

Temporal Characteristics of Large IP Traffic Flows

Henrik Abrahamsson and Bengt Ahlgren
Swedish Institute of Computer Science

{henrik,bengta}@sics.se

May 5, 2004

SICS Technical Report T2003:27

ISSN 1100-3154

ISRN:SICS-T-2003/27-SE

Abstract

Several studies of Internet traffic have shown that it is a small percentage of the flows that dominate the traffic. This is often referred to as the mice and elephants phenomenon. It has been proposed that this might be one of very few invariants of Internet traffic and that this property could somehow be used for traffic engineering purposes. The idea being that one in a scalable way could control a major part of the traffic by only keeping track of a small number of flows. But for this the large flows must also be stable in the meaning that they should be among the largest flows during long periods of time. In this work we analyse packet traces of Internet traffic and study the temporal characteristics of large aggregated traffic flows defined by destination address prefixes.

1 Introduction

Several studies of Internet traffic have shown that a small number of flows account for the majority of the traffic while most of the flows are small and together only stand for a small fraction of the packets and bytes. This is sometimes referred to as the *mice and elephants phenomenon*. This property of Internet traffic has been found on many different levels of flow granularity. It has for instance been observed for flows between ports on pairs of hosts (e.g TCP or UDP flows) [4, 12], for aggregated traffic flows between networks and autonomous systems [5], for flows defined by destination address prefixes [2, 3], as well as in point-to-multipoint traffic demands in a large IP backbone network [6]. It has been suggested that this property of Internet traffic might be used for implementing scalable mechanisms for traffic engineering and load balancing. The idea being that one could control a large portion of the traffic by only keeping track of a small number of flows. But for this the large elephant flows need to be persistent. The fact that a few flows account for most of the bytes says only something about their volume or average rate

during the investigated time interval. But what are the temporal characteristics of these elephant flows? Are they large all the time? Do they together always stand for a majority of the traffic? Is it common that the elephant flows are classified as small in smaller subintervals? These are the issues that we consider in this report and we are looking at different methods for trying to answer these questions.

We analyse packet traces of Internet traffic and study flows defined by destination address prefixes. We identify the largest flows and investigate their behaviour. The study shows that the large flows can have very different temporal behaviours. In the traffic traces that we have looked at the elephant-flows together usually stand for a major part of the traffic but it is not uncommon for individual flows to be classified as small in many subintervals.

The remainder of the report is organized as follows. Section 2 gives a background on flow measurements and discusses related work. Section 3 describes the methods we use to characterise the temporal behaviour and stability of large flows. The results when applying these methods on traces of Internet traffic are presented in Section 4. The report is concluded with a discussion in Section 5.

2 Background on traffic flow characterisation

There are many different definitions of the term flow and there are several different ways to characterize flows. In general a flow is just a set of packets that have some common properties. This could be micro-flows where all packets have the same source IP address, destination IP address, source port, destination port and protocol id. The common properties could also define much larger flows like all packets with destination addresses that map to the same entry in a routing table. Other ways that flow definitions sometimes differ is whether a flow is defined as all packets (with common properties) that passes a single observation point in the network or if the flow is defined as packets between end-points, if the flow is bidirectional or if we are looking at traffic in only one direction, and whether the flow is everlasting or if it is timed-out after being inactive for a certain time. Depending on how flows are defined they can then be characterised by their volume in number of bytes or packets, by duration in seconds, or by their rate.

Several studies of Internet traffic have shown that there is a small percentage of the flows that dominate the traffic. This is often referred to as the *mice and elephants phenomenon*. This property has been found on many different levels of flow granularity and it has even been proposed to be one of few invariants of Internet traffic. From the mice and elephants somewhat of a zoological trend has followed with researchers naming different types of flows after animals. For instance Brownlee and Claffy [4] classify flows as *dragonflies* and *tortoises* when investigating flow lifetimes.

In this work we start out from the mice and elephants phenomenon and we focus on the characteristics of large aggregated traffic flows.

2.1 Related work

Traffic measurements has been a hot research topic for several years now and a lot of work has been done in this area. Here we present some of the related work that is most similar to ours in that they all to some extent deal with the characteristics and stability of large aggregated traffic flows. The different studies have a lot in common but they also differ from each other and from our work in several ways. Different datasets are investigated. The definitions of flows and of large flows (elephants) are not always the same, and different methods are used for investigating the stability of large flows.

Fang and Peterson [5] analyse packet traces from research and commercial networks and use routing table information to map the packet source and destination IP-addresses to networks and autonomous systems (ASes). They study active flows between pairs of hosts, pairs of networks and pairs of ASes and use a threshold value of 32 seconds to timeout inactive flows. They show that there are a highly non-uniform distribution of traffic among the flows on all aggregation levels. In one hour that they use as an example 9% of the flows between ASes comprise 87% of the traffic. Using a low-pass filter they also study the average sending rate of the individual flows. The average rate is updated every minute for each active flow and the top 10% of the flows with the highest sending rates are identified. For the AS to AS traffic the top 10% of the flows stand for 80-90% of the bytes and packets in each minute. They also investigate the average number of active flows and the change rate (i.e, the number of flows that becomes active or inactive in each second) across all flows and for the top 10% only. From this they draw the conclusion that flows which are sending at high rates tend to last long as well.

Feldmann *et al.* [6] study traffic demands in the AT&T IP backbone network. These demands are point-to-multipoint volumes because a given destination is typically reachable from multiple edge routers and which edge router that is used depend on the intra-domain routing. The authors find that a very few of these point-to-multipoint demands contribute to the majority of the traffic. They investigate how the top demands vary over the day and find that they can have quite different time-of-day behaviour. They also investigate the stability of the traffic demands. This is done by ranking the demands after their size and sorting them into 20 groups where the first group contains the largest 5%, the second group the next 5% and so on. They then investigate how large portion of the demands that change rank and move between the different groups from one time period to another. The results indicate that most rank changes are small and that the largest flows are the most stable.

