TRƯỜNG ĐẠI HỌC ĐÀ LẠT

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO QOS**

**NHÓM: 1**

Giáo viên hướng dẫn: Thầy Nguyễn Văn Huy Dũng

Sinh viên thực hiện: 2111903 – Nguyễn Nhật Trường

2111860 – Nguyễn Hoàng Bảo Long

2111886 – Nguyễn Quốc Thắng

***Đà Lạt, tháng 11 năm 2024***

**MỤC LỤC**

[**CHƯƠNG I: TỔNG QUAN LÝ THUYẾT** 3](#_Toc183760019)

[**1.1.** **QoS là gì?** 3](#_Toc183760020)

[**1.2.** **QoS hoạt động như nào?** 3](#_Toc183760021)

[**1.3.** **Các thành phần của QoS** 4](#_Toc183760022)

[**1.4.** **Các kỹ thuật đảm bảo QoS** 5](#_Toc183760023)

[**1.4.1.**  **Tránh mất gói tin (Avoiding Packet Loss)** 5](#_Toc183760024)

[**1.4.2.** **Các công cụ QoS** 6](#_Toc183760025)

[**1.4.3.** **Phân loại và đánh dấu** 6](#_Toc183760026)

[**1.4.4.** **Đánh dấu ở lớp 2** 7](#_Toc183760027)

[**1.4.5.** **Đánh dấu ở lớp 3** 7](#_Toc183760028)

[**1.4.6.** **Loại dịch vụ và trường lớp lưu lượng truy cập** 8](#_Toc183760029)

[**1.4.7.** **Giá trị DSCP** 9](#_Toc183760030)

[**1.4.8.** **Bit chọn lớp** 9](#_Toc183760031)

[**1.4.9.** **Ranh giới tin cậy** 10](#_Toc183760032)

[**1.4.10.** **Tránh tắc nghẽn** 10](#_Toc183760033)

[**1.4.11. Định hình và chính sách** 11](#_Toc183760034)

[**1.4.12.** **Chính sách QoS** 12](#_Toc183760035)

[**CHƯƠNG 2: BÀI TẬP QOS** 13](#_Toc183760036)

[**2.1.**  **QoS Policing Configuration Ver\_1** 13](#_Toc183760037)

[**2.2.**  **QoS Traffic Shaping** 13](#_Toc183760038)

# **CHƯƠNG I: TỔNG QUAN LÝ THUYẾT**

## **1.1. QoS là gì?**

- Chất lượng dịch vụ (QoS) là việc sử dụng các cơ chế hoặc công nghệ hoạt động trên mạng để kiểm soát lưu lượng và đảm bảo hiệu suất của các ứng dụng quan trọng có dung lượng mạng hạn chế. Nó cho phép các tổ chức điều chỉnh lưu lượng mạng tổng thể của mình bằng cách ưu tiên các ứng dụng hiệu suất cao cụ thể.

A circular object with arrows and speech bubbles

Description automatically generated

- Các dịch vụ phổ biến yêu cầu QoS bao gồm IPTV, VOD, VoIP, trò chơi trực tuyến, nền tảng streaming, họp hội nghị.

- Việc sử dụng QoS, cho phép các tổ chức có khả năng tối ưu hóa hiệu suất của nhiều ứng ụng trên mạng của họ và đảm bảo được khả năng hiển thị tốc độ bit, độ trễ và tốc độ gói tin trong mạng nhằm tránh độ trễ truyền tải.

- Mục tiêu chính khi sử dụng QoS cung cấp băng thông chuyên dụng, độ trễ được kiểm soát và độ trễ thấp hơn. Nâng cao hiệu suất của các ứng dụng kinh doanh, mạng WAN và mạng của các nhà cung cấp dịch vụ.

