Noname manuscript No.
(will be inserted by the editor)

On the complexity of Internet traffic dynamics on its topology

Steve Uhlig

the date of receipt and acceptance should be inserted later

Abstract Most studies of Internet traffic rely on observations from a single link. The corresponding traffic dynamics has been studied for more than a decade and is well understood. The study of how traffic on the Internet topology, on the other hand, is poorly understood and has been largely limited to the distribution of traffic among source-destination pairs inside the studied network, also called the traffic matrix. In this paper, we make a first step towards understanding the way traffic gets distributed onto the whole topology of the Internet. For this, we rely on the traffic seen by a transit network, for a period of more than a week. As we are still at the stage of understanding the topological traffic distribution, we do not try to model the traffic dynamics. Rather we concentrate on understanding the complexity of describing the traffic observed by a transit network, how it maps onto the AS-level topology of the Internet and how it changes over time. For this, we rely on well-known tools of multi-variate analysis and multi-resolution analysis.

Our first observation is that the structure of the Internet topology highly impacts the traffic distribution. Second, our attempts at compressing the traffic on the topology through dimension reduction suggests two options for traffic modeling: 1) to ignore links on the topology for which we do not see much traffic, or 2) to ignore time scales smaller than a few hours. In either case, important properties of the traffic might be lost, so might not be an option to build realistic models of Internet traffic.

Realistic models of Internet traffic on the topology are not out of reach though. In this paper, we identify two properties such models should have: 1) use a compact representation of the dependencies of the traffic on the topology, and 2) be able to capture the complex multi-scale nature of traffic dynamics on different types of links.

Keywords Internet traffic, topology, PCA, multi-resolution analysis

Technische Universität Berlin/Deutsche Telekom Laboratories, Ernst-Reuter-Platz 7, 10587 Berlin, Germany
E-mail: steve@net.t-labs.tu-berlin.de

1 Introduction

After more than a decade of work on understanding the properties of Internet traffic [16, 15, 19, 29, 18, 33], there is a consensus that scaling models provide a concise and relevant description of traffic dynamics. However relevant, those models are limited by the fact that they describe traffic observed on an individual link, i.e. an access link connecting a University to its Internet provider or even a high-speed backbone network link. These works give us a view of Internet traffic completely focused on time dynamics, i.e. the traffic is seen as a mathematical function or a signal that varies over time.

Network traffic however, is generated by hosts distributed over the Internet topology. How the traffic is exchanged between Internet hosts, and which paths this traffic takes is as relevant as how the traffic varies over time. Several studies have addressed the question of how Internet traffic is distributed across different hosts of the Internet. For example, [12] studied the traffic distribution on the ARPANET as early as in 1974. An important finding of [12] was that a few sites (i.e. networks) were responsible for the majority of the traffic. Similar results were obtained in the 90's on the NSFNet backbone [6] and at universities and a commercial backbone [9]. Recent studies have observed the same trends on a large transit ISP [10] and several non-transit networks [31, 32].

In this paper, we combine two viewpoints, time dynamics and topological distribution. Contrary to [14, 26, 25] who also take a time-space approach to analyze network traffic, we do not limit the network topology to a single network.

Rather, we rely on the whole network-level path that crosses the studied network until it reaches the destination. We study Internet traffic by attributing traffic to paths on its network-level topology, and analyze the dynamics of the traffic on the networks crossed by those paths. Monitoring the physical paths, for instance with traceroute [17], taken by Internet traffic is hardly feasible. Instead, we use the Internet topology at the autonomous system (AS)-level. An AS can be seen as a network, like an Internet Service Provider (ISP) or a University network. ASs are of two types: transit or stub. Transit ASs (e.g. an ISP) allow traffic that is not sent or received by hosts located in their network, to cross their network. Stub networks (e.g. a university network) on the other hand do not allow traffic to cross their network unless the source or the destination of the traffic belongs to their network. The abstraction of the AS topology, while not completely capturing the geographic properties of the physical Internet, still gives an accurate view of the networks crossed by traffic to reach their destination.

We apply multi-variate and multi-resolution analysis, to analyze the traffic variability over the different links of the AS-level topology and across time scales. Our analysis leads to the following new observations about Internet traffic:

1. Most of the variance of the traffic is localized on a subset of the Internet topology. However, to account for most of the variance of the traffic, a large number of AS-level edges must be considered.
2. Most of the variance of Internet traffic on the topology corresponds to time scales of hours or more. However, different AS-level edges have different burstiness structures, mostly depending on how much traffic these edges carry.
3. To reduce the complexity of a traffic model on the AS topology, we find two options: 1) to ignore links on the topology for which we do not see much traffic, or 2) to ignore time scales smaller than a few hours. In either case, important properties of the traffic might be lost, like representativeness or burstiness.
4. When building a concise description of Internet traffic dynamics on its topology, a better information compression can be obtained by leveraging topological information, than by trying to condense time information. This observation may help building traffic models that reproduce topological traffic dynamics while keeping the model complexity low.

The first observation is consistent with [32] that showed significant churn in the paths capturing most of the traffic seen by stub networks. Even though the second observation states that time scales smaller than hours can be ignored, traffic dynamics on the Internet topology is harder to describe than pure time dynamics. Time dynamics can be concisely described with scaling models, as only time scales are related to each other. When traffic is attributed to the path it follows on the other hand, the dependencies between different parts of the topology have to be considered, which probably considerably complexifies the traffic model.

We now summarize the remainder of this paper. Section 2 defines our terminology and describes our methodology. Section 3 analyzes the distribution of the traffic on AS-level edges. Section 4 studies how to simplify the data while retaining most of the variance and the traffic volume. Section 5.2 studies the multi-scale components of the traffic dynamics on AS-level edges. Section 6 concludes and discusses further work in building models of Internet traffic on the topology.

2 Data and methodology

We obtained traffic and routing information from the GÉANT network. GÉANT is the pan-European research network and it is operated by DANTE. It carries research traffic from the European National Research and Education Networks (NRENs) connecting universities and research institutions. GÉANT has a point of presence in each European country. At the time of the study, GÉANT was composed of 23 routers interconnected using 38 bi-directional links. In addition, GÉANT had 53 links with other networks.

To properly attribute traffic to the right path it followed when traffic was observed, we need to build a model of the routing of GÉANT [21]. To compute paths inside its network, GÉANT uses the ISIS routing protocol. We obtained a trace of the ISIS messages. With these messages, we keep an up-to-date view of the internal state of GÉANT and compute the shortest paths from any router to any other router inside the GÉANT network during the whole time of the study. Once we know the internal path followed by the traffic inside the GÉANT network, we can find out the exit router of GÉANT that forwarded traffic outside the network.

Then, we rely on information from the BGP routing protocol to determine the global AS-level paths taken by traffic observed by GÉANT to reach its destinations. BGP [23] is the current routing protocol used between ASs. With BGP, each AS learns the paths to reach each destination in the Internet. In GÉANT, the BGP routes are collected using a dedicated workstation running GNU Zebra [1], a software implementation of different routing protocols including BGP. The workstation has an iBGP session with all the border routers of the network. Using this technique, it is possible to collect all the BGP routes selected by the border routers of GÉANT and thus find out the global AS-level path followed by traffic entering GÉANT towards any destination in the Internet. With this, we know the set of ASs crossed by traffic entering GÉANT towards any destination, at any time instant of the study [21].

We also obtained Netflow [5] traces collected from all external links of the GÉANT network, i.e. all the traffic en-

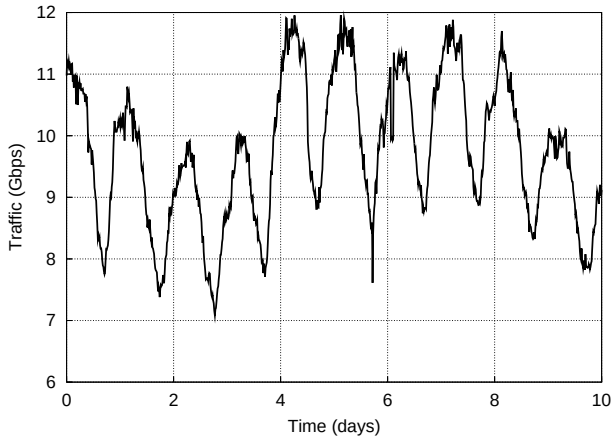

Fig. 1 Traffic evolution of GÉANT.

tering the network was recorded. Netflow provides the aggregated information of the layer-4 flows, by recording the starting time, the ending time and the total volume in bytes for each unidirectional TCP and UDP flow. Netflow was configured with a 1/1000 packet sampling rate. With this sampling, only one out of 1000 is considered by Netflow. In a large network such as GÉANT, the amount of traffic prohibits to use low sampling rates as it is unsafe for the proper operation of the routers. Given that the aim of this paper is not to study the small time scales, the decision was made to use a granularity of 15 minutes for the finest time scale.

