RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

TOULOUSE INP N7

QUALITY OF SERVICE

CLASS 2

QOS MECHANISMS

NETWORK CONGESTION

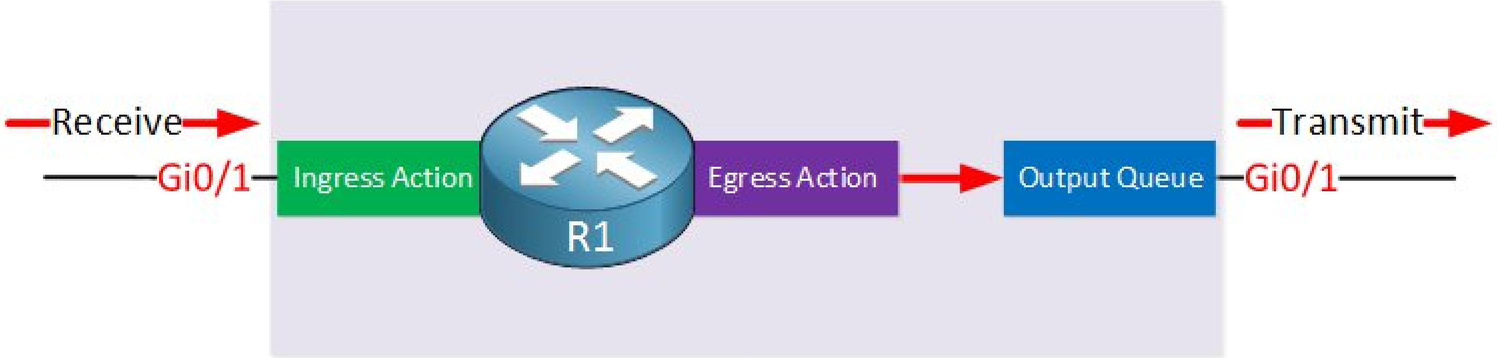
Network Congestion

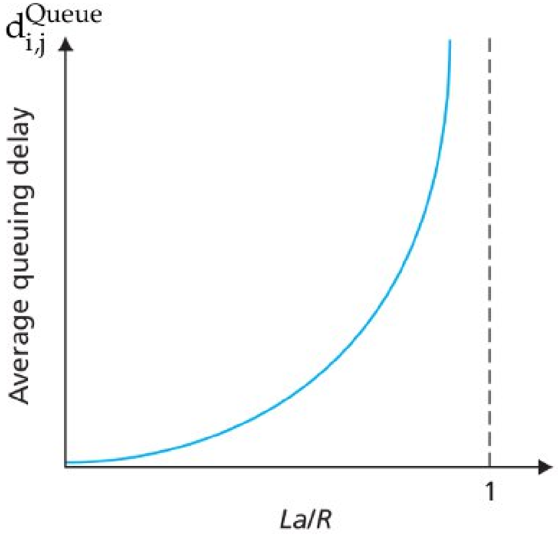
- **Definition**: Congestion is a state of a network component where it experiences more input traffic than it is able to handle with the available (shared) resources.
- Network nodes handle congestion differently, depending on the OSI layer it is associated with.

Network Congestion

Congestion in Layer 3

- Congested routers have unstable output queues

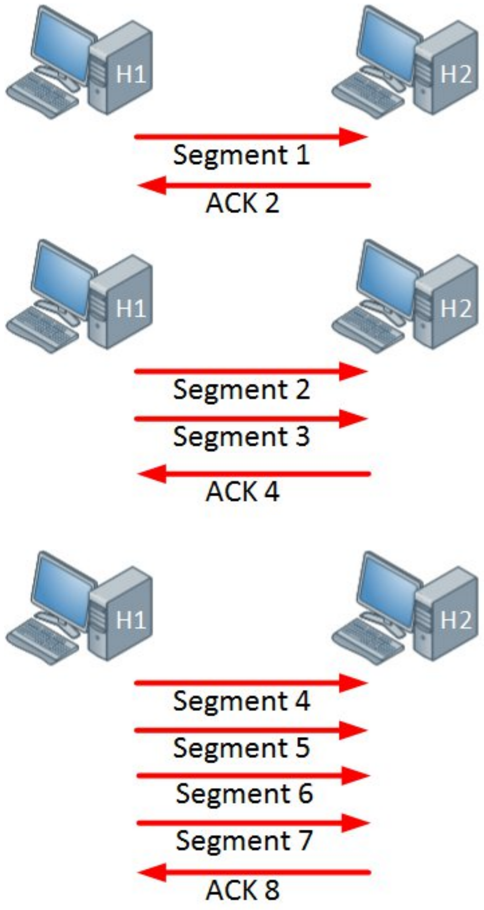




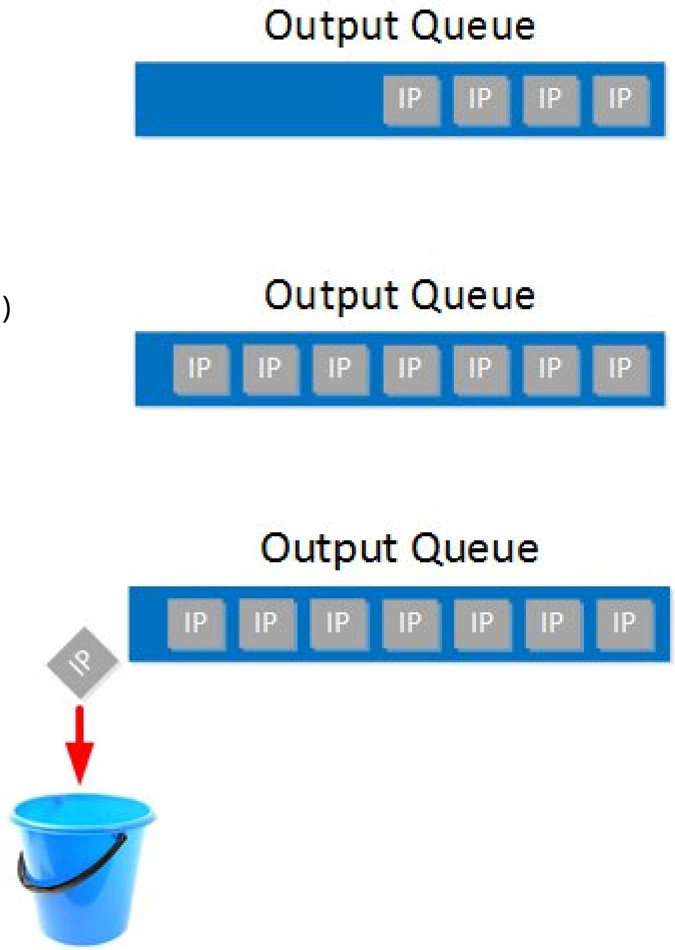
- Network congestion causes (i) Queuing Delay, (ii) Packet dropping, and (iii) Blocking

