

A System for Detecting Network Anomalies based on Traffic Monitoring and Prediction

Pere Barlet-Ros, Helena Pujol, Javier Barrantes, Josep Solé-Pareta, Jordi Domingo-Pascual

Abstract—SMARTxAC is a passive monitoring and analysis system for high-speed links. This report describes its anomaly detection module and proposes the usage of adaptive traffic prediction as an effective technique to detect network anomalies in real-time.

Index Terms—Anomaly detection, Passive measurement, Traffic prediction

I. INTRODUCTION

SO far, network managers have been detecting anomalies either by interpreting raw traffic data like packet traces and NetFlow data, or by looking at traffic graphs generated by well-known traffic monitoring tools, such as MRTG, CoralReef or ntop. This tough process is not always effective since some network anomalies may go unnoticed, camouflaged in the rest of the traffic. Moreover, many of these monitoring systems are only confined to report plain statistics related to network usage, and therefore they require high-demanding human intervention following closely their evolution in a graphical interface in order to manually detect anomalies.

Our aim is to develop a system to capture, analyze and also interpret Internet traffic in real-time, based on SMARTxAC [1]. SMARTxAC is an always-on passive measurement system for high-speed links with full-traffic capture and real-time analysis capabilities, provided with a web-based graphical interface that presents the traffic-analysis results online. This system has been developed at the Technical University of Catalonia (UPC) and it is being used for the continuous monitoring of the Catalan academic and research network (Anella Científica).

Currently, SMARTxAC is reporting detailed information about the usage of the Anella Científica, so it has become a very useful and powerful tool for its

network managers to manually detect irregular usage or anomalies once they occur.

In order to facilitate to network managers the daily task of reviewing the network data looking for anomalies, we propose to add some intelligence to our system, so it can come to some conclusions by itself after having extracted and processed the information contained in the captured packets.

If the regular traffic of all the institutions that are connected to the network was known a priori, to detect anomalies would be as easy as comparing the regular traffic of an institution to its actual measured traffic. Nevertheless, because of inherent variations of Internet traffic by itself (e.g. burstiness, day/night, workday/weekend, etc.), to know a priori the regular traffic pattern of an institution is not an easy task.

Anomalies might be seen as unexpected changes on the regular traffic pattern of an institution and unexpected changes may be easily detected using traffic prediction. Using prediction, anomalies would be detected when the actual measured traffic differed significantly from the predicted one for the same interval. Therefore, regular traffic profiles for each institution would not have to be explicitly known. Moreover, its implementation is lightweight enough to be used in a real-time measurement system like SMARTxAC.

Consequently, anomaly detection in SMARTxAC is based on using adaptive traffic prediction in conjunction with other simpler techniques, like threshold-based detectors, in order to solve some limitations of adaptive prediction in the presence of anomalies that imply progressive and long-term traffic changes.

The rest of this extended abstract is organized as follows: Section II briefly reviews related work. Section III gives an overview of the SMARTxAC system and the measurement scenario. Section IV discusses our approach for anomaly detection. Section V defines the prediction algorithm. Section VI presents some preliminary results. Finally, in Section VII we introduce the results that we expect to include in the final paper.

This work is supported in part by CESCA (SMARTxAC agreement) and MCyT (TIC2002-04531-C04-02).

Authors are with the Advanced Broadband Communications Center (CCABA), Computer Architecture Department, Technical University of Catalonia (UPC), Barcelona, Catalunya, Spain. e-mail: {pbarlet, hpujol, jbarranp, pareta, jordid}@ac.upc.es

II. RELATED WORK

There are many techniques that can be applied to detect an abnormal behavior in Internet traffic, some of which are based on examining packet contents, whereas others study the temporal traffic-evolution.

Since anomalies are often caused by deliberate attacks or intrusions into users' systems, many solutions focus on discovering data patterns within packet payloads such as the ones set out in [2] and [3].

Other proposals are based on applying wavelet transforms in traffic analysis either in the temporal domain [4], [5] or the spatial domain [6]. The wavelet analysis allows to divide a time series into different frequency components, so the slow-varying trends of the series and the fine-grained details can be studied independently.