Once we have a model of the routing of GÉANT, we compute for each Netflow entry the corresponding AS path the traffic takes to reach its destination, and attribute the traffic seen to each AS-level link along the path. We call an *edge*[1] e of the AS graph G, a pair $ASX - ASY$ appearing as two consecutive and distinct ASs in the AS path computed by our model of GÉANT. We attribute to each *edge* e the amount of traffic t_{ie} it carries during time interval i. We denote by $\mathbf{X}$ the matrix whose columns contain the traffic for each *edge* e and whose rows give the traffic for all *edges* during time interval i.

We study a contiguous 10 days period between May 5 2005 and May 15 2005, corresponding to 1000 15-minutes time intervals. Figure 1 provides the evolution over time of the traffic volume as seen in the Netflow statistics. On Figure 1, we multiplied the traffic volume seen in the Netflow statistics by 1000 to account for the 1/1000 Netflow sampling performed by GÉANT.

[1] We use the terms edge and link interchangeably in this paper, but they always refer to an AS-level edge. An AS-level edge does not correspond to a physical link of the router-level graph, but may correspond to several physical links on the topology.

3 Traffic distribution on the AS topology

Before going into details about traffic dynamics, let us first better understand how traffic seen by GÉANT is distributed over the AS-level topology of the Internet. Previous works have studied the distribution of the traffic, but from the viewpoint of stub networks [31,32]. Stub networks do not allow traffic that is not sent or received by hosts located within their networks. As a transit network, most of the traffic carried through GÉANT is not produced nor has as destination hosts from GÉANT itself. The traffic it carries is exchanged between hosts within the academic institutions it connects.

The structure of the Internet is believed to be hierarchical [35,27], with a few tens of large ASs in the core, surrounded by a few hundreds of smaller transit ASs, and finally thousands of sub ASs at the periphery. Due to the centrality of a few of those core ASs, we expect to see a lot of traffic concentrating onto a few central edges in the topology, while edges at the periphery should see a rather limited amount of traffic.

3.1 AS hop distance

In previous work [31,32], it has been shown that the distribution of the traffic on the AS topology follows a tree-like structure. Previous work has shown this behavior for stub networks. In this section, we show that a transit AS like GÉANT observes a similar distribution of its traffic over the AS topology. We conjecture that this behavior is an invariant in the Internet and largely a consequence of the choice of interconnections by ASs.

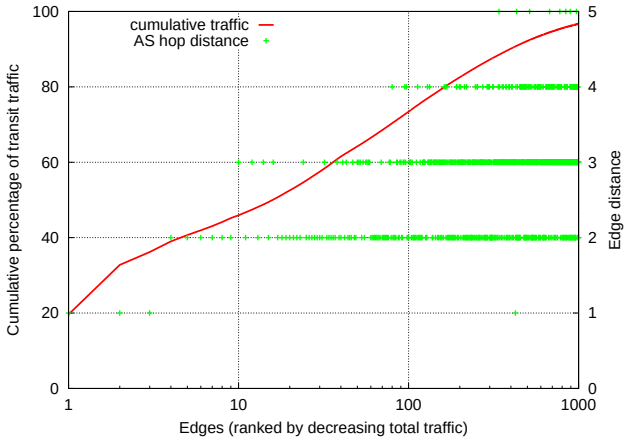

Fig. 2 Transit traffic on AS topology.

On Figure 2, we ordered edges by decreasing amount of traffic seen over the whole studied period. We then plot the cumulative fraction (left of y-axis) of the traffic seen over all edges. Note that the sum of the traffic seen on all edges

is larger than the total traffic seen by GÉANT, as some traffic is seen over several edges along the path from GÉANT towards the destination AS. We observe that a few edges see a significant fraction of the traffic. However, as observed in [31,32], a significant number of edges need to be considered to capture a large fraction of the total traffic. To help understand where the edges are located with respect to GÉANT, we also plot on Figure 2 the distance in AS hops from GÉANT (right of y-axis). A distance of 1 means that the edge is directly connected to GÉANT. The largest three edges in traffic are one hop away from GÉANT. Most of the largest ten edges are located at one or two AS hops away from GÉANT. These edges account for 45% of the total transit traffic. The largest 50 edges go one hop further: 11 of these are at a distance of three AS hops, 46 at two AS hops, and 3 at one AS hop. Most of these larger edges at a distance of two AS hops connect large transit ASs.

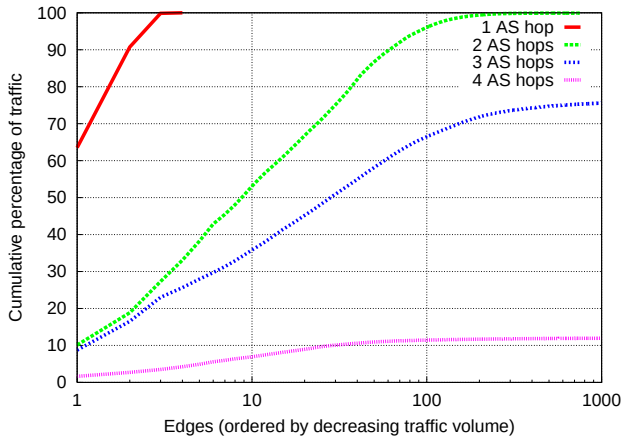

Fig. 3 Traffic on AS topology as a function of AS hop distance from GÉANT.

Another way of visualizing the importance of a few edges at a given AS hop distance from GÉANT is presented on Figure 3. For each AS hop distance, we plot the (cumulative) fraction of the total traffic seen by the largest edges. Note that contrary to Figure 2 where we added the traffic seen by all edges, here we count the total traffic that is received by GÉANT and see what fraction of this traffic is seen by edges at a given AS hop distance. At a distance of one (direct peers), only 3 edges capture almost all the traffic. At a distance of 2 AS hops, the largest edge sees only 10% of the traffic. We already observe at an AS hop distance of 2 how Internet traffic is split across many edges, with the largest 10 edges accounting for more than 50% of the traffic, the largest 100 95% of the traffic. At increasing AS hop distances, more edges are required to capture a given fraction of the traffic, due to the less and less traffic being aggregated on these edges.

On Figure 3, we can also see how much traffic is sent towards destinations located at a given AS hop distance. The difference at the right-end of the cumulative distribution between consecutive curves i and $i+1$ tells what fraction of the total traffic has as destinations ASs located at AS hop distance i. Almost no traffic is sent towards destinations at one AS hop, as neighboring ASs of GÉANT are transit providers, not stub ASs. About 25% of the traffic is sent to destination ASs at a distance of 2, more than 60% to ASs at a distance of 3, and more than 10% at a distance of 4 AS hops. Other studies [22,31,32] have observed a similar distribution of the traffic as a function of AS hop distances.

3.2 AS strength

Section 3.1 showed how traffic is split over edges as a function of AS hop distance, which shows a perspective highly centered on GÉANT. In Section 3.1, important edges were those closer to GÉANT as they concentrate all its traffic. In this section, we want to take a different perspective, more biased towards the ASs that are important in terms of the amount of traffic that transits through them. For this, we rely on a metric called the *strength* s_i of a node i. Strength combines the information about the connectivity of a node and the traffic carried by its *edges* [34,3]:

$$s_i = \sum_j w_{ij} + w_{ji} \quad (1)$$

where w_{ij} denotes the amount of traffic carried by the edge between AS i and AS j. As edges are directed, we sum in equation 1 the traffic for both inbound and outbound edges of vertex i. The $strength$ thus corresponds to all the traffic that enters AS i plus the traffic that leaves it. An AS that receives an amount of traffic x but does not forward it to another AS will have half the strength of an AS that transits the same amount x of traffic. The rationale for this metric is that an AS that transits traffic carries traffic over both inbound and outbound edges, while an AS that only received traffic carries this traffic only on its inbound edges.

Figure 4 plots the (cumulative) distribution of the *strength* as a function of AS degree (dashed curve). On this figure, we do not consider ASs that do not carry traffic from GÉANT. AS degree thus means the number of neighboring ASs with which a given AS exchanges traffic, as seen by GÉANT. To relate strength and AS hop distance, we also plot on Figure 4 the AS hop distance for all ASs of a given degree (crosses).