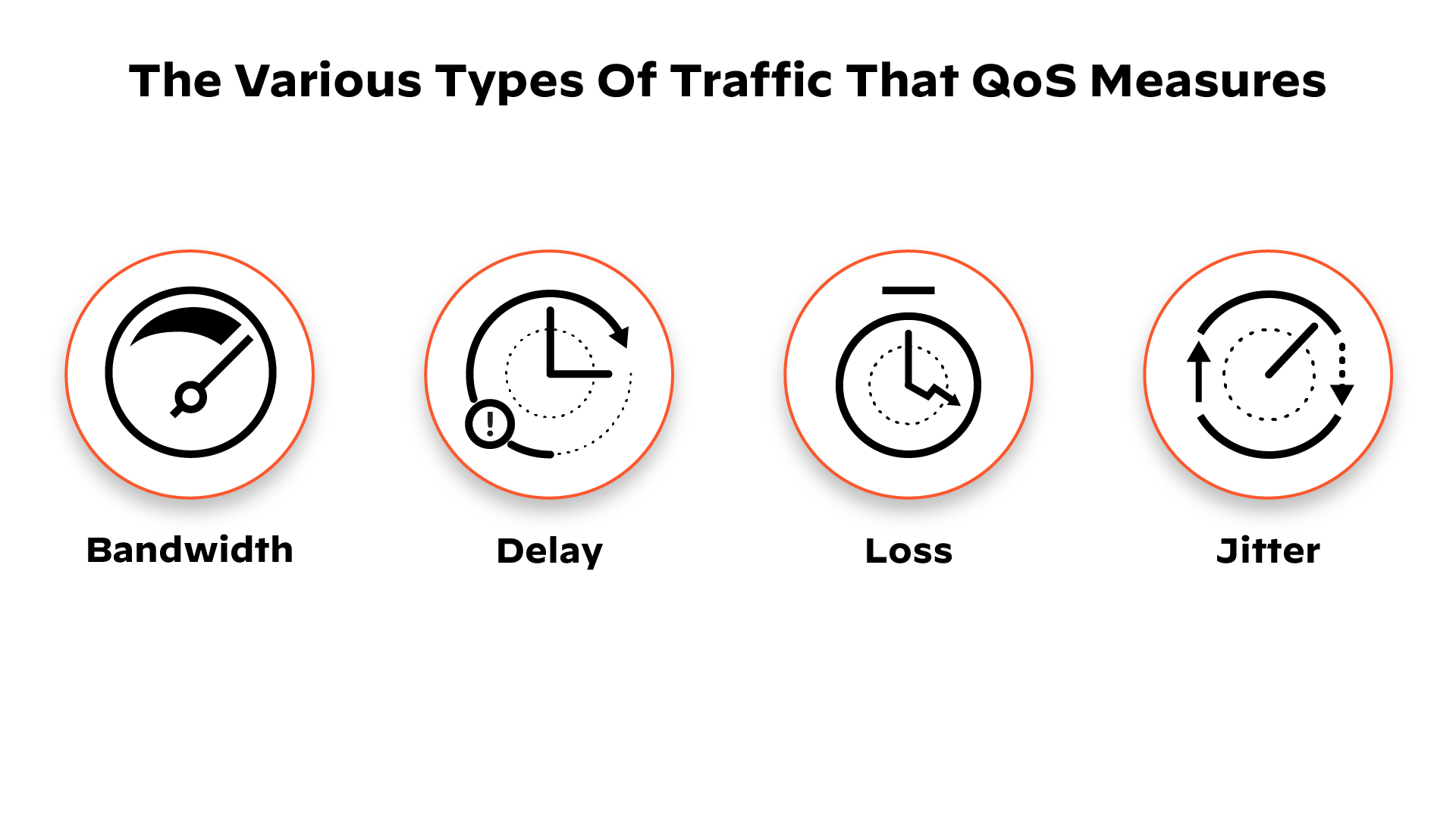
## **1.2. QoS hoạt động như nào?**

- QoS hoạt động bằng cách đánh dấu các gói tin để xác định loại dịch vụ, sau đó cấu hình bộ định tuyến để tạo hang đợi ảo riêng cho từng ứng dụng, dựa trên mức độ ưu tiên của chúng. Do đó, băng thông được dành riêng cho các ứng dụng quan trọng hoặc website đã được chỉ định quyền truy cập ưu tiên.

A diagram of a wire with a couple of wires

Description automatically generated with medium confidence

## **1.3. Các thành phần của QoS**



- Băng thông (Bandwidth): tốc độ của liên kết. QoS có thể cho bộ định tuyến biết cách sử dụng băng thông. Ví dụ, chỉ định một lượng băng thông nhất định cho các hàng đợi khác nhau cho các loại lưu lượng khác nhau. Trong môi trường mạng WAN thực tế, sẽ có hàng trăm gói đang chờ để đi qua liên kết. Bộ định tuyến trên WAN có thể được cấu hình QoS để có thể dự trữ 50% băng thông chó các lưu lượng dữ liệu rất quan trọng hoặc khẩn cấp, 10% cho thoại và phần còn lại cho các lưu lượng khác.

