SMARTxAC: Sistema de monitorización y análisis de tráfico para la Anella Científica

SMARTxAC: A System for Monitoring and Analysing the Traffic of the Anella Científica

P. Barlet, J. Solé y J. Domingo

Resumen

El presente artículo describe un sistema de monitorización y análisis de tráfico en tiempo real para redes troncales de alta velocidad. Esta herramienta proporciona información detallada sobre el uso de la red, que es de gran utilidad para su gestión y dimensionado, así como para la optimización de los recursos y la detección de usos irregulares y ataques. El proyecto SMARTxAC tiene como objetivo instalar una versión de este sistema para monitorizar de forma permanente el tráfico de la *Anella Científica*, adaptada a las necesidades de sus gestores (CESCA).

Palabras clave: monitorización, captura de tráfico, análisis de tráfico.

Summary

This paper describes a system for monitoring and analysing the traffic of high-speed networks in real-time. This tool gives detailed information about network usage, which can be very useful for the network management and dimensioning, resource optimization and for detecting irregular usage and network attacks. The main objective of the SMARTxAC project is to install a version of this system for monitoring permanently the traffic of the *Anella Científica*, adapted to its manager (CESCA) requirements.

Keywords: monitoring, traffic accounting, traffic analysis.

El proyecto SMA_TxAC es un sistema de monitorización y análisis de tráfico en tiempo real, para redes troncales de alta velocidad

1.- Introducción

En los últimos años, en la red académica española (RedIRIS) se han llevado a cabo diferentes proyectos relacionados con la monitorización y la caracterización del tráfico Internet, como por ejemplo los proyectos CASTBA, MEHARI [1] y MIRA [2]. Estos proyectos se realizaron de forma conjunta entre la Universidad Politécnica de Madrid, la Carlos III de Madrid, la Politècnica de Catalunya (UPC), y con la participación como EPOs de RedIRIS, Telefónica Investigación y Desarrollo, el Centre de Supercomputació de Catalunya (CESCA) y el Institut Català de Tecnologia.

Una vez finalizados estos proyectos y basándose en la experiencia adquirida en su participación, el Centre de Comunicacions Avançades de Banda Ampla (CCABA) de la UPC, desarrolló un prototipo propio de un sistema completo de monitorización que permite el análisis de tráfico en tiempo real en enlaces de alta velocidad. Este prototipo proporciona información detallada sobre el uso que se hace de la red monitorizada, información que puede ser de gran ayuda para el dimensionado y la optimización de recursos, además de ser útil para detectar usos irregulares y ataques.

El funcionamiento de este prototipo se probó en el troncal de Cataluña de RedIRIS (*Anella Científica*), que constituye la principal vía de salida a Internet de las universidades y centros de investigación catalanes. Los resultados de estas primeras pruebas fueron muy satisfactorios y animaron a los gestores de la *Anella Científica* (CESCA) a encargar al CCABA-UPC el desarrollo de una versión mejorada de dicho sistema, que ha dado lugar al proyecto SMARTxAC.

El proyecto SMARTxAC es un acuerdo de colaboración entre el CESCA y la UPC, que se inició en julio de 2003, con el objetivo de instalar una versión del prototipo desarrollado en el CCABA-UPC para la monitorización permanente de la *Anella Científica*. Este nuevo sistema deberá proporcionar información útil que ayude al CESCA en las tareas diarias de gestión de la red.

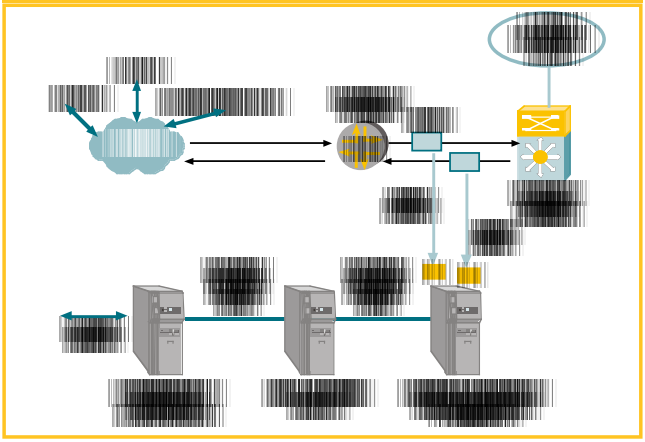
2.- Descripción del sistema SMARTxAC

El sistema SMARTxAC está dividido en tres sistemas, por razones de eficiencia, instalados en equipos de proceso diferentes

La funcionalidad principal del sistema SMARTxAC es la monitorización y el análisis de tráfico en tiempo real en enlaces troncales de alta velocidad. Estos enlaces son compartidos por gran cantidad de redes, que a su vez están dando servicio a miles de usuarios. Estos usuarios tienen necesidades y perfiles muy diferentes, y pueden acceder a una gran variedad de servicios. Debido a esta heterogeneidad, no sólo el volumen de tráfico presente en estas redes es muy elevado, sino que también lo es el número de sesiones establecidas simultáneamente. Precisamente, la gran cantidad y variedad de datos a capturar y analizar es la principal dificultad a abordar en el desarrollo de un sistema de estas características. El equipo de captura debe ser capaz de capturar todo el tráfico, sin perder ningún paquete, pero la dificultad principal está en el sistema de análisis, que debe ser lo suficientemente ligero y eficiente para poder tratar toda esta información en tiempo real, y resumirla para que sea viable su almacenamiento de forma permanente.