We observe three different regions in the strength distribution. ASs with a small degree (less than 4) capture a very small fraction of the total strength, while they account for almost 90% of the ASs. Most of these ASs are at a distance of 3 and 4 AS hops from GÉANT. These ASs are typically only destination ASs, i.e. they do not carry transit traffic having as destination another AS. Important ASs in strength do

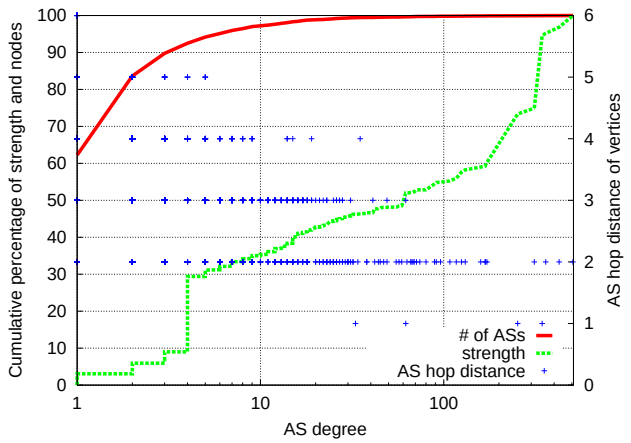

Fig. 4 AS strength as a function of AS degree.

not always have high degree, for example the most popular destination ASs: ASs of degree 4 alone account for about 20% of the total $strength$. For higher degrees, we observe a smoother increase in the $strength$ for degrees from 5 to 200, which account for 30% of the total $strength$. These ASs are at 2 or 3 AS hops away from GÉANT. Finally, a small number (about 10) of high-degree ASs capture about 40% of the strength. These high-degree ASs are the largest providers in the Internet, called tier-1 providers. These tier-1 ASs are relatively close to GÉANT, at a distance of 1 or 2 AS hops. Note that there are about 20 such tier-1 ASs, among which only ten carry a significant amount of traffic sent by GÉANT. The common belief that tier-1 providers are central in the AS topology is confirmed when looking at traffic. Note that our observations do not constitute a proof that the traffic observed by GÉANT is representative of traffic observed by other transit ASs, but they give evidence that the traffic observed by GÉANT reflects the hierarchical structure of the Internet [27].

4 Traffic on the AS topology

After having introduced basic properties related to the global distribution of the traffic of GÉANT on the AS topology, we turn to the study of the dynamics of the traffic. To the best of our knowledge, no model of the dynamics of traffic on the Internet topology exists. Existing Internet traffic models do not take time into account, like the gravity model [13], or do not consider the topology [16,15,19,29,18,33].

Note that several papers in the networking literature [14, 26,25] have relied on a space-time analysis of network traffic, similarly to this paper. Those studies are similar to our work in that they use PCA or Kalman filtering in order to identify abnormal patterns in large traffic datasets. Our work is different as we want to find a way to characterize the dynamics of traffic on all paths that go from the observation network, e.g. GÉANT, towards all destinations to which the observation network sends traffic. [14,26,25] focus on the paths of the traffic observed by some network, but only the part that crosses that network. Our work is the first to try to characterize the space-time dynamics of traffic as seen by a large network provider as it crosses the whole Internet topology.

To understand the dynamics of the traffic on the AS topology, we rely on principal component analysis. Principal Component Analysis (PCA) [11] is now a commonly used multivariate data analysis technique for reducing the dimensionality of a dataset with a large number of interrelated variables. PCA does that by defining a new set of axes that span the whole dimensionality of the data. The axes are ordered by decreasing fraction of the variations they capture from the data. In our context, variations spans the time dimension, while the related variables are the *edges* on the AS topology. We use PCA in order to have a concise description of the dynamics of the traffic on AS edges, that tells us where most of the variability is coming from. Knowing that edges relate to paths, we consider *edges* as dependent variables in PCA while time intervals are independent. Time intervals are in practice not independent, as patterns like the time of the day create dependencies. We will see later that the assumption about the independence of time intervals is partly justified. We leave for further work the study of the dependence between time intervals.

4.1 Background on PCA

We use a matrix representation of the data. Let us call this matrix $\mathbf{X}$. The p variables to be analyzed (e.g. edges) are columns of $\mathbf{X}$: $x_i, i = 1,\ldots,p$. Each variable has n elements (e.g. time intervals), hence $\mathbf{X}$ is a $n \times p$ matrix. In this paper, $n = 1000$ (time intervals) and $p = 9580$ (edges). PCA performs a rotation of this matrix[2] $\mathbf{X}$ such that

$$\mathbf{Y} = \mathbf{A}'\mathbf{X}' \tag{2}$$

where $\mathbf{A}$ is an orthonormal matrix[3]. $\mathbf{Y}$ is the matrix of the rotated data, it is a square matrix of order n. $\mathbf{A}$ is found by constraining the covariance matrix of $\mathbf{Y}$, $\mathbf{C}_Y = \frac{1}{n-1}\mathbf{Y}\mathbf{Y}'$, to be diagonalized. A symmetric matrix can be diagonalized by the orthogonal matrix of its *eigenvectors* so that

$$\mathbf{C}_Y = \frac{1}{n-1}\mathbf{A}\mathbf{\Lambda}\mathbf{A}' \tag{3}$$

where $\mathbf{\Lambda} = \mathbf{X}\mathbf{X}'$. $\mathbf{A}$ is selected so that its columns are the eigenvectors of $\mathbf{\Lambda}$ and the *principal components* of $\mathbf{X}$.

[2] Note that $\mathbf{X}$ must be normalized so as to have zero-mean columns to capture the actual variance.

[3] $\mathbf{A}'\mathbf{A} = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix.

The diagonal elements of $\mathbf{C}_Y$ give the variance of $\mathbf{X}$ along each principal component. We assume in the remainder of this paper that the eigenvectors a_k are always ordered by decreasing value of their eigenvalue λ_k. Successive eigenvectors capture decreasing amounts of the total variance of the dataset. The term *principal component* (PC) refers to the eigenvectors having the largest eigenvalues, hence also capturing most of the variance. Note that we use the terms eigenvalue and eigenvector interchangeably in this paper. We mainly use *eigenvector* when referring to the column of the matrix $\mathbf{A}$.

4.2 Contribution of AS edges to variance

Let us represent traffic on the topology with matrix $\mathbf{X}$. Each column $x_{.j}, j = 1, \ldots, p$ of $\mathbf{X}$ contains the traffic for a given edge j. Each row $x_{i.}, i = 1, \ldots, n$ of $\mathbf{X}$ contains the traffic seen during time interval i. When applying PCA on $\mathbf{X}$, we obtain PCs that are a linear combination of the columns, i.e. edges. Each PC is a linear combination of the edges, i.e. of whole columns of $\mathbf{X}$. If one thinks in terms of the traffic dynamics on edges, a single PC combines all the time information of the edges having a non-zero coefficient. PCs hence do not ignore the dynamics of traffic along time, but rather show the algebraic basis that would capture as much of the variations of the data in as few edge-based axes as possible.

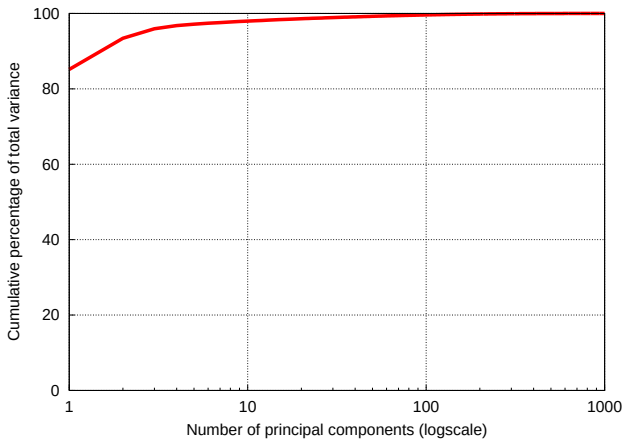

Fig. 5 Percentage of variance captured by edge-based PCs.

Figure 5 shows the cumulative variance captured by the principal components computed on $\mathbf{X}$. The first PC alone captures more than 85% of the total variance contained in the data. The first ten PCs capture 98% of the variance. As each PC is a linear combination of all edges, having few PCs that capture most of the variance does not mean that the data can be simplified. Rather, it suggests that the dependence between edges can be used to summarize the variance of the data.

The coefficients a_{kl} of an arbitrary eigenvector a_k do not allow for a direct interpretation of the edges' contribution to the variance captured by the corresponding PC, since edges are correlated variables. However, the coefficients a_{kl} give an idea of the contribution of each edge to the total *mass* of the PC. We define the *mass* of PC k as the sum of its coefficients $mass_k = \sum_{l=1,\ldots,p} |a_{kl}|$. Some coefficients (i.e. edges) may for instance contribute to a large fraction of the total mass of a PC that captures a large fraction of the variance of the data. In that case, we know that certain edges are more important to explain the variations in the traffic.