Bhattacharyya *et al.* [2] study traffic demands in the Sprint IP backbone. From packet traces collected at one Point of Presence (PoP) they investigate traffic streams between (ingress link, egress PoP) pairs. They consider destination address prefixes of different length as the basis for aggregating traffic into streams. All packets that have the same first N bits in their destination IP-address belong to the same pN-stream. They investigate p8, p16 and p24 streams and show that the traffic is dominated by a few high-volume streams and that this elephants and mice phenomenon exists on all three levels of granularity. Further, they study the stability of the streams by investigating how they change in size during the day and by examining the frequency and size of rank changes among all the streams. This is done by dividing the traces into subintervals (of for instance 30 minutes), ranking the streams according to their average bandwidths in each in-

terval and examine the change in rank from one interval to another. They also examine the top 15 elephants and their ranking throughout the day. The results indicate that most rank changes are small and that in general the elephants remain elephants and the mice remain mice. The authors argue that the stability of the elephants makes them well-suited as a basis for load balancing in the backbone.

Papagiannaki *et al.* [9, 10] start out from the mice and elephants phenomenon and point out that the large elephant flows also need to be persistent in time in order to be useful for traffic engineering and load balancing. They investigate different classification schemes for identifying those flows that contribute to a large fraction of the traffic consistently over time. They analyse packet traces from links in the core of the Sprint IP backbone network and they study flows on the granularity of BGP destination network prefixes. This means that all packets with destination addresses that map to the same routing prefix are aggregated into a single flow. They propose two different methods to distinguish the elephant-flows from the mice. The first is called α -constant load where α denote the fraction of total traffic that the elephants should stand for. For instance a 0.8-constant load scheme means that a flow is classified as an elephant if it is among the largest flows that together stand for 80% of the total traffic. The second technique, called *aest*, relies on the fact the flow size distribution is known to be heavy-tailed. A flow is classified as an elephant if it is in the tail of the flow size distribution. These techniques are used to identify the elephants in each 5-minute interval. For each elephant flow the holding time in elephant state is calculated; that is the number of consecutive intervals where it was classified as an elephant. The average holding time is then used as a metric for stability. It turns out that the elephants are quite volatile. There are a lot of flows that are classified as elephants in only a single interval and the average holding times are 20-40 minutes. The authors propose a way of improving this by combining the above methods with a moving average scheme called *latent heat*. This takes the past behaviour of a flow into account and can avoid reclassification due to short bursts and dips. In recent work by Papagiannaki *et al.* [11] these classification schemes are further investigated.

Cecilia Borg [3] analyse traffic data from a link between Japan and the USA and also a packet trace taken at the network at SICS. This is the same dataset that is further investigated in this report. She studies active traffic flows defined by destination address prefixes. The largest flows that together carry 80% of the traffic are defined as elephants. For different interval sizes she investigate the probability that a flows that has been large for x intervals stays large for $x+k$ intervals.

3 Methodology

We analyse packet traces of Internet traffic and study large aggregated traffic flows defined by destination address prefixes. All packets that have the first N bits in the destination address in common are aggregated into the same pN -flow. For instance all packets to 193.x.y.z can be put together into one $p8$ -flow. This is the same type of flow aggregation as used in [2, 3].

The starting point for this work is the mice and elephants phenomenon and the fact that there are a small number of flows that dominate the traffic. Here we define a flow to be an elephant-flow if it is among the largest flows that together stand for more than 80% of the traffic.

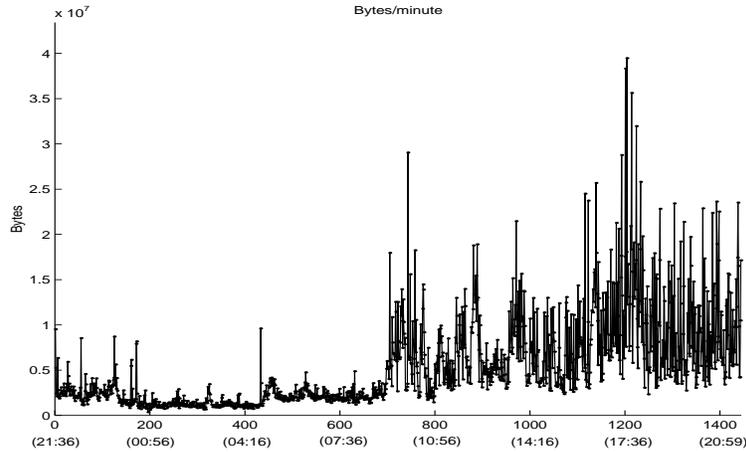


Figure 1: Bytes per minute in the trace taken at SICS 1999-04-14

We identify the elephant flows over a certain time interval, for instance one hour. We then study the behaviour and stability of these large flows by looking at their size in smaller subintervals. We look at how the individual flows vary in size over time and we look at how large fraction of the traffic that the elephant flows stand for together in each subinterval.

We also identify the largest flows in each subinterval. We here continue the zoological trend of naming flows after animals and call these flows *hippos*. This is just to be able to separate them from the elephant-flows (that are the largest when the whole interval is considered). For each subinterval we look at the number of flows (hippos) needed to add up to 80% of the traffic and the relationship between these flows and the elephant-flows.

Finally we look at how common it is that elephant-flows are considered small in the subintervals. For each elephant-flow we count the number of intervals where it is not among the largest flows (the hippos) and present this in a frequency plot.

3.1 Traffic data

Two sets of traffic data are analysed. The first is a 24-hour packet trace taken at SICS. Only external traffic, conversations between machines at SICS and the outside world, was captured using tcpdump. The packet trace was taken in April 1999 and includes more than 21 million packets. Figure 1 shows how the traffic varies during the day. For more information about this trace and the traffic characteristics see [1, 3]. The second trace comes from a traffic archive maintained by the MAWI Working Group of the WIDE Project [8]. This trace was taken in May 1999 on a link between Japan and the USA. The IP addresses in this trace has been anonymised with a modified version of the tool tcpdriv. For more information about this trace and the traffic characteristics see [3, 8].

