

Modelado de Tráfico

Una estrategia inteligente contra la congestión de una red

Daphne I. Bouche B.

Universidad Tecnológica de Panamá, estudiante

Resumen - Al momento de implementar una red de telecomunicaciones, no sólo se necesita elegir el protocolo adecuado según el tipo de datos y la velocidad de transmisión que ha de manejar, sino que también se debe tomar en cuenta qué mecanismos de control de tráfico llevará a cabo la red. De hecho, este aspecto es crucial a la hora de aplicar la calidad de servicio que se ofrecerá a los usuarios de la misma.

El modelado de tráfico (Traffic Shaping) es una de las técnicas de control de tráfico que ha dado grandes resultados en disminuir la congestión en redes WAN. Ésta hace uso de algoritmos como el Leaky Bucket y el Token Bucket. La clave de su efectividad en el control de la congestión consiste en retrasar los paquetes antes de introducirlos en la red, con la finalidad de mantener la tasa de bits dentro de un perfil establecido. Aunque requiere de un procesamiento adicional con respecto a otras técnicas que recurren al descarte de paquetes, presenta ventajas que la posicionan como uno de los mecanismos más efectivos contra la congestión.

Palabras claves – Congestión, calidad de servicio, modelado de tráfico, Traffic Shaping.

1. Introducción

La congestión del tráfico de datos es uno de los problemas más relevantes para los administradores de redes, ya que deben buscar soluciones que sean lo suficientemente eficientes para prevenirla o combatirla cuando ya existe.

Existen un sinnúmero de definiciones sobre congestión, pero todas coinciden en que es un fenómeno que ocurre en situaciones de alta carga de tráfico en la red. En las redes de conmutación de paquetes sin reserva de recursos para una conexión emisor-receptor, la velocidad a la que arriban los paquetes a un nodo puede exceder la velocidad de servicio (que depende del coste del procesamiento y la velocidad del enlace de salida). Como consecuencia, los paquetes se tardarán en llegar a su destino al ser puestos en una cola o búfer. Llegará un momento en que la capacidad de la cola del nodo superará su propia capacidad y se perderán paquetes que no puedan ser acomodados en la cola. El retardo excesivo o la pérdida de paquetes, producirá que los nodos emisores retransmitan los paquetes que se tardaron o perdieron en el camino. Esta situación agrava aún más el problema al aumentar la congestión debido a las retransmisiones. Es por ello que se hace necesario dotar a los nodos de la red de un mecanismo que controle el flujo del tráfico entrante para mejorar su desempeño, y que se vea menos afectada por los cuellos de botella.

Para comprender mejor el comportamiento de la red ante el tráfico que se le ofrece, observe la Figura 1. En ella, la curva 1

representa el comportamiento ideal de la red que indica una relación directamente proporcional entre el tráfico cursado y el ofrecido, y se mantiene así hasta llegar a la capacidad máxima nominal de la red, donde empieza la congestión. La curva 2 representa el comportamiento real de la red, donde se observa que, cuando la red llega a la congestión y el tráfico ofrecido sigue aumentando, habrá menos tráfico en curso, lo que indica una degradación en el servicio. La curva 3 representa el comportamiento de la red cuando se lleva a cabo un control de congestión. Puede observarse que dicha curva se aproxima al comportamiento ideal, y se resuelve el problema de la degradación en las prestaciones de la red.

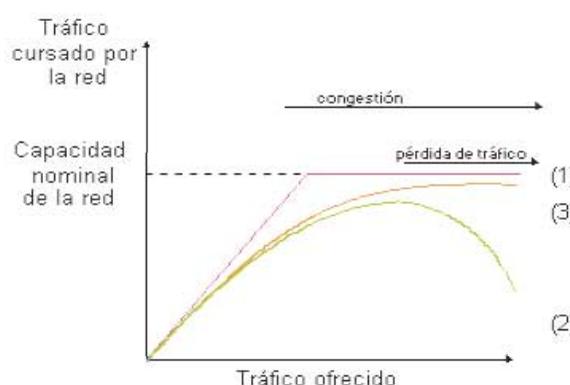


Figura 1: Caudal de la red en función del tráfico ofrecido [1]

Son varios los mecanismos de control de tráfico que tratan de solventar el problema de la congestión. Estos mecanismos pueden ser clasificados, según el momento en que actúan contra la congestión, en dos tipos: soluciones en bucle abierto y soluciones en bucle cerrado.

