Propuesta de ponencia para las XIII jornadas Telecom. I+D, noviembre 2003.

Titulo de la ponencia:

El control de conmutadores de circuitos en el núcleo de la Internet por medio de flujos de usuario

Autores:

Pablo Molinero Fernández (Stanford University y NetSpira Networks)
Nick McKeown (Stanford University)

Dirección de contacto:

Pablo Molinero Fernández
C/Batalla de Garellano, 26
28023 Madrid
+34-91-307 9194 (a partir del 25 de junio)
+1-650-315 3376 (hasta el 25 de junio)
molinero@stanfordalumni.org

Áreas temáticas:

Área II: Sistemas y Tecnologías de Red
Área III: Internet

El control de conmutadores de circuitos en el núcleo de la Internet por medio de flujos de usuario

Hoy en día, el núcleo de la Internet contiene numerosas regiones que usan conmutación de circuitos para transmitir la información. Estas regiones forman una red de transporte compuesta por conmutadores SDH/SONET y DWDM que interconectan los grandes encaminadores centrales (core routers) de los proveedores de servicio de Internet. Actualmente, esta red de transporte no está integrada con el resto de la Internet, que esta basada en IP - un protocolo de conmutación de paquetes. Usualmente, una unidad de negocio distinta de la que provee servicios de Internet o de telefonía opera la red de transporte. En consecuencia, los circuitos del núcleo son dimensionados manualmente basándose en previsiones sobre el volumen y crecimiento del tráfico en la Internet. Para cambiar la capacidad de un circuito o crear uno nuevo se requiere la intervención de varias personas y el proceso suele tardar días, e incluso semanas. Así, la red de transporte reacciona muy lentamente y con retraso frente a eventos imprevisibles y externos, como por ejemplo las avalanchas de tráfico creadas por una noticia trágica e inesperada, por un fallo en un elemento de la red, por una nueva versión de un popular programa o por un nuevo "hit" de un grupo musical famoso. Esta enorme inercia del sistema hace que la red esté muy sobredimensionada para evitar atascos y que, por tanto, el uso de la red sea altamente ineficiente.

Frente a este escenario actual, numerosos investigadores y compañías han propuesto dimensionar los circuitos del núcleo de manera más dinámica para que reaccionen frente a cambios significativos en los patrones de tráfico. Las cuestiones a resolver son, primero, cómo determinar la capacidad necesaria para cada circuito y, segundo, como señalizar a los conmutadores los cambios requeridos. Actualmente, varios grupos están estudiando esta segunda cuestión. Estos grupos, como el IETF, el OIF o la ITU-T, están estandarizando la señalización entre los conmutadores de circuitos, pero no especifican los algoritmos de dimensionado la red y dejan total libertad a los fabricantes para determinarlos. Esta ponencia se centra en como dimensionar la red de transporte en tiempo real.

Los conmutadores de circuitos del núcleo tienen dos limitaciones que hay que tener en cuenta. La primera limitación es que la mayoría de los conmutadores sólo conmutan circuitos que son múltiplos exactos de una granularidad mínima (típicamente 51 Mbit/s para conmutadores SDH-SONET, y 2.5 Gbit/s para conmutadores DWDM). La segunda limitación es que los conmutadores tienen una cierta latencia para el establecimiento de circuitos que puede llegar a ser de varios cientos de milisegundos. Esta latencia puede ser por varias causas como el uso de tecnologías de conmutación que requieren el movimiento de partes electromecánicas (como espejos MEMS) o protocolos de señalización que requieren una confirmación de todos los nodos intermedios para garantizar el establecimiento del canal.