Having presented the data sets it should as always be noted that one need to be cautious about drawing general conclusions from traffic measurements. The traffic characteristics depends on both when and where we observe the traffic. Backbone links might have other characteristics than access links and the traffic on an access-link that connects a corporate network to the Inter-

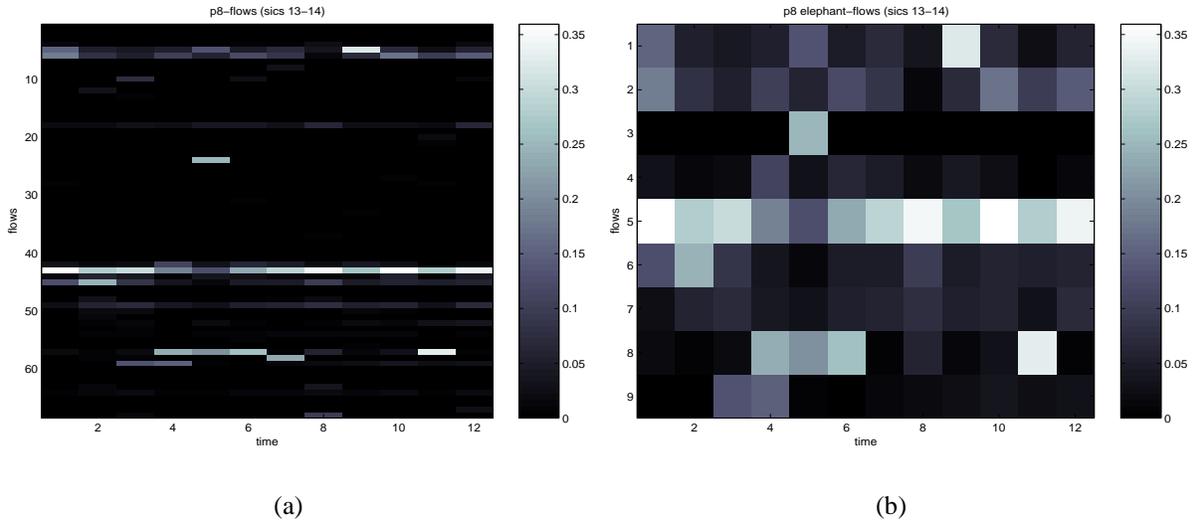


Figure 2: Flow size variations between 13:00 and 14:00 in the SICS trace. (a) All flows. (b) The elephant-flows only.

net might have different characteristics than a link that connects modem users or web servers to the Internet. The traffic also changes with time both in daily and weekly cycles and more unpredictable with new user behavior, new applications and infrastructure. For a further discussion about these challenges see Floyd and Paxson [7].

4 Results

Here we present results from investigating a one-hour interval and a six-hour interval in the SICS trace as well as an example hour in the MAWI traffic data. We have investigated several other time periods with similar results.

4.1 SICS 13:00-14:00

This section shows some results from investigating the hour between 13:00 and 14:00 in the trace taken at SICS. Only destinations outside SICS were considered and all traffic to destinations with addresses that have the same first 8 bits were aggregated into a single p8-flow. The total number of flows during this hour were 68. The largest flows during this hour were identified and using the 80 percent limit for identifying elephants there were 9 elephant-flows. These 9 largest flows (13 percent of the flows) stood for 82% of the total number of bytes during this hour. The largest flow stood for more than 27% of the bytes and the smallest elephant-flow (that is needed to add up to 80%) stood for just above 3% of the traffic.

The investigated hour was also divided into 5 minute intervals and then further analyzed. Figure 2 shows how the relative sizes of the flows vary during this hour. Figure 2(a) shows all 68

p8-flows and figure 2(b) shows only the nine elephant-flows. There is one row for each flow and the x-axis shows time where the hour has been divided into twelve five-minute intervals. The color-scale represents the relative size of the flows. As an example, in figure 2(b) flow number 5 stands for 36% of the bytes during the first 5 minutes, just under 30% in the next interval and so on. Figure 2(b) shows that the elephant-flows can have very different temporal behaviours. This is also shown in more detail for three of the elephant-flows in figure 3. The graphs to the left in this figure show the number of bytes sent in each five minute interval and the graphs to the right show percentage of the total number of bytes in each interval. The top flow has a very smooth rate during this hour while the middle one fluctuates a lot, sometimes contributing to 20-30% of the traffic and in other periods only standing for a few percent of the bytes. The bottom flow is almost zero most of the time but it has one large burst of traffic where it stands for 25% of the bytes. In this case that is also enough to be among the largest flows when the whole hour is considered.

We have now seen that the individual elephant-flows can vary a lot in size. The next question is how large part of the traffic that the elephant-flows stand for together during the investigated hour. Figure 4 shows the total traffic sent by the elephants-flows. Figure 4(a) shows the percentage of the total traffic that the elephant-flows stand for and figure 4(b) shows the number of bytes sent by the elephant-flows in each interval. Each bar represents a 5 minute subinterval and each colour represents an individual elephant-flow. The total contribution from the elephant-flows varies between 62% and 92% of the traffic in the 5 minute intervals during this particular hour. Note that figure 4(a) only shows the relative size of the elephant contributions and that the average of the twelve bars does not have to be 80%. One could imagine an extreme example where the contribution from the elephant-flows are zero in all but one subinterval and then just above 80% in the last interval.