Las soluciones en bucle abierto, también llamadas soluciones pasivas, se implementan durante el diseño de la red y están enfocadas a combatir la congestión antes de que ocurra. Para implementarlas, el administrador de la red debe jugar con ciertos parámetros que influirán en el comportamiento de la misma ante situaciones de congestión. Entre estas variables están el tiempo límite de cuenta de los temporizadores en el emisor que controlan la retransmisión, el tiempo de vida de los paquetes (TTL-time to live), las políticas de manejo de tráfico, como el de descarte de paquetes, el control de flujo, el establecimiento de prioridades, etc. También es importante la elección de un algoritmo de enrutamiento adecuado.

Las soluciones en bucle cerrado, también llamadas soluciones reactivas o activas, son las que actúan cuando se detecta congestión. Estas clases de mecanismos se llevan a cabo en tres fases:

- Fase de monitorización: En la que se detecta dónde y cuándo ocurre congestión. Para ello, se vigila la ocupación de los enlaces, los tiempos de retardos, la variación en la latencia del arribo de paquetes (jitter), la cantidad de paquetes retransmitidos y descartados, además de otros parámetros que indiquen congestión.
- Fase de reacción: Se envía información a los nodos donde se pueda combatir la congestión. Para esto, se alerta sobre el problema a dichos nodos mediante paquetes especiales que se saltan las colas (un ejemplo es Frame Relay).

- Fase de ajuste de la red: Se llevan a cabo las operaciones necesarias para controlar el problema, como la reducción de la velocidad de transmisión de los nodos emisores, el descarte de paquetes que llegan a una cola llena o la prohibición de nuevas conexiones.

Dentro de los mecanismos de control de congestión en bucle cerrado, se pueden mencionar los de descarte de paquetes y paquetes reguladores, y de los mecanismos en bucle abierto, está la técnica de Traffic Shaping. Vamos a centrar el tema en esta última técnica, puesto que, además de ayudar a combatir la congestión en la red de una forma eficiente, permite al Proveedor de Servicios de Internet la aplicación de Servicios diferenciados para cada tipo de usuario, con el fin de garantizar una calidad de servicio deseada [1].

2. Modelado de Tráfico

Como se mencionó anteriormente, el modelado de tráfico (*Traffic Shaping*) es un mecanismo que busca resolver el problema de congestión en la red antes de que éste se produzca.

Permite a la red tener conocimiento de las características del tráfico que se desea transmitir, para que sea capaz de decidir si puede transmitirlo según se encuentre o no dentro de un determinado perfil de tráfico. El emisor debe informar a la red sobre las características del flujo que transmite, así como la calidad de servicio que desea recibir, según la aplicación de la cual se trate (video, voz, datos, etc.).

El perfil de tráfico se refiere a si el flujo de datos entrante a un nodo se encuentra dentro de un límite de tasa de transferencia preestablecido, previo acuerdo entre el usuario y el proveedor de servicios. Si este flujo se pasa del límite de la tasa especificada, pueden darse una de dos situaciones para controlar la tasa de bits: descartar los paquetes o ser procesados por un modelo de tráfico para colocar el flujo en el perfil correcto.

Podría confundirse la operación del modelado de tráfico con el control de flujo inherente que lleva a cabo TCP, pero existe la diferencia en que TCP es un protocolo basado en ventana que limita la cantidad de paquetes en tránsito, pero no la velocidad a la que entran en la red. *Traffic Shaping* sí regula la tasa de transferencia a la que los paquetes son introducidos en la red. Este mecanismo puede ser implementado a través de los algoritmos *Leaky Bucket* y *Token Bucket*. También se pueden mencionar los modelos *D-BIND*, *Double Leaky Bucket* o modelos multiparámetros, que son variaciones de los dos primeros [2].