We now study the importance each edge plays in the mass of different PCs, as it will give us insight into which parts of the topology are more important to model traffic dynamics on the AS topology. For each PC, only a relatively small number of edges capture a significant fraction of its mass. For each PC, we order edges by decreasing fraction of its mass they capture. For each PC, we compute the fraction of the mass that the top coefficients capture. This allows us to look at the relationship between the largest coefficients (and the corresponding edges) and the fraction of total variance captured. We introduce the contribution of edge l to the total variance, when only the n largest coefficients in absolute value of each PC are considered:

$$contrib_{top_n}(l) = \sum_{k=1,\ldots,n} \frac{\lambda_k}{\sum_k \lambda_k} \times |a_{kl}| \times \mathbf{I}_{top_n(kl)} \qquad (4)$$

where $\sum_k \lambda_k$ is the weighting with respect to PC k (see Section 4.1), a_{kl} is the weighting proportional to the relative contribution to the total *mass* of the PC, and $\mathbf{I}_{top_n(kl)}$ is an indicator function whose value is 1 if coefficient a_{kl} is among the n largest ones in absolute value for eigenvector a_k

$$\mathbf{I}_{top_n(kj)} = \begin{cases} 1, \ if \ \mathrm{a}_{kl} \in top_n(k) \\ 0, \ otherwise \end{cases}$$

and $top_n(k)$ refers to the largest n coefficients of PC k in absolute value. $contrib_{top_n}(l)$ represents the contribution of *edge* l to the *mass* of all PCs, weighted by the relative variance of each PC, counting only those PCs for which the edge has a coefficient among the largest n ones in absolute value. We ignore the contribution to the variance of edges when they do not appear among the largest coefficients of the PCs in absolute value. The reason to introduce such a complex notion of the contribution to variance is that we want to show how different edges contribute to different parts of the total variance.

Figure 6 provides the cumulative value of $contrib_{top_n}(l)$ for the top coefficients $n = 10, 50, 100, 1000, 9580$. We ordered edges on the x-axis of Figure 6 by decreasing traffic

volume to better visualize the largest edges in traffic volume. We use a log-scale for the x-axis because most of the edges only account for a very small fraction of the total variance. The curve for the 10 largest coefficients ("top 10 coefficients" Figure 6) shows that the largest edges in traffic also have the largest coefficients. These edges also appear in the first few PCs, that capture most of the variance. If we restrict ourselves to those 10 largest coefficients, we only capture a bit more than 40% of the traffic variability. Taking into consideration the 100 largest edges in traffic, we capture an additional 20% of the traffic variability. Capturing an increasing fraction of the variance requires considering smaller and smaller edges in traffic, and thus also more and more coefficients of the PCs. Edges having a small amount of traffic do not simply appear in the top coefficients of any PC.

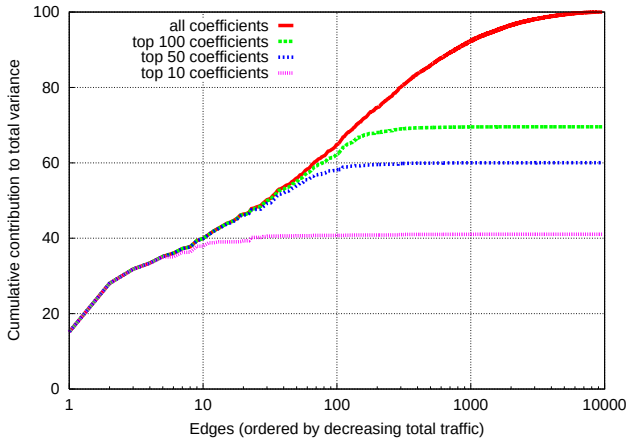

Fig. 6 Contribution of edges to total variance.

We now better understand how PCA was able to drastically reduce the dimensionality of the data: by taking into account almost all edges of the topology. The first few PCs are not sparse, as they must have non-zero coefficients for the small edges to be able to capture such a large fraction of the variance. PCA cannot reduce the size of the topology too much if most of the variance has to be captured. Models that reproduce the traffic on the AS topology will have to consider a relatively large topology, with hundreds or thousands of edges.

4.3 Contribution of time to variance

In Section 4.2, we assumed that the correlated variables to be analyzed were edges. PCA can also be performed on $\mathbf{X}'$, the transpose of $\mathbf{X}$. In that case, the dependent variables of $\mathbf{X}'$ are the time intervals, and the elements of each variable are edges. The eigenvectors a_k are then linear combinations of the time intervals. Similarly to Section 4.2, PCs computed on $\mathbf{X}'$ do not ignore the traffic over edges. Rather, the PCs will try to build an algebraic basis of the data that captures as much of the variations as possible using as few time interval-based axes as possible.

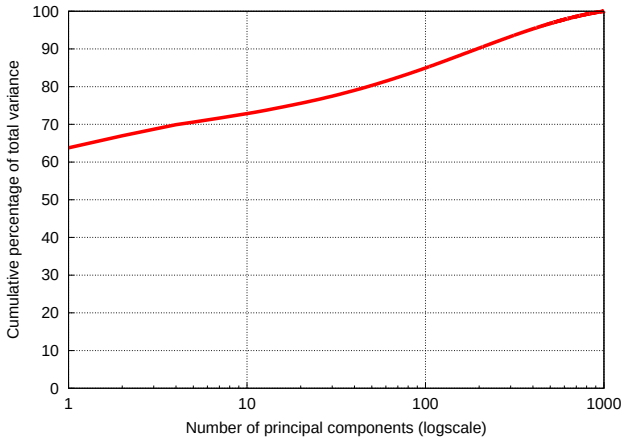

Fig. 7 Percentage of variance captured by time intervals-based PCs.

Figure 7 shows the amount of variance of the dataset captured by the largest PCs when PCA is applied on $\mathbf{X}'$. The first PC captures 64% of the total variance. Additional PCs capture a limited additional fraction of the total variance of the dataset. The first 100 PCs account for 85% of the variance, while when PCA is performed on $\mathbf{X}$ the first PC alone captured this fraction of the variance. This indicates that strong correlations also exist among time periods. However, similarly to Section 4.2, the first few PCs capturing a large fraction of the variance does not mean that a small subset of the time intervals account for much of the variance, as PCs computed from $\mathbf{X}'$ are a linear combination of **all** time intervals.

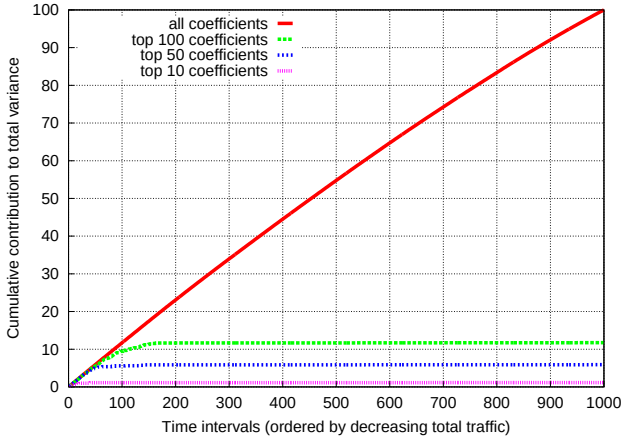

Fig. 8 Contribution of time intervals to total variance.

As $\mathbf{X}'$ is the same data as $\mathbf{X}$, it is not obvious why PCA has more trouble to compress variance into a few PCs compared to Section 4.2. The explanation can be found on Figure 8, the counterpart of Figure 6. Figure 8 provides the values of $contrib_{top_n}(l)$ for the transpose of $\mathbf{X}$. The x-axis of Figure 8 gives the time intervals sorted by decreasing contribution to the total traffic. The coefficients refer to time intervals in this case, so each PC contains 1000 coefficients. As we can readily observe on the curve showing the cumulative variance for all coefficients, each time interval contributes more or less linearly to the total variance. Selecting the top coefficients also does not allow to capture much more than a proportional share of the variance. The ten largest coefficients of any PC capture only about 1% of the total variance. This implies that all time intervals are equally important to capture the variance of the process, i.e. it is not possible to compress the time dimension to capture variations in the data.

Note that the results in the potential sparsity of a_k computed on $\mathbf{X}$ and the impossibility for making the a_k computed on $\mathbf{X}'$ are not contradictory. The possibility of making coefficients of the eigenvectors a_k sparser when PCA is performed on $\mathbf{X}$ is rooted in the fact that a large fraction of the edges do not contribute to a significant fraction of the total traffic (and thus of the variance). The a_k's computed on $\mathbf{X}'$ on the other hand do not allow PCA to select subset of time intervals while still retaining most of the variance. The total number of coefficients that account for most of the variance is pretty much the same in the two cases, about $1,000,000$, compared to the original $n \times p = 1000 \times 9580 = 9,580,000$.

4.4 Variance-traffic volume trade-off

The previous two sections taught us that traffic dynamics on the AS topology can be approximated by reproducing the traffic of the largest edges, while time on the other hand cannot be compressed. In Section 4.2, we did not study in details how edges contribute to individual PCs. Better understanding the contribution of edges within PCs will tell us how to trade-off variance and traffic volume in a model of Internet traffic.