So far we have only been looking at the flows that are among the largest when the whole hour is considered (the elephants). We also identified the largest flows in each 5-minute interval. The largest flows that together send more than 80 percent of the traffic in a 5-minute subinterval are here called *hippos*. Figure 5(a) shows the number of flows (hippos) needed to add up to 80% of the traffic in each 5-minute subinterval in the hour between 13:00 and 14:00 in the SICS trace. For each subinterval we then compared the largest flows (the hippos) with the flows that are among the largest when the whole hour is considered (the elephants). The result of this comparison is also shown in figure 5(a). The *hippos that are not elephants* are flows that dominate the traffic in a subinterval but are not among the largest when the whole hour is considered. The graph labeled *elephants that are not hippos* shows the number of elephant-flows that are not among the largest flows in a certain 5-minute subinterval. For example, in the first 5-minute interval there were 4 flows that together stood for more than 80% of the bytes (hippos). These 4 flows were all also among the largest flows during the whole hour (elephants). Since there were 9 elephant-flows in this hour this also means that 5 of these were not among the largest flows in the first subinterval.

Finally we looked at how common it is that elephant-flows are considered small in the subintervals. For each elephant-flow we counted the number of intervals where it was not among the largest flows (the hippos). The frequency is shown in figure 5(b). The first bar means that there was only one elephant-flow that never was small (always a hippo). The second bar means that