2.1. Modelos de Tráfico para el Control de Congestión

El modelado de tráfico, básicamente, puede ser implementado a través de dos algoritmos de control en Bucle Abierto: el *Leaky Bucket* y el *Token Bucket*, además de unas variaciones de estos. Veamos en qué consisten estos algoritmos a continuación [3].

2.1.1 Modelo Leaky Bucket. Este modelo es el más utilizado para describir tráfico (definido por el ATM Forum). Su nombre, en inglés, significa "cubeta agujereada" (*Leaky*-que gotea, *Bucket*-cubeta) debido a que su funcionamiento se basa en la analogía de pasar el tráfico por una cubeta con goteo, como se presenta en la Figura 2.

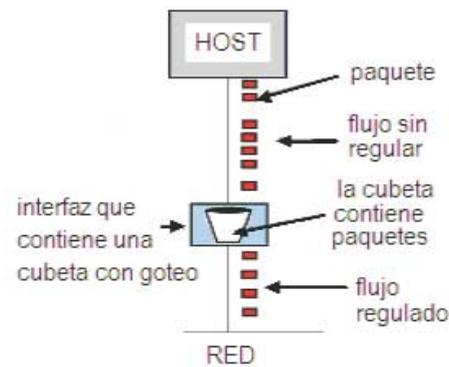


Figura 2: Funcionamiento del algoritmo Leaky Bucket [3]

Este algoritmo tiene como objetivo moderar las ráfagas de flujo de datos que un terminal envíe a la red, convirtiendo el flujo desigual de tráfico en un flujo continuo. Dos parámetros son importantes para describir la operación de este modelo: la capacidad de la cubeta s (bits) y la tasa de drenaje r (bits/s). Para comprender mejor su funcionamiento, piense en una cubeta agujereada; aunque veríamos poco o mucho caudal de líquido en la cubeta, la salida del agua estará limitada por los agujeros, y el caudal de salida será distinto al de entrada, siendo menor. Si llenamos muy de prisa la cubeta con agua, se llenará y llegará un punto en que el agua se derramará por arriba. Esto es precisamente lo que ocurre cuando se utiliza *Leaky Bucket*. Si se tiene una ráfaga lista para transmitir, el algoritmo no transmite todo de seguido. El proveedor del servicio garantiza al usuario que se cursará todo el tráfico si se transmite a una velocidad determinada, y si se sobrepasa, se descartarán paquetes (análogo a que se desborde el agua por arriba de la cubeta). De esta forma, se limita la tasa de transmisión del tráfico al valor r , que representa la tasa media de transmisión del emisor. El valor de s (capacidad de la cubeta) se debe determinar de manera que no se descarten paquetes. Básicamente el *Leaky Bucket* consiste en una cola finita (cubeta). Mientras existan paquetes en la cubeta, se transmite un paquete en cada pulso de reloj. Para que no se descarten paquetes, la fuente no deberá transmitir en el periodo $[0, t]$ más de $s+rt$ bits. Este modelo es utilizado principalmente en redes ATM.

2.1.2 Modelo Token Bucket. La desventaja que tiene *Leaky Bucket* es que impone una tasa promedio de transmisión, que no puede ser superada. Esta es la ventaja que tiene el modelo *Token Bucket* (definido por la IETF – *Internet Engineering Task Force*), que permite la transmisión de picos de tráfico durante periodos pequeños.

El nombre del algoritmo significa, en Inglés, "cubeta de fichas" debido a la manera en que funciona. Ésta es presentada en la Figura 3.

El funcionamiento del modelo *Token Bucket* se basa en la idea de una cubeta que contiene fichas (*tokens*), que se insertan en ésta cada r segundos. La cubeta permite un máximo de b fichas. Para poder transmitir, el emisor debe tomar una ficha y consumirla. Mientras existan fichas en la cubeta, la fuente podrá transmitir tráfico a una velocidad deseada, hasta un límite de tasa máxima de b . Si la cubeta se vacía, se debe esperar a que lleguen nuevas fichas, y la velocidad de transmisión vuelve a ser r , que representa

la tasa de tráfico continua que sostiene el sistema. La regla general de *Token Bucket* dicta que la cantidad de información transmitida no debe exceder al valor $rt+b$, para cualquier periodo de tiempo t . Algunas veces se establece un límite de velocidad de transmisión, llamada tasa pico p . De esta forma, el tráfico no puede ser superior a $\min [pt, rt+b]$, lo que significa que, si $rt + b$ es mayor que pt , ésta última impondrá la tasa de transmisión a la cual se enviarán los datos.