Network Congestion

Layer 4: Sliding Windows in TCP



Congestion Window (CWND)
Receiver window (RWND)



CLASSIFICATION

Classification

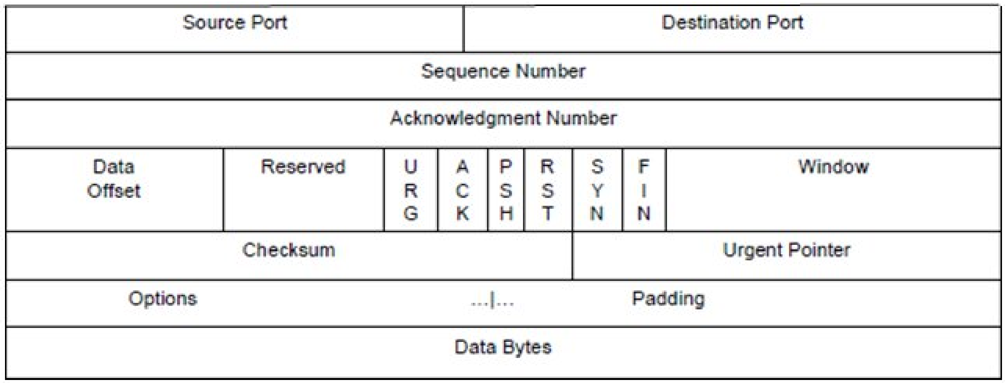
Definition

- **Definition**: Classification is the practice of inferring application’s nature based on the inspection of specific characteristics
- Examples of applications natures:
  - web browsing
  - streaming
  - voice calls
- Hard-coded or automatically performed by routers
- Inspection types:
  - Header inspection
  - Payload inspection

Classification

Header Inspection

Layer 4 (Transport Layer): “Source Port” and “Destination Port”



Source Port | Destination Port

Sequence Number

Acknowledgment Number

Data Offset | Reserved | URG | ACK | PSH | RST | SYN | FIN | Window

Checksum | Urgent Pointer

Options ...|... Padding

Data Bytes

Classic ports:

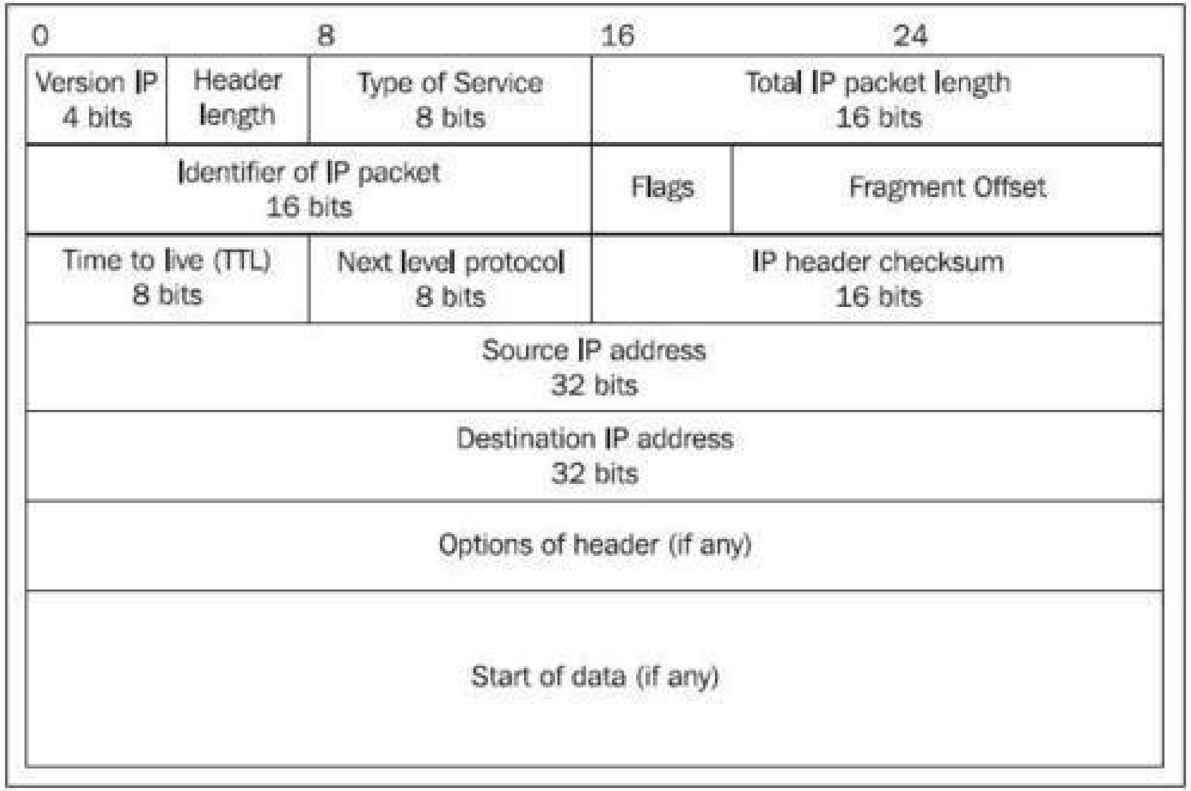
- 80 HTTP - Nature: web
- 443 SSL - Nature: playback/browsing (streaming)
- 22 SSH - Nature: interactive
- 5060 VoIP - Nature: voice calls

Is there a problem with this approach?