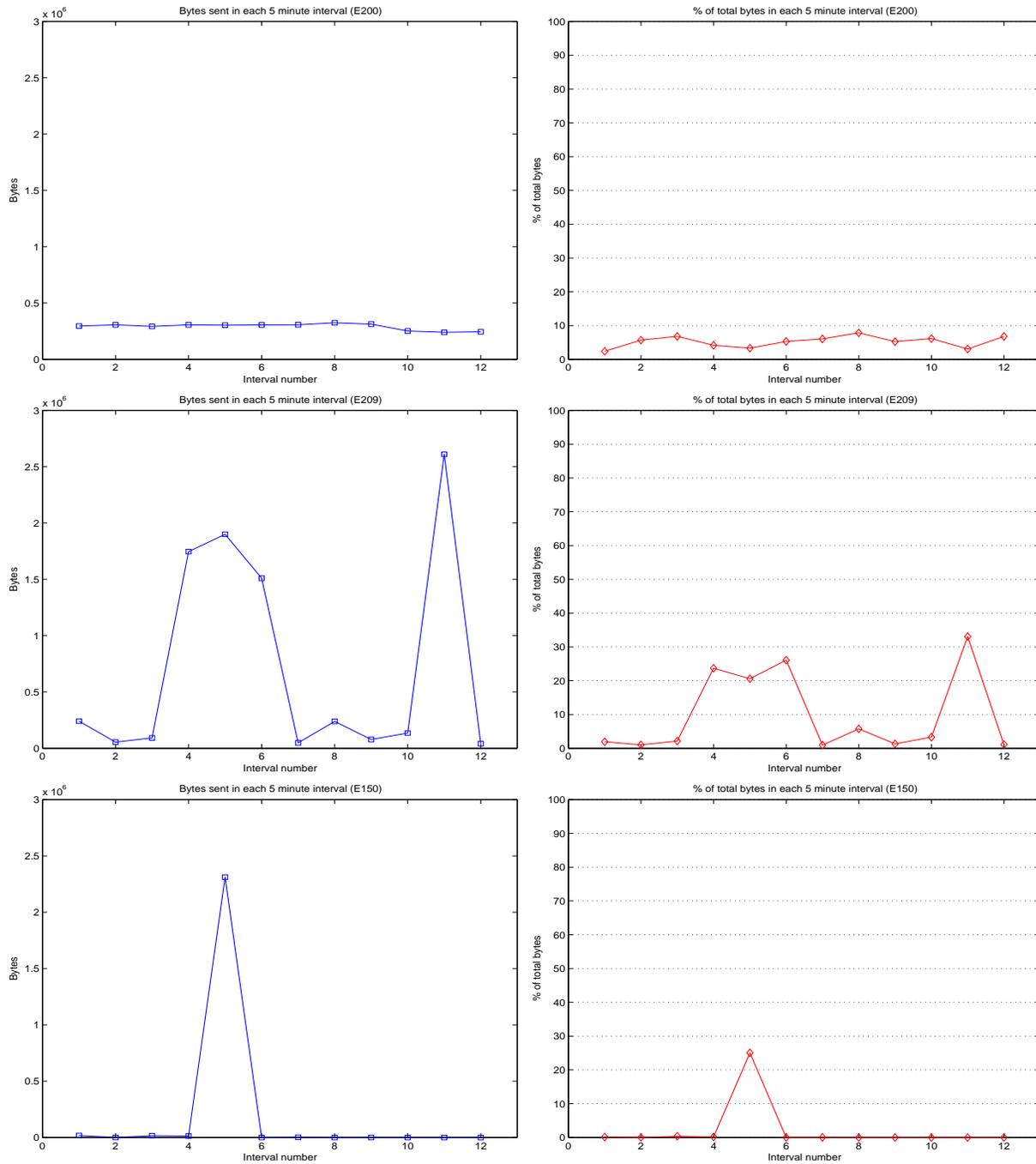


Figure 3: Example of three different temporal behaviours

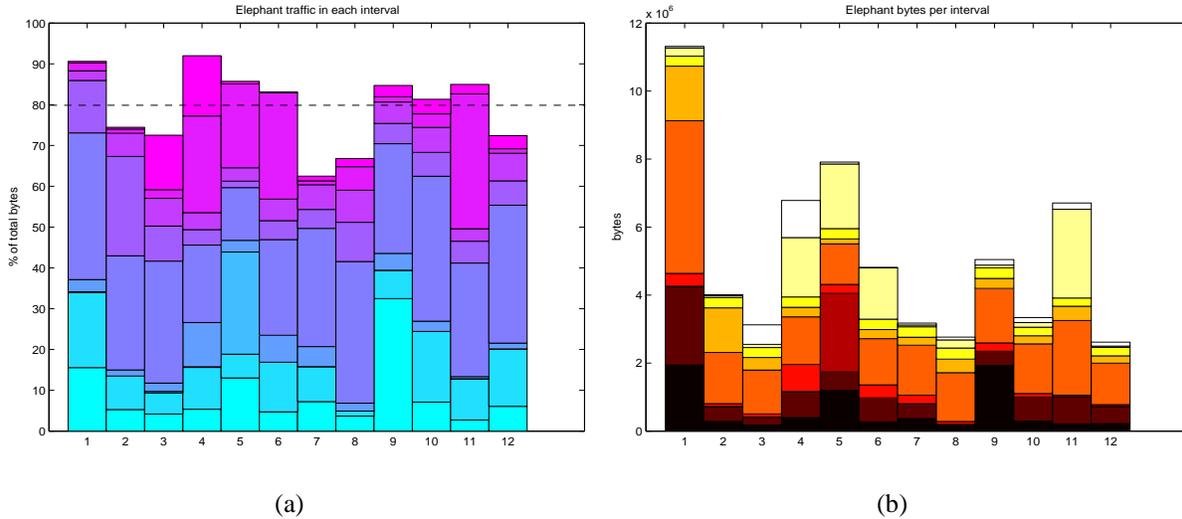


Figure 4: The SICS trace 13:00-14:00. Total elephant-flow traffic in 5 minute intervals. (a) Percentage of all traffic. (b) Bytes.

there were two elephants that were small in one interval and so on.

4.1.1 SICS 13:00-14:00 p16-flows

We also looked at the characteristics of p16-flows during the hour between 13:00 and 14:00 in the SICS trace. Here all traffic to destinations with addresses that have the same first 16 bits in common are aggregated into a single flow. There were a total of 574 p16-flows during this hour and the largest 24 flows stood for just above 80% of the traffic. So 4% of the flows stood for 80% of the traffic.

Some characteristics of the elephant-flows are presented in figure 6. Figure 6(a) shows the size of each elephant-flow as fraction of total traffic and figure 6(b) shows that the total contribution from the elephant-flows varies between 67% and 90% of the traffic during this particular hour. We also identified the largest p16-flows in each 5-minute interval. Figure 6(c) shows the number of flows (hippos) needed to add up to 80% of the traffic in each subinterval and the relationship between these flows and the elephant-flows. The frequency plot in figure 6(d) shows that, as for the p8-flows, it is common that elephant-flows are classified as small in many subintervals. There are only a few flows that always are among the largest.

4.2 SICS 09:00-15:00

In the previous section we identified the elephant-flows in a one-hour interval and then analysed the traffic flows further in 5-minute subintervals. Here we look at a six hour period and analyse the hours between 09:00 and 15:00 in the SICS trace. As before only destinations outside SICS were considered and all traffic to destinations with addresses that have the same first 8 bits were

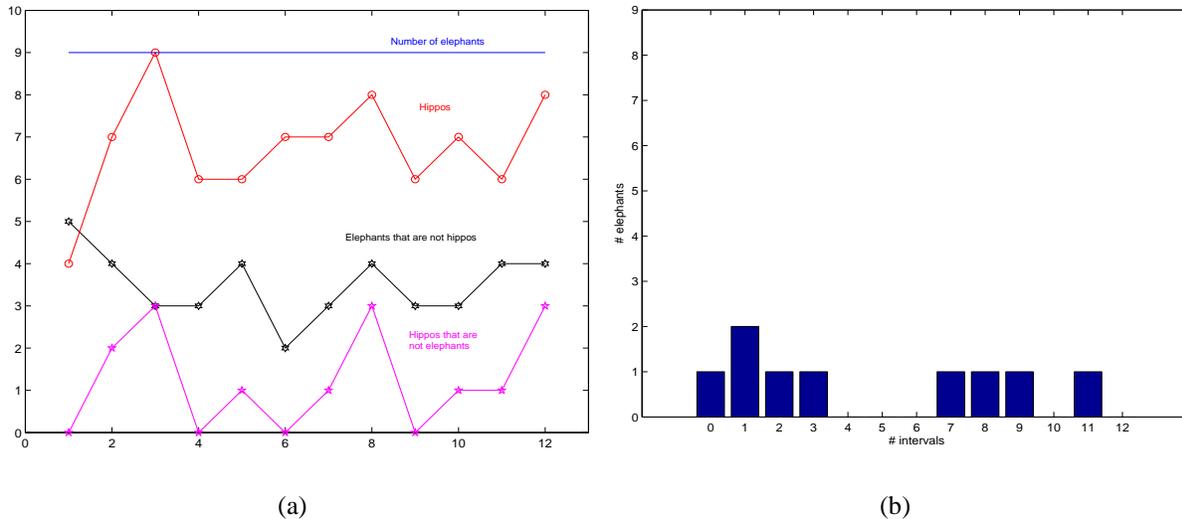


Figure 5: The SICS trace 13:00-14:00. (a) Number of flows (hippos) needed to add up to 80% of the traffic in each 5-minute subinterval and the relationship between these flows and the elephant-flows. (b) Frequency plot displaying how often the elephant-flows were classified as small.

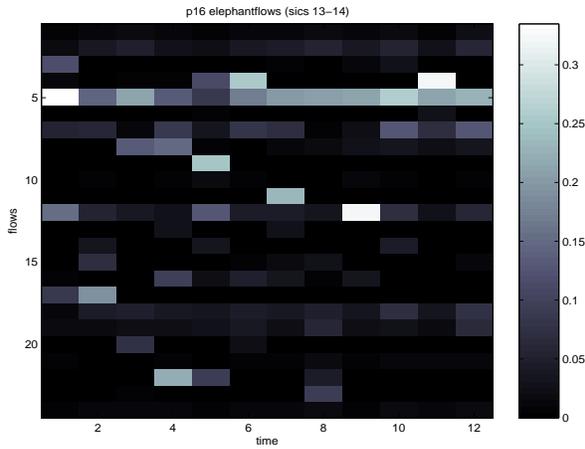
aggregated into a single p8-flow. The total number of flows during this time period were 83 and with an 80 percent limit for identifying elephants there were 8 elephant-flows. Here 10% of the flows stood for 81% of the bytes. The six-hour period was also divided into 30 minute subintervals and further analyzed. Some characteristics of the large traffic flows are presented in figure 7. Figure 7(a) shows the size of each elephant-flow as fraction of total traffic in each 30-minute interval. There is one p8-flow that dominates the traffic in this time period sometimes contributing to up to 55% of the bytes. Figure 7(b) shows how large part of the traffic that the elephant-flows stood for together during the investigated hours. We also identified the largest flows in each 30-minute interval. Figure 7(c) shows the number of flows (hippos) needed to add up to 80% of the traffic in each 30-minute subinterval. Figure 7(d) shows a frequency plot displaying in how many intervals the elephant-flows were not among these largest flows.

