Traffic analysis for anomaly detection and traffic engineering

S. Farraposo, P. Owezarski, E. Monteiro

Abstract—**The robustness and reliability of the Internet are highly dependent on efficient, stable connections and routing among provider networks. An analysis of real world trends in routing behavior across the Internet will have direct implications for the next generation of networking hardware, software and operational policies. This is why, observations of macroscopic traffic patterns provide insights into the legitimacy of the traffic, and how it affects network performance.**

***Index Terms*— Traffic anomalies, anomaly profile, routing**

I. Introduction

NEW connections among core Internet backbones are established continuously, ranging in capacity from 155 Mb/s to 10 Gb/s fiber optic pipes. This physical structure supports a myriad of new technologies and products, including live (or 'streaming') audio and video, distance education, entertainment, peer-to-peer, telephony and video-conferencing, as well as numerous new and often still evolving communications protocols. With no central authority serving to regulate and quality check, nor any feedback structure to throttle unfriendly practices or products, the Internet will continue its unbounded growth.

This kind of environment is an adequate one to the presence of anomalies, which can range from a single network failure to a complex security attack, being orchestrate through a thousand of separate machines. Trying to stop these anomalies is something that is very difficult to accomplish, due to several factors, as the size of the Internet, and the unpredictable behavior of those anomalies. However trying to control the extent and harshness of these anomalies, is one point where major contributions can arise.

The approach presented in this paper, sustain that through a tomography like process it is possible to define traffic profiles, particularly, anomaly traffic profiles and to decide how and when to act over traffic packets, whenever those profiles are found among traffic. Moreover, we intend to act at overlay networks routing level, based on these anomalies traffic profiles. Then, for example, when an unexpected increase of traffic would occur, we expect being able to differentiate between a flash crowd (due to the 50 millions euros prize in the on-line lottery), and a Denial of Service (DoS) attack, and in the first case to offer other routes to traffic, and in the second case, at least do not permit malefic traffic to overload the network resources, discarding these bad packets.

The remaining of this paper has the following structure: section II presents other works related with the definition of traffic and anomaly profiles to improve performance. Section III presents some of the work being accomplished in defining anomaly profiles using statistical analysis, and finally section IV presents future work.

II. Improve Routing with Traffic Analysis

Overlay networks are virtual communications structures that are logically 'laid over' a physical network such as the Internet. Manually configured static overlays are nothing new; however, a salient, modern feature of overlay networks is that they can be made to autonomically self-organize which provides great ease of deployment. Autonomic overlays have been used in two main areas: i) to alleviate the effects of slow or sporadic deployment of new services in the Internet (e.g. application-level multicast); and ii) to directly provide application-level functionality that is out-of-scope for the underlying network (e.g. proxies/caches in CDNs). Overlay distribution structures are now being considered in many application domains such as content distribution, conferencing, gaming, etc.

The increase in the usefulness of overlay networks to the all new kind of applications requires solutions at several layers, including routing. Our proposal is to use our tomography based traffic analysis, to act in real-time on routing decisions. However, this framing limits the way traffic analysis can be performed – only anomalies that occur at low frequencies (i.e., that span several minutes to hours) are of interest.

The majority of traffic analysis works being accomplished until now only consider small time scales, eliminating an all group of anomalies. This is the case of the works conduced by Katzela and Schwartz which focuses on methods for isolating failures in networks [1], by Feather *et al.* which shows that faults can be detected by statistical deviations from regularly observed behavior [2], and by Brutlag which applies

Manuscript received December 22, 2005. This work has been partly supported by the European Union under the E-NEXT NoE FP6-506869.

Sílvia Farraposo is with the Department of Informatics Engineering of Escola Superior de Tecnologia e Gestão de Leiria, Morro do Lena - Alto do Vieiro 2411-901 Leiria, Apartado 4163, Portugal (e-mail: silvia@estg.ipleiria.pt).

Philippe Owezarski is with the LAAS – CNRS, 7, Avenue du Colonel Roche, 31077 Toulouse, France (e-mail: owe@laas.fr).

Edmundo Monteiro is with the Department of Informatics Engineering of the University of Coimbra, Pólo II – Pinhal de Marrocos, 3030-290 Coimbra, Portugal (e-mail: edmundo@dei.uc.pt).

thresholds to time series models to detect aberrant network behavior [3]. An exception is the work developed by Lakhina *et al.* [4], which considers several time scales, when identifying anomalies in aggregate traffic at the campus border router.

III. Defining Traffic Profiles

Network anomalies typically refer to circumstances when network operations deviate from normal network behavior. These anomalous events will disrupt the normal behavior of some measurable network data. And, this is the basis of our approach – to detect through traffic measurements deviations from normality.

A. Methodology Definition

To correctly detect the occurrence of an anomaly and act accordingly, it is necessary to recognize it as soon, as possible. Assuming that, each anomaly has its own signature, the first stage of our approach is to define all these signatures, based on statistical analysis. Since our goal is to apply this knowledge at the routing level, we can only use the available information at that same level, which is: the number of bytes, the number of packets, and the flow dynamic (i.e., flow that are established and removed).