Figure 9 provides the cumulative distribution of the coefficients (i.e. edges) within the mass of the individual PCs. The x-axis of Figure 9 shows edges ordered by decreasing fraction of the mass of the PC they capture. Note that the ordering of edges is different for each PC. Figure 9 gives an idea of the minimal number of edges to be considered to capture a given fraction of the mass of different PCs. The three main PCs have a cumulative curve slightly different from the others. A single edge captures about 20% of the mass of the largest PC. Only ten edges are required to account for about 50% of the mass of the first PC. As already inferred in Section 4.2, several hundreds of edges make the mass of the first

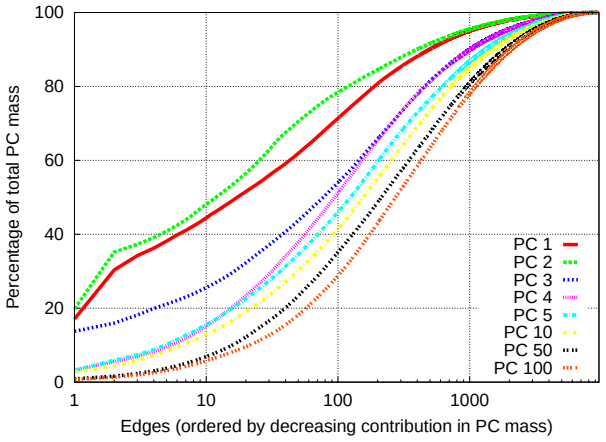

Fig. 9 Fraction of the PC mass captured by edges.

PC. Even though PCA tries to find the linear combination of edges (PC) that captures most of the variance using as PC's as possible, hundreds of edges are necessary to account for a significant fraction of the variance of the traffic.

Assume now that our goal is not to build a model that reproduces traffic variability only, but a traffic-weighted version of variability. For example, instead of reproducing a given fraction of traffic variability on edges, we want to reproduce the variability of most of the traffic. In that case, what we can do is consider only the largest PCs, and find out how many edges among the largest ones in traffic volume will account for a significant fraction of a traffic-weighted mass of the largest PCs. One may have good reasons to use a particular trade-off between capturing variability of the traffic and a given fraction of the total traffic. Choosing to concentrate only on traffic variance as done so far is arbitrary, and may not be desirable for particular traffic models. We show the outcome of such a weighting of the mass of

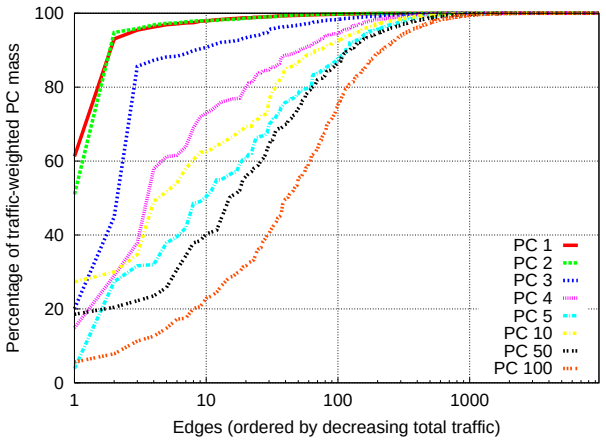

Fig. 10 Fraction of the PC mass captured by edges, weighted by edge traffic.

the PCs by traffic volume on Figure 10. On the x-axis of

Figure 10, edges are ordered by decreasing amount of total traffic, hence the ordering is the same across all PCs. The coefficients a_{kl} of each PC have been multiplied by the total traffic seen by the corresponding edge. The curves of the largest PCs give us a far more optimistic picture for traffic modeling than the previous Figures. The few largest edges already capture a large fraction of the traffic-weighted mass of the largest PC. A similar behavior is observed for the first few PCs. It is thus possible to build models that are both representative of most of the traffic volume, and at the same time capture a significant fraction of the variability of the traffic on the AS topology. For example, if we select the coefficients corresponding to the largest 10 edges in traffic volume, Figure 6 tells us that we will capture 40% of the total variance, and at the same time 45% of the transit traffic on the AS topology. With the largest 100 edges, we obtain about 60% of the total variance and more than 70% of the transit traffic. Capturing larger fractions of the variance or the traffic, as already pointed out earlier, requires to take into account several hundreds of edges.

The curves of Figure 10 exhibit the same kind of behavior as those of Figure 9, with smaller PCs that require more and more edges to capture a given fraction of their traffic-weighted mass. Something that was not apparent on Figure 9 is that the largest edges in terms of traffic consistently capture a large fraction of the traffic-weighted mass of the first PCs, while their contribution is smaller for PCs that rely on smaller edges to build up their mass. This confirms that in order to better approximate both the variance and the total volume of the traffic, edges that carry a small amount of traffic must be taken into account in a model of the traffic on the AS topology.

The observations made so far point out the rather fundamental difference between total traffic volume and dynamics of the traffic on the AS topology. The networking literature mainly takes the viewpoint of capturing most of the traffic [6,9,24,10,31,22]. This is why we often read that popular destinations do provide a view representative of the traffic. While this might be true from the viewpoint of *traffic percentage*, we show in this paper that this view is not valid anymore if the variance of the data is to be taken into account. Even though edges that see a small amount of traffic do not contribute significantly to the total fraction of the traffic, PCA tells us that they are important to faithfully capture the whole variance of the traffic.

5 Traffic over time

In the previous section, we have been looked at whether PCA was able to compress the data into a smaller set of edges or time intervals, while keeping most of the variance and the amount of traffic. In this section, we study explicitly the dynamics of the traffic over time. In Section 5.1 we look back at PCA to understand the impact of the first PCs on the traffic dynamics on edges. We then turn to an explicit multi-resolution analysis of the traffic in Section 5.2.

5.1 Traffic variance and PCA

By building a set of eigenvectors that capture as much of the variance as possible, PCA is approximating the original data. Each successive PC is complementing the previous ones by adding part of the variance of the original data. The successive approximations can be studied by projecting the data onto the first few PCs, which is equivalent to spectral filtering [11]. We thus expect that projecting the original data onto the first PCs affects the traffic on edges.

To project the data on a subset of the eigenvectors, we invert PCA by multiplying the rotated data $\mathbf{Y}$ to the left by $\mathbf{A}$ in which we set to 0 the columns so as to keep only the first m eigenvectors

$$\mathbf{X}_{proj}{}' = \mathbf{A}_{sub}\mathbf{Y} = \mathbf{A}_{sub}\mathbf{A}'\mathbf{X}' \tag{5}$$

where $\mathbf{A}_{sub}$ denotes the matrix $\mathbf{A}$ from which all columns besides those corresponding to the m selected eigenvectors are set to 0. $\mathbf{X}_{proj}$ is the reconstructed matrix of the traffic on edges after the projection on the first m PCs. Note that we work on the original matrix $\mathbf{X}$, so that the PCs on which we project the data are composed of edges. We have seen on Figure 5 that the first few PCs capture most of the variance.

As already mentioned, projecting the data onto the first m eigenvectors is equivalent to spectral filtering [11]. The mean of the traffic over each edge is unchanged after the projection, but the variance of traffic over edges is changed. Figure 11 shows the impact of the projection onto the first $m = 1, ..., 10$ eigenvectors. For each edge and each value of m, we compute the variance of the traffic over each edge, and plot it on Figure 11 as a percentage of the original variance of the traffic on the edge. The top graph of Figure 11 shows the percentage of edges (y-axis) as a function of the variance reduction on edges (x-axis). The bottom graph of Figure 11 shows the percentage of traffic (y-axis) as a function of the variance reduction on edges (x-axis). On the bottom graph, we simply multiply each edge by the total traffic it sees, so that the x-axes of the two graphs refer to the same edges. The top graph shows the view of the fraction of edges, while the bottom graph the view of the fraction of the transit traffic.