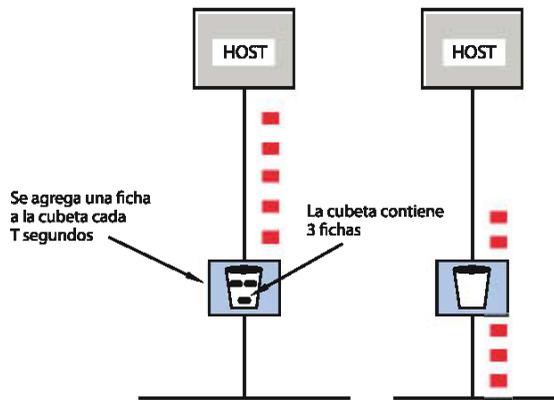


Figura 3: Funcionamiento del algoritmo Token Bucket [3]

2.2. Importancia de la Aplicación del Modelado de Tráfico en Internet

Como hemos visto, la principal razón por la que se utiliza la técnica del perfilado de tráfico es para prevenir el problema de la congestión, con lo que se lograría brindar una calidad de servicio requerida por los usuarios de la red.

Uno de los aspectos que deben tener en cuenta los administradores de una red de Internet es la manera en que se asignarán prioridades a algunas aplicaciones sobre otras, o a ciertos usuarios en detrimento de otros. Se dice entonces que la red debe ser capaz de diferenciar servicios, es decir, dar a ciertos paquetes un mejor o especial trato con respecto a otros. Las aplicaciones en tiempo real necesitan una mayor calidad de servicio en cuanto al ancho de banda y al retraso, puesto que no son tolerantes a los retardos como lo son las aplicaciones de navegación de Internet o transferencia de archivos, por ejemplo. La aplicación de calidad de servicio no sólo es necesaria para garantizar un rendimiento eficiente de la red, sino también para brindar a los clientes una variedad de planes de contrato, de acuerdo a la velocidad ofrecida. Pero, ¿cómo puede el Proveedor de Servicios controlar la asignación de ancho de banda para los diferentes tipos de usuarios de los recursos de la red? Una de las maneras más eficientes es, precisamente, mediante el uso de *Traffic Shaping* [4]. Este es parte de la arquitectura *DiffServ* (*Differentiated Services*), que es un modelo destinado a proveer de diferentes niveles de calidad de servicio a las aplicaciones o usuarios que utilizan los recursos de la red. Esta técnica es una de las que se utiliza dentro del componente de condicionamiento, que es el responsable de clasificar, marcar, medir y modelar el tráfico que ingresa a la red. El funcionamiento de *Traffic Shaping* dentro de este componente se presenta en la Figura 4.

Básicamente lo que hace *Traffic Shaping* dentro de este

componente es clasificar primero los paquetes de acuerdo a un perfil específico de tráfico, indicado en las tramas IPv4 en un campo denominado ToS (*Type of Service*), o en tramas IPv6, indicado en el campo DSCP (*Differentiated Services Code-Point*). Luego, encola los paquetes clasificados en colas de tipo *Leaky Bucket* o *Token Bucket*, para darles el perfil requerido para la transmisión. Se extraen los paquetes encolados normalmente a una tasa de bits inferior a la que tiene el ancho de banda de la interfaz de salida. Si hay algún tráfico que no se encuentre dentro del perfil, no es sometido al proceso de clasificación y pasa directamente a la última cola de salida, que puede ser del tipo FIFO (*First in First out*), WFQ (*Weighted Fair Queuing*), CQ (*Custom Queuing*), PQ (*Priority Queuing*), entre otras.