Conversely, the suitability of many prediction algorithms has also been studied for other purposes. As discussed in [7], simple solutions as linear prediction algorithms, like the linear fit or extrapolation polynomials of degree higher than one, have a significant predictive power. The solution adopted in [8], applied in particular for dynamic bandwidth reservation in video transmission, is the usage of adaptive and non-adaptive least mean square linear predictors, while both [9] and [10] analyze the possibility of using short-memory models like the autoregressive moving average (ARMA) model or the Markov-modulated Poisson process (MMPP) model.

On the other hand, in [5] network traffic anomalies have been classified into three general categories according to their cause. *Network operation* anomalies are those caused by device outages, configuration changes or traffic reaching bandwidth limit. *Flash crowd* anomalies are distinguished by a quick increase in traffic flows of a particular type or directed to a well-known destination with a gradual drop off over time. Finally, *network abuse* anomalies are usually recognized visually because they are associated to very high temporary traffic peaks.

III. WORKING SCENARIO

Since July 2003, SMARTxAC is being used for monitoring the Anella Científica, which connects about fifty universities and research centers in Catalonia.

Traffic collection is performed by standard PC hardware equipped with an Endace 4.3GE measurement card, which is synchronized via GPS. The tapped link is built from a pair of Gigabit Ethernet links, one for each traffic direction, which connect the Anella Científica to

RedIRIS (the Spanish academic and research network) and to the global Internet. Fig. 1 shows the three components of the SMARTxAC platform: the capture system, the traffic analysis system and the result visualization system.

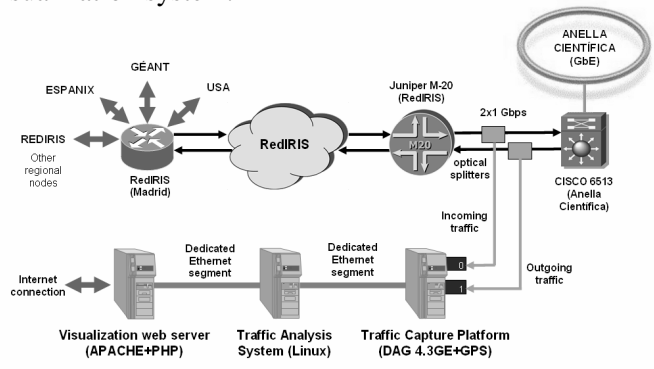


Fig. 1. SMARTxAC platform overview.

In order to analyze all the traffic in real-time, only packet headers are captured and aggregated into a special kind of flows that we have called *classified flows*. This aggregation is performed by translating the identifying values of the traditional 5-tuple flows (source/destination IP addresses, ports and protocol) into more general and meaningful values (origins, destinations and applications, respectively). *Origins* are the institutions connected to the monitored network, whereas *destinations* are considered as the external networks that Anella Científica is connected to (see Fig. 1). However, the system can be easily configured to support other network scenarios.

IV. DISCUSSION

Traditional intrusion detection systems (IDS) are usually designed to protect end-user systems or small networks. Most of their detection techniques are based on payload or system log analysis, since they assume that system log information can be easily accessed and traffic volume to be analyzed will be low enough to perform real-time payload analysis.

On the contrary, SMARTxAC is focused on the analysis of large networks and the measurement of high-speed links. Therefore, an anomaly detection module for SMARTxAC can not use end-system data and it has to be designed to operate with strong real-time constraints. To accomplish this, only packet headers are captured instead of full packets.

The threshold method is the simplest technique we have evaluated to detect anomalies. It consists of establishing an upper and a lower traffic limit manually, so that the system will consider as an anomaly those

situations in which traffic exceeds a preconfigured value. The threshold method has multiple usages. On the one hand, it is useful to check whether an organization is exceeding its allowed traffic. On the other hand, it is a simple technique that can detect most of the anomalies causing an important performance degradation of the monitored network. It is worth noting that SMARTxAC is focused on the entire network protection, so its network managers are usually interested on those anomalies that can degrade the quality of the service they offer to their users, more than protecting end-user systems from malicious attacks, like most IDS do. Under this assumption, threshold-based methods would be a simple, effective and easy-to-implement solution for detecting severe anomalies in large networks.

However, the fact that traffic of most networks follows a clear weekly pattern, where traffic volume increases during the working hours from Monday to Friday and decreases at night as well as on weekends (or vice versa in some kinds of networks), implies that these fixed thresholds must be much more complex than a single couple of values, and must include hour and traffic direction flexibility in order to become really meaningful.

