

Understanding the topological properties of Internet traffic: a view from the edge

Juan Antonio Cordero, Olivier Bonaventure
ICTEAM, Université catholique de Louvain (Belgium)

{juan.cordero|olivier.bonaventure}@uclouvain.be

Abstract—Traffic between an edge network and the rest of the Internet can be represented as a dynamic loop-free graph. Understanding in depth the dynamics in time and space (spatial structure, topological breadth, destination persistency, traffic dominating paths) of this graph provides significant insight on the Internet internal architecture and capabilities. This paper analyzes inter-domain traffic from a large campus network based on one month by way of Netflow measurements. Our analysis reveals the topological properties and structure of the traffic graph (breadth, depth, volume), the stability of contacted destinations and the relationship between their popularity and their path length. Based on the observed traffic, we explore the suitability of a simple mathematical model to describe the structure of the outgoing traffic graph.

I. INTRODUCTION

Achieving a complete understanding of the Internet properties, dynamics and behavior is a highly challenging task. This is due to the huge size of the Internet as a complex system, on one side, and its continuous growth and rapid, decentralized evolution, on the other. It is however necessary to gain insight about the way that data flows across the Internet, information is demanded by final users and provided by content providers.

In the last years, a significant amount of research has been performed to describe and model the key aspects of the Internet topology at different levels, and to characterize the structure and volume of Internet traffic. Less research efforts, to the best of our knowledge, have been dedicated to measure, understand and model the topology of the interdomain traffic graph and its evolution, more in particular between edge networks and the Internet.

The goal of the present paper is to contribute to the latter objective. We focus on the traffic exchanged between an edge campus network, connected to the Internet by means of a border router (BR), and the rest of the Internet. We propose a simple methodology to

capture the evolution of the traffic graph at the router level. We present in this paper the first, preliminary observations from one month measurements of interdomain traffic, and describe the main research worklines to be followed in the future.

Literature provides several examples of analysis of traffic in the Internet or in edge networks connected to the Internet. Paxson (1994) performs an extensive analysis of TCP wide-area traces and characterizes analytically different applications running on top of TCP [5]. Thompson [7] and Fraleigh [14] analyze the structure, main applications and transport properties of wide-area Internet traffic through commercial backbones of MCI and Sprint, respectively. Other studies focus on the analysis of application-specific traffic (such as Youtube [19] [24]) in campus networks, mobility, link-layer aspects and traffic structure in wireless campus networks [16] or traffic structure in broadband Internet traffic [25]. More recently, Mikians *et al.* [31] [32] have studied the statistical properties of Interdomain Traffic Matrices (ITMs) measured from transit networks, and proposed a modeling tool for them. While most of these studies address Internet traffic from the edge, none of them explores the topology and graph characteristics of the edge-to-Internet traffic.

In parallel, many research efforts have been deployed to describe the Internet topology – mostly at AS and IP router level. Govindan *et al.* [8] draw AS-level conclusions on the topology and growth of the global Internet (diameter, sparsity, inter-AS route stability) based on BGP traces. Faloutsos *et al.* [9] proposed a power laws model for characterizing AS-level interdomain topology in the Internet. Uhlig *et al.* [12], [15] explored the interaction of Internet traffic and AS-level topology, and examines the distribution of traffic, stability and topological properties of interdomain paths towards destination ASes: presented results show that a relatively small number of destination ASes concentrate most of the observed traffic, confirming previous find-

ings, but this traffic traversed a significant number of intermediate ASes. Later interdomain topology studies such as Gill *et al.* [20], Dhamdere *et al.* [29] and Labovitz [27] reported a shift in the structure of interdomain Internet from a hierarchical architecture to a more horizontal and “flat” Internet in which more traffic is sent from content providers to consumer networks.

These AS-level studies are complemented with router-level topology analysis. One of the main tools for performing such analysis is `traceroute` [4], extended in 2006 to the `paris-traceroute` tool. Paxson [6] was the first to examine properties of end-to-end inter-domain paths (stability, routing conditions, etc.) by using `traceroute`. Later, CAIDA’s Archipelago project has collected and maintained topology information about the Internet via `traceroute`-like tools; the resulting traces have been used, among other purposes, for comparing the different `traceroute` methods [21].

The contribution of this paper is two-fold. First, it presents a new, simple methodology for measuring inter-domain traffic from an edge network, based on the combination of Netflow measurements and `paris-traceroute` to reconstruct the outgoing IP-level traffic graph. Second, it presents a preliminary analysis of the traffic collected with this methodology at the Border Router of a large campus network during 31 days. The paper focuses on two aspects of interdomain network traffic: the study of destinations characteristics and evolution (popularity, dynamics, interdomain paths) and the properties and structure of the network outgoing traffic graph.