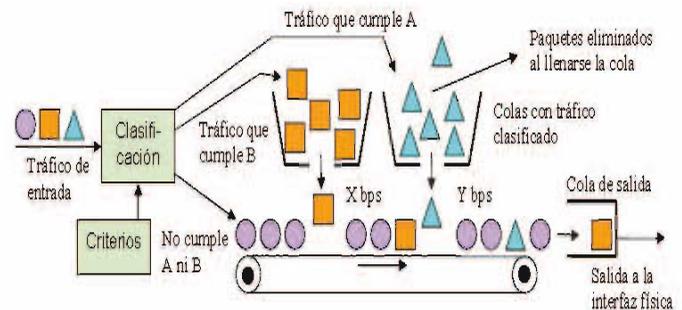


Figura 4: Utilización del perfilado de tráfico en la aplicación de calidad de servicio mediante la arquitectura *DiffServ* [4].

2.3. Vigilancia de Tráfico o Modelado de Tráfico, ¿cuál es mejor?

Además de *Traffic Shaping*, existen otros mecanismos utilizados para controlar los tráficos a ráfagas típicos de Internet. Uno de los más comúnmente utilizados es el de vigilancia de tráfico, conocido como *Policing*. Antes de realizar una comparación de esta técnica con la del modelado de tráfico, se explicará brevemente cómo funciona [5][6].

Policing es un ejemplo de estrategia de control de tráfico que se utiliza también para limitar ancho de banda de ciertas aplicaciones, según sus prioridades, pero, a diferencia de *Traffic Shaping*, no utiliza colas para almacenar los paquetes sino que los descarta (o marca) si superan la tasa máxima permitida.

Un ejemplo de la estrategia de *Policing* es la de Tasa de Acceso Comprometido (CAR - *Committed Access Rate*). Esta técnica (desarrollada por CISCO) está basada en el algoritmo *Token Bucket*, con una variación adicional denominada Sistema de Deudas, que consiste en que, si hay fichas en la cubeta, el paquete sigue, sino, es descartado. Debido a esto, el tráfico de salida tendrá una forma de diente de sierra. Al descartar paquetes, se logra limitar la tasa de envío de datos en una red que use TCP, puesto que el protocolo, al detectar pérdidas de paquetes, lleva a cabo un control de flujo de datos a través de regulación por ventana (en alerta de congestión, hay reducción de su ventana de congestión).

Traffic Shaping, por el contrario, trata de evitar el descarte de paquetes, almacenando el tráfico que no puede ser transmitido y enviándolo más tarde. Se comporta como un *Leaky Bucket*. Debido a esto, la tasa de salida es constante y el tráfico es "suavizado". El comportamiento del tráfico de salida al utilizar estas dos técnicas se presenta en los gráficos de la Figura 5.

En ella se puede observar cómo *Traffic Shaping* suaviza la curva

del tráfico de salida para mantener la tasa de transferencia por debajo del límite máximo establecido, mientras que *Policing* recorta abruptamente los picos de tráfico que sobrepasan este límite.

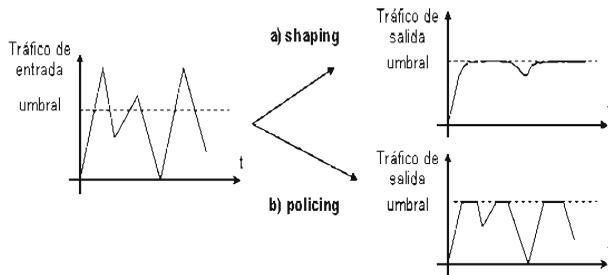


Figura 5: Comparación del tráfico de salida al usar las estrategias de control de tráfico *Traffic Shaping* y *Policing* [5].

En la Tabla 1 se presenta un resumen detallado de las características de las técnicas de *Policing* y *Traffic Shaping*, y las ventajas y desventajas de cada una.