La figura muestra que el sistema SMARTxAC está dividido en tres sistemas, por razones de eficiencia, instalados en equipos de proceso diferentes y son: la plataforma de captura, el sistema de análisis y el sistema de visualización de resultados.

FIGURA 1: VISION GENERAL DEL SISTEMA SMARTXAC



2.1.- Plataforma de captura

Llamamos plataforma de captura a la parte hardware y software dedicada a la captura de tráfico. Al igual que en el proyecto MIRA [2], se realiza una captura pasiva (no intrusiva) del tráfico, utilizando divisores de fibra pasivos (splitters), que permiten enviar una copia íntegra del tráfico a un PC, que se encarga de la captura y el procesado de los paquetes. Al contrario de lo que sucede con otras técnicas de captura, como Cisco NetFlow [3] o los basados en SNMP, nuestro sistema no afecta en absoluto al rendimiento de la red monitorizada, ya que la captura no se realiza directamente en los equipos dedicados a la interconexión de redes, ni tampoco se genera tráfico adicional.

Se ha desestimado utilizar la plataforma de captura desarrollada en el proyecto MIRA, debido a que los requisitos actuales difieren de forma considerable con los que se plantearon en dicho proyecto. El sistema MIRA realizaba una captura estadística del tráfico (aproximadamente un 10% del tráfico real) debido a que se capturaba el contenido de los paquetes. En el sistema SMARTxAC se ha preferido capturar únicamente las cabeceras de los paquetes, y conseguir así una captura completa del tráfico, utilizando el software de libre distribución *CoralReef* [4]. De esta forma, las cabeceras capturadas pueden agregarse en forma de flujos, y reducir así el volumen de datos a tratar por el sistema de análisis. Pero además, la captura de contenidos también presenta varias limitaciones, como la posible infracción de confidencialidad o la imposibilidad de analizar los paquetes cifrados mediante técnicas de encriptación.

2.2.- Análisis de tráfico

El sistema de análisis de tráfico se encarga de procesar los flujos IP capturados por la plataforma de captura y transformarlos en un nuevo tipo de flujos que denominamos *flujos clasificados*. Esta transformación consiste en la traducción de los valores que identifican un flujo IP (direcciones IP, puertos y protocolo) a valores más generales (origen, destino y aplicación) y por tanto más útiles para conocer el uso que se hace de la red. Al reducir el número de posibles valores que identifican un flujo, se consigue una reducción considerable del número de flujos a almacenar, ya que en un mismo flujo clasificado quedan agregados varios flujos IP. La reducción aún es mayor ya que los flujos clasificados son bidireccionales, mientras que los flujos IP son unidireccionales.

El sistema de análisis de tráfico se encarga de procesar los flujos IP capturados por la plataforma de captura y transformarlos en un nuevo tipo de flujos que denominamos *flujos clasificados*

Entendemos como *origen* los diferentes rangos de direcciones IP definidos en la red monitorizada. Por ejemplo, en nuestro caso son las instituciones conectadas a la *Anella Científica*, pero más adelante se considerarán también sus puntos de acceso (varias instituciones comparten un mismo punto de acceso), aunque también podrían ser los departamentos dentro de cada institución, etc. En general, el número de orígenes definidos dependerá del nivel de detalle del análisis que requiera el gestor de la red.

Análogamente llamamos *destino* a los diferentes rangos de direcciones IP de interés fuera de la red monitorizada. Por ejemplo, en nuestro caso se han definido cuatro grupos de destinos, dependiendo de las diferentes conexiones que dispone RedIRIS con el exterior (Géant, Espanix e Internet Global) y con los otros troncales de RedIRIS.

Una vez realizada la transformación de flujos, se acumulan los datos obtenidos en períodos diarios, semanales y mensuales, con lo que se consigue una reducción aún mayor del volumen de datos a almacenar. Adicionalmente, se mantiene un registro de los flujos que no se han podido clasificar, porque contienen direcciones falsas, puertos o protocolos desconocidos, etc. Esta información puede ser de gran valor para la detección de usos irregulares o ataques, y en general para conocer más detalladamente el uso de la red.