Classification

Header Inspection

Layer 4 (Transport Layer): “Source IP address”, “Destination IP address”, “Next-level protocol”

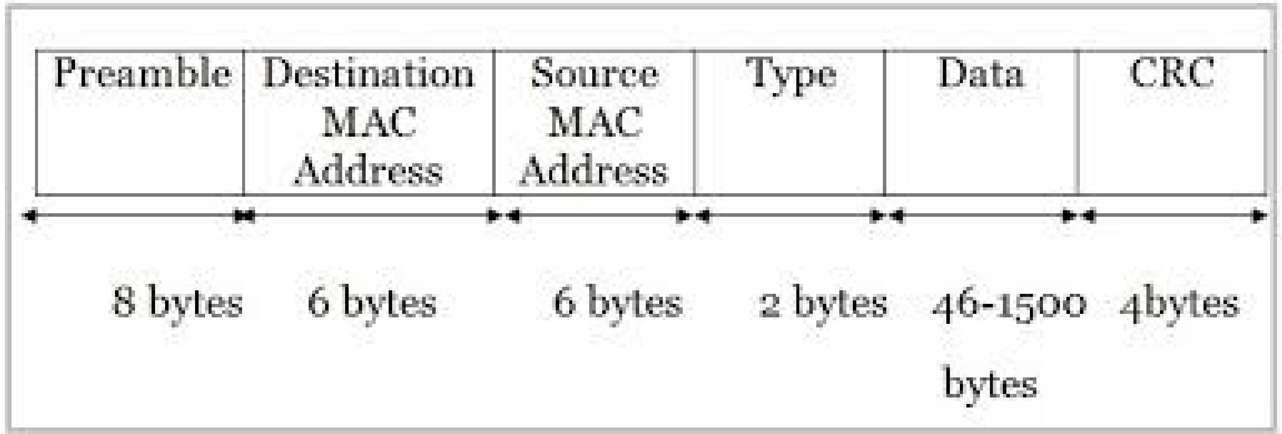


0		8	16	24
Version IP 4 bits	Header length	Type of Service 8 bits	Total IP packet length 16 bits	
Identifier of IP packet 16 bits			Flags	Fragment Offset
Time to live (TTL) 8 bits		Next level protocol 8 bits	IP header checksum 16 bits	
Source IP address 32 bits				
Destination IP address 32 bits				
Options of header (if any)				
Start of data (if any)				

Classification

Header Inspection

- Layer 2 (Link Layer): “Destination MAC address” and “Source MAC address”



Preamble	Destination MAC Address	Source MAC Address	Type	Data	CRC
8 bytes	6 bytes	6 bytes	2 bytes	46-1500 bytes	4bytes

Classification

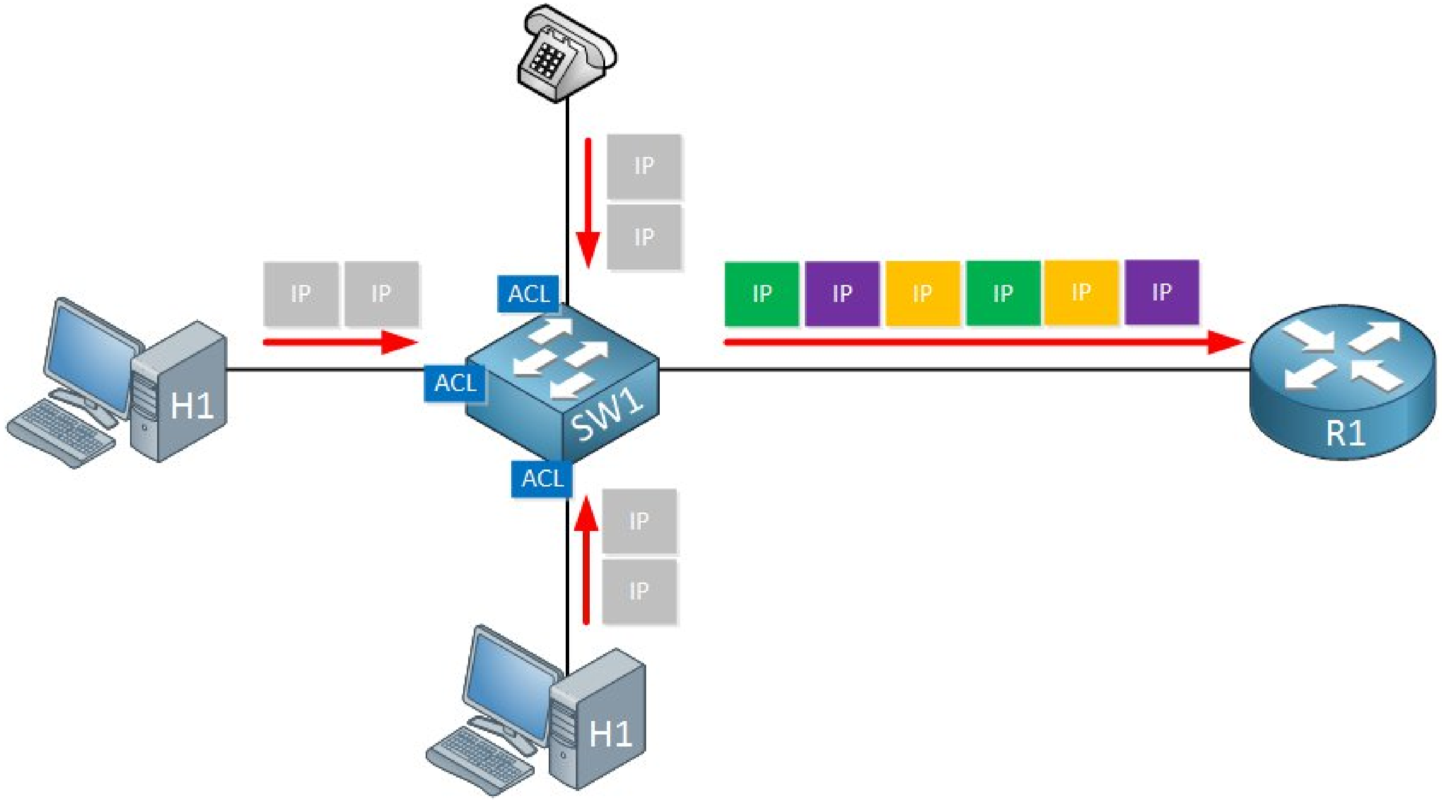
Payload Inspection – Network-Based Application Recognition (NBAR)

- **Definition**: Network-Based Application Recognition (NBAR) is a classification method that is able to identify application information from the segment's payloads.
- Must be enabled at a given NIC
    - May create overhead..
- Is able to identify:
    - URL
    - MIME-type (zip file, image, etc.)
    - User-agent (Mozilla, Opera, etc.)
- Can be used to block websites!