- Độ trễ (Delay): thời gian cần thiết để một gói tin đi từ nguồn đến đích (one-way delay). Thời gian để gói tin đi từ nguồn đến đích xong gửi ngược về lại (round-trip delay). Nhưng việc này có thể bị ảnh hưởng bởi việc xếp hàng do trong thời gian tắc nghẽn và một gói tin phải chờ trong hàng đợi trước khi được truyền đi. QoS cho phép tạo ra một hàng đợi ưu tiên cho một số loại lưu lượng nhất định. Trong mạng sẽ có rất nhiều loại độ trễ ví dụ như độ trễ xử lý (Processing delay), độ trễ xếp hàng (Queuing delay), độ trễ số sê-ri (serialization delay) và độ chậm lan truyền (Propagation delay).

- Mất (Loss): lượng dữ liệu bị mất do mất gói tin, việc này xảy ra là do tắc nghẽn mạng. QoS cho phép quyết định loại bỏ gói tin nào trong trường hợp này. Nếu gửi đi 100 gói tin và chỉ có 95 gói đến được đích, thế là đã mất đi 5% gói tin. Việc mất gói tin là việc luôn luôn có thể xảy ra. Khi có tắc nghẽn, các gói sẽ được xếp hàng đợi để truyền việc này sẽ làm các gói bị mất đi. Ít nhất QoS có thể quyết định được gói nào bị mất khi điều này xảy ra.

- Jitter: tốc độ không đều của các gói tin trên mạng do tắc nghẽn có thể dẫn tới các gói tin đến trễ và không theo trình tự. Điều này có thể gây ra hiện tượng méo tiếng hoặc khoảng cách trong âm thanh hoặc video được truyền tải. Giả sử như điện thoại IP gửi một luồng gói thoại ổn định. Vì việc tắc nghẽn trong mạng, một số gói tin bị trì hoãn. Độ trễ giữa gói 1 và 2 là 20ms, độ trễ giữa gói 2 và 3 là 40ms, …. Người mà nhận được các gói thoại này phải xử lý tình trạng chập chờn, đảm bảo các gói có độ trễ ổn định nếu không bạn sẽ gặp phải việc méo tiếng.

## **1.4. Các kỹ thuật đảm bảo QoS**

### **1.4.1. Tránh mất gói tin (Avoiding Packet Loss)**

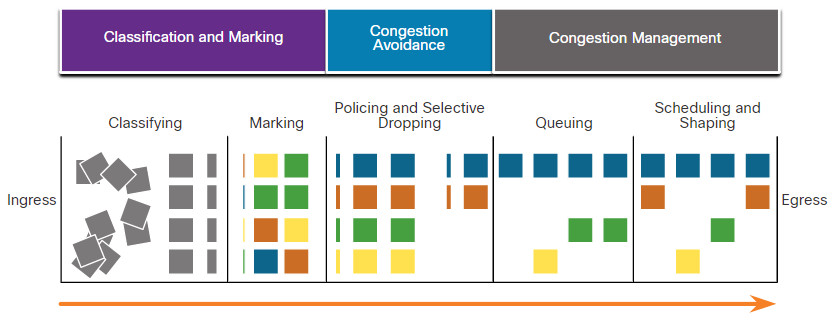
- Các phương pháp sau đây có thể ngăn ngừa giảm mất gói tin trong các ứng dụng:

* Tăng khả năng liên kết để giảm bớt hoặc ngăn chặn tắc nghẽn.
* Đảm bảo đủ băng thông và tăng không gian đệm để đáp ứng sự bùng nổ lưu lượng truy cập từ các luồng yếu. WFQ, CBWFQ và LLQ có thể đảm bảo băng thông và cung cấp chuyển tiếp ưu tiên cho các ứng dụng.
* Bỏ các gói ưu tiên thấp hơn trước khi tắc nghẽn xảy ra. Cisco IOS QoS cung cấp các cơ chế xếp hàng, chẳng hạn như phát hiện sớm ngẫu nhiên có trọng số (WRED), bắt đầu giảm các gói ưu tiên thấp hơn trước khi tắc nghẽn xảy ra.