4.3 MAWI 20:00-21:00

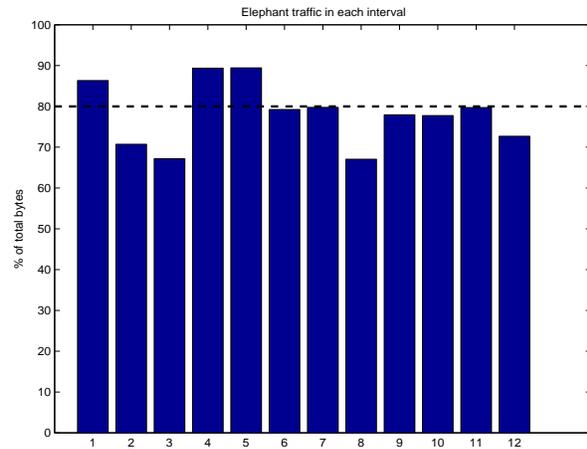
This is some example results from the analysis of the MAWI trace¹. The first hour of the trace (20:00-21:00) was divided into 5-minute intervals and then analysed. As before the traffic was aggregated into p8-flows based on destination address-prefixes. The total number of flows during this hour were 132. As many as 57 flows were needed to add up to the 80% limit used for identifying the elephants. This means that 43% of the flows stands for 80% of the traffic.

Figure 8(a) shows the size of each elephant-flow as fraction of total traffic and figure 8(b) shows the total contribution from the elephant-flows. The number of flows (hippos) needed to

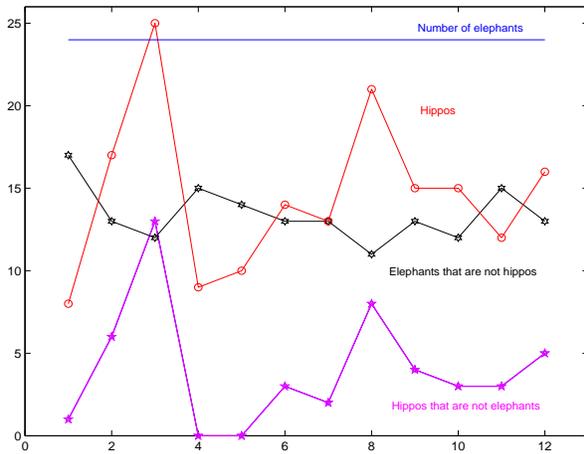
¹<http://tracer.csl.sony.co.jp/mawi/samplepoint-B/1999/>.



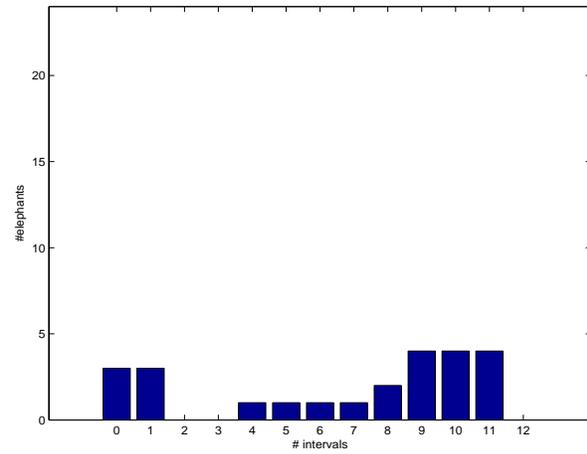
(a)



(b)

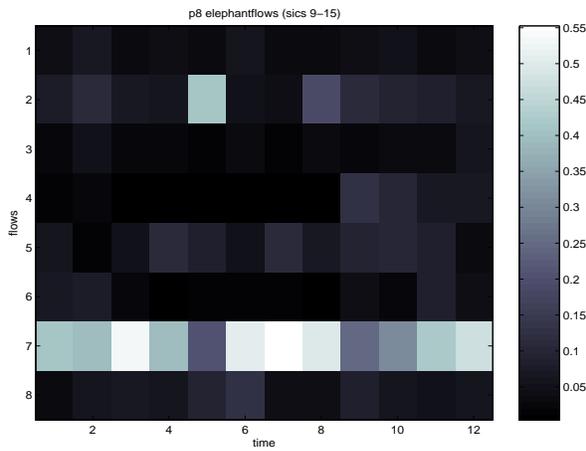


(c)

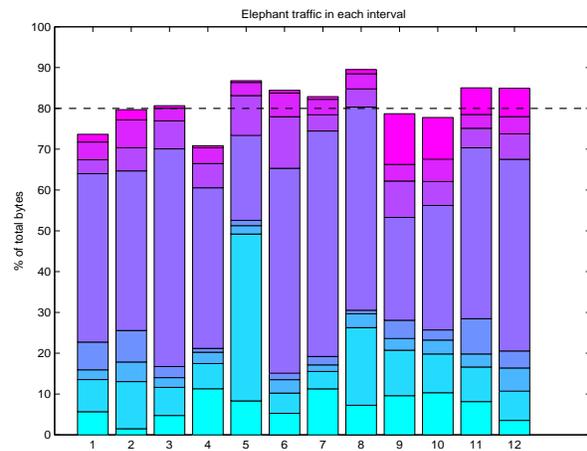


(d)