MARKING

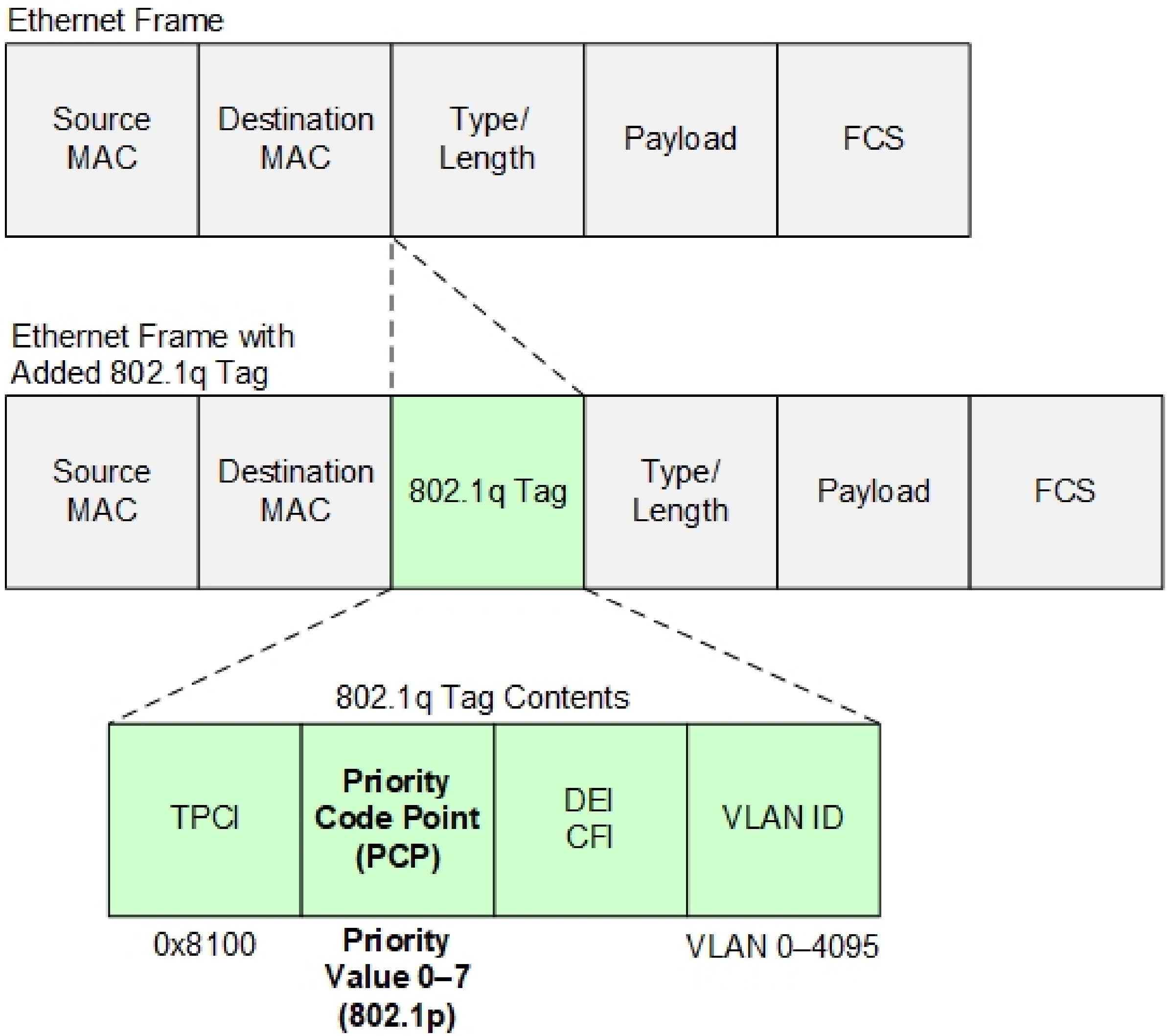
Marking

Definition: Marking is the act of changing one or more header fields in the packet to reflect the classification result



Marking

Ethernet (IEEE 802.11Q)



Ethernet Frame

Source MAC	Destination MAC	Type/ Length	Payload	FCS

Ethernet Frame with Added 802.1q Tag

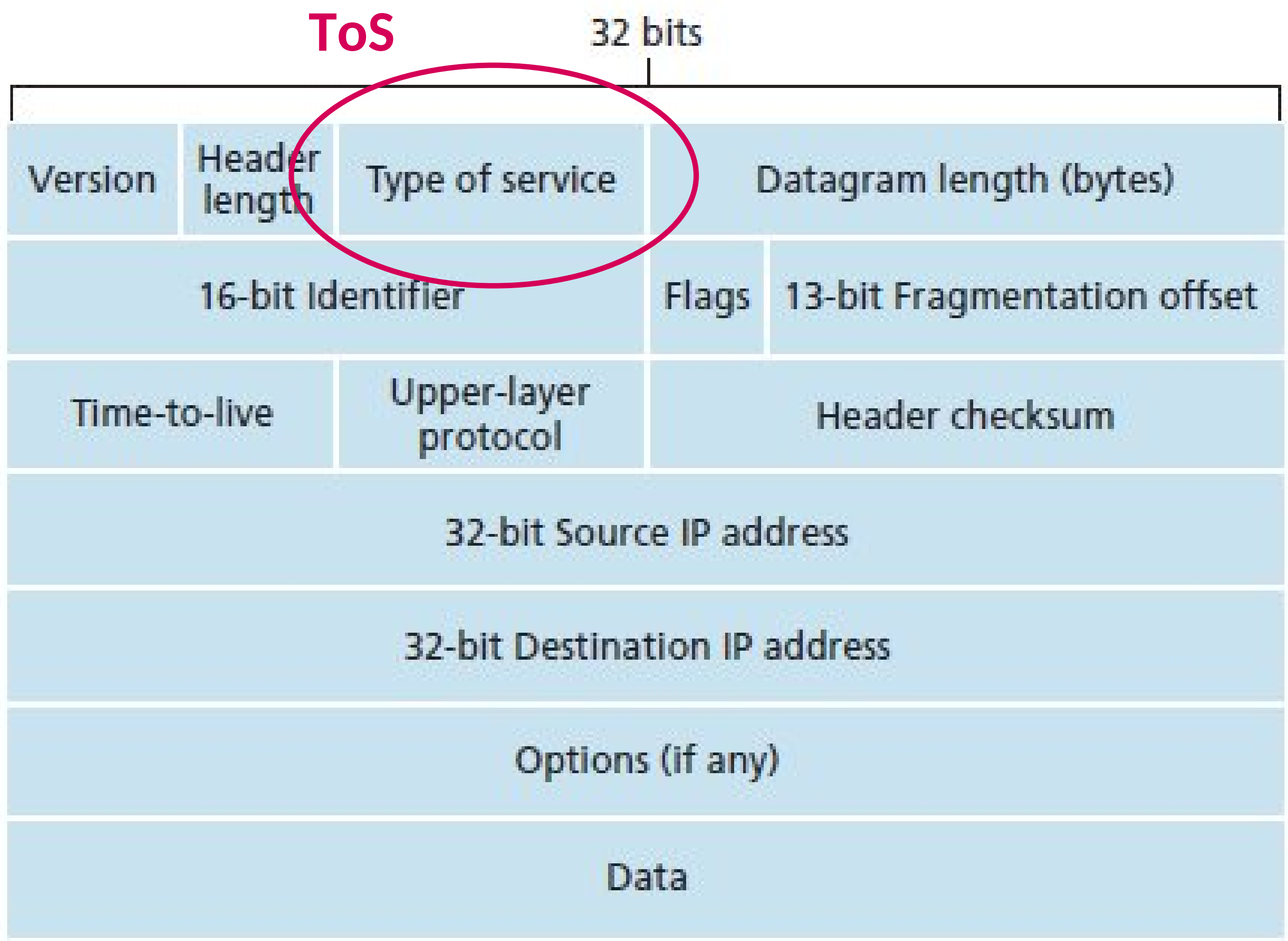
Source MAC	Destination MAC	802.1q Tag	Type/ Length	Payload	FCS

