SMARTxAC: A Passive Monitoring and Analysis System for High-Speed Networks

Pere Barlet-Ros, Josep Solé-Pareta, Javier Barrantes, Eva Codina, Jordi Domingo-Pascual
Advanced Broadband Communications Center (CCABA), Computer Architecture Department
Universitat Politècnica de Catalunya (UPC)
Jordi Girona 1-3, 08034 Barcelona, Catalunya, Spain
{pbarlet, pareta, jbarranp, ecodina, jordi.domingo}@ac.upc.edu

Keywords: *Traffic monitoring, Traffic analysis, Network anomaly detection*

During the last years, the Advanced Broadband Communications Center (CCABA) of the UPC has been involved in several projects related to Internet traffic monitoring and analysis in the Spanish National Research and Education Network (RedIRIS), namely CASTBA, MEHARI [1] and MIRA [2, 3]. As a result of such an experience, a new traffic monitoring and analysis system, called SMARTxAC, has been developed at CCABA in joint work with the Supercomputing Center of Catalonia (CESCA).

SMARTxAC is an always-on passive monitoring and analysis system that operates at gigabit speeds without packet loss. It also integrates a web-based graphical interface that offers numerous traffic statistics online. The main difference between SMARTxAC and other passive measurement infrastructures and systems for high-speed networks, such as those instrumented by NLANR [4] and Sprint [5], lies in that SMARTxAC has been explicitly designed for online traffic analysis, whereas in other approaches traffic analysis tasks are usually postponed to an offline stage.

Measurement scenario: Since July 2003, SMARTxAC is being used for monitoring the Anella Científica network, which is managed by CESCA and connects about fifty universities and research centers in Catalonia. SMARTxAC is continuously monitoring the link that connects the Anella Científica to RedIRIS, which constitutes its main connection to the global Internet. This link is built from a pair of full-duplex Gigabit Ethernet links by using load balancing techniques. The average load of this link during working hours is about 1.5 Gbits/sec (270 Kpackets/sec). A GPS-synchronized and anonymized IP header trace of this link was collected for the NLANR-PMA project and can be downloaded at [6].

Fig. 1 shows the three main components of the SMARTxAC platform that are described next: the capture system, the traffic analysis system and the result visualization system. Due to performance reasons, each of them is executed on a different computer, although in other scenarios under lower traffic loads, they can run on a single computer.

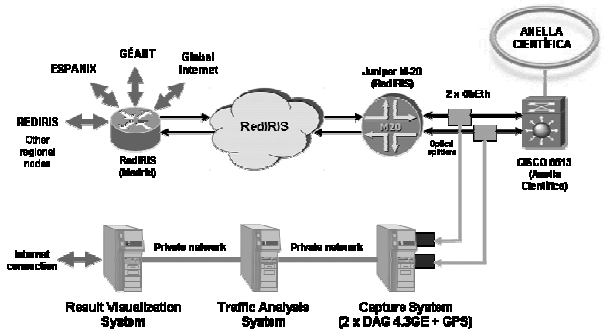

Fig. 1. SMARTxAC platform overview

Capture system: Both Gigabit Ethernet links are tapped by using four optical splitters (one per link and direction). A copy of the traffic is passively sent to an off-the-shelf computer equipped with a pair of Endace DAG 4.3GE cards [7]. The DAG 4.3GE card is dual port, so that only two cards are needed for monitoring both links. The internal clocks of both DAG cards are synchronized via an external GPS source. The traffic measurement software running on the capture system basically collects packet headers without loss and performs a first aggregation of the traffic into traditional 5-tuple flows.

Traffic analysis system: The main task of the traffic analysis system is to aggregate the collected flows into a special kind of flows we have called *classified flows*, while computing several traffic statistics. This aggregation is performed online by translating the identifying values of the 5-tuple flows (source/destination IP addresses, ports and protocol) into more general and meaningful values (origins, destinations and applications, respectively). We refer to *origins* as 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 configured to support any other network scenario.

This classification not only allows us to compute detailed statistics per institution and destination network, but also provides us valuable information about the nature of the traffic. For instance, P2P traffic that is routed to a commercial network (e.g. Espanix) is probably less academic than mail traffic routed to an academic or research network (e.g. RedIRIS or GÉANT). This knowledge can be useful for cost-sharing and billing purposes. In SMARTxAC, a pricing matrix can be defined by setting a price or weight per byte according to its origin, destination and application as described in [8], but this feature is not used in our scenario.