### **1.4.2. Các công cụ QoS**

- Phân loại và đánh dấu:

* Các phiên, hoặc luồng, được phân tích để xác định chúng thuộc về lớp lưu lượng truy cập nào.
* Khi lớp lưu lượng được xác định, các gói được đánh dấu.



- Công cụ tránh tắc nghẽn:

* Các lớp lưu lượng được phân bổ các phần của tài nguyên mạng, như được xác định bởi chính sách QoS.
* Chính sách QoS cũng xác định cách một số lưu lượng truy cập có thể bị giảm, trì hoãn hoặc đánh dấu lại có chọn lọc để tránh tắc nghẽn.
* Công cụ tránh tắc nghẽn chính là WRED và được sử dụng để điều chỉnh lưu lượng dữ liệu TCP theo cách hiệu quả băng thông trước khi giảm đuôi do tràn hàng đợi xảy ra.

- Công cụ quản lý tắc nghẽn:

* Khi lưu lượng truy cập vượt quá tài nguyên mạng có sẵn, lưu lượng truy cập được xếp hàng đợi để chờ tài nguyên sẵn có.
* Các công cụ quản lý tắc nghẽn dựa trên Cisco IOS phổ biến bao gồm các thuật toán CBWFQ (**Class-Based Weighted Fair Queuing**) và LLQ (Low Latency Queuing).

### **1.4.3. Phân loại và đánh dấu**

- Trước khi áp dụng chính sách QoS cho gói dữ liệu, gói dữ liệu phải được phân loại. Phân loại và đánh dấu cho phép chúng ta xác định hoặc “đánh dấu” các loại gói dữ liệu. Phân loại xác định lớp lưu lượng mà các gói hoặc khung thuộc về. Chỉ sau khi lưu lượng được đánh dấu, các chính sách mới có thể được áp dụng cho nó.

- Cách một gói dữ liệu được phân loại phụ thuộc vào việc triển khai QoS. Các phương pháp phân loại luồng lưu lượng ở Lớp 2 và 3 bao gồm sử dụng cổng, ACL và bản đồ lớp. Lưu lượng cũng có thể được phân loại ở Lớp 4 đến 7 bằng cách sử dụng Nhận dạng ứng dụng dựa trên mạng (NBAR).

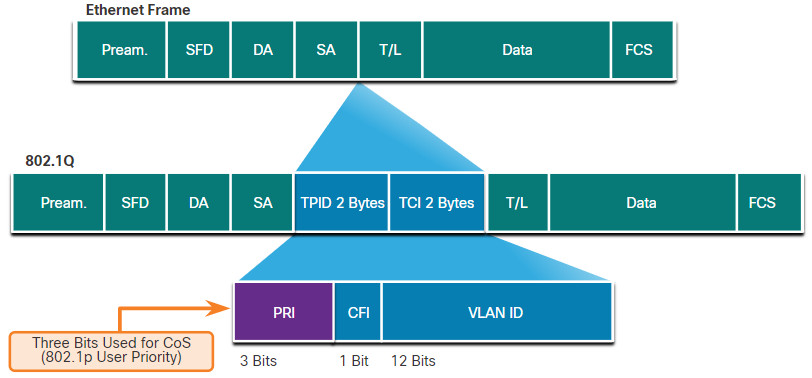
- Đánh dấu có nghĩa là thêm một giá trị vào tiêu đề gói. Các thiết bị nhận gói dữ liệu sẽ xem trường này để xem nó có khớp với một chính sách đã định hay không. Đánh dấu nên được thực hiện càng gần thiết bị nguồn càng tốt.

- Cách đánh dấu lưu lượng thường phụ thuộc vào công nghệ. Quyết định có nên đánh dấu lưu lượng ở Lớp 2 hoặc 3 (hoặc cả hai) không phải là đơn giản và nên được đưa ra sau khi xem xét các điểm sau:

* Đánh dấu khung Lớp 2 có thể được thực hiện cho lưu lượng không phải IP.
* Đánh dấu khung Lớp 2 là tùy chọn QoS duy nhất có sẵn cho các bộ chuyển mạch không “nhận biết IP”.
* Đánh dấu Lớp 3 sẽ mang thông tin QoS từ đầu đến cuối.