802.1q Tag Contents

TPCI	**Priority Code Point (PCP)**	DEI CFI	VLAN ID
0x8100	**Priority Value 0–7 (802.1p)**		VLAN 0–4095

Marking

IPv4



ToS

32 bits

Version

Header length

Type of service

Datagram length (bytes)

16-bit Identifier

Flags

13-bit Fragmentation offset

Time-to-live

Upper-layer protocol

Header checksum

32-bit Source IP address

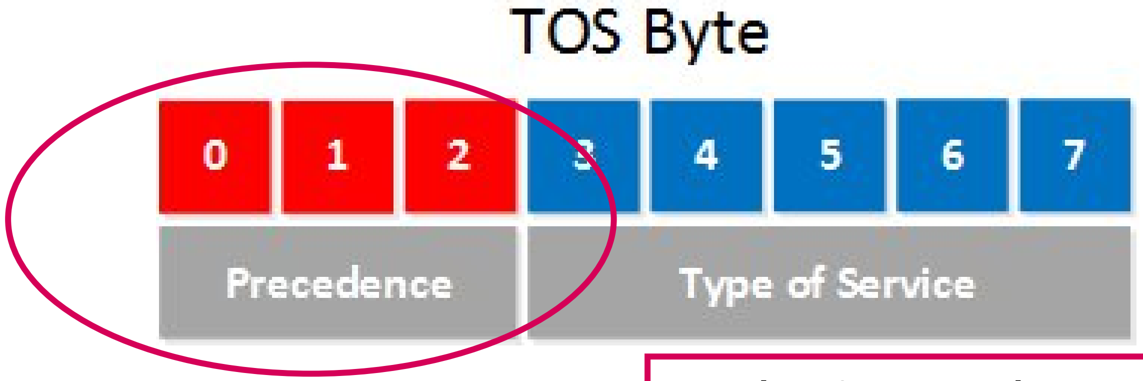
32-bit Destination IP address

Options (if any)

Data

Marking

IPv4 – IP Precedence: First design, RFC791 (1981)

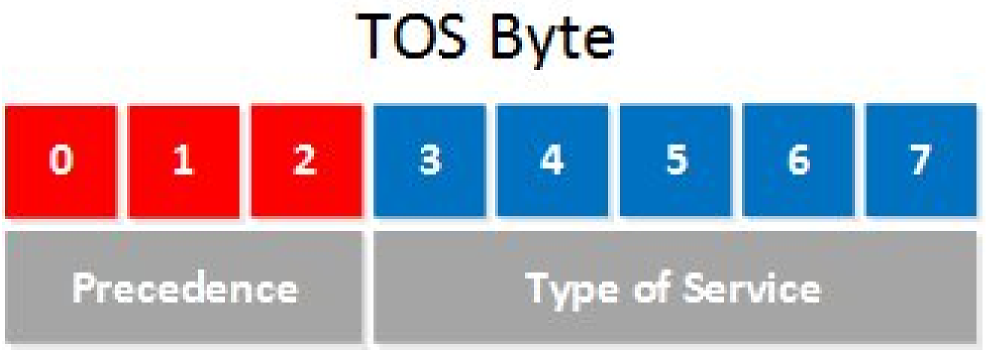


Rank of precedence (3 bits)

0	000	Routine	(lowest priority)
1	001	Priority	
2	010	Immediate	
3	011	Flash	
4	100	Flash Override	
5	101	Critic/Critical	
6	110	Internetwork Control	
7	111	Network Control	(highest priority)

Marking

IPv4 – IP Precedence: First design, RFC791 (1981)

