

Tutorial en video: Técnicas de implementación de QoS (9 min)

Podemos categorizar la implementación de las herramientas de QoS en tres categorías generales: herramientas de clasificación y marcado, herramientas de prevención de congestión y herramientas de administración de congestión. En las herramientas de clasificación y marcado, el tráfico de sesión se clasifica en diferentes paquetes y agrupaciones de prioridad que aparecen marcados. En las herramientas de prevención de congestión, se asignan recursos de la red a las clases de tráfico. Parte del tráfico se puede descartar de manera selectiva, se puede demorar o se puede volver a marcar para evitar la congestión. La detección temprana aleatoria y ponderada regula el tráfico de datos del protocolo TCP de manera eficaz según el ancho de banda antes de que las caídas de cola ocasionadas por la cola se desborden y, básicamente, se produzca la congestión.

En las herramientas de administración de congestión, cuando el tráfico excede los recursos de la red disponibles, o se produce congestión, el tráfico se coloca en cola para esperar la disponibilidad de dichos recursos. Las herramientas de administración de congestión basadas en CISCO IOS incluyen la puesta en cola de espera equitativa y ponderada basada en clases y la puesta en cola de baja latencia, o los algoritmos de cola de CBWF y de cola de LLQ. Si observamos este diagrama, puede ver los tres conjuntos de herramientas utilizados en secuencia. Desde la interfaz de ingreso en un router hasta la salida de la interfaz del router, puede ver que el tráfico primero se clasifica y luego se marca.

Observe en el gráfico cómo los paquetes, antes de que se clasifiquen, se muestran aquí en color gris. A medida que se clasifican en diferentes clases, se marcan en el encabezado IP indicado aquí con los colores de los diferentes paquetes. La prevención de congestión es donde se realiza la vigilancia y el descarte selectivo del tráfico. Aquí es donde se asignan diferentes prioridades a los recursos de la red. Los paquetes se colocan en distintas colas según diferentes pesos o tamaños. La herramienta principal para la prevención de congestión que se utiliza en los dispositivos de Cisco es la detección temprana aleatoria y ponderada o W-RED. Si los búferes de cola se llenan, los paquetes pueden descartarse de manera proactiva para evitar la congestión. Por último, tenemos la sección de administración de congestión donde los paquetes se colocan en sus colas, y se programan y se modelan en función de las prioridades o políticas de la red que se hayan aplicado.

Si observamos el marcado de tráfico de QoS un poco más de cerca, podemos ver que el marcado de tráfico se produce en la capa dos y la capa tres. En la capa dos, tenemos una clase de servicio, o CoS, que se aplica con los protocolos de calidad de servicio de etiquetado de VLAN 802.1Q, y 802.1P. El marcado de tráfico en la capa 2 se produce en la red de área local, y se limita a la LAN mediante los switches de la capa dos. En Wi-Fi, el marcado de tráfico se produce con el campo identificador de tráfico de Wi-Fi, o TID. Con el MPLS, también cuenta con el campo experimental o EXP, también en la capa dos. En la capa 3 con IPv4 e IPv6, utilizamos los campos de TOS y de clase de tráfico mediante la prioridad IP o IPP, y el DSCP, el punto de código de servicios diferenciados. Con el marcado de tráfico de la capa tres, el marcado puede ir de un router a otro y, por lo tanto, el marcado de QoS puede realizarse de forma completa en toda la red.

En esta diapositiva, puede ver cómo se produce el marcado de QoS en la capa dos. Tenemos una trama de Ethernet, y el protocolo 802.1Q puede insertarse entre la dirección de origen y el tipo de campo de longitud. Puede ver los dos campos introducidos en la trama de Ethernet, el campo TPID y el campo TCI, o campo de información de control de etiqueta, que contiene un campo de prioridad de PRI de tres bits que identifica la clase de servicio o el marcado de CoS. Los marcados de la clase de servicio tienen seis valores diferentes: mejor esfuerzo, datos de prioridad media, datos de alta prioridad, señalización de llamada, videoconferencia y portador de voz para el tráfico de voz. Los valores de bits seis y siete se reservan para la información de control de la red.

En la capa tres, la QoS se implementa en los encabezados de paquete de IPv4 e IPv6 en el campo de tipo de servicio y el campo de clase de tráfico. En la implementación anterior, la clase de servicio en la capa tres se implementaba utilizando la IPP y el TOS. La prioridad IP es de tres bits y el tipo de servicio es de cinco bits. Puede verlos allí en los ocho bits. Las clasificaciones tienen ocho niveles, que se asignan a la clase de servicio de la capa dos. Allí puede ver la rutina, la prioridad, el Flash inmediato, la anulación de Flash y la información fundamental. Nuevamente, los valores de seis y siete se reservan para la información de control de la red.

La implementación actual de la QoS utiliza los campos de punto de código de servicios diferenciados y de notificación de congestión explícita, o DSCP y ECN. El DSCP y la ECN se asignan directamente en los ocho bits de los campos de IPP y ToS. El campo de DSCP es retrocompatible con el campo de IPP. En este ejemplo, puede ver que una clasificación de IPP de cinco incluye este valor de aquí y este de aquí. Si tiene IPP y tiene tres bits, entonces sería un valor de cuatro y uno, que correspondería a un cinco, lo que indicaría un flujo crítico. Dentro de la especificación DSCP, tenemos seis bits. Uno, dos, tres, cuatro, cinco, seis. Este es un 32 y este un ocho. 32 más 8 es 40, por lo que puede ver cómo se asignaría un valor de cinco en el IPP mediante tres bits o un valor de 40 con el DSCP utilizando seis bits. Dado que el DSCP utiliza seis bits, hay 64 posibles clasificaciones de tráfico.

El marcado de QoS se realiza en los límites de la confianza. El marcado debe realizarse lo más cerca posible del origen. Los terminales como teléfonos IP de voz, puntos de acceso inalámbricos y gateways de videoconferencia tienen la capacidad de marcar paquetes de QoS, como se indica en el número uno. El tráfico también puede marcarse mediante los switches de la capa dos, como en el número dos, o en la capa tres utilizando los switches y routers de la capa tres. Como los paquetes viajan a través de la red, el remarcado del tráfico puede ser necesario, el remarcado de los valores de la clase de servicio, o los valores de prioridad IP o valores de DSCP.

El mecanismo principal para la prevención de congestión en el QoS en los routers y switches de Cisco es la detección temprana aleatoria ponderada o W-RED. Con la detección temprana aleatoria y ponderada, un porcentaje de los paquetes se descarta una vez que hayamos alcanzado el umbral mínimo en una cola. Si la cola supera el umbral máximo, los paquetes se descartan automáticamente. En esta imagen, puede ver la diferencia entre la vigilancia y el modelado del tráfico. En la vigilancia de tráfico, cuando la velocidad de tráfico alcanza la velocidad máxima configurada, se descarta el tráfico excedente. El resultado es una velocidad de salida que tiene una apariencia similar a la de los dientes de una sierra, con crestas y canales, como se muestra en el gráfico en la parte superior. Con el modelado del tráfico, los paquetes en exceso se conservan en una cola y luego se reprograman para la transmisión posterior en incrementos de tiempo. El resultado del modelado del tráfico es una velocidad de salida fluida de paquetes, como se muestra en el gráfico en la parte inferior.