### **1.4.4. Đánh dấu ở lớp 2**

- 802.1Q là tiêu chuẩn IEEE hỗ trợ gắn thẻ VLAN ở Lớp 2 trên mạng Ethernet. Khi 802.1Q được triển khai, hai trường được thêm vào Ethernet Frame sau trường địa chỉ MAC nguồn.

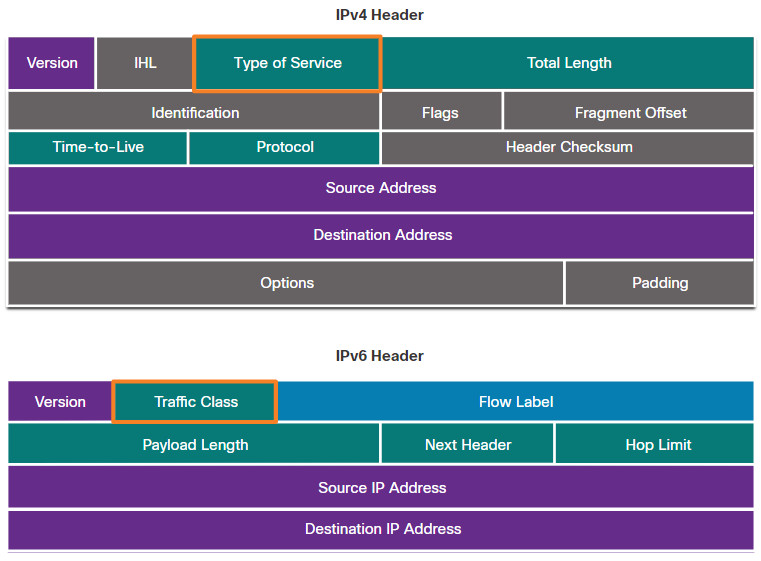


- Tiêu chuẩn 802.1Q cũng bao gồm sơ đồ ưu tiên QoS được gọi là IEEE 802.1p. Tiêu chuẩn 802.1p sử dụng ba bit đầu tiên trong trường Tag Control Information (TCI).

- Được biết đến là trường Ưu tiên (PRI), trường ba bit này xác định đánh dấu Class of Service (CoS). Ba bit có nghĩa là một khung Ethernet Lớp 2 có thể được đánh dấu bằng một trong tám mức ưu tiên (giá trị từ 0 đến 7).

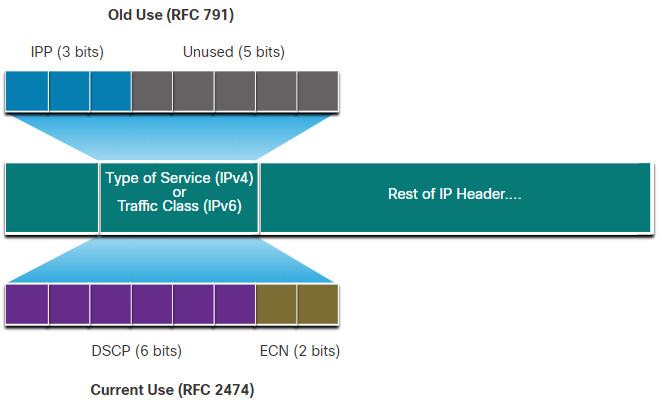
### **1.4.5. Đánh dấu ở lớp 3**

- IPv4 và IPv6 chỉ định một trường 8 bit trong tiêu đề gói của chúng để đánh dấu các gói. Cả IPv4 và IPv6 đều hỗ trợ một trường 8 bit để đánh dấu: trường loại dịch vụ (ToS) cho IPv4 và trường lớp lưu lượng cho IPv6 (Traffic Class).



### **1.4.6. Loại dịch vụ và trường lớp lưu lượng truy cập**

- Type of Service (IPv4) và Traffic Class (IPv6) mang theo đánh dấu gói như được chỉ định bởi các công cụ phân loại QoS. Trường này sau đó được tham chiếu bởi các thiết bị nhận chuyển tiếp các gói dựa trên chính sách QoS được chỉ định phù hợp.