From the top graph, we see that more than 80% of the edges see a reduction of their variance of a factor of more than 5, for all values of $m \leq 10$. PCA manages to capture a large fraction of the variance by filtering aggressively the traffic over a large fraction of the edges. The lower graph shows that the large number of edges that undergo a significant reduction of the variance do not account for a large

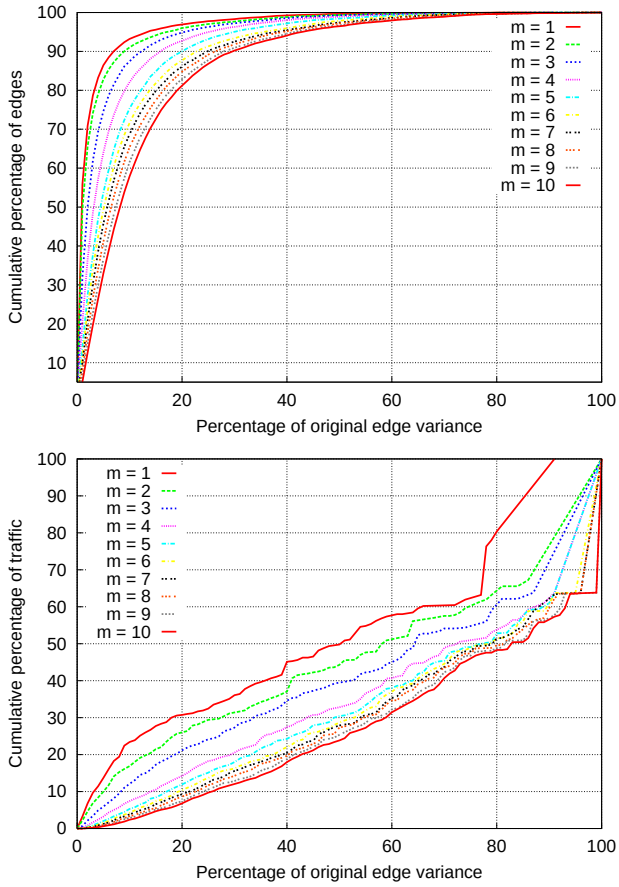

Fig. 11 Edge variance reduction by PCs.

fraction of the traffic. The very few edges that do not suffer from a significant reduction of their variance account for 40% to 50% of the total traffic. The region in the middle of the graphs, corresponding to a few percent of the edges undergo variable reduction of their variance after projection onto the first eigenvectors. These edges account for a significant fraction of the total traffic, which cannot be ignored if most of the traffic has to be taken into account.

We better understand why PCA can hardly compress the data into a small subset of components (edges and time intervals) that would faithfully represent the original data. The first PCs filter out the dynamics of the traffic on the edges that have a limited amount of traffic, and retain traffic dynamics on edges that have a large fraction of the traffic. As already shown on Figure 9, PCs that capture a small fraction of the total variance have most of their mass composed by the contribution from the many edges that do not have much traffic.

5.2 Multi-resolution analysis

Spectral filtering, as performed by PCA, modifies the traffic pattern in order to obtain a compact representation of the data that retains most of its variance. The goal of our work is to reach an understanding of the behavior of observed traffic, that would help to build a model of the traffic dynamics over the Internet topology. While we are still far from this point, we have already made progress in the previous section where PCA told us the limits in summarizing traffic over edges and time. In this section, we break down the variance of the traffic on edges across time scales.

Traffic is known to exhibit periodic patterns on time scales of hours and more [16, 15, 19, 29, 18, 33]. Shorter time scales exhibit scaling properties [2] that we assume, in this paper, irrelevant for a coarse model of the traffic on the AS topology. We leave for further work a more detailed study of those short time scales. The short-term dynamics of Internet traffic stems from the behavior of individual traffic sources and their interactions. These interactions between individual sources may have an effect of the global traffic distribution. However, at the time scales we are considering (more than 15 minutes), it is unclear whether these should be considered at all.

On the other hand, we know that to build parsimonious models of the traffic dynamics on the AS topology, we have to describe the burstiness of the traffic on individual edges of the graph. As the behavior of individual edges might span different types of processes, from smooth periodic ones to ON/OFF types of sources, we need to rely on traffic descriptors that capture a range of processes and are as lightly biased as possible towards a specific kind of behavior. Spectral analysis [20] relies on the assumption of stationarity, which might not be valid on time scales as days [30]. To minimize the risk of bias from non-stationary sources, we rely on wavelet analysis [8].

5.2.1 Wavelet basics

Discrete wavelet signal decomposition [8] consists in analyzing a signal $X(t)$ through a bandpass oscillating function $\psi_{j,k}$ where j represents the time scale and k the time instant. Throughout this paper, small values of j represent the smallest time scales and large values represent the coarsest time scales. By scaling and shifting this function ψ, it is possible to break the signal into its time scale components (at time scale j) and within each time scale along the time axis (at time k):

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k) \tag{6}$$

where the $\psi_{j,k}(t)$ form an orthonormal basis of square integrable functions. Hence any square integrable signal can be approximated by a finite linear combination of the $\psi_{j,k}(t)$.

A scaling and shifting operation by a factor (j, k) moves the original central frequency f_0 of the function ψ to frequency $2^{-j}f_0$ while shifts it in time by a factor of $2^j k$. The

discrete wavelet transform algorithm performs a dyadic tree decomposition (multi-resolution analysis) of the signal

$$X(t) = \sum_k c_X(j_0,k)\phi_{j_0,k} + \sum_{j\leq j_0}^{J}\sum_k d_X(j,k)\psi_{j,k}(t) \quad (7)$$

where the first term represents an approximation of the signal while the second term represents the details, j_0 being the resolution depth, i.e. the coarsest resolution at which the signal is analyzed. The $c_X(j_0,k)$ are called the scaling coefficients and ϕ the scaling function. The $d_X(j,k)$ are called the wavelet coefficients and are of special interest because they are used as a substitute for the increments of the process over the dyadic tree of the time-frequency plane. The orthogonal basis property of ϕ and ψ allows the easy computation of the coefficients by simple inner products with the signal. The nice properties of the wavelet transform make the wavelet coefficients better suited for statistical analysis in comparison to the original increments of the process. In addition, the wavelet transform has a low time complexity in $O(n)$, making it efficient. We are interested in two properties of the wavelet coefficients in this paper:

1. Robustness of the mother wavelet against non-stationarity: the mother wavelet (ψ_0) has a number $N \geq 1$ of vanishing moments $\int t^k\psi_0(t)dt \equiv 0$, $k = 0, ..., N-1$, allowing the wavelet to remove any polynomial trend of order up to $N-1$.
2. Almost decorrelation of the wavelet coefficients: under the assumption of $N \geq H+1/2$, global LRD among the increments of the process can be turned into short-range dependence among the wavelet coefficients [28].

The first property ensures that non-stationarity from polynomial trends is removed from the signal, so that the decomposition of the signal is not biased by long-term trends like those occurring in network traffic. The second property allows to consider the different scales of the analyzed signal as statistically uncorrelated, so that analyzing them separately is possible.

5.2.2 Multi-scale analysis of traffic on edges

Thanks to the de-correlation of the wavelet coefficients across scales, we can analyze the traffic on edges across different time scales. In Section 5.1 we have observed the impact of the first PCs on the variance of traffic on edges. With multi-scale analysis, we can observe how PCA affects the traffic dynamics on edges.

As a reference, we show on Figure 12 the decomposition of the variance of total traffic among time scales. Most of the variance appears at time scales of several hours. Note that as less and less samples are available at larger time scales, it is unclear how much the relative distribution of the variance among the largest time scales (e.g. more than 8 hours) is a true property of the traffic, or rather an artifact of too limited a number of samples. We checked with longer time series of total traffic and found a similar variance decomposition for time scales smaller than 8 hours.

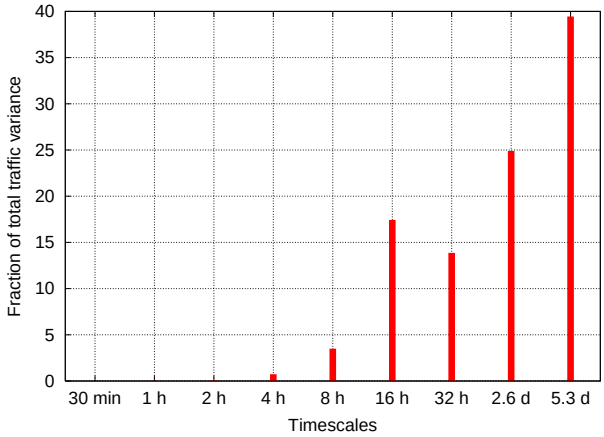

Fig. 12 Decomposition of total traffic variance among time scales.

Figure 13 provides the breakdown of the traffic variance within each edge across the different time scales as computed through the wavelet coefficients. We computed the forward wavelet transform for each edge and computed the fraction of the variance contained at each time scale. The scales go from 1 (30 min) to 9 (about 5 days). For each edge, we stack the relative contribution of each time scale to the total variance of the traffic of this edge, by starting from the smallest time scale and successively stacking larger ones.

The top graph of Figure 13 shows this decomposition for the original traffic (top), for the projection of the traffic on the first 10 PCs (middle) and on the first PC (bottom). The x-axis of Figure 13 gives the edges, ordered by decreasing amount of traffic. Let us first focus on the original data. We have 9 time scales on Figure 13, from 30 minutes (lowest) to 5 days (highest). On the original data, only the largest time scale does not account for a significant fraction of the traffic variance for most edges. All other time scales play a role, which differs from one edge to another. We observe that edges having most traffic (left of the graph) tend to be less bursty than edges that have less traffic. The lowest three time scales (between 30 minutes and 2 hours) account for almost 30% of the variance of edges having low traffic. The same time scales account for less than 10% of the variance of large edges. Large edges have most of their variance at time scale 6 (16 hours). The burstiness of the traffic varies much across edges. A variety of behaviors occur on different edges. This poses a serious challenge for models of the traffic that would have to capture the burstiness of real traffic. Note that observing such a variety of behaviors is not surprising, as many works have debated on the traffic vari-

ability on different types of links. Studies of large backbone links have typically concluded that traffic burstiness tends to a non-stationarity Poisson process as link capacity increases [4]. Studies of smaller links and networks on the other hand have found that scaling processes better describe traffic [16, 15, 19, 29, 18, 33]. The top part of Figure 13 shows that the process that best describes traffic burstiness a given link has much to do with the amount of traffic observed on this link.