Una posible solución que rápidamente viene a la mente para dimensionar los circuitos es el monitorizar las llegadas de paquetes o el tamaño de las colas en los encaminadores. La mayor parte de los fabricantes ya tienen soporte, al menos parcial, para estas dos soluciones. Sin embargo, las llegadas de paquetes tienen numerosas dependencias a corto y largo plazo debido al multiplexado de flujos y al control de congestión de TCP. Estas dependencias hacen que las medidas oscilen mucho y que contengan mucho ruido, lo que dificulta enormemente la observación de tendencias en el tráfico. Estas oscilaciones además incrementan significativamente la señalización al estar reajustando continuamente la capacidad de los circuitos. Por último, su interacción con los algoritmos de control de congestión de TCP no se comprende del todo y está todavía por estudiar. Indudablemente, es posible aplicar filtros temporales para reducir el ruido y estabilizar la medida. Sin embargo, estos filtros retrasan la toma de decisiones y ralentizan la respuesta del sistema, algo que es preferible evitar.

Esta ponencia aborda una manera distinta de dimensionar los circuitos del núcleo. Proponemos monitorizar los flujos de IP como una manera sencilla y precisa de determinar las capacidades necesarias para la red de transporte. Ciertamente, el proceso de llegadas de flujos es mucho menos variable que las llegadas de los paquetes y tiene autocorrelaciones muy débiles. Una ventaja de nuestra propuesta es que típicamente hay una serie de intercambios de mensajes antes de que un flujo de IP comience a transmitir los datos de la aplicación. Así, los flujos de IP dan un aviso por anticipado de los incrementos de tráfico, en lugar de observar los atascos cuando ya está sucediendo, cuando las acumulaciones y perdidas de información son ya inevitables. Nuestra propuesta se basa en que nuestro uso de la Internet está muy orientado a la conexión, como demuestra el que consistentemente más del 90% del tráfico use TCP.

El monitor de los flujos de IP determina cuando empiezan y terminan los flujos (utilizando un contador de actividad) y estima la tasa máxima de transferencia del flujo. Esta tasa parece difícil de estimar *a priori*, pero, en general, está determinada por el enlace de acceso (p.ej. un MODEM de 56 Kbit/s o una línea de DSL a 2 Mbit/s), que es muy estático y tarda varios meses o años en ser cambiado. Por eso, proponemos utilizar una pequeña base de datos local para estimar la tasa máxima de transferencia basándonos en nuestro conocimiento de la red de acceso y la historia pasada. El soporte físico (hardware) necesario para monitorizar los flujos de IP es equivalente a los clasificadores utilizados en los conmutadores de Ethernet, pero utilizando un ancho de 64 bits (dos direcciones IPv4 concatenadas) en lugar de 48 bits. De hecho, ya hay clasificadores similares para paquetes de IP que están disponibles comercialmente para velocidades de 10 Gbit/s [1,2].

Como mencionábamos anteriormente, los conmutadores pueden tener latencias que no son despreciables. Para compensar estas latencias, proponemos usar unas salvaguardas que incrementan la capacidad de los circuitos de manera que los posibles incrementos en el tráfico puedan ser absorbidos con una alta probabilidad mientras que creamos un nuevo circuito. En esta ponencia, estimamos el tamaño de estas salvaguardas usando trazas reales del núcleo de Sprint, uno de los tres mayores proveedores de servicios de Internet en los EEUU. Además, proponemos un modelo

matemático basado en la teoría de colas para estimar las salvaguardas usando tan sólo unos pocos parámetros estadísticos del tráfico. Concretamente, sólo requerimos la tasa de llegada de nuevos flujos de IP y la distribución de la duración y el ancho de banda de estos. La Figura 1 muestra la probabilidad de sobrecarga al usar las salvaguardas que proponemos para distintas latencias para la creación-ampliación de circuitos. La figura muestra los resultados al usar las trazas de Sprint (línea continua) y nuestro modelo (línea discontinua).