Although this classification process needs to perform some expensive operations, such as computing twice per flow the longest prefix match algorithm, it can be done online. The main advantage of this approach lies in drastically reducing the volume of data to be processed and stored onto disk. It makes the permanent storage of historical data feasible without losing valuable information about the network usage. However, if more detailed information is needed, for instance by network researchers or for intrusion detection purposes, the system can be configured to collect header or full packet traces as well.

Moreover, some statistics (e.g. Top-N origin and destination IP addresses) are computed before this second aggregation, mainly because are bounded by N and do not significantly increase the volume of data to be stored. In addition, when a network anomaly is detected, finer-grained information related to the anomaly is kept, since can be useful to analyze the causes of the detected anomaly.

The total amount of data stored in our scenario usually does not exceed 30 MBytes/day, so that historical data about the Anella Científica's usage is being permanently stored since July 2003.

During the classification process, several statistics per institution are computed and can be consulted online by using the SMARTxAC web-based interface. In order to preserve institution privacy, institutions only can see their own traffic statistics. A complete list of the graphs available from the web-based interface can be found at [9], and include daily, weekly and monthly traffic profiles per applications (Fig. 2), destinations, destinations per applications (Fig. 3), Top-N IP addresses, ports and protocols, detected network anomalies, etc.

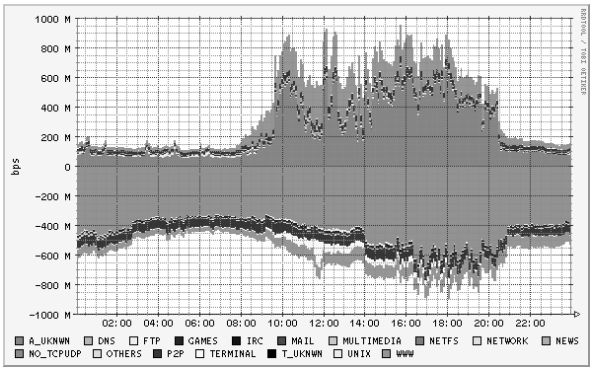

Fig. 2. Traffic per application (bits/sec)

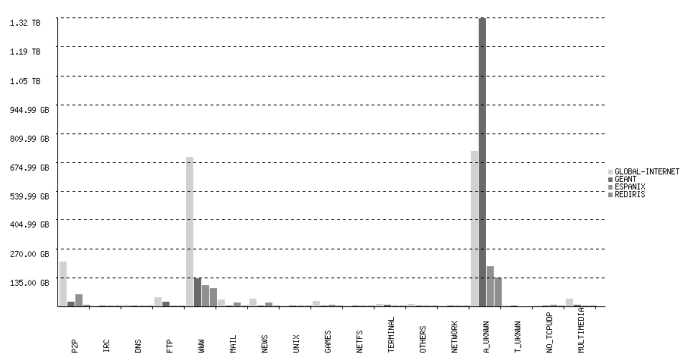

Fig. 3. Incoming traffic per application and destination (bytes)

Anomaly detection: 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 assume that log information can be easily accessed or the traffic volume to be analyzed will be low enough to perform online payload analysis. On the contrary, SMARTxAC is focused on the analysis of large networks and the measurement of high-speed links. Therefore, the anomaly detection method used by SMARTxAC must be lightweight enough to operate with strong real-time constraints. Under this assumption, SMARTxAC integrates a simple anomaly detection method based on traffic thresholds. It consists of establishing an upper and lower traffic limit per institution, so that the system will consider as anomalies those situations in which the traffic of an institution exceeds a predefined threshold. These thresholds can be configured manually or automatically by training the system with historical data, and can be independently set for both traffic directions in bits/second, packets/second and flow/second units. This simple technique is effective to detect those anomalies causing an important degradation of the network performance, although its detection power for other anomalies is relatively limited. For this reason, we are currently working on more effective, but still lightweight, techniques for detecting and classifying network anomalies based on traffic prediction techniques. Preliminary results of this work are described in [10].

Ongoing work: Currently, besides of improving the anomaly detection system, we are working on integrating into our system sampling measurement techniques, IPv6 support, and some recent proposals for detecting network applications based on heuristic techniques (e.g. [11]) and by analyzing the first bytes of the application payload. Moreover, CCABA is collaborating with the Intel's CoMo project [12] in developing a general-purpose network monitoring system.

Acknowledgements

This work is supported in part by the Supercomputing Center of Catalonia (CESCA), under the SMARTxAC agreement, and by the Spanish Ministry of Science and Technology, under contract TIC2002-04531-C0402.