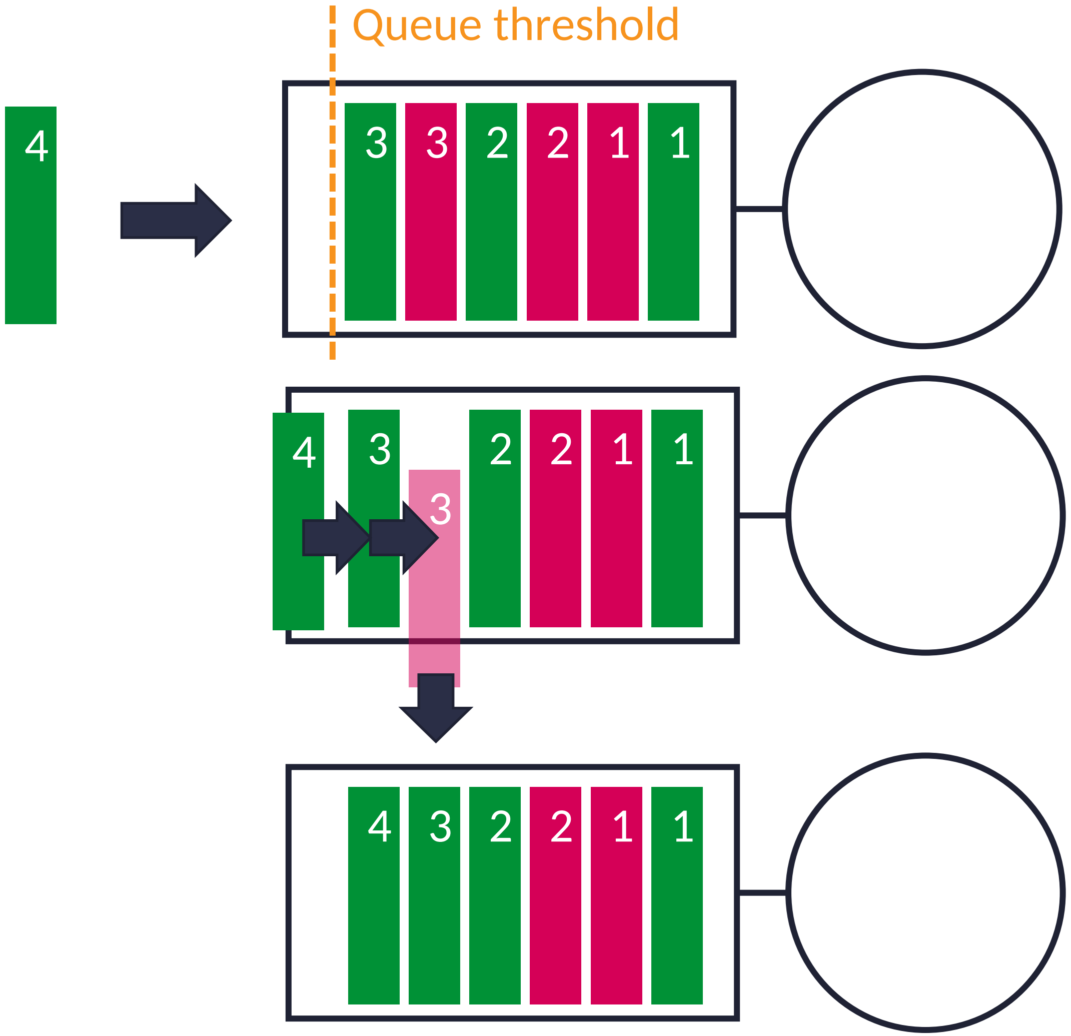


TOS Byte

0	1	2	3	4	5	6	7
Precedence			Type of Service				

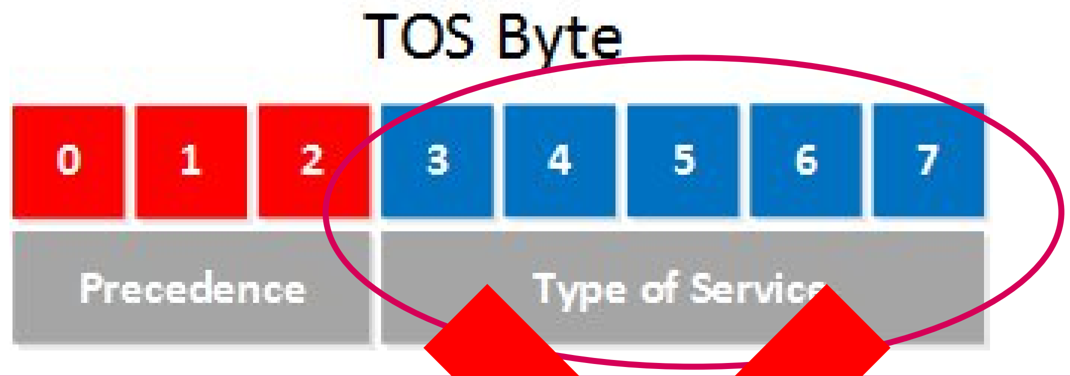
Rank of precedence

0	**000**	**Routine**
1	001	Priority
2	010	Immediate
3	011	Flash
4	100	Flash Override
5	101	Critic/Critical
6	110	Internetwork Control
7	**111**	**Network Control**



Marking

IPv4 – IP Precedence: First design, RFC791 (1981)

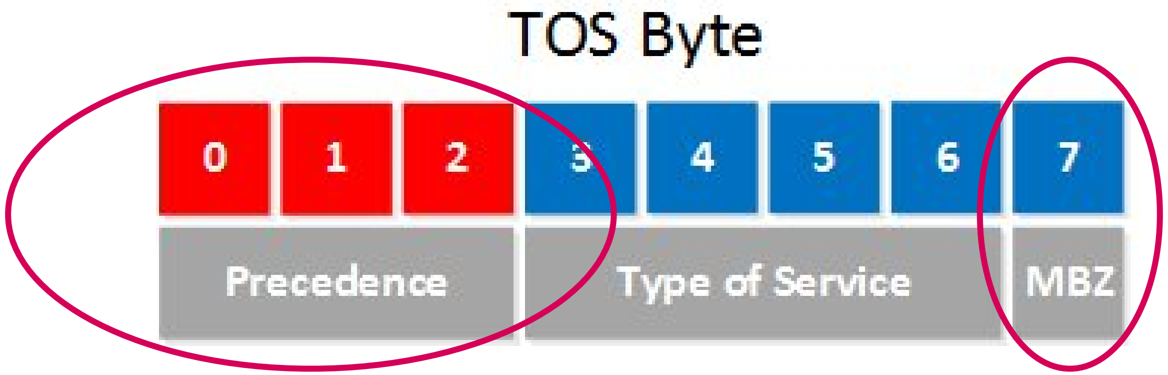


Type of Service (7 bits)

Bit 3:	0 = normal delay	1 = low delay
Bit 4:	0 = normal throug	1 = high throughput
Bit 5:	0 = normal relia	1 = high reliability
Bit 6-7:	Reserved for [illegible] use	

Marking

IPv4 – IP Precedence: Second design, RFC1349 (1992)



TOS Byte

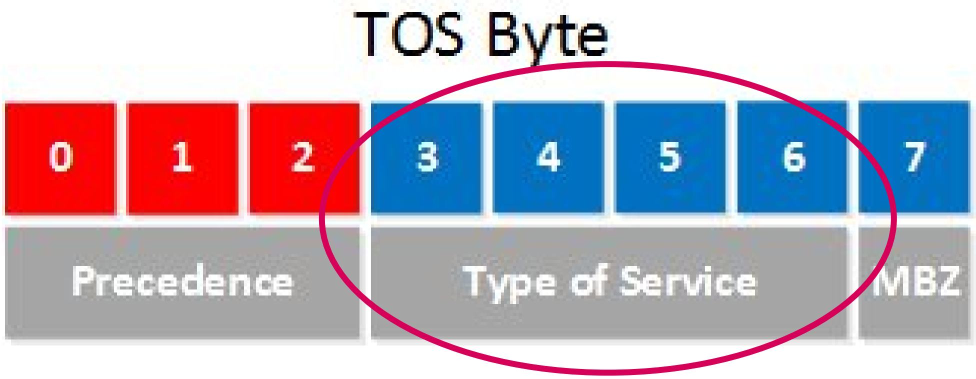
0	1	2	3	4	5	6	7
Precedence			Type of Service				MBZ

Precedence is unchanged!