Sería incorrecto decir que una técnica es mejor que la otra, puesto que no son mecanismos excluyentes, y dependen de las aplicaciones que se están manejando. La técnica de vigilancia de tráfico o *Policing* se aplica tanto a tráfico de entrada como de salida de una interfaz, en tanto que la de modelado de tráfico es aplicada sólo a flujos de salida. Además, es importante tener conocimiento de las funcionalidades de los protocolos que se utilicen en la red, para saber si la estrategia que se implementará es la adecuada. Por ejemplo, las aplicaciones que utilicen protocolos UDP (*User Datagram Protocol*) que no cuentan con control de flujo (entrega no confiable de paquetes), no podrán retransmitir paquetes descartados si se utiliza *Policing*; en este caso la misma aplicación tendría que recuperar los datos perdidos. En el caso de TCP, este sí cuenta con un control de flujo, permitiendo retransmisión de paquetes descartados. De esta forma, a una aplicación FTP (soportada por TCP) no le interesa la pérdida de paquetes, sino la velocidad a la que se transmiten.

Tabla 1. Comparación de las estrategias *Policing* y *Shaping* [6]

Criterio	Traffic Shaping	Policing
Objetivo	Almacenar temporalmente los paquetes que superan velocidades límites.	Eliminar los paquetes que superan velocidades límites.
Refresco	Las tasas de bits se evalúan en intervalos. Se configura en bits por segundo.	Funcionamiento continuo. Se configura en bytes.
Colas	CQ, PQ, WFQ, etc.	No se usan
Efecto sobre ráfagas	Suaviza los cambios de tráfico tras varios intervalos	No se alteran las ráfagas.
Ventajas	Si no hay exceso de tráfico, no elimina paquetes.	Evita los retardos de paquetes en colas.
Desventajas	Puede introducir retardos en los paquetes, sobre todo con colas grandes.	Al descartar paquetes, TCP disminuye su ventana, lo que disminuye el rendimiento.

3. Comentarios Finales

La utilización de técnicas de control de congestión es un requisito fundamental para una red de recursos compartidos como lo es Internet, ya que los Proveedores de Servicio deben garantizar a los usuarios una buena calidad de servicio según el acuerdo establecido previamente en cuanto a la tasa de transmisión de los datos. Si bien es cierto que es deseable contar con una alta velocidad de transmisión que nos permita utilizar todo el ancho de banda posible del enlace, esto podría ser causa de un problema más grave de congestión, llevando a la red a situaciones de retardos mayores que si se cursaran los paquetes a una velocidad moderada, pero con continuidad en la transmisión. De esta forma, *Traffic Shaping* intenta controlar el tráfico para lograr un mejor rendimiento en las prestaciones de la red, aunque implique un retardo de procesamiento, el cual es casi imperceptible.

Referencias

- [1] (1998) Sitio web de la Universidad Carlos III, Redes y Servicios de Comunicaciones. Apuntes, "Tema 6. Control de Congestión". [Online]. Disponible en: <http://www.it.uc3m.es/~prometec/rsc/apuntes/Conges/conges.html>
- [2] J. L. Marzo, "Control de Tráfico en redes de altas prestaciones: ATM e Internet de nueva generación", Lecture Notes in Computer Science, Departamento de Electrónica Informática y Automática, Universidad de Girona, España. [Online]. Feb. 2005. Disponible en: http://eia.udg.es/~marzo/doctorat/ctav_v00.pdf
- [3] C. Vega, D. Arvizu, A. García, "Algoritmos de Encriptación de Datos", ser. Libros y Manuales: Informática, Unidad Multidisciplinaria CIET. España. 2008. [Online]. Disponible en: <http://www.eumed.net/>
- [4] A. J. Vargas O. "Aplicación de Calidad de Servicio (QoS) para Transmitir Voz sobre IP en Redes de Área Amplia (WAN's)", Trabajo de grado, Ciudad Universitaria Rodrigo Facio, Costa Rica, Dic. 2008.
- [5] S. Avellanet e Ing. R. Ingver, "CAR vs Shaping", Curso de Evaluación de Performance. 2004. [Online]. Disponible en: http://ie.fing.edu.uy/ense/assign/perfredes/.../CARvsShaping/CAR_Shaping.ppt
- [6] Cisco, "Comparing Traffic Policing and Traffic Shaping for Bandwidth Limiting". [Online]. Disponible en: http://www.cisco.com/en/US/tech/tk543/tk545/technologies_tech_note09186a00800a3a25.shtml.com