The remainder of this paper is organized as follows. Section II describes the analyzed internetwork, the tools and the methodology used to measure interdomain traffic. Section III examines some relevant characteristics of contacted destinations, in particular their persistency in time and the relationship between path length and destination popularity. Section IV studies the topological structure of the outgoing traffic graph and its daily evolution, and studies the similarity of the obtained graphs with a tree. Section V concludes the paper.

II. METHODOLOGY AND SETTING

This section describes the methodology and tools used to collect and analyze information from the measured inter-domain traffic (section II-A); and describes the main characteristics of the examined edge network and the extracted dataset (section II-B).

A. Methodology

Netflow v7 was enabled on the campus border router (BR) to collect traffic statistics. We used the `flow-tools` collector [11] to aggregate Netflow reports in 5min periods. Each Netflow record corresponds to one transport-level flow. It contains the source and destination addresses and port numbers, as well as the flow duration and the number of bytes and packets transmitted and received [26]. We combine the Netflow collector with a script that analyzes the external IP addresses reported inside each report.

The Netflow collector is combined with two daemons: `retriever` and `merge30`. `retriever` parses the Netflow report every 5min to detect new external destinations. For each IP address belonging to a new /24 prefix, `retriever` performs a `traceroute` to collect the path towards this prefix. We run the exhaustive algorithm of `paris-traceroute` with ICMP [17] for computing IP paths from the campus network to the destination. Each 5min interval is thus represented by a list of IP external destinations (hereafter, `iplist`) and a list of IP paths towards these destinations (`pathlist`). This tool is an extension of standard ICMP-based `traceroute` tool; the use of the *exhaustive* algorithm with ICMP enables it to identify load-balancers along the path and has proven to achieve a good performance in terms of destination discovery [30].

`merge30` enables offline measurements over the extracted `iplist` and `pathlist` files. It operates on 30min intervals, that is, it waits until 6 consecutive Netflow `iplist/pathlist` reports are ready, merges the corresponding files, generates the traffic graph from the BR to the rest of Internet and computes other parameters related to the destination prefix (contacting sources, traffic exchanged, hop distance...). Rationale for this 30min merging is related to the Netflow recording behavior: as flows have a maximum duration of 30min and are recorded only at their end [11], using finer-grained reports (*i.e.*, with intervals smaller than 30min) for analysis may be misleading.

B. Campus Network Characteristics and Trace Summary

The UCL campus internetwork is assigned the 130.104.0.0/16 prefix. The internetwork contains a set of fixed wired prefixes with around 13k machine connections available (hosts and servers), and a set of wireless prefixes (see Table I), available via Wifi.

Wi-fi prefixes
130.104.88.0/24
130.104.120.0/22
130.104.124.0/22
130.104.160.0/20
130.104.202.128/25
130.104.203.0/25

TABLE I
UCL WI-FI NETWORK PREFIXES [1].

This study only considers traffic exchanged with the Internet. Traffic from wired networks is dominant during weekdays (Fig. 1), but wireless traffic is always significant. *A priori*, particularities of the observed network (*e.g.*, the fact that a substantial part of the network, mostly in its wired region, is used for scientific and research purposes) are expected to have an impact in the examined traffic that needs to be taken into account; note however that traffic through the wifi network (accessible to students and university staff) is likely to show a less specific profile. As in many campus networks, traffic is *asymmetrical* with more incoming than outgoing (Fig. 2), but since TCP is the dominant protocol, most external destinations appear in both the incoming and the outgoing traffic. In the following, the term *source* will refer to IP addresses in the local campus internetwork, and *destination* to IP addresses in the rest of the Internet. The terminology is used to indicate the main sense of the traffic, although both directions of traffic are considered together.

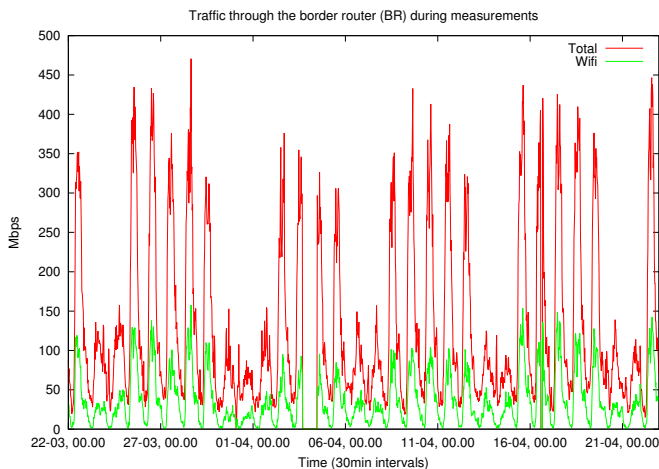


Fig. 1. Total traffic (bytes) through the BR between 2013/03/22 and 2013/04/22, wireless and wired traffic.