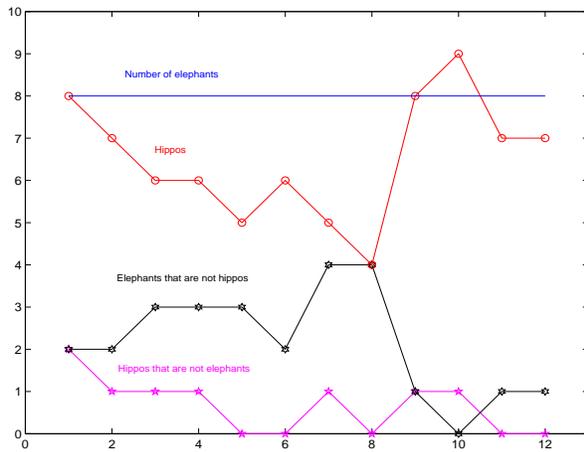
Figure 6: P16-flows in the SICS trace 13:00-14:00. (a) Size of each elephant-flow as fraction of total traffic in each 5-minute interval. (b) Total elephant-flow traffic as percentage of all traffic in 5-minute intervals. (c) Number of flows (hippos) needed to add up to 80% of the traffic in each 5-minute subinterval and the relationship between these flows and the elephant-flows. (d) Frequency plot displaying how often the elephant-flows were classified as small.



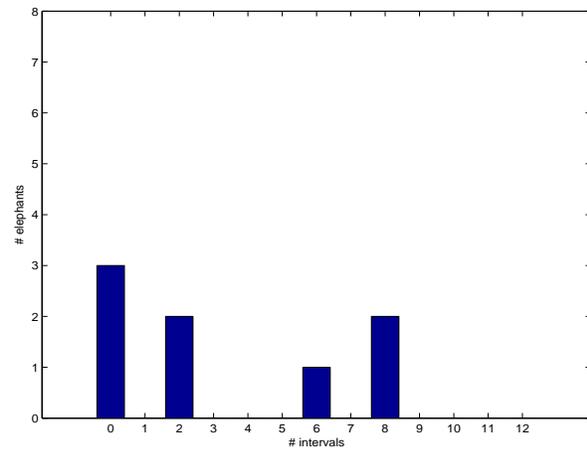
(a)



(b)



(c)



(d)

Figure 7: The SICS trace 09:00-15:00.(a) Size of each elephant-flow as fraction of total traffic in each 30-minute interval. (b) Total elephant-flow traffic as percentage of all traffic in 30-minute intervals. (c) Number of flows (hippos) needed to add up to 80% of the traffic in each 30-minute subinterval and the relationship between these flows and the elephant-flows. (d) Frequency plot displaying how often the elephant-flows were classified as small.

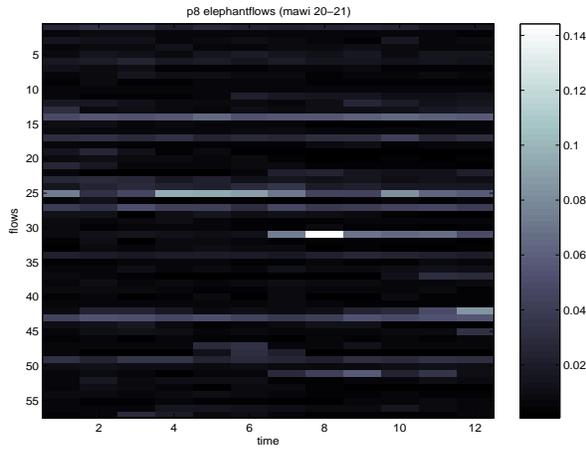
add up to 80% of the traffic in each subinterval and the relationship between these flows and the elephant-flows is shown in figure 8(c). The frequency plot in figure 8(d) shows that most of the elephant-flows are among the largest flows in all subintervals.

This traffic data does not really show the mice and elephants phenomenon since 43% of the flow are needed to get 80% of the traffic. The results are similar for other investigated timeperiods in this trace. It could be the case that the traffic on this link has other characteristics than what we have seen before. But another possible explanation could be that this is a consequence of the anonymisation of the addresses in the packet trace.

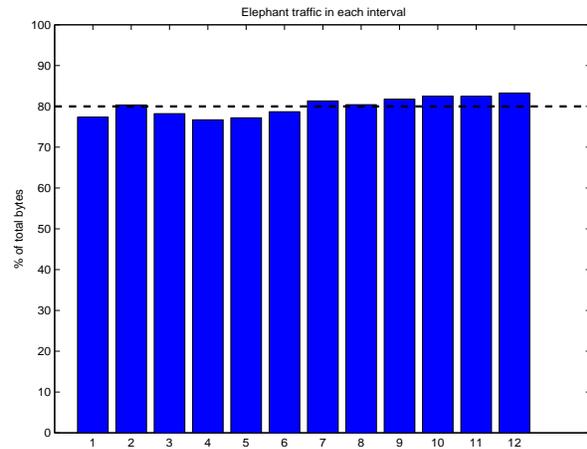
5 Discussion

Starting from the mice and elephants phenomenon we have in this work looked at some temporal characteristics of large traffic flows defined by destination address prefixes. We defined a flow to be an elephant if it is among the largest flows that together stand for more than 80% of the traffic and we have looked at how the elephant-flows vary in size over time both individually and all together. In the traffic data from SICS the mice and elephants phenomenon is evident with a few flows dominating the traffic. Here the elephant-flows together usually stand for a major part of the traffic but the individual flows can have very different temporal characteristics and it is not uncommon for individual flows to be classified as small in many subintervals.