References

[1] P. J. Lizcano, A. Azcorra, J. Solé-Pareta, J. Domingo-Pascual and M. Álvarez-Campana. MEHARI: A System for Analyzing the Use of the Internet Services. *Computer Networks*, 31(21):2293-2307, 1999.

[2] C. Veciana-Nogués, J. Domingo-Pascual and J. Solé-Pareta. Servers Location and Verification Tool for Backbone Access Points. *Proc. of 13th ITC Specialist Seminar: IP Traffic Measurement, Modeling and Management*, Monterey, USA, Sep. 18-20, 2000.

[3] C. Veciana-Nogués, J. Domingo-Pascual and J. Solé-Pareta. Cost-Sharing and Billing in the National Research Networks: the MIRA Approach. *Terena Networking Conference*, Limerick, Ireland, June 3-6, 2002.

[4] NLANR: National Laboratory for Applied Network Research. http://www.nlanr.net

[5] C. Fraleigh, S. Moon, B. Lyles, C. Cotton, M. Khan, R. Rockell, D. Moll, T. Seely and C. Diot. Packet-Level Traffic Measurements from the Sprint IP Backbone. *IEEE Network*, 17(6):6-16, Nov./Dec. 2003.

[6] NLANR Special Traces Archive: CESCA-I Data Set. http://pma.nlanr.net/Special/cesc1.html

[7] Endace Measurement Systems. http://www.endace.com

[8] B. Stiller, P. Barlet-Ros, J. Cushnie, J. Domingo-Pascual, D. Hutchison, R. Lopes, A. Mauthe, M. Popa, J. Roberts, J. Solé-Pareta, D. Trcek, C. Veciana and L. Wolf. Pricing and QoS. Quality of Future Internet Services: COST Action 263 Final Report, 2856:263-291, Springer-Verlag, Dec. 2003.

[9] The SMARTxAC project. http://www.ccaba.upc.edu/smartxac

[10] P. Barlet-Ros, H. Pujol, J. Barrantes, J. Solé-Pareta and J. Domingo-Pascual. A System for Detecting Network Anomalies based on Traffic Monitoring and Prediction. Technical report, UPC-DAC-RR-CBA-2005-5, June 2005.

[11] A. W. Moore and D. Zuev. Internet Traffic Classification Using Bayesian Analysis Techniques. *Proc. of the ACM SIGMETRICS*, Banff, Canada, June 2005.

[12] The CoMo Project, Intel Research Cambridge. http://como.intel-research.net

Vitae

Pere Barlet-Ros received the M.Sc. degree in Computer Science from the Universitat Politècnica de Catalunya (UPC) in 2003. He is Assistant Professor and Ph.D. student at the Computer Architecture Department of the UPC. In 2004, he was visiting Ph.D. student at the National Laboratory for Applied Network Research (NLANR) for two months and intern at Intel Research Cambridge for three months. His research interests are in the fields of network measurement, network forensics and network performance evaluation.

Josep Solé-Pareta was awarded his Master's degree in Telecommunication Engineering in 1984, and his Ph. D. in Computer Science in 1991, both from the Universitat Politècnica de Catalunya (UPC). In 1984 he joined the Computer Architecture Department of UPC, where currently is Full Professor in Computer Science and Communications. He did a Postdoc stage (summers of 1993 and 1994) at the Broadband and Wireless Networking Lab. of the Georgia Institute of Technology. His current research interests are in broadband Internet and high-speed and optical networks, with emphasis on traffic engineering, traffic characterization, traffic management, QoS provisioning and MAC protocols for legacy and optical metro networks.

Javier Barrantes received the M.Sc. degree in Computer Science from the Universitat Politècnica de Catalunya (UPC) in 2003. He is currently a Projects Graduate Scholarship Holder and Ph.D. student in the

Computer Architecture Department, UPC. His research interests are in the fields of traffic measurements, modeling and prediction.

Eva Codina received the Bachelor degree in Computer Science from the Universitat Politència de Catalunya (UPC) in 2004. She is currently a Projects Graduate Scholarship Holder in the Computer Architecture Department, UPC. Her research interests are in the fields of traffic measurements, modeling and prediction.

Jordi Domingo-Pascual is Full Professor of Computer Science and Communications at the Universitat Politècnica de Catalunya (UPC) in Barcelona. There, he received the engineering degree in Telecommunication (1982) and the Ph.D. Degree in Computer Science (1987). In 1983 he joined the Computer Architecture Department. He was visiting researcher at the International Computer Science Institute in Berkeley (California) for six months. His research topics are broadband communications and applications, IP/ATM integration, QoS management and provision, traffic engineering, IP traffic analysis and characterization, group communications and multicast.