Traces analyzed in this paper were collected between 2013/03/22 and 2013/04/22 (both included), that is,

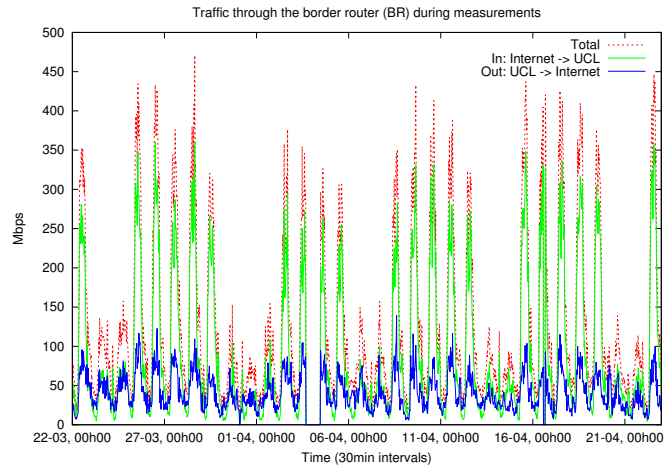


Fig. 2. Total traffic (bytes) through the BR between 2013/03/22 and 2013/04/22, incoming and outgoing traffic.

31 days. Figs. 1 and 2 show the amount of traffic monitored in the measurement interval. There are three periods in which no data was collected via Netflow: between 2h and 2h30 on 2013/03/31 (2 timeslots), between 16h30 on 2013/04/03 and 11h on 2013/04/04 (38 timeslots), and between 14h and 16h on 2013/04/16 (5 timeslots). These empty periods correspond to failures at the `netflow` router due to external events (electricity shortages, network breakdowns).

For fairness, results about graph topology daily evolution are restricted to the 27 days for which complete data is available. The study of load-balancers (LB) impact in paths towards IP destinations is not addressed: reported LB subpaths towards the same destination prefix are collapsed into a single *IP path class* for which length is the length of the longest LB subpath.

III. DESTINATIONS AND PATHS

This section studies the paths from the border router to external *destinations* (that is, Internet prefixes outside the edge network). Section III-A discusses a notion of destination popularity and examines the relationship between popularity and path length. Section III-B introduces and discusses the Persistency Index (PI) of destinations, which describes the presence of a destination in the traffic graph; this section also uses the PI parameter to characterize the variability of the set of contacted destinations during the measured time interval.

A. Destination Popularity and Path Length

Recent studies [20] [27] [29] indicate that the Internet is suffering in the last years a fundamental change

in the users traffic demands and the interdomain architecture. Traffic exchanged with consumer networks is increasingly dominated by a few content providers (e.g. Google, Youtube or Akamai [28]). Moreover, deployment and co-location of CDNs [13] close to ASes from service providers, may be in the origin of a shift towards a “flatter” Internet, in which growing portions of the traffic are not necessarily carried by transit AS, as in the standard hierarchical Internet, but directly from content providers to consumer networks [20] [29]. From the consumer point of view, such a trend implies that popular content is reachable *closer* from the edge.

This section examines this trend from the observed campus network. We measure the *destination popularity during a time interval* as the number of sources that contacted the destination within the considered interval. Figs. 3 and 4 show the relationship between the popularity of external IP destinations and the length of the IP paths towards them, for a sample day (point cloud, Fig. 3) and for traffic exchanged during the whole measurement interval (histogram, Fig. 4). The sample day (2013/03/22) is the first day with complete measures; it was selected for illustration purposes, as the shape of the displayed point cloud is representative of other daily samples. According to these definition, and not surprisingly (even in a partly research-oriented campus network) the most daily-popular destination prefixes correspond to Autonomous Systems belonging to Google (AS15169), Facebook (AS32934) or CDNs such as Akamai (AS20940).

Note that each single source requesting a destination at different times in the day (different 30min intervals, given the granularity of the processed Netflow reports) is counted several times, once for each 30min period in which the destination was contacted. This corresponds to a notion of *popularity* not only based on the absolute number of users contacting a destination, but also the frequency of contacts – a destination contacted many times a day by the same source is more *popular* than another destination contacted only once *per* day by a source.