More sophisticated methods to detect anomalies include prediction algorithms that forecast the next traffic value to be observed, so network traffic can be estimated in the short term. Using prediction, anomalies would be detected when the actual measured value of one parameter (e.g. bandwidth usage per institution) differed significantly from the predicted one for the same time interval.

One of the main strengths of prediction-based detectors is their capacity to adapt to inherent traffic variations. Moreover, they have been successfully used in other kinds of applications, e.g. dynamic bandwidth reservation for video transmission [8].

The simplest prediction algorithms have proved to be very efficient when it comes to identify anomalies in time series. They do not add much additional computing requirements and they are easy to implement. This simplicity is crucial for a system like SMARTxAC, which has strong real-time constraints and is expected to be used for detecting anomalies in large and high-speed networks.

Nevertheless, this technique is unable to detect those anomalies that imply progressive and long-term traffic changes, precisely due to its adaptive nature. In order to solve this limitation, our approach to detect anomalies is based on using adaptive linear predictors in conjunction with threshold-based detectors, which can efficiently

detect long-term changes on traffic patterns.

V. PREDICTION ALGORITHM

The prediction algorithm that we have chosen is the minimum mean square error linear filter as in [8], which estimates the value of $x(n+k)$ using a linear combination of the current and previous values of $x(n)$. Through the feedback of the differences between the predicted and the real values, $e(n)$, the p prediction coefficients of the filter, $w(n)$, are continuously adapted to minimize mean square error.

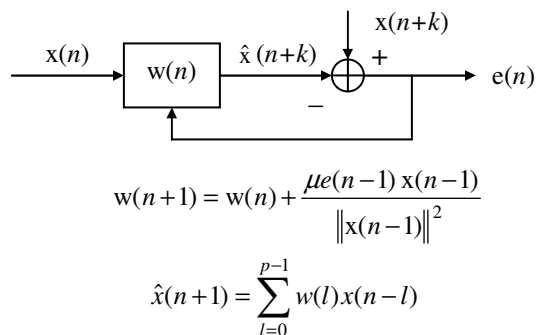


Fig. 2. Minimum mean square error linear filter.

Using a large value of μ , close to 2, results in a quicker response and adaptation to changes, whereas a small μ , close to 0, gives less relevance to former observations and brings about a slower convergence.

VI. PRELIMINARY RESULTS

An implementation of the threshold-based anomaly detector has already been integrated into SMARTxAC. Thresholds can be either manually preconfigured by the network managers, or automatically configured by training the system with historical data. Maximum and minimum thresholds for each institution can be defined for both traffic directions in bits/second, packets/second and flows/second units at the same time. Moreover, all thresholds can also be associated to a particular interval, so that it is possible to configure different thresholds for different periods of time (e.g. day/night, workday/weekend). Once a threshold is exceeded, an alarm is triggered and notified through the SMARTxAC web-based graphical interface, and a mail is sent to the network manager. When more than one alarm is triggered at the same time, the system is able to group all of them and show them as a unique alarm, to keep their number as low as possible. At this point, the system also stores some additional data concerning the anomaly. Thus, it is possible to carry out an off-line analysis of stored data to infer the anomaly causes.

Fig. 3 shows the number of flows/second per application for a particular institution of the Anella Científica for September 23th, 2004. Three alarms were raised at 8:06, at 9:36, and at 11:06, respectively, because the upper flows/second threshold was exceeded. After reviewing the collected information, it was concluded that the last one was due to a TCP port scanning attack against port 445.

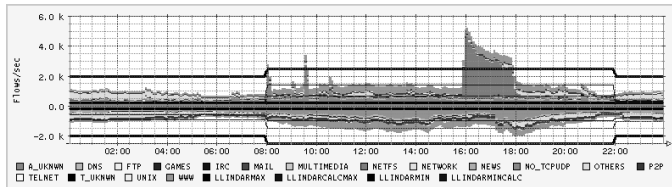


Fig. 3. Flows/second threshold-based anomaly detector.

The adaptive prediction implementation has been also developed and tested using real packet traces. Currently, we are porting this code to run as a module of SMARTxAC. First results using a preliminary version of this module point out a very good performance working in real-time in high-speed links.

Fig. 4 shows the link utilization in packets/second and bits/second respectively, for a particular institution of the Anella Científica for September the 25th and the 26th. The solid lines represent the actual incoming and outgoing traffic, whereas the dotted lines correspond to the predicted traffic.