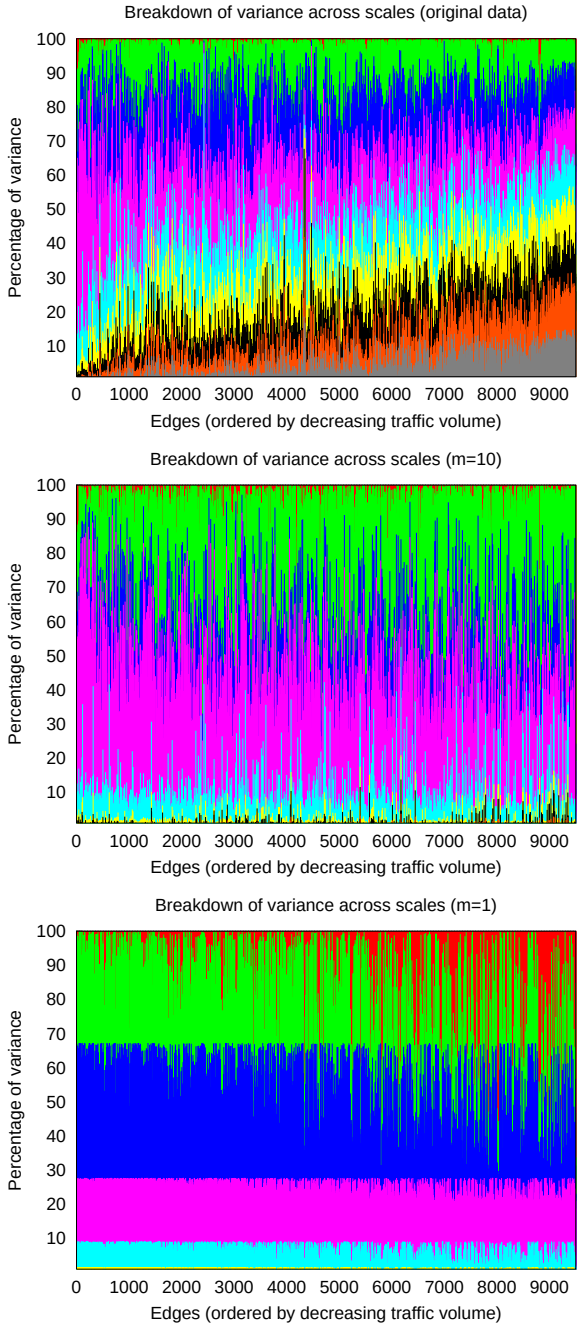

Fig. 13 Decomposition of traffic variance among time scales for different PCs: original data (top), first ten PCs (middle), and first PC (bottom).

To provide evidence about the relationship between the total traffic carried by an edge and its burstiness, we show on Figure 14 the evolution of traffic for three different types of edges: large, medium and small. We observe on Figure 14 that a large edge has relatively smooth traffic, with limited burstiness. Large edges typically have a traffic variability similar to the one observed for total traffic (see Figure 1), with very strong time of the day and day of the week trends. Medium edges are more bursty than large ones, with more variability on time scales of hours. Finally, small edges that do not aggregate the traffic between a large number of end-hosts will exhibit a very irregular traffic pattern, with ON/OFF periods similar to LAN traffic [16, 15]. By comparing different types of edges, we can understand why traffic aggregation makes different types of traffic models relevant. A large edge for instance looks like a non-stationary process, where the time of the day or day of the week play the most important role in traffic variability. Medium edges on the other hand look more similar to a LRD or self-similar process, with highly irregular patterns that are similar at several time scales. Finally, small edges may exhibit any kind of behavior. Some small edges do even consist of a single burst spanning a very small amount of time. Modeling such heterogeneous behaviors is clearly a challenge for any traffic model that wants to look realistic and parsimonious at the same time.

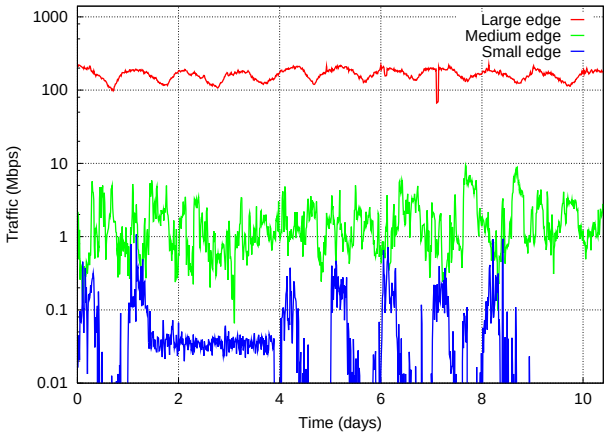

Fig. 14 Variability of traffic on different edges.

If we turn to the breakdown of traffic variance after it has been projected on PCs, as on the middle and bottom parts of Figure 13, we see that PCA is modifying the variance structure of the traffic. The lowest four time scales (from 30 minutes until 4 hours) have almost completely disappeared. PCA is indeed doing spectral filtering of the data, removing the high-frequency components. Projecting data onto the first PCs hence fundamentally modifies the original burstiness of the traffic. Depending on the purpose of a traf-

fic model, reproducing the burstiness of real Internet traffic may or may not be an issue. Projecting the traffic on the first PC (bottom of Figure 13) leads to a variance decomposition that is very similar for most edges. This variance decomposition is not realistic at all, but allows PCA to capture most of the original variance in a simple way.

Using a traffic model that would simplify the variance decomposition in the same way as PCA might not be so problematic. Indeed, one might not be interested in reproducing the full burstiness of the traffic, but only most of the burstiness for the largest fraction of the traffic. If the variance, weighted by the traffic carried on each edge, does not change too much with the projection on the first PCs, we can still claim that the approximation performed by PCA does not impede on the realism of the traffic.

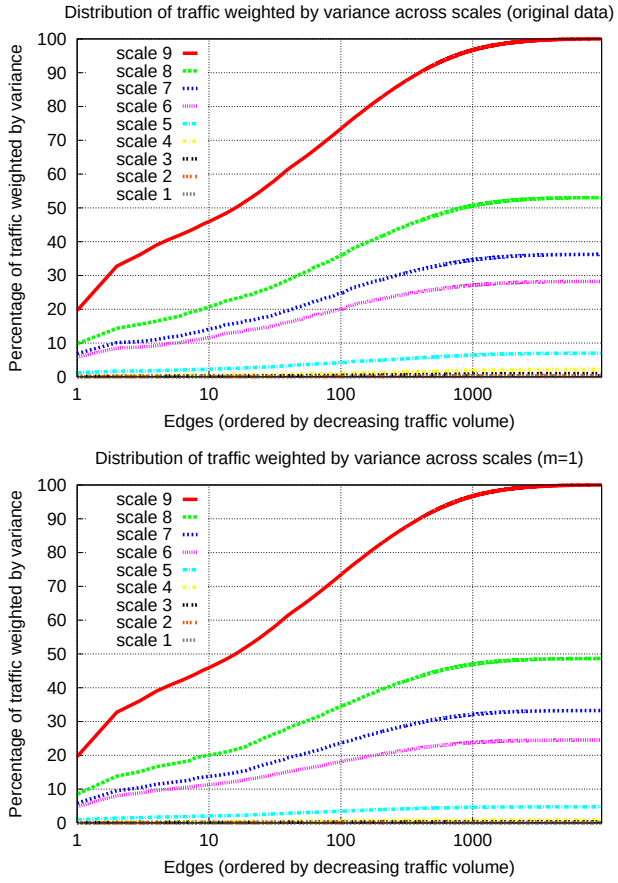

Fig. 15 Traffic-weighted distribution of the variance: original data (top) and projection on first PC (bottom).

Thanks to wavelet decomposition, we can attribute to each time scale the fraction of the variance it represents for each edge. For each time scale, we show on Figure 15 the fraction of the variance, weighted by the traffic of each edge, that is captured by the largest edges. The curves for each scale contain the variance of smaller scales, to make the plots cumulative both in scales and edges. As usual, edges are ordered by decreasing amount of traffic on the x-axis. We observe on Figure 15 that the lower four time scales do not contribute to a significant fraction of the total traffic-weighted variance. For those time scales, even the approximation of the first PC (bottom of Figure 15) does not change much the variance of the traffic pattern, as long as variance is weighted by traffic. Similarly to what we noted in Section 4.4, a pure variance-oriented view of the traffic leads to completely different conclusions as to the realism of an approximation of the traffic dynamics. If we forget about the amount of traffic and want to reproduce burstiness and variance of the traffic on the AS topology, we can hardly reduce the number of edges to be considered. In this section, we are confronted to a similar situation with regard to different time scales. A model that would have to reproduce the burstiness of all edges would require to model several time scales. Weighting variance at each scale by the amount of traffic on the other hand makes it possible to ignore the burstiness at time scales smaller than a few hours.

