A Pragmatic Definition of Elephants in Internet Backbone Traffic

K. Papagiannaki, N. Taft, S. Bhattacharyya, P. Thiran, K. Salamatian, C. Diot

I. INTRODUCTION

Studies of the Internet traffic at the level of network prefixes, fixed length prefixes, TCP flows, AS's, and WWW traffic, have all shown that a very small percentage of the flows carries the largest part of the information. This behavior is commonly referred to as "the elephants and mice phenomenon".

Traffic engineering applications, such as re-routing or load balancing, could exploit this property by treating elephant flows differently. In this context, though, elephants should not only contribute significantly to the overall load, but also exhibit sufficient persistence in time. The challenge is to be able to examine a flow's bandwidth and classify it as an elephant based on the data collected across all the flows on a link. In this paper, we present a classification scheme that is based on the definition of a separation threshold, that elephants have to exceed. We introduce two single-feature classification schemes, and show that the resulting elephants are highly volatile. We then propose a two-feature classification scheme that incorporates temporal characteristics and show that this approach is more successful in isolating elephants that exhibit consistency - thus making them more attractive for traffic engineering applications.

II. METHODOLOGY

We use packet traces and BGP tables collected in the core of the Sprint's Tier-1 IP backbone network. We present results from two different OC-12 links, from one PoP on the east and the west coast of the USA respectively. Similar results have been obtained on other OC-12 and OC-48 links. The links used are two hops away from the periphery of the network, and traffic is captured on its way towards the core. Therefore, traffic towards specific destinations should exhibit a sufficient level of aggregation.

Since our intended application is intra-domain traffic engineering, we chose as the flow granularity the one of the BGP destination network prefix. Our methodology focuses on the identification of those flows that contribute high volumes of traffic consistently over time. Let i denote the index of a network prefix flow, i.e., a BGP routing table entry. Let τ denote

K. Papagiannaki, N. Taft, S. Bhattacharyya, and C. Diot are with Sprint ATL, {dina, nina, supratik, cdiot}@sprintlabs.com. P. Thiran is with EPFL, Switzerland, Patrick.Thiran@epfl.ch. K. Salamatian is with LIP6, France, K.Salamatian@lip6.fr. K. Papagiannaki is also with UCL, UK.

the length of the time interval over which measurements are taken. Time is discretized into these intervals, and n is the index of time intervals. We define $X_i(n)$ to be the average bandwidth of the traffic destined to a particular network prefix i during the nth time slot of length τ. We use 5 minutes as our default measurement interval. Similar results were obtained for $\tau = 1$ min, and $\tau = 30$ mins.

Single Feature Classification Our methodology consists of two phases: 1) threshold detection phase, and 2) threshold update phase. In the first phase, we calculate a bandwidth value $v(n)$ that separates the high volume flows at interval n. This value $v(n)$ is likely to change with time always isolating those flows that contribute the highest amount of information in each time interval. In order to use this threshold to detect elephants in the next interval $n + 1$ we calculate a new threshold value $\hat{v}(n + 1)$, according to $\hat{v}(n + 1) = (1 - \beta) * \hat{v}(n) + \beta * v(n)$. We found that a value of $\beta = 0.2$ leads to a sufficiently smooth $\hat{v}(n + 1)$.

We propose two different techniques to identify the initial threshold value, namely the "aest" and the "α-constant load" approach. The "aest" approach takes into account the heavy-tail nature of the flow-bandwidth distribution, as observed in the collected data. It sets the threshold value in such a way so that a flow is characterized as an elephant, only if it is located in the tail of the flow bandwidth distribution. Using the *aest* test [1] we identify the points in the flow bandwidth distribution that follow a power-law, and we set $v(n)$ equal to the first point after which such behavior can be witnessed. The "α-constant load" technique requires the setting of an input parameter α corresponding to the fraction of total traffic we would like to place in the elephant class. The threshold is set in such a way that all the flows exceeding it account for the chosen fraction of total traffic. For more details please refer to [2].

The length of time that an elephant remains an elephant is both a function of the flow itself and of the classification. It is a function of the classification in the sense that a particular high-volume flow will remain in the elephant class as long as the continually adjusting threshold stays lower that its average bandwidth. Note that the classification scheme proposed induces the following underlying two-state process on each flow; $I_i(n) = 1$ if $i \in C_1$ (elephant class), and $I_i(n) = 0$ if $i \in C_2$ (mouse class). At each classification time interval, the process either transitions to the other state or stays in the same state.

Based on this process $I_i(n)$, we compute for each i the average holding time in the elephant state during the five hour busy period. Our results indicate that elephant flows maintain their state for surprisingly short periods of time; their average holding time is 20-40 minutes. Moreover, more than 1000 flows in

