-Edge (E2E) information about network transfers can be inferred. The motivation for this work came from the need of a statistically significant set of E2E throughput samples, allowing us to perform TCP throughput prediction in a monitored network based on historical measurements [6]. The network wide availability of SNMP data and the limited throughput performance samples from Netflow data motivated us to explore different approaches for increasing our sample data set.

Our method is the result of two main observations. Looking at the time series data of a link's usage, we observe events where the usage of the link increases (or decreases) to a different level, deviating from the link's normal behavior up to that point. These events could be considered as starting (or ending) points of high throughput transfers. Checking all other links of the same router, we can identify the exit point of that specific event, which leads to the next router. Following this path, we can then infer the actual route that the specific event followed, tracking down its entrance and exit points in the monitored network.

Figure 1 presents a simple example. In this diagram, $R1$ is the edge router for autonomous system $AS1$ and $R2$ is the edge router for autonomous system $AS2$. Each router has several input and output interfaces, connecting to other routers. The path $R1 - R2 - R3 - R4$ connects the two ASes, carrying all traffic exchanged between them. The graphs above each link show the actual traffic transferred through that link during the specific observation period, as it would be available through SNMP. Each link caries traffic for a variety of source and destination pairs. Link $R1 - R2$ experiences three different events during the observation period. The first and last events are not visible after $R2$. The middle event continues from $R2$ over the link $R2 - R3$ and from there to $R3 - R4$. This event and its transfer route can be attributed to a network transfer initiated at $AS1$ and destined to $AS2$. The magnitude of the deviation on the utilization time series provides information about the throughput this transfer achieved. The time of the increase and decrease events provide information regarding the starting and ending times of the transfer.

Figure 2 shows a more realistic example. On a real-world network, several events can be traced simultaneously. The figure shows two different events, traversing different paths of the network. We believe that the extra knowledge these events would provide will be helpful in several network management appli-

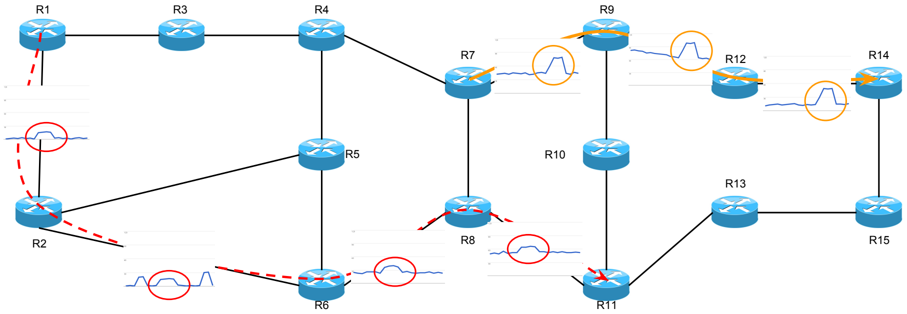

Fig. 2. In a real-world network several traffic utilization increase and decrease events can be identified in each observation period.

cations such as Throughput Prediction, Traffic Matrix Estimation [7], Anomaly Detection techniques [1], and more.

The work presented in this paper is, to the best of our knowledge, the first that suggests the possibility of inferring Edge-to-Edge information from aggregated link utilization measurements. Using our observations we propose a methodology for identifying traffic utilization increase and decrease events and tracking down the links these events traverse in the network. Preliminary results, over publicly available SNMP data from ESnet, a large National Research and Educational Network (NREN), suggest that we can identify and track up to 80% of the events that appear in a network link. Furthermore, the magnitude of the identified events seems not to be limited to the high throughput ones.

References

1. V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3):15:1–15:58, July 2009.
2. B. Choi and S. Bhattacharyya. On the accuracy and overhead of cisco sampled netflow. In *Proceedings of ACM SIGMETRICS Workshop on Large Scale Network Inference (LSNI)*, 2005.
3. S. Coull, M. Collins, C. Wright, F. Monrose, and M. Reiter. On web browsing privacy in anonymized netflows. In *Proceedings of 16th USENIX Security Symposium on USENIX Security Symposium*, page 23. USENIX Association, 2007.
4. C. Estan and G. Varghese. New directions in traffic measurement and accounting. In *ACM SIGCOMM Computer Communication Review*, volume 32, pages 323–336. ACM, 2002.
5. M. Foukarakis, D. Antoniades, S. Antonatos, and E. Markatos. Flexible and high-performance anonymization of netflow records using anontool. In *Security and Privacy in Communications Networks and the Workshops, 2007. SecureComm 2007. Third International Conference on*, pages 33–38. IEEE, 2007.
6. Q. He, C. Dovrolis, and M. Ammar. On the predictability of large transfer tcp throughput. In *ACM SIGCOMM Computer Communication Review*, volume 35, pages 145–156. ACM, 2005.

7. A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot. Traffic matrix estimation: existing techniques and new directions. In *Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications*, SIGCOMM '02, pages 161–174, New York, NY, USA, 2002. ACM.