The histogram in Fig. 4 shows the average length of the IP path towards destinations with respect to destinations’ popularity. Ten histogram classes are defined; the bounds of those including destinations with several number of sources are selected so that the number of destinations *per* histogram class is as similar as possible (see Table II). The i -th histogram class, with bounds $x_{\min}(i)$ and $x_{\max}(i)$, contains all destinations

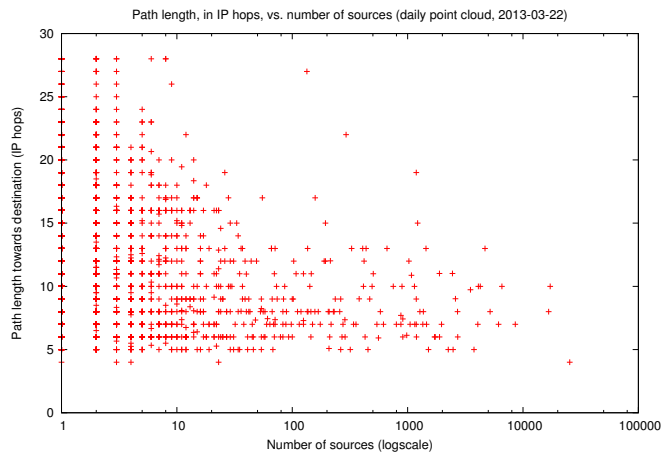


Fig. 3. Example of daily point cloud of path length (in IP hops) vs. number of sources.

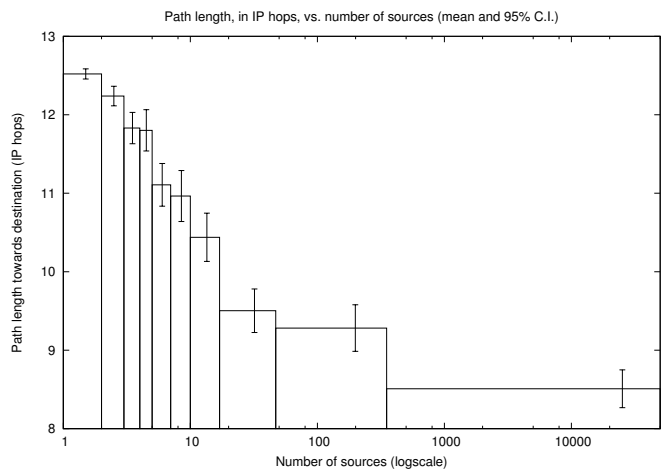


Fig. 4. Path length towards destination vs. number of sources.

with number of sources $|S|$ satisfying $x_{\min}(i) \leq |S| < x_{\max}(i)$; the height of the corresponding bar is the average path length towards all contained destinations. Results are presented with the 95% confidence interval.

Correlation shown in Fig. 4 between destination popularity, in number of contacting sources, and IP path length clearly indicates that more popular destinations are typically reachable through a smaller number of hops. This is consistent and provides additional evidence to support the trends described in previous studies [20] [27].

B. Persistency Index of Destinations

Let D_n be the set of destinations contacted at timeslot n ($n = 0$ corresponds to 2013/03/22, 0h, time difference between two consecutive timeslots is 30min), and let I_n^p be the set of destinations that are contacted in (at least) p consecutive slots starting (and including)

Sources	Destinations
1	18898
2	5181
3	1961
4	1140
[5,7)	1088
[7,10)	756
[10,17)	697
[17,47)	666
[47,351)	667
[351,50000)	497

TABLE II
DESTINATIONS *per* HISTOGRAM CLASS.

n , that is, $n, n + 1, \dots, n + p$. That is:

$$I_n^p = D_n \cap D_{n+1} \cap \dots \cap D_{n+p} = \bigcap_{i=n}^{n+p} D_i$$

Then, the topological p -**Persistence Index** (PI) is defined with the following quotient:

$$(PI^t)_{n,p} = \frac{|I_n^p|}{|D_n|}$$

This corresponds to the fraction of destination prefixes (with respect to the number of destination prefixes at timeslot n) contacted at timeslot n and also contacted at timeslots $(n + 1), \dots, (n + p)$. The definition extends in the obvious way to the traffic-weighted PI.

Fig. 5 shows the evolution of $(PI^t)_{n,p}$ along the measured month, for different values of p . The proportion of persistent destinations oscillates on a daily basis (and reaches minimums in the early morning). Beyond this daily oscillation, around 40% of the destination prefixes are 1-persistent, that is, they are contacted at $(n + 1)$ if they were contacted at n ; this percentage decreases as persistency is larger (that is, bigger values of p are considered).

The Traffic-weighted PI at Fig. 6 shows the fraction of the total traffic in each timeslot that is exchanged with 1-persistent destinations. It turns out that, while around 40% of destination prefixes are renewed between two consecutive timeslots (1-persistent, see Topological PI in Fig. 5), these destination prefixes are “responsible” for the majority (above 70%, in average) of exchanged traffic of the considered timeslot. In other words, the set of destination prefixes is highly volatile, meaning that most of them are not present in two consecutive timeslots. This volatility, however, only affects a reduced fraction of exchanged traffic: the 40%

