Contribution of anomalies detection and analysis on traffic engineering

Sílvia Farraposo
DEI
ESTG – IPL
Leiria, Portugal
silvia@estg.ipleiria.pt

Philippe Owezarski
LAAS – CNRS
Toulouse, France
owe@laas.fr

Edmundo Monteiro
DEI – UC
University of Coimbra
Coimbra, Portugal
edmundo@dei.uc.pt

Abstract—**In this paper we present a methodology for detecting traffic anomalies. To accomplish that, and as a demarcation from similar works, we combine multi-scale and multi-criteria analysis with a tomography process analysis. With a complete knowledge of traffic anomalies, we intend to define anomalies signatures that could be used in a large range of scopes as traffic engineering, routing and intrusion detection systems, and much more.**

Traffic anomalies, anomaly profile, overlay network routing

I. Introduction

Assuring Quality of Service (QoS) in a network requires more and more a deep knowledge of traffic behavior. If, at the beginning the main concern of QoS frameworks was to reserve enough resources to assure an accurate data flowing, nowadays concerns are directed to traffic connections interactions.

Because traffic is not well behaved, i.e., always with the same pattern, the knowledge and characterization of traffic irregularities seems to be an important research field – this was the starting point for this work.

Traffic irregularities or traffic anomalies can be described as the result of one or more occurrences that changes the normal flow of data over a network. Such occurrences can be triggered by a diversity of things, such as DoS attacks, flash crowds or management operations.

Because traffic anomalies might occur at any point of the Internet, have unpredictable behaviors, and can range from a single network failure to a complex security attack, being orchestrate through a thousand of separate machines, stopping these anomalies is something that is very difficult to accomplish. However, trying to control the extent and harshness of these anomalies is one point where major contributions can arise – and this is the main goal of this work.

Analyzing traffic and its anomalies, is not a new topic, since several traffic analysis works were accomplished. However, none of the previous studies were based in tri-dimensional measurement structure, as our work. Most of them, only consider one time scale for their study, which is usually a small time scale, eliminating the detection of 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 thresholds to time series models to detect aberrant network behavior [3]. In contrary, this work intends to use a multi-scale analysis, ranging from 1second granularities to several minutes, to analyze the behavior of the following criteria's: number of bytes, packets and flows. Like this, we want to overview all types of anomalies.

To obtain more accurate results, we intend to combine the multi-scale and multi-criteria analysis, with a network tomography process – i.e., to analyze how traffic anomaly characteristics varies with different levels of prefix aggregation. The use of packet features, such as the IP addresses and ports is an idea being used by others, such as Lakhina *et al.* [4] and K. Xu *et al.* [5], however, their study only considers one level of network aggregation, while we are interested at several levels, ranging from /0 (i.e., all packets from all sources, to all destinations) to /32 (i.e., selected flows from a source to a destination).

The remaining of this paper has the following structure: section II presents our methodology and some preliminary results. Section III concludes this paper, and presents some areas where our approach might be of interest.

II. Anomaly Detection Methodology

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. Several measurements can be accomplished to extract traffic characteristics, however since one of our achievements is to define an as simple as possible methodology, that can be applied easily at the routing level, for example, we only use basic data parameters.

A. Measurement data

One of our main intentions is to assure that our approach has the simplest structure, and to accomplish that, manipulated

This work has been partly supported by the European Union under the E-NEXT NoE FP6-506869.

data must be as simple as possible. So, all our analysis is based in the relationship of the following data per flow, per unit of time:

- Number of packets
- Number of bytes
- Number of flows
- Source and destination IP addresses
- Source and destination ports

From several studies, it has been showed that most of anomalies can be detected through the observation of the parameters presented above. While the variation of number of packets and bytes (known as volume parameters) has already been stressed by several studies, the utilization of IP addresses and ports is a recent approach [5][6].

B. Methodology Definition

The base idea of the approach being developed is that any traffic anomaly is responsible for variability in one, or more, of the parameters presented above, and that we want to discover the flows being responsible for those variabilities.

With this in mind, we are developing an iterative approach that at each layer analyses parameters variabilities, and reduces the number of flows, among all the possible ones.

To accomplish the flow selection process self-made filters are being defined. At the moment, filters being defined consider the mean of variation of one of the first three parameters above, and a "potential" anomaly occurs when at some point, the value of the parameter exceeds K times the standard deviation.

So, the starting point of our approach is to detect slots in time, where a significant parameter variation occurs – this step is executed per parameter at several time scales. Then, at those points the tomography process is applied. This consists in applying several levels of aggregation (ranging from /0 to /32) to extract all the flows. Then, filtering is again applied to select the flows responsible for traffic anomalies.

We expect that the consideration of several layers of aggregation will permit to detect anomalies that otherwise would not be detected. As an example, are some types of DoS, that when considered isolated are presented as normal small flows, but when aggregated are presented as variable ones, with some IP addresses characteristics.

C. 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 available in the Passive Measurement and Analysis (PMA) Project [6].

To process the trace, the methodology presented was applied – and some time slots where extracted as "potentially" anomalous. As an example, Figure 1 presents one result obtained after applying our methodology. Analyzing the figure suggests the occurrence of DoS, 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.

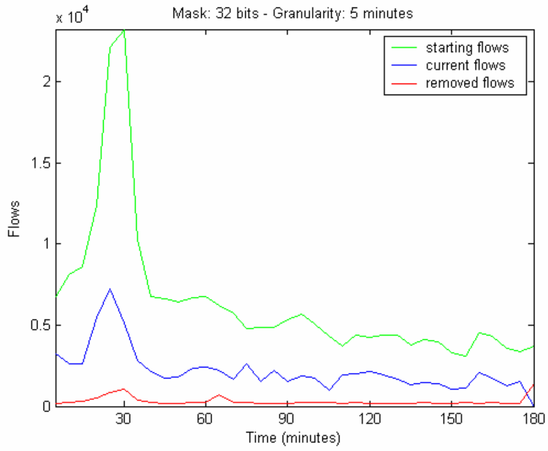

Fig. 1. Flow evolution with a granularity of /32 at the destination addresses.

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.

III. Conclusion

The approach presented in this paper, sustain that through a combined multi-scale, multi-criteria processes and a tomography like process it is possible to extract traffic anomalies and to define traffic profiles, particularly, anomaly traffic profiles.

Moreover, we intend to define traffic anomalies signatures that could be used to speedup the anomaly detection process, and permit to act over the anomaly accordingly.

The conjugation of anomalies detection and signatures constitute a pair that can be applied at several areas, ranging from overlay networks to network security. The identification of an anomaly, as a legitimate or illegitimate one, is important to decide which action must be taken. For instance, open other communication paths, or remove all the "inconvenient" packets and backtrack its origin

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 14th System Administration Conference LISA XIV, New Orleans, LA, December 2000.

[4] A. Lakhina, M. Crovella, C. Diot, "Mining Anomalies Using Traffic Feature Distributions", in Proceedings of ACM SIGCOMM'05, Philadelphia - USA, August 2005.

[5] K. Xu, Z. Zhang, S. Bhattacharyya "Profiling Internet Backbone Traffic: Behavior Models and Applications", in Proceedings of ACM SIGCOMM'05, Philadelphia - USA, August 2005.

[6] Available at pma.nlanr.net.