Also important, is the definition of "flow" at the routing level. For this, we have assumed the packet train model [5], which reflects the fact that much of network communication involves many packets spaced closely in time between the same two endpoints. Particularly, with such a flow definition it is possible to define several timeout parameters, and granularities for the source and destination addresses.

Considering both assumptions, several flows granularities were considered, ranging from flows with the same /24 destinations addresses to the same /32 destination addresses. Since we are analyzing traffic at an Internet link, coarser granularities do not improve our knowledge about traffic.

B. Some Results

The results presented refers to the analysis of a packet trace, captured with a DAG card, over 14 days, at the Internet link access of an academic network – the Auckland VIII trace [6].

To process the trace, at a first level a high-pass filter was defined (using mean and standard deviation), in such a way that only parts of the trace with abnormal behavior (presenting high variability) were considered. Then, over these time periods under observation, a more concise analysis was accomplished.

As an example, Figures 1 and 2 presents some of the results after applying our methodology. Analyzing both figures permit to detect the occurrence of a SYN Flood attack, which is characterized by a significant increase in the number of flows being established per unit of time, when compared with other periods of time.

It is interesting to notice, that if considering only an analysis over the /24 granularity, this type of occurrence is not visible, because of TCP behavior.

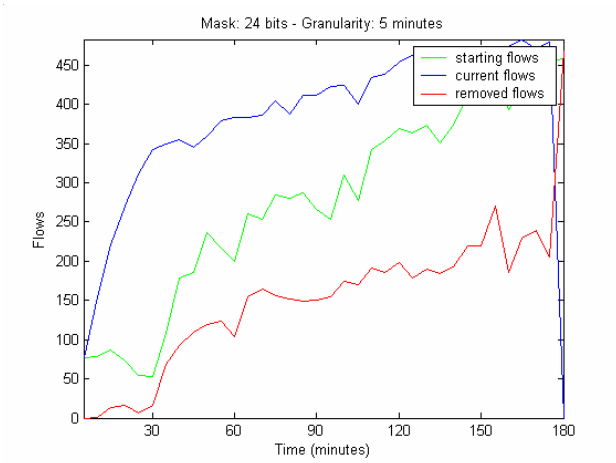

Fig. 1. Flow evolution with a granularity of /24 at the destination addresses.

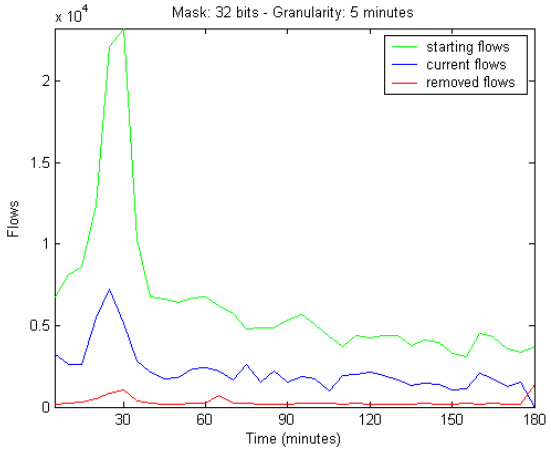

Fig. 2. Flow evolution with a granularity of /32 at the destination addresses.

IV. Conclusion and Future Work

This paper presents an approach that we have defined to take some accurate decisions when acting over the network routing protocol.

When considering large networks, such as overlay networks, only traffic anomalies that could impact all the remaining traffic are to be considered. So, anomalies that stand less than 5 minutes in a network (acceptable value, if we consider the time needed to propagate a change in a link over all the routers in the same AS), can be considered as "absorbed" ones, while all the others must be considered. There, another judgment must me taken into account: what is the cause of this anomaly, a legitimate or illegitimate event? With an appropriate anomaly signature, this question can be correctly answered, and the appropriate action taken. For instance, open other communication paths, remove all the "inconvenient" packets, and backtrack its origin, and …

References

[1] I. Katzela and M. Schwartz, "*Schemes for fault identification in communications networks*", IEEE/ACM Transactions on Networking, vol. 3(6), pp. 753–764, December 1995.

[2] F. Feather, D. Siewiorek, and R. Maxion, "*Fault detection in an Ethernet network using anomaly signature matching*", in Proceedings of ACM SIGCOMM '00, San Francisco, CA, September 2000.

[3] J. Brutlag, "*Aberrant behavior detection in time series for network monitoring*", in Proceedings of the USENIX Fourteenth System

Administration Conference LISA XIV, New Orleans, LA, December 2000.

[4] A. Lakhina, M. Crovella, C. Diot, "*Characterization of Network-Wide Anomalies in Traffic Flows*", in Proceedings of IMC'04, Sicily – Italy, October 2004.

[5] K. C. Claffy, H. Braun, G. Polyzos, "A parameterizable methodology for Internet traffic flow profiling", CAIDA, 1995.

[6] Available at: http://pma.nlanr.net/