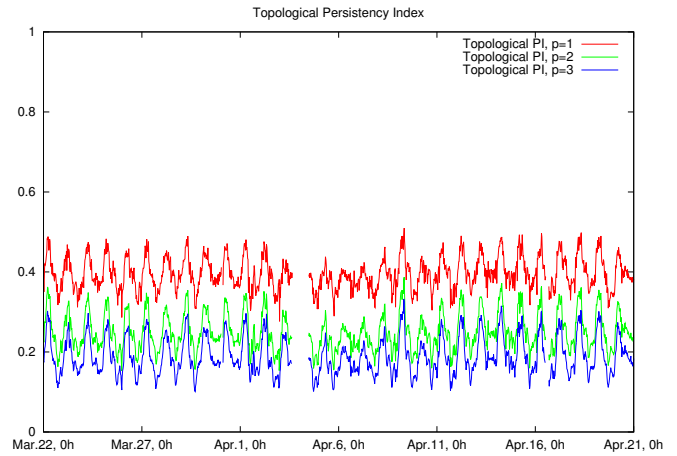


Fig. 5. Topological PI, for $p = 1, 2, 3$.

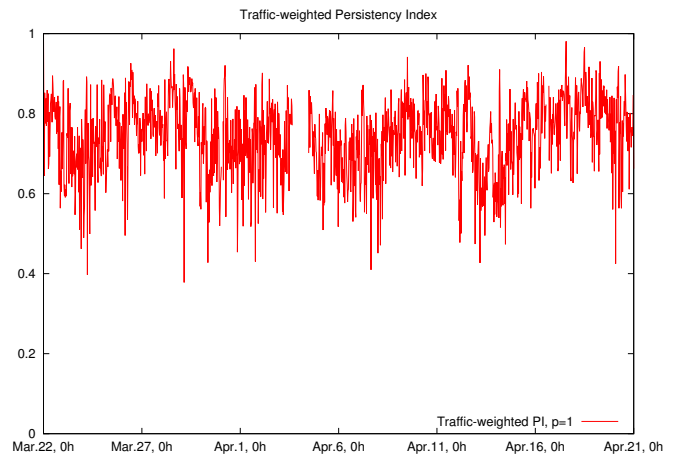


Fig. 6. Traffic-weighted PI, for $p = 1$.

1-persistent destinations are “responsible” for most of the exchanged traffic in each timeslot.

IV. TRAFFIC GRAPH

This section examines the structure of the interdomain traffic graph. Section IV-A describes the depth and breadth of the traffic graph and its daily evolution. Section IV-B introduces and motivates the Tree Similarity Index (TSI) of a traffic graph, and discusses the tree similarity of the measured traffic graphs.

In the following, let $G_n = (V_n, E_n)$ be the traffic graph for timeslot n , where V_n is the set of vertices and E_n is the set of edges, and let $w : E \rightarrow \mathbb{N}$ be the weight function that maps an edge $e \in E$ to the traffic $w(e)$ (in bytes) traversing e . Figure 7 shows a simplified graphical representation of this traffic graph at a particular date and time.

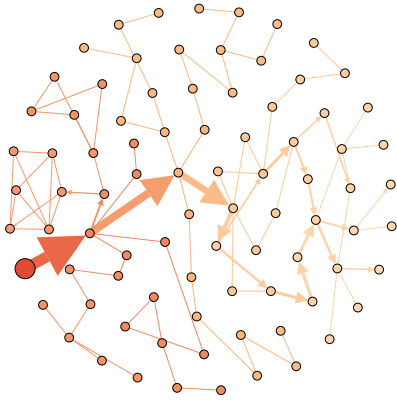


Fig. 7. Simplified representation of the (directed) traffic graph on March 3rd, 2013, at 16h (edges weighed according to amount of exchanged traffic, in bytes; the initial vertex corresponds to the border router).

A. Graph Depth and Breadth

Fig. 8 shows the daily evolution of average topologic breadth and length of the traffic graph. The *length* of the graph is the maximum number of consecutive IP hops towards a destination present in the graph. The *breadth* of the graph is the maximum number of different routers (vertices) at the same distance of the BR.

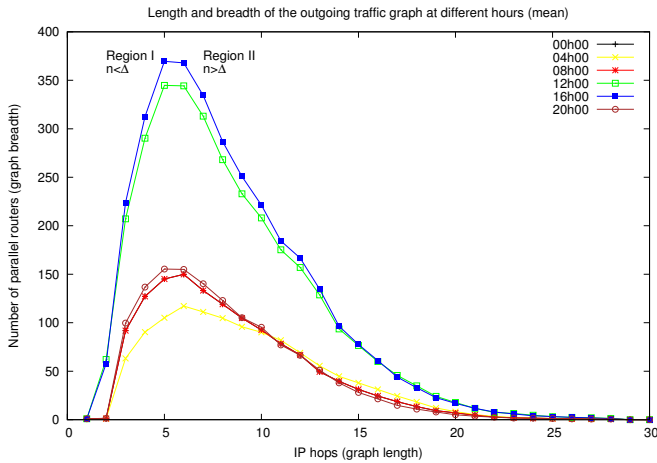


Fig. 8. Topological depth and average breadth of the outgoing traffic graph, daily evolution.