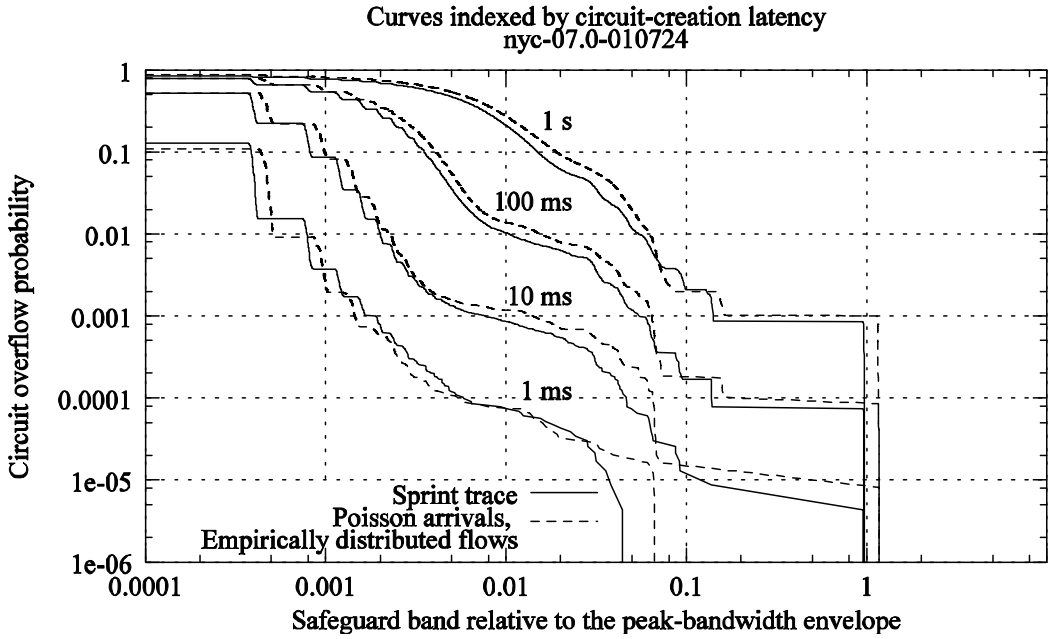

Figure 1.- Probabilidad de sobrecarga de los circuitos frente a distintas salvaguardas para las latencias de creación y ampliación de circuitos de 1ms, 10ms, 100ms y 1s.

Nuestro algoritmo de control es muy genérico y puede ser utilizado con varios mecanismos de señalización que están siendo estandarizados en estos momentos, como GMPLS del IETF [3], UNI del OIF [4] o LCAS y ASTN de la ITU-T [5,6]. Una posible ampliación de nuestra propuesta es el muestreo de tan sólo los flujos de mayor tamaño (en términos del tráfico generado), puesto que estos representan menos del 10% de los flujos, pero transportan más del 90% de los bytes transmitidos.

Este trabajo forma parte de la tesis de doctorado que Pablo Molinero Fernández ha realizado en la universidad de Stanford en California, EE.UU..

Referencias

[1] Applied Micro Circuits Corporation, “AMCC nPX5700, 10 Gbps Traffic Manager/ Switch Fabric”, 2003. http://www.amcc.com/cardiff/docManagement/displayProductSummary.jsp?prodId=nPX5700.

[2] EZchip, Technologies. “NP-1, OC-192 Network Processor”, 2003. http://www.ezchip.com/html/pr_np-1.html.

[3] A. Banerjee, J. Drake, J. Lang, B. Turner, D. Awduche, L. Berger, K. Kompella, y Y. Rekhter. “Generalized Multiprotocol Label Switching: An overview of signaling enhancements and recovery techniques”. IEEE Communications Magazine, 39(1):144–150, enero 2001.

[4] G. Bernstein, B. Rajagopalan, y D. Spears. “OIF UNI 1.0 — controlling optical networks: A White paper”. Optical Internetworking Forum, 2001.

[5] ITU Telecommunication Standardization Sector. “Link capacity adjustment scheme (LCAS) for virtual concatenated signals”. Unión Internacional de Telecomunicaciones, Recomendación G.7042/Y.1305, febrero 2003.

[6] ITU Telecommunication Standardization Sector. “Architecture for the Automatic Switched Transport Network (ASTN)”. Unión Internacional de Telecomunicaciones, Recomendación G.807/Y.1302, noviembre 2001.