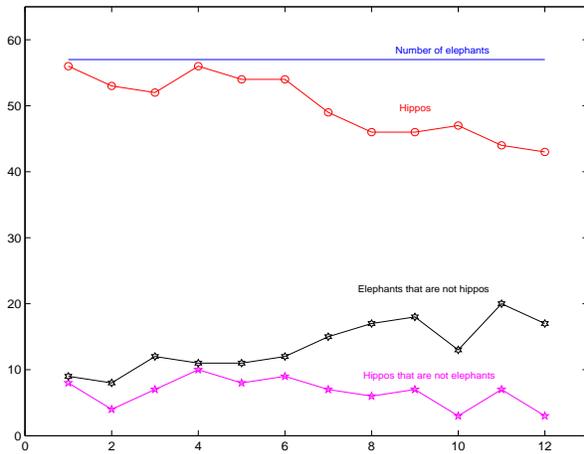
As was pointed out in section 2 there are several different ways to study the characteristics and stability of large traffic flows. Different definitions of flows and of elephant-flows can be used, there are different methods for describing the stability of large flows and the results might differ depending on what datasets that are investigated. The Internet is a moving target and the traffic characteristics depend on both when and where you measure. It would therefore be interesting as future work to apply our methods to more recent data sets and also packet traces from backbone links. Related to this is also access to routing tables and the definition of flows. In this work we defined flows based on artificial destination address prefixes. All packets that had the first N bits in the destination address in common were aggregated into the same pN -flow. If routing tables are collected together with the traffic traces it would be better to define a flow as all packets that map to the same BGP routing table prefix. For future work it might also be interesting to consider other ways of classifying flows as elephants. The 80% limit for identifying elephant flows is simple and frequently used but it is also somewhat arbitrary. A value of 70% or 90% would have been equally good. Defining the elephants as a fraction of total traffic also sometimes make the results and the notion of stability harder to interpret. For instance, a flow can have a constant rate but yet be unstable in the meaning that it fluctuates between being classified as an elephant and as a mouse, because other traffic flows vary in size. Therefore it could be interesting to look at definitions that are independent of other traffic and always classify a flow as an elephant if it has a rate larger than x Kbps or uses $y\%$ of the link capacity over a certain time interval. From a traffic engineering point of view it also makes sense to somehow take the link capacity or utilisation into account. If there is very little traffic on a link then it is of course of less importance to identify the largest flows that stand for 80% of this traffic.



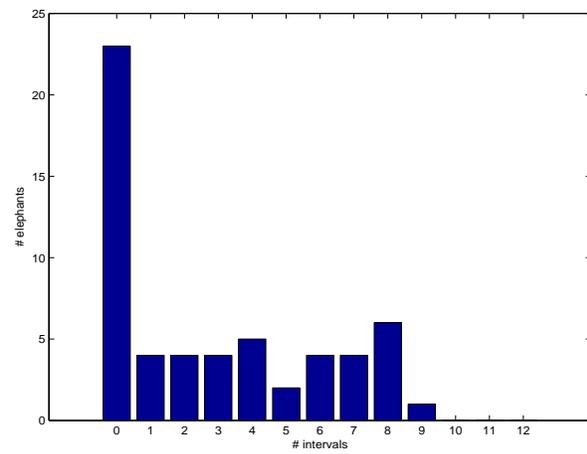
(a)



(b)



(c)



(d)

Figure 8: The Mawi-trace 20:00-21:00. (a) Size of elephant-flows as fraction of total traffic in each 5-minute interval. (b) Total elephant-flow traffic as percentage of all traffic in 5-minute intervals. (c) Number of flows (hippos) needed to add up to 80% of the traffic in each 5-minute subinterval and the relationship between these flows and the elephant-flows. (d) Frequency plot displaying how often the elephant-flows were classified as small.

References

- [1] H. Abrahamsson. Traffic measurement and analysis. Technical Report T99:05, SICS – Swedish Institute of Computer Science, September 1999.
- [2] S. Bhattacharyya, C. Diot, J. Jetcheva, and N. Taft. Pop-level and access-link-level traffic dynamics in a tier-1 PoP. In *Proceedings of ACM SIGCOMM Internet Measurement Workshop*, San Francisco, USA, November 2001.
- [3] Cecilia Borg. Existence, identification and stability of elephant flows in IP traffic. Technical Report T2002:13, SICS – Swedish Institute of Computer Science, August 2002. Master thesis.
- [4] N. Brownlee and kc Claffy. Understanding internet traffic streams: Dragonflies and tortoises. *IEEE Communications Magazine*, 40(10):110–117, October 2002.
- [5] W. Fang and L. Peterson. Inter-AS traffic patterns and their implications. In *Proceedings of IEEE Globecom: Global Internet*, Rio de Janeiro, Brazil, December 1999.
- [6] A. Feldmann, A. Greenberg, C. Lund, N. Reingold, J. Rexford, and F. True. Deriving traffic demands for operational IP networks: Methodology and experience. In *Proceedings of ACM SIGCOMM'00*, Stockholm, Sweden, August 2000.
- [7] S. Floyd and V. Paxson. Difficulties in simulating the Internet. *IEEE/ACM Transactions on Networking*, 9(4):392–403, August 2001.
- [8] Mawi working group traffic archive. <http://tracer.csl.sony.co.jp/mawi/>.
- [9] K. Papagiannaki, N. Taft, S. Bhattacharyya, P. Thiran, K. Salamatian, and C. Diot. On the Feasibility of Identifying Elephants in Internet Backbone Traffic. Sprint ATL Research Report RR01-ATL-110918, Sprint ATL, November 2001.
- [10] K. Papagiannaki, N. Taft, S. Bhattacharyya, P. Thiran, K. Salamatian, and C. Diot. A Pragmatic Definition of Elephants in Internet Backbone Traffic. In *ACM SIGCOMM Internet Measurement Workshop*, Marseilles, France, November 2002.
- [11] K. Papagiannaki, N. Taft, and C. Diot. Impact of flow dynamics on traffic engineering design principles. In *Proceedings of IEEE Infocom*, Hong Kong, March 2004. to appear.
- [12] A. Shaikh, J. Rexford, and K. G. Shin. Load-sensitive routing of long-lived IP flows. In *Proceedings of ACM SIGCOMM'99*, Cambridge, Massachusetts, USA, 1999.