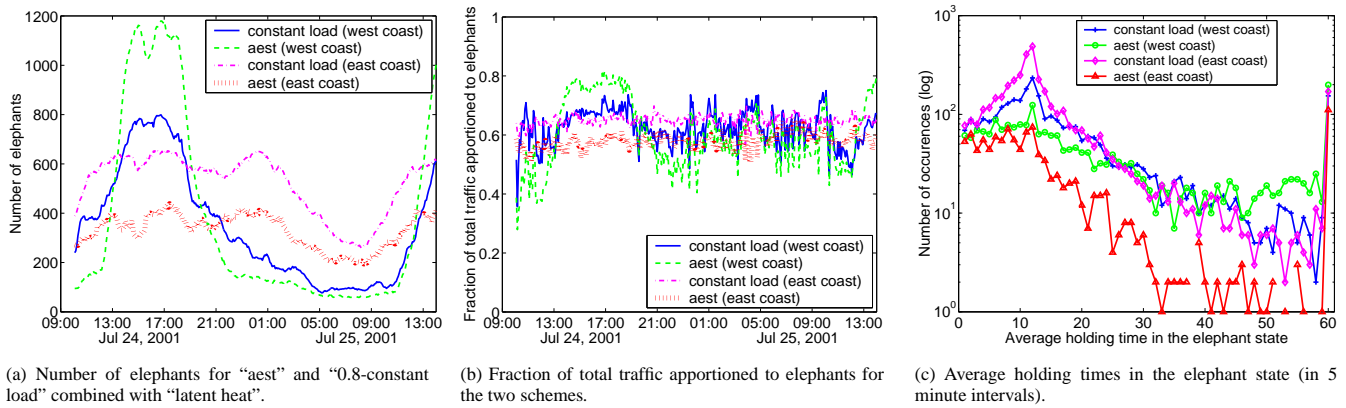| (a) Number of elephants for "aest" and "0.8-constant load" combined with "latent heat". | (b) Fraction of total traffic apportioned to elephants for the two schemes. | (c) Average holding times in the elephant state (in 5 minute intervals). |

Fig. 1. Elephant statistics for two-feature classification scheme.

each link become elephants for just a single interval.

Two Feature Classification Short-lived elephant flows are due to low-volume flows bursting beyond the threshold $\hat{v}(n)$ for small periods of time. In order to allow flows to experience short-term transitions across the threshold we define a new metric, which we call "latent heat". At each time interval we calculate the distance between the bandwidth achieved by a flow and the corresponding threshold value, derived using the proposed approaches "aest", and the "constant load". We define the "latent heat" of a flow as the sum of those distances in the past 12 timeslots, i.e. the previous hour. $LH_i(n) = \sum_{t=n-12}^{t=n}(X_i(t) - \hat{v}(t))$, thus incorporating a flow's persistence in time as a second feature. In each classification interval $n+1$, if $LH_i(n) \geq 0$, $i \in C_1$, otherwise $i \in C_2$.

The "latent heat" metric takes into account how much above or below the threshold a flow has been transmitting through time and reacts to transient moves above/below $\hat{v}(n)$ with sufficient latency. As a result, short transient bursts or dips are filtered avoiding unnecessary reclassification of flows.

III. RESULTS

Our classification approach leads to a small number of elephant flows, accounting for a substantial amount of the overall traffic, while exhibiting sufficient persistence in time. Results are presented for both links and for both schemes in Figure 1. Recall that under the "0.8-constant load" scheme, $v(n)$ is selected so that 80% of the total traffic is apportioned to elephant flows. Under the "aest" scheme, $v(n)$ is selected as a cut-off point in the flow bandwidth distribution.

Figure 1(a) presents the number of elephants identified at each time interval. The west coast link corresponds to a link experiencing a high burst in its utilization during the working hours. The east coast link exhibits smoother utilization levels during the day. As a consequence, the number of identified elephants for the west coast link under both schemes exhibits a similar burst during the working hours. For the east coast link, the number of identified elephants evolves in a smoother fashion during the day. The average number of elephants is 600 for the west coast link, and 500 for the east coast link.

The fraction of traffic apportioned to elephants for both links and under both schemes exhibits less fluctuation and is approximately 0.6 (Figure 1(b)). Even though the target of the "0.8-

constant load" scheme is to classify elephants so that they account for 80% of the total traffic, incorporating the "latent heat" metric leads to a smaller elephant load, since flows classified as elephants during the initial classification step turn out to exhibit insufficient persistence in time.

Indeed after incorporating the "latent heat" metric, the average holding time of an elephant flow increases to approximately 2 hours, while the number of flows classified as elephants for one interval dramatically decreases to approximately 50 (Figure 1(c)). We believe that classification schemes such as this one, that avoid reclassification for short-term fluctuations, identifies the type of long-lived elephant flows that are good candidates for traffic engineering applications.

Initial observations on the characteristics of elephants reveal that they correspond to networks with prefix lengths between /12 and /26, belonging to other Tier-n ISP providers. Although 100 /8 networks became active during the day, only three received traffic at a rate sufficiently high to place them in the elephant class. Therefore, there is little correlation between the size of a network prefix and its ability to act as an elephant.

IV. CONCLUSIONS

The idea of isolating elephants for traffic engineering purposes has been widely proposed, but there has been no prior effort on assessing the feasibility and issues involved in doing so. We propose a way to define "elephant" flows within a traffic engineering context. According to the proposed definition flows are characterized as "elephants" based on both their volume and their persistence in time. We show that while single feature classification schemes are attractive in their simplicity, they are insufficient for most traffic engineering applications, in that they lead to short-lived elephant flows. We show that our "latent heat" classification scheme is capable of detecting elephants characterized by the desired properties. Nevertheless, we conclude that identifying elephants imposes challenges despite their heavy-hitter nature.

REFERENCES

[1] M. Crovella and M. Taqqu. Estimating the Heavy Tail Index from Scaling Properties. *Methodology and Computing in Applied Probability*, 1999.
[2] K. Papagiannaki et. al. On the feasibility of identifying elephants in internet backone traffic. Sprint ATL Technical Report TR01-ATL-110918, Sprint Labs, November 2001.