Results are collected for particular times of the day, averaged over 27 days. Two typical traffic topology patterns are observed during the day:

- 1) during the central hours of the day (12pm-4pm), the graph breadth reaches its maximum, with 345 simultaneous branches (different paths to different IP destination prefixes), 5 hops away from the BR;

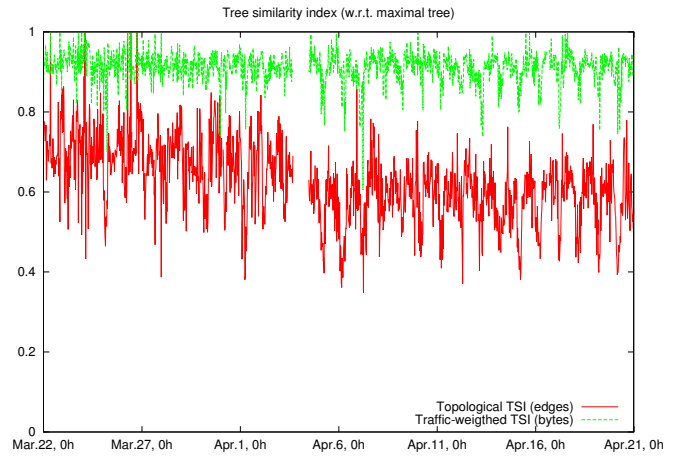


Fig. 9. Estimated Tree similarity index.

- 2) during the late evening, night and early morning, the figure shows the background traffic graph shape, with a minimum router breadth of 117 simultaneous branches, on average, 6 hops away from BR.

Two regions are clearly distinguished in Fig. 8 along the x axis. In region I, the number of simultaneous routers keeps increasing until it reaches a maximum, 5 or 6 hops away from the BR. The traffic graph is there dominated by the underlying Internet topology, meaning that the number of destinations to be reached through the graph is smaller than the number of available links towards these destinations. As traffic moves away from the BR, the number of available links increases dramatically. In region II (beyond 5-6 hops from BR), this is no longer a dominating factor on the graph structure.

B. “Treerization” and Tree Similarity

We explore the topological characteristics of the measured traffic graph. In particular, we focus on the similarity of obtained traffic graphs to trees (that is, graphs in which there is only a single path between any pair of vertices) with root in the border router (BR). Motivation for measuring tree similarity is two-fold.

Firstly: in ideal conditions, a traffic graph having a tree structure would indicate a *fully consistent* routing policy (at the IP level) in the Internet. In practice, uncoordinated routing policies, routing transitions, load balancing and other phenomena may lead to redundant or partly-overlapping IP-level routes. Hence, the tree similarity brings an indicator of router-level routing inconsistencies perceived in the Internet. Secondly, tree-structured graphs are tractable objects that can be

used for modeling purposes – in particular, modeling of inter-domain traffic dynamics. Determining the similarity of real traffic graphs and trees allows to assess the ability of tree-based models (e.g., stochastic branching processes such as Galton-Watson [2], [3]) to capture key aspects of real edge-Internet inter-domain traffic graphs and their dynamics.

The **Tree Similarity Index** (TSI) of G_n measures the difference between the graph (not considering LB subpaths) and the *maximal tree* contained in G_n . This *maximal tree*, denoted by $T(G_n) \subseteq G_n$, corresponds to the subgraph tree with maximum traffic weight. Computation of the maximal tree of a given traffic graph is sketched in Algorithm 1.

Algorithm 1 “Treerization” algorithm.

Require: $G_n = (V_n, E_n)$

Ensure: $T(G_n) = (V_T, E_T)$

$E_T \leftarrow E_n; V_T \leftarrow V_n$

for $k = 1$ to 30 **do**

$V_k = \{v \in V_T : d(v, \text{BR}) = k\}$

find $S \in V_k : \text{nparents}(S) = n > 1$

$\text{parents}(S) = \{P_1, \dots, P_n\}$ ▷ sorted by

increasing traffic

for $i = 1$ to $n - 1$ **do**

$R \leftarrow P_i; X \leftarrow S$

if $\text{nparents}(R) = 1$ **then**

if $\text{nchildren}(\text{parent}(R)) > 1$ **then**

Break

else

$X \leftarrow R; R \leftarrow \text{parent}(R)$

$E_T \leftarrow E_T \setminus \{R\bar{X}\}; V_T \leftarrow V_T \setminus \{X\}$

Fig. 9 shows an estimation for the tree similarity of measured traffic graphs. Two values of TSI are displayed: the estimated Topological TSI, $TSI^t(G_n)$, which indicates the proportion of edges of G_n that are kept in $T(G_n)$; and the estimated Traffic-weighted TSI, $TSI^{tw}(G_n)$, which indicates the proportion of traffic (in bytes) carried by edges kept at $T(G_n)$. Due to limitations in the algorithm implementation and the traceroute output (mostly related to routers unresponsiveness), estimations are conservative, meaning that examined traffic graphs are *at least* as similar to a tree (not considering LB subpaths) as indicated in the TSI estimation.