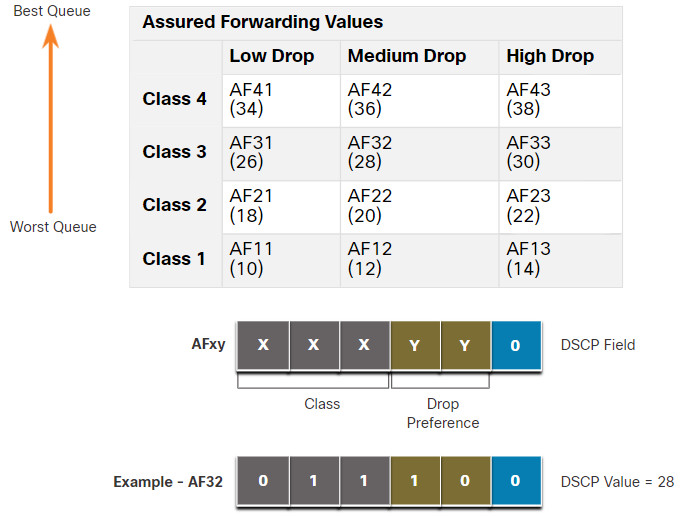
- RFC 2474 thay thế RFC 791 và định nghĩa lại trường ToS bằng cách đổi tên và mở rộng trường IPP. Trường mới, như được hiển thị trong hình, có 6 bit được phân bổ cho QoS. Được gọi là Differentiated Services Code Point (DSCP), sáu bit này cung cấp tối đa 64 lớp dịch vụ có thể có. Hai bit IP Extended Congestion Notification (ECN) còn lại có thể được sử dụng bởi các bộ định tuyến nhận biết ECN để đánh dấu các gói thay vì bỏ chúng. Đánh dấu ECN thông báo cho các bộ định tuyến hạ nguồn rằng có tắc nghẽn trong luồng gói.

### **1.4.7. Giá trị DSCP**

- Best-Effort (BE) – Đây là mặc định cho tất cả các gói IP. Giá trị DSCP là 0. Hành vi mỗi bước nhảy là định tuyến bình thường. Khi một bộ định tuyến gặp tắc nghẽn, các gói này sẽ bị bỏ rơi. Không có kế hoạch QoS nào được triển khai.

- Expedited Forwarding (EF) – RFC 3246 định nghĩa EF là giá trị thập phân DSCP 46 (nhị phân 101110). Ba bit đầu tiên (101) ánh xạ trực tiếp vào giá trị CoS Lớp 2 là 5 được sử dụng cho lưu lượng thoại. Ở Lớp 3, Cisco khuyến nghị chỉ sử dụng EF để đánh dấu các gói thoại.

- Assured Forwarding (AF) – RFC 2597 định nghĩa AF để sử dụng 5 bit DSCP quan trọng nhất để chỉ định hàng đợi và ưu tiên bỏ. Định nghĩa của AF được minh họa trong hình

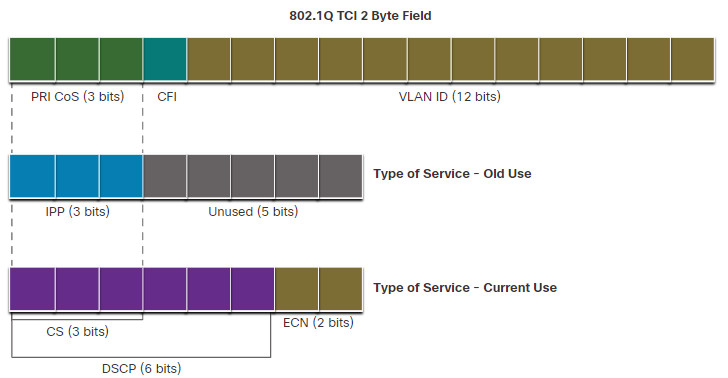


- Công thức AFxy được quy định như sau:

* Ba bit quan trọng nhất đầu tiên được sử dụng để chỉ định lớp.
* Lớp 4 là hàng đợi tốt nhất và Lớp 1 là hàng đợi tệ nhất.
* Hai bit quan trọng nhất thứ 4 và thứ 5 được sử dụng để chỉ định ưu tiên bỏ.
* Bit quan trọng nhất thứ 6 được đặt thành 0.