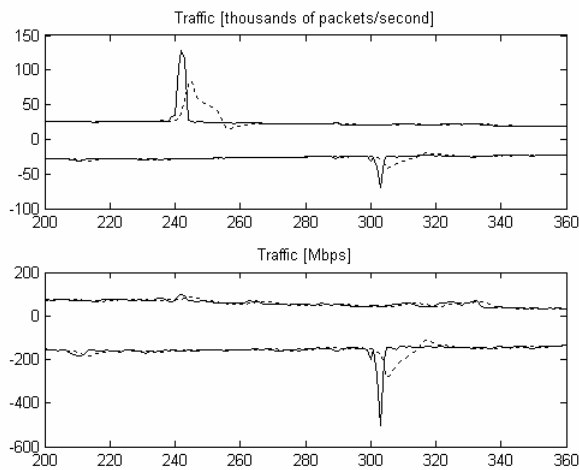


Fig. 4. Actual vs. predicted traffic for a particular institution.

According to the algorithm, an alarm was raised on the 25th at 20:05, when the number of incoming packets got multiplied by 5 unexpectedly, although no irregular situation was detected in relation to the number of incoming bytes.

At 1:15 on the next day, two alarms concerning the same anomaly were raised because not only the number of outgoing bytes, but also the amount of packets, increased much higher than the predicted value.

VII. FUTURE WORK

The feasibility and performance of our prediction-based anomaly detection module for SMARTxAC will be evaluated using real traffic collected from the Anella Científica. We will test it in real-time to acquire at least one month of data for our final study. During this data-acquirement period, we will compare our results to the anomalies reported by our network managers. Moreover, we expect to demonstrate that all kind of anomalies mentioned in Section II and presented in [5] can be easily detected using our proposal based on traffic prediction.

Another aspect that we still have to study is the worthiness of adapting the parameters of prediction algorithms depending on previous stored traffic statistics in order to obtain more accurate predictions and, consequently, to get a lower false alarm rate.

So far, the system is triggered to store some additional data concerning detected anomalies. The exact amount of extra information that should be kept depends on the kind of phenomena that we will be looking for. The experience of our network managers will be used as a very valuable input to know which kind of information should be gathered depending on the alarm nature.

REFERENCES

- [1] P. Barlet and J. Domingo, "SMARTxAC: A passive monitoring and analysis system for high-speed links", *RIPE 49 Meeting*, Manchester, UK, September 2004.
- [2] M. V. Mahoney and P.K. Chan, "Learning Rules for Anomaly Detection of Hostile Network Traffic", in *Proceedings of the Third IEEE International Conference on Data Mining (ICDM)*, pp. 601-4, 2003.
- [3] R. A. Maxion and K. M.-C. Tan, "Benchmarking Anomaly-Based Detection Systems", in *Proceedings of the International Conference on Dependable Systems and Networks*, New York, NY, June 2000.
- [4] P. Huang, A. Feldmann, and W. Willinger, "A non-intrusive, wavelet-based approach to detecting network performance problems," in *Proceedings of ACM SIGCOMM Internet Measurement Workshop*, San Francisco, CA, November 2001.
- [5] P. Barford and D. Plonka, "Characteristics of Network Traffic Flow Anomalies" in *Proceedings of ACM SIGCOMM Internet Measurement Workshop*, San Francisco, CA, November 2001.
- [6] M. Crovella and E. Kolaczyk, "Graph Wavelets for Spatial Traffic Analysis" in *Proceedings of IEEE INFOCOM 2003*, San Francisco, CA, April 2003.

- [7] Y. Baryshnikov, E. Coffman, D. Rubenstein and B. Yimwadsana, "Traffic Prediction on the Internet". Technical Report EE200514-1, Computer Networking Research Center, Columbia Univ., May 2002.
- [8] A. Adas. "Supporting Real Time VBR Video Using Dynamic Reservation Based on Linear Prediction" in *Proceedings of IEEE INFOCOM '96*, San Francisco, CA, March 1996.
- [9] A. Sang and S. Li, "A Predictability Analysis of Network Traffic" in *Proceedings of IEEE INFOCOM 2000*, Tel Aviv, Israel, March 2000.
- [10] M. Ghaderi, "On the Relevance of Self-Similarity in Network Traffic Prediction". Technical Report CS-2003-28, Computer Science, University of Waterloo, October 2003.