$$TSI^t = \frac{|E(T(G_n))|}{|E(G_n)|}$$

$$TSI^{tw} = \frac{\sum_{e' \in E(T(G_n))} w(e')}{\sum_{e \in E_n} w(e)}$$

It can be observed that the maximal tree contains between a 60% and a 70% of edges of the complete graph, but it carries more than 90% of the total traffic exchanged between the campus network and the Internet. Since examined traffic graphs are computed on a 30min basis, and result from merging 6 individual 5min traffic graphs, some routing inconsistencies may be due to the presence of different routes towards the same destination prefix valid on different times within the corresponding 30min timeslot. According to our observations, though, this is a relatively rare case; more likely to occur for low impact (in number of contacts and in exchanged traffic) prefix destinations.

In this context, the tree approximation preserves most of the exchanged traffic. From a topological perspective, however, the tree reduction implies the exclusion of a significant amount of edges from the original traffic graph.

V. CONCLUSION AND FUTURE WORK

The study of interdomain traffic dynamics in the edge of the Internet is essential to understand the situation and leading trends of Internet evolution. The paper describes a new methodology for edge interdomain traffic measurements based on the combination of Netflow and traceroute. This is used to measure the interdomain traffic at the *Université catholique de Louvain* during one month.

Based on these measurements, this paper presents a characterization of the observed interdomain traffic based on three observations. First, the relationship between IP path length and different measures of destination popularity is examined: our study shows that IP paths are smaller towards destinations contacted more frequently, which provides additional evidence supporting the notion of “Internet flattening” already suggested in previous work. Second, the observed interdomain traffic is topologically volatile in time, meaning that most destination prefixes are not present in the graph in consecutive timeslots; traffic exchanged with persistent prefixes is however dominant. Third, the paper examines the structure of the traffic graph and its similarity to a tree: first results indicate that, although topology does not correspond to a tree, the reduction

to a maximal tree graph (denoted as “*treerization*”) preserves most of the exchanged traffic; the graph can thus be modeled as a tree carrying most interdomain traffic combined with additional edges with limited traffic significance. This preliminary result suggests that the tree approximation is reasonable and traffic modeling based on stochastic branching processes [3] can be explored in further work.

Described methodology and results enable several other directions for future research related to Internet inter-domain traffic characterization. LB impact is not addressed, but should be taken into account in finer analysis. Presented results correspond to observations over the total amount of exchanged traffic. Topology, load distribution and dynamics of the observed graph may however show important differences for specific types of traffic; for instance, depending on the involved application (Web, mail, etc.) or depending on the type of local network (wireless or fixed wired inside the edge internetwork).

ACKNOWLEDGEMENTS

This work has been partly supported by the Belgian Science Policy Office, by way of the IAP BESTCOM (Belgian Network on Stochastic Modelling, Analysis and Optimization of Communication Systems) program; and by the European Community’s Seventh Framework Programme (FP7/2007-2013) Grant No. 317647 (Leone).