6 Conclusions and further work

Throughout this paper, we have analyzed the behavior of the traffic observed by a transit network, GÉANT, for a period of more than a week. We have first shown that the distribution of the traffic on the topology seen by a transit network is similar to what has been observed for non-transit networks. This contribution might be helpful to build models of the global distribution of the traffic on the topology.

Then, we have studied the question of how to obtain a concise description of the traffic dynamics on the topology. For this, we have relied on a well-known data analysis technique: principal component analysis. We have found that if we take a viewpoint purely centered on the variance of the data, we could hardly reduce the complexity of the data. All time intervals are equally important, hence cannot be summarized. Edges on the other hand do contain strong correlations. Correlations among edges however do not allow to reduce much the complexity of the data, unless a significant fraction of the variance, of the traffic, or of both variance and traffic, is lost. If variance and traffic are combined by weighting the variance by the amount of traffic on the other hand, it is possible to approximate the traffic dynamics on the topology.

Finally, we have relied on multi-resolution analysis to understand the importance of traffic burstiness at several time scales. We have found that Internet traffic on the topology has a complex signature: traffic on different edges have different burstiness behaviors at different time scales. Traffic burstiness on a given edge seems to be mostly related to the amount of traffic the edge carries: the more traffic on the edge, the more regular the traffic pattern on the edge.

When combining PCA and multi-resolution analysis, we concluded that two simplifications of traffic dynamics can be made: 1) by not taking into account edges that do not capture much traffic, or 2) by ignoring time scales smaller than a few hours. However, those simplifications come at the cost of loosing the original burstiness structure of the traffic or some fraction of the original traffic.

Our findings suggest that a model of Internet traffic on the topology is not out of reach. However, a compact description of the traffic on the topology, and of its dynamics, will require representations of the dependencies between edges on the topology, as well as of the evolution of traffic on different types of edges. To our knowledge, such representations do not exist, but might be built upon existing work like graph wavelets [7].

The findings of this paper are based on a single transit network. Even though we are confident that our findings will apply to other large transit providers, different networks like small stub networks may exhibit a different traffic pattern with different properties. We hope that our work will stimulate more studies of the space-time properties of Internet traffic.

Acknowledgments

The author thanks Nicolas Simar from DANTE and Richard Gass from Intel research for the collection and access to the data from GÉANT. The author also wishes to thank Bingjie Fu from Delft University of Technology and Matthew Roughan from the University of Adelaide for several discussions on the topic. Part of this work was carried while the author was a Postdoctoral Fellow of the Belgian National Fund for Scientific Research (F.N.R.S.) and visiting the University of Adelaide.

References

1. GNU Zebra Routing Suite. `http://www.zebra.org`, 2003.
2. P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, and D. Veitch. The Multiscale Nature of Network Traffic: Discovery, Analysis, and Modelling. *IEEE Signal Processing Magazine*, May 2002.
3. A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. In *Proc. Natl. Acad. Sci. U.S.A*, March 2004.
4. J. Cao, W. Cleveland, D. Lin, and D. Sun. On the nonstationarity of Internet traffic. In *Proc. of ACM SIGMETRICS*, June 2001.
5. Cisco. NetFlow services and applications. White paper, available from `http://www.cisco.com/warp/public/732/netflow`, 1999.
6. K. Claffy, H. Braun, and G. Polyzos. Traffic characteristics of the T1 NSFNET backbone. In *INFOCOM93*, 1993.
7. M. Crovella and E. Kolaczyk. Graph wavelets for spatial traffic analysis. In *Proc. of IEEE INFOCOM*, April 2003.
8. I. Daubechies. *Ten Lectures on Wavelets*. Number 61 in CMBS-NSF Series in Applied Mathematics, SIAM, Philadelphia, 1992.
9. W. Fang and L. Peterson. Inter-AS traffic patterns and their implications. In *IEEE Global Internet Symposium*, December 1999.
10. N. Feamster and J. Rexford. Network-Wide BGP Route Prediction for Traffic Engineering. In *Proc. of ITCOM 2002, Boston, MA*, August 2002.
11. I. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New-York, 2002.
12. L. Kleinrock and W. Naylor. On measured behavior of the ARPA network. In *AFIS Proceedings, 1974 National Computer Conference*, pages 767–780, 1974.
13. J. Kowalski and B. Warfield. Modeling traffic demand between nodes in a telecommunications network. In *Proceedings of ATNAC'95*, 1995.
14. A. Lakhina, M. Crovella, and C. Diot. Diagnosing Network-wide Traffic Anomalies. In *Proc. ACM SIGCOMM*, August 2004.
15. W. Leland, M. Taqqu, W. Willinger, and D. Wilson. On the Self-similar Nature of Ethernet Traffic (Extended Version). IEEE/ACM *Transactions on Networking*, 1994.
16. W. Leland and D. Wilson. High time-resolution measurement and analysis of LAN traffic: Implications for LAN interconnection. In *Proc. of* IEEE INFOCOM'91, 1991.
17. G. Malkin. Traceroute Using an IP Option. RFC 1393, Jan. 1993.
18. K. Park and W. Willinger. *Self-Similar Network Traffic and Performance Evaluation*. Wiley-Interscience, 2000.
19. V. Paxson and S. Floyd. Wide-Area Traffic: The Failure of Poisson Modeling. IEEE/ACM *Transactions on Networking*, 3(3):226–244, 1995.
20. M. Priestley. *Spectral Analysis and Time Series*. Academic Press, 1981.
21. B. Quoitin and S. Uhlig. Modeling the Routing of an Autonomous System with C-BGP. *IEEE Network*, 19(6), November/December 2005.
22. B. Quoitin, S. Uhlig, C. Pelsser, L. Swinnen, and O. Bonaventure. Interdomain traffic engineering with BGP. *IEEE Communications*, May 2003.
23. Y. Rekhter, T. Li, and S. Hares. A Border Gateway Protocol 4 (BGP-4). Internet draft, draft-ietf-idr-bgp4-22.txt, work in progress, October 2003.
24. J. Rexford, J. Wang, Z. Xiao, and Y. Zhang. BGP Routing Stability of Popular Destinations. In *Proc. of ACM SIGCOMM Internet Measurement Workshop*, November 2002.
25. H. Ringberg, A. Soule, J. Rexford, and C. Diot. Sensitivity of pca for traffic anomaly detection. In *Proc. of ACM SIGMETRICS*, pages 109–120, 2007.
26. A. Soule, A. Lakhina, N. Taft, K. Papagiannaki, K. Salamatian, A. Nucci, M. Crovella, and C. Diot. Traffic matrices: balancing measurements, inference and modeling. In *Proc. of ACM SIGMETRICS*, 2005.
27. L. Subramanian, S. Agarwal, J. Rexford, and R. Katz. Characterizing the Internet hierarchy from multiple vantage points. In *INFOCOM 2002*, June 2002.
28. A. Tewfik and A. Kim. Correlation structure of the discrete wavelet coefficients of fractional brownian motion. IEEE *Trans. Inform. Theory*, 38:904–909, 1992.
29. K. Thompson, G. Miller, , and R. Wilder. Wide-area internet traffic patterns and characteristics. IEEE *Network*, 11(6):10–23, 1997.
30. S. Uhlig. Non-stationarity and high-order scaling in tcp flow arrivals: a methodological analysis. *ACM SIGCOMM Comput. Commun. Rev.*, 34(2):9–24, April 2004.
31. S. Uhlig and O. Bonaventure. Implications of Interdomain Traffic Characteristics on Traffic Engineering. *European Transactions on Telecommunications*, January 2002.
32. S. Uhlig, V. Magnin, O. Bonaventure, C. Rapier, and L. Deri. Implications of the Topological Properties of Internet Traffic on Traffic Engineering. In *Proc. of ACM SAC'04*, March 2004.

33. W. Willinger, V. Paxson, R. H. Riedi, and M. S. Taqqu. Long-Range Dependence and Data Network Traffic. *"Long range Dependence : Theory and Applications", edited by Doukhan, Oppenheim and Taqqu*, 2001.
34. S. Yook, H. Jeong, A. Barabasi, and Y. Tu. Weighted evolving networks. *Phys. Rev. Lett.*, 86:5835–5838, 2001.
35. E. Zegura, K. Calvert, and S. Bhattacharjee. How to model an internetwork. In *Proc. of IEEE INFOCOM*, March 1996.