### **1.4.8. Bit chọn lớp**

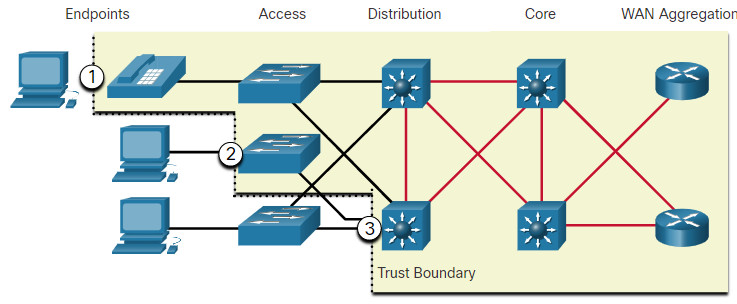
- Vì 3 bit quan trọng nhất đầu tiên của trường DSCP chỉ định lớp, nên các bit này cũng được gọi là các bit Class Selector (CS). 3 bit này ánh xạ trực tiếp vào 3 bit của trường CoS và trường IPP để duy trì khả năng tương thích với 802.1p và RFC 791, như được hiển thị trong hình.Bảng trong hình cho thấy cách các giá trị CoS ánh xạ vào các Class Selectors và giá trị 6 bit DSCP tương ứng.



### **1.4.9. Ranh giới tin cậy**

- Đánh dấu nên xảy ra ở đâu? Lưu lượng truy cập nên được phân loại và đánh dấu càng gần nguồn của nó càng khả thi về mặt kỹ thuật và hành chính. Điều này xác định ranh giới tin cậy, như được hiển thị trong hình.

* Các điểm cuối đáng tin cậy có khả năng và thông minh để đánh dấu lưu lượng ứng dụng thành các giá trị CoS Lớp 2 và / hoặc DSCP Lớp 3 phù hợp.
* Các ví dụ về các điểm cuối đáng tin cậy bao gồm điện thoại IP, điểm truy cập không dây, cổng và hệ thống hội nghị truyền hình, trạm hội nghị IP và hơn thế nữa. Các điểm cuối bảo mật có thể có lưu lượng được đánh dấu ở bộ chuyển mạch Lớp 2.
* Lưu lượng cũng có thể được đánh dấu ở các bộ chuyển mạch / bộ định tuyến Lớp 3.



- Đánh dấu lại lưu lượng, ví dụ, đánh dấu lại các giá trị CoS thành các giá trị IP Precedent hoặc DSCP, thường là cần thiết.

### **1.4.10. Tránh tắc nghẽn**

- Quản lý tắc nghẽn bao gồm các phương pháp xếp hàng và lập lịch nơi lưu lượng dư thừa được đệm hoặc xếp hàng trong khi chờ được gửi ra khỏi giao diện thoát. Chúng giám sát tải trọng lưu lượng mạng để cố gắng dự đoán và tránh tắc nghẽn tại các điểm nghẽn mạng và internet thông thường trước khi tắc nghẽn trở thành vấn đề. Các công cụ này có thể giám sát độ sâu trung bình của hàng đợi. Khi hàng đợi thấp hơn ngưỡng tối thiểu, sẽ không có lượt bỏ. Khi hàng đợi được lấp đầy đến ngưỡng tối đa, một tỷ lệ nhỏ các gói sẽ bị bỏ qua. Khi vượt quá ngưỡng tối đa, tất cả các gói sẽ bị bỏ qua.

- Một số kỹ thuật tránh tắc nghẽn cung cấp ưu tiên cho các gói nào sẽ bị bỏ rơi. Ví dụ, Cisco IOS QoS bao gồm phát hiện sớm ngẫu nhiên có trọng số (WRED) như một giải pháp tránh tắc nghẽn có thể. Thuật toán WRED cho phép tránh tắc nghẽn trên các giao diện mạng bằng cách cung cấp quản lý bộ đệm và cho phép lưu lượng TCP giảm hoặc giảm tốc trước khi bộ đệm cạn kiệt. Sử dụng WRED giúp tránh thả đuôi và tối đa hóa việc sử dụng mạng và hiệu suất ứng dụng dựa trên TCP. Không có tránh tắc nghẽn cho lưu lượng dựa trên Giao thức Datagram Người dùng (UDP), chẳng hạn như lưu lượng thoại. Trong trường hợp lưu lượng dựa trên UDP, các phương pháp như xếp hàng và kỹ thuật nén giúp giảm và thậm chí ngăn chặn mất gói UDP.