REFERENCES

- [1] UCL/SRI: *Réseau UCL et intranets UCL : réseaux et nombres de raccordements* (UCL network and UCL intranets: networks and number of plugs). <http://www.sri.ucl.ac.be/reseauUCL.html>. Last access on March 17th, 2014.
- [2] H. W. Watson, F. Galton: On the Probability of Extinction of Families. In *Journal of the Anthropological Institute of Great Britain*, vol. 4, pp. 138–144, 1875.
- [3] T. E. Harris: *The Theory of Branching Processes*. Berlin: Springer, 1963.
- [4] V. Jacobson: “traceroute”, February 1989. Available at <ftp://ftp.ee.lbl.gov/traceroute.tar.gz> (last access on March 17th, 2014).
- [5] V. Paxson: Empirically-Derived Analytic Models for Wide-Area TCP Connections. In *IEEE/ACM Transactions on Networking*, vol. 2, num. 4, pp. 316–336, Aug 1994.
- [6] V. Paxson: End-to-End Routing Behavior in the Internet. In *ACM SIGCOMM Computer Communication Review*, vol. 16, num. 4, Oct. 1996.
- [7] K. Thompson, G. J. Miller, R. Wilder: Wide-Area Internet Traffic Patterns and Characteristics. In *IEEE Network*, Nov.-Dec. 1997.
- [8] R. Govindan, A. Reddy: *An analysis of Internet inter-domain topology and route stability*. Proc. IEEE Int. Conf. on Computer Communications (INFOCOM’97), Apr. 7-12, 1997.
- [9] M. Faloutsos, P. Faloutsos, C. Faloutsos: *On Power-Law Relationships of the Internet Topology*. Proc. ACM SIGCOMM Conf., Sept. 1999.
- [10] W. Fang, L. Peterson: *Inter-AS Traffic Patterns and Their Implications*. Proc. IEEE Global Internet Symp., Dec. 1999.
- [11] M. Fullmer, S. Romig: *The OSU Flow-Tools Package and Cisco Netflow Logs*. Proc. 14th Systems Administration Conf. (2000 LISA XIV), Dec. 2000.
- [12] S. Uhlig, O. Bonaventure: Implications of Interdomain Traffic Characteristics on Traffic Engineering, *ETT Journal, Special Issue on Traffic Engineering*, 2002.
- [13] J. Dille et al.: Globally Distributed Content Delivery. In *IEEE Internet Computing*, vol. 6, no. 5, pp. 50–58, 2002.
- [14] C. Fraleigh, S. Moon, B. Lyles et al.: Packet-Level Traffic Measurements from the Sprint IP Backbone. In *IEEE Network*, Nov.-Dec. 2003.
- [15] S. Uhlig, O. Bonaventure et al.: *Implications of the Topological Properties of the Internet Traffic on Traffic Engineering*. Proc. 18th Systems Administration Conf. (SAC’04), Mar. 2004.
- [16] T. Henderson, D. Kotz, I. Abyzov: *The changing usage of a mature campus-wide wireless network*. Proc. 10th ACM Annu. Int. Conf. on Mobile Computing and Networking (MOBICOM), Sept. 2004.
- [17] B. Augustin et al.: *Avoiding traceroute anomalies with Paris traceroute*. Proc. ACM Internet Measurement Conf. (IMC’06), Oct. 2006.
- [18] B. Donnet, T. Friedman: Internet Topology Discovery – A Survey. In *IEEE Communications Surveys*, vol. 9, no. 4, 4th quarter 2007.
- [19] P. Gill, M. Arlitt, Z. Li, A. Mahanti: *Youtube traffic characterization: a view from the edge*. Proc. ACM Internet Measurement Conf. (IMC’07), Oct. 2007.
- [20] P. Gill, M. Arlitt, Z. Li, A. Mahanti: *The Flattening Internet Topology – Natural Evolution, Unnoticed Barnacles or Contrived Collapse?*. Proc. Passive and Active Measurement Conf. (PAM’08), Apr. 2008.
- [21] M. Luckie, Y. Hyun, B. Huffaker. *Traceroute probe method and forward IP path inference*. Proc. ACM Internet Measurement Conf. (IMC’08). Oct. 2008.
- [22] H. Haddadi et al.: Network Topologies – Inference, Modeling and Generation. In *IEEE Communications Surveys*, vol. 10, no. 2, 2nd quarter 2008.
- [23] M. Luckie, Y. Hyun, B. Huffaker: *Traceroute Probe Method and Forward IP Path Inference*. Proc. ACM Internet Measurement Conf. (IMC’08), Oct. 2008.
- [24] M. Zink, K. Suh, Y. Gu, J. Kurose: *Watch Global, Cache Local: Youtube Network Traffic at a Campus Network – Measurements and Implications*.
- [25] G. Maier, A. Feldmann, V. Paxson, M. Allman: *On Dominant Characteristics of Residential Broadband Internet Traffic*. Proc. ACM Internet Measurement Conf. (IMC’09), Oct. 2009.
- [26] M. W. Lucas: *Network Flow Analysis*, No Starch Press, 2010.
- [27] C. Labovitz et al.: *Internet Inter-Domain Traffic*. Proc. ACM SIGCOMM Conf., Sept. 2010.
- [28] E. Nygren, R. K. Sitaraman, J. Sun: The Akamai Network: A Platform for High-Performance Internet Applications. In *SIGOPS Operating Systems Review*, vol. 44, issue 3, July 2010.
- [29] A. Dhamdhere, C. Dovrolis: *The Internet is Flat: Modeling the Transition from a Transit Hierarchy to a Peering Mesh*. Proc. ACM 6th Int. Conf. on emerging Networking EXperiments and Technologies (CoNEXT’2010), Nov.-Dec. 2010.
- [30] B. Huffaker, M. Fomenkov, kc claffy: *Internet Topology Data Comparison*, Tech. Rep., CAIDA, May 2012.

- [31] J. Mikians *et al.*: *Towards a Statistical Characterization of the Interdomain Traffic Matrix*. Proc. IFIP Networking 2012, May 2012.
- [32] J. Mikians *et al.*: *ITMgen – A First-principles Approach to Generating Synthetic Interdomain Traffic Matrices*. Proc. IEEE Int. Conf. on Communications (ICC'13), Jun. 2013.