2.3.- Visualización de resultados

El prototipo dispone de una interfaz gráfica, basada en un entorno *web*, que permite consultar gráficamente todos los resultados de análisis de tráfico. Estas graficas son generadas dinámicamente por el servidor *web* bajo demanda.

Las gráficas de análisis que pueden ser consultadas por los gestores de la red son: la evolución temporal del tráfico por aplicaciones (ver como ejemplo la figura 2); la comparativa por orígenes de la evolución temporal del tráfico por aplicaciones; el tráfico cruzado por aplicaciones/destinos (ver figura 3) y orígenes/destinos; el tráfico por orígenes, destinos, aplicación y protocolos de transporte; el tráfico IP no TCP/UDP; el tráfico con puertos, direcciones y protocolos de transporte desconocidos, y las gráficas de tarificación (basada en el uso de la red).

Cada una de estas gráficas puede representarse para el total de tráfico, o de forma desglosada para cada uno de los orígenes (instituciones y puntos de acceso a la *Anella Científica*), en unidades de bits por segundo, paquetes por segundo y octetos o número de paquetes totales.

3.- Estado actual del proyecto SMARTxAC y tareas pendientes

En primer lugar ha sido necesaria la adaptación del prototipo a la tecnología Gigabit Ethernet, después de que en mayo de 2003 la *Anella Científica* cambiara de ATM a esta tecnología. Para

Actualmente se están obteniendo los primeros resultados en la monitorización de enlaces GbEth, y está previsto que para febrero de 2004 ya se instale en el CESCA la primera versión del sistema SMARTxAC

conseguir la captura completa en este nuevo escenario, se ha modificado la plataforma de captura, ya que el volumen de tráfico en los dos enlaces GbEth monitorizados puede crecer hasta los 2 Gbps. Por este motivo no es posible utilizar tarjetas de red estándar GbEth para la captura, ya que no son capaces de realizar una captura completa con estos volúmenes de tráfico. Actualmente se ha instalado una tarjeta GbEth especializada para la captura (DAG 4.2GE de Endace [5]) que consigue capturar todo el tráfico de dos enlaces GbEth sin perder ningún paquete.

FIG. 2: EVOLUCIÓN DIARIA DEL TRÁFICO POR APLICACIONES (ENTRADA/SALIDA *ANELLA*)

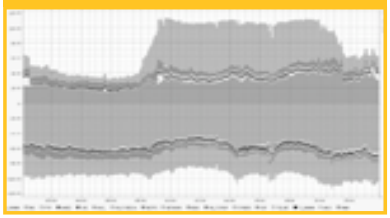


FIG. 3: TRÁFICO DIARIO POR APLICACIONES Y DESTINOS (ENTRADA *ANELLA CIENTÍFICA*)

Actualmente se están obteniendo los primeros resultados en la monitorización de enlaces GbEth, y está previsto que para febrero de 2004 ya se instale en el CESCA la primera versión del sistema SMARTxAC para analizar de forma estable y permanente el tráfico de la Anella Científica.

El siguiente paso será la modificación del sistema de análisis según las necesidades que pueda tener el CESCA como gestor de la Anella. Éste se concreta en el desarrollo de un módulo de detección automática de situaciones irregulares, como pueden ser cambios repentinos en los patrones habituales de tráfico de alguna de las instituciones conectadas a la red, ataques (DoS, DDoS, Spoofing, etc.), uso de aplicaciones peer-to-peer o equivalentes, etc. Durante los períodos en que se detecte alguna situación irregular, el sistema avisará al administrador mediante una alarma y guardará información adicional que pueda ser útil para descubrir las causas, o si es posible el causante.

Referencias

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

[2] Veciana, C.; Domingo, J.; Solé, J., "Cost-sharing and Billing in the National Research Networks: the MIRA Approach". Presentada en: Terena Networking Conference. 2002. Limerick (Irlanda).

[3] CISCO SYSTEMS. "NetFlow Services Solutions Guide". 2001. http://www.cisco.com/univercd/cc/td/doc/cisintwk/intsolns/netflsol/nfwhite.pdf

[4] Moore, D; Keys, K.; Koga, R.; Lagache, E.; Claffy, K., "The CoralReef SW Suite as a Tool for System and Network Administrators". 2001. http://www.caida.org/outreach/papers/2001/CoralApps/CoralApps.pdf

[5] Endace Measurements Systems. "DAG 4.2GE Dual Gigabit Ethernet". 2003. http://www.endace.com/dag4.2ge.htm

Trabajo financiado parcialmente por el CESCA (convenio SMARTxAC) y por el MCyT en el marco del proyecto TIC2002-04531-C04-02.

Pere Barlet Ros, Josep Solé Pareta
(pbarlet@ac.upc.es), (pareta@ac.upc.es)
Jordi Domingo Pascual
(jordi.domingo@ac.upc.es)
CCABA -Dept. Arquitectura de Computadors – UPC
(http://www.ccaba.upc.es)