### **1.4.11. Định hình và chính sách**

- Traffic shaping và traffic policing là hai cơ chế được cung cấp bởi phần mềm Cisco IOS QoS để ngăn ngừa tắc nghẽn.Traffic shaping giữ lại các gói tin dư thừa trong hàng đợi và sau đó lên lịch truyền chúng sau một khoảng thời gian nhất định.

Kết quả của việc định hình lưu lượng là tốc độ gói tin đầu ra được làm mịn, như thể hiện trong hình.

- Shaping (định hình) ngụ ý sự tồn tại của hàng đợi và đủ bộ nhớ để lưu trữ các gói tin bị trì hoãn, trong khi policing (kiểm soát) thì không. Hãy đảm bảo rằng bạn có đủ bộ nhớ khi bật chức năng shaping. Ngoài ra, shaping yêu cầu một chức năng lập lịch để truyền các gói tin bị trì hoãn sau đó. Chức năng lập lịch này cho phép bạn sắp xếp hàng đợi shaping thành các hàng đợi khác nhau.



- Ví dụ về các chức năng lập lịch là CBWFQ và LLQ:

* Shaping là một khái niệm hướng ra ngoài; các gói tin đi ra khỏi một giao diện sẽ được xếp hàng đợi và có thể được định hình. Ngược lại, policing được áp dụng cho lưu lượng truy cập hướng vào trong trên một giao diện. Khi tốc độ lưu lượng đạt đến tốc độ tối đa đã cấu hình, lưu lượng vượt quá sẽ bị loại bỏ.
* Policing thường được các nhà cung cấp dịch vụ triển khai để thực thi tốc độ thông tin khách hàng theo hợp đồng (CIR).

- Tuy nhiên, nhà cung cấp dịch vụ cũng có thể cho phép vượt quá CIR nếu mạng của họ hiện không bị tắc nghẽn.

### **1.4.12. Chính sách QoS**

- Chính sách QoS của bạn phải xem xét toàn bộ đường dẫn từ nguồn đến đích. Nếu một thiết bị trên đường dẫn sử dụng chính sách khác với mong muốn, thì toàn bộ chính sách QoS sẽ bị ảnh hưởng.

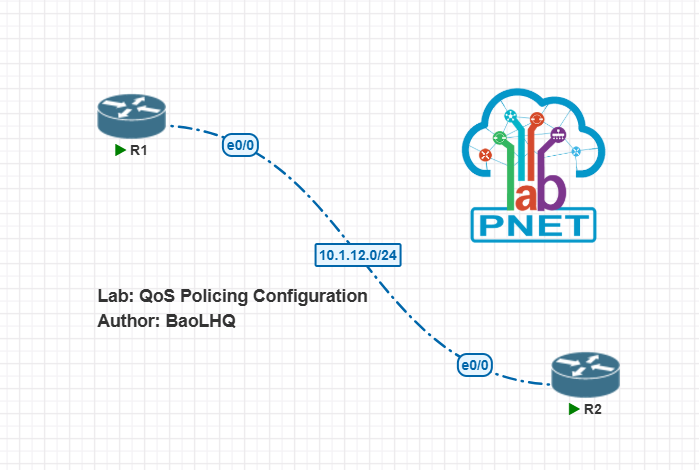
- Ví dụ, hiện tượng giật hình trong khi phát video có thể là do một switch trên đường dẫn không được đặt giá trị CoS phù hợp.

- Một số hướng dẫn giúp đảm bảo trải nghiệm tốt nhất cho người dùng cuối bao gồm:

* Bật chức năng xếp hàng đợi ở mọi thiết bị trên đường dẫn giữa nguồn và đích.
* Phân loại và đánh dấu lưu lượng càng gần nguồn càng tốt.
* Định hình và kiểm soát luồng lưu lượng càng gần nguồn càng tốt.

# **CHƯƠNG 2: BÀI TẬP QOS**

## **2.1. QoS Policing Configuration Ver\_1**



## **2.2. QoS Traffic Shaping**

A diagram of a cloud with a logo and text

Description automatically generated with medium confidence