Must Be Zero (MBZ) bit

Marking

IPv4 – IP Precedence: Second design, RFC1349 (1992)

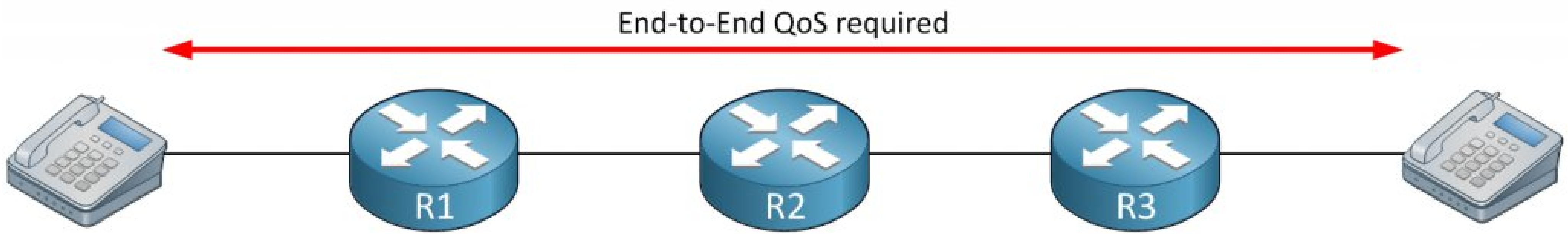


Type of Service (7 bits)

8	1000	minimize delay
4	0100	maximize throughput
2	0010	maximize reliability
1	0001	minimize monetary cost
0	0000	normal service

Marking

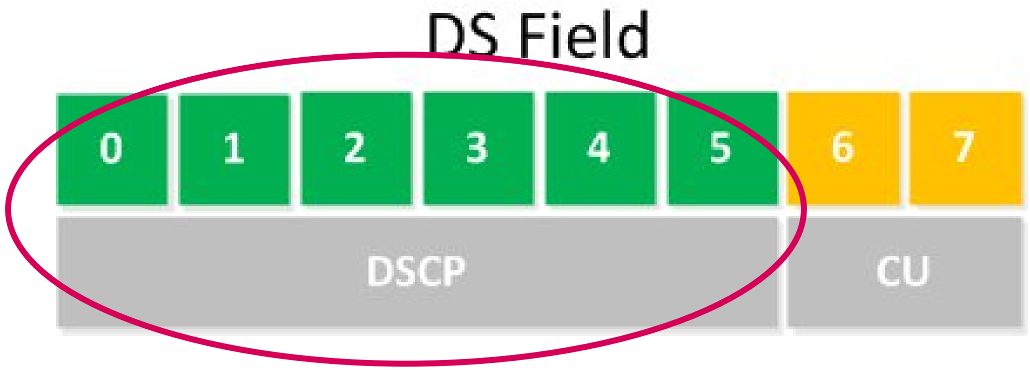
Per-Hop Behavior (PHB)



Each type of service is implemented as the same “behavior” throughout every router in the data flow.

Marking

IPv4 – Differentiated Services (DiffServ)

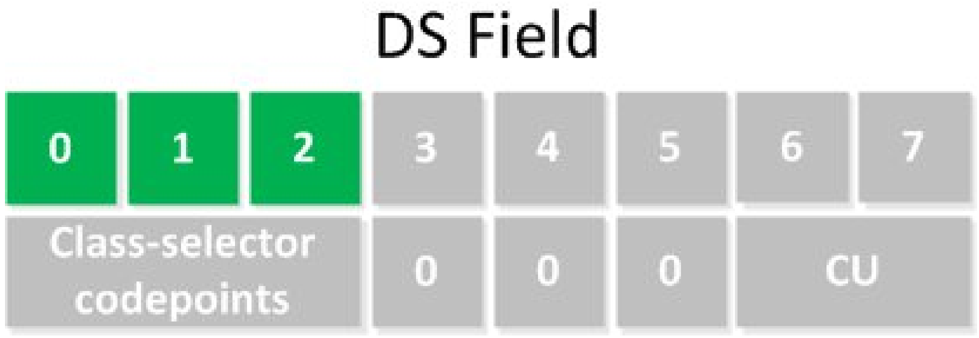


Differentiated Service CodePoint (DSCP) → Per-Hop Behavior (PHB)

Default PHB:
000000 Best Effort

Marking

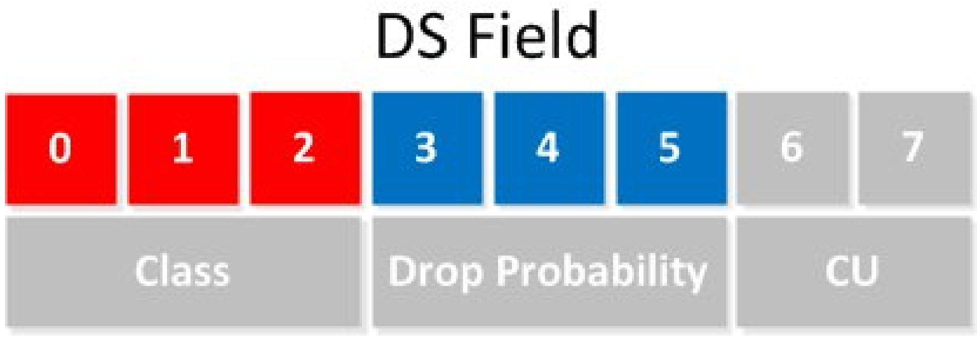
IPv4 – DiffServ: Class-Selector PHB, RFC 2474 (1998)



Class selector name	DSCP value	IP Precedence value	IP Precedence name
Default / CS0	000000	000	Routine
CS1	001000	001	Priority
CS2	010000	010	Immediate
CS3	011000	011	Flash
CS4	100000	100	Flash Override
CS5	101000	101	Critic/Critical
CS6	110000	110	Internetwork Control
CS7	111000	111	Network Control

Marking

IPv4 – DiffServ: Assured Forwarding (AF) PHB, RFC 2597 (1999)



DiffServ-AF PHB has two functions:

1. Queuing
2. Congestion Avoidance

For a packet marked with a specific class:

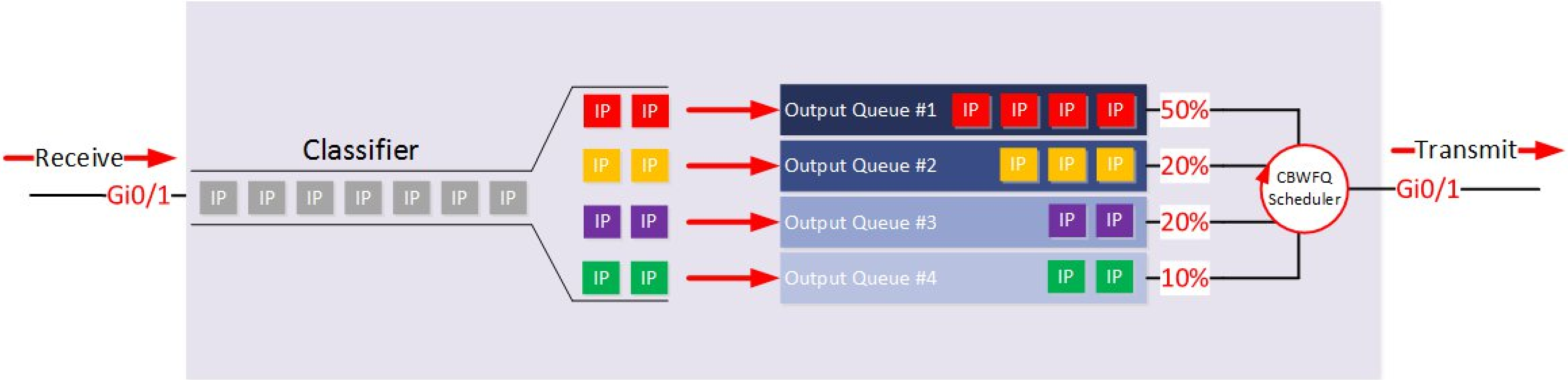
- forwarding is independent of other classes
- there are dedicated resources (capacity and buffer)
- spare resources may be used, even if it is more than the required amount.

Drop	Class 1	Class 2	Class 3	Class 4
Low	001 010 AF11	010 010 AF21	011 010 AF31	100 010 AF41
Medium	001 100 AF12	010 100 AF22	011 100 AF32	100 100 AF42
High	001 110 AF13	010 110 AF23	011 110 AF33	100 110 AF43

Conversion: “Class name” -> decimal
AFxy = $(8x + 2y)_D$

Marking

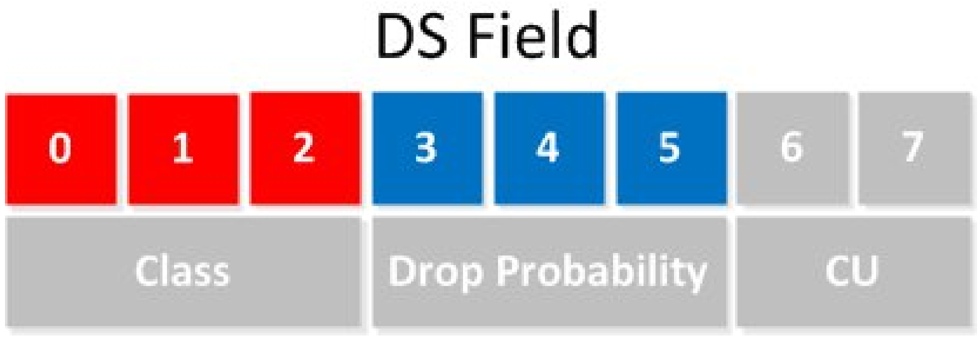
IPv4 – DiffServ: AF PHB – Class-Based Weighted Fair Queue (CBWFQ)



What does it mean "percentage of capacity"?

Marking

IPv4 – DiffServ: Expedited Forwarding (EF) PHB, RFC 2597 (1999)



DSCP name:	EF
DSCP binary:	$(101\ 110)_B$
DSCP decimal:	$(46)_D$

DiffServ-EF PHB has two functions:

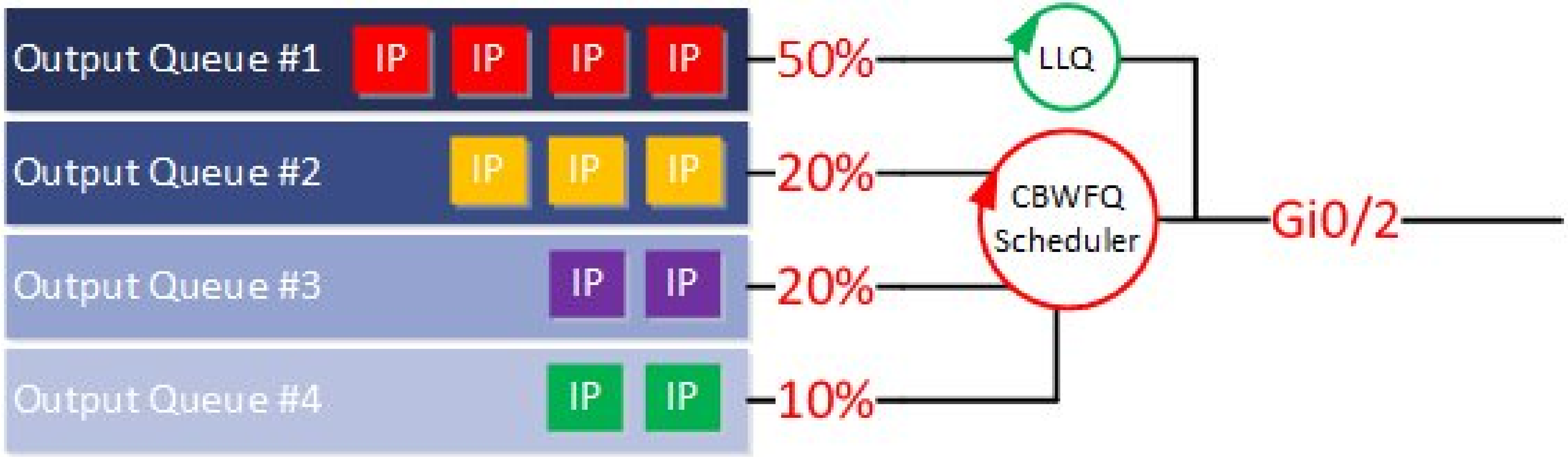
1. Queuing - priority queue
2. Policing - non-blocking policies

For a packet marked as EF:

- same rules applied to AF
- output has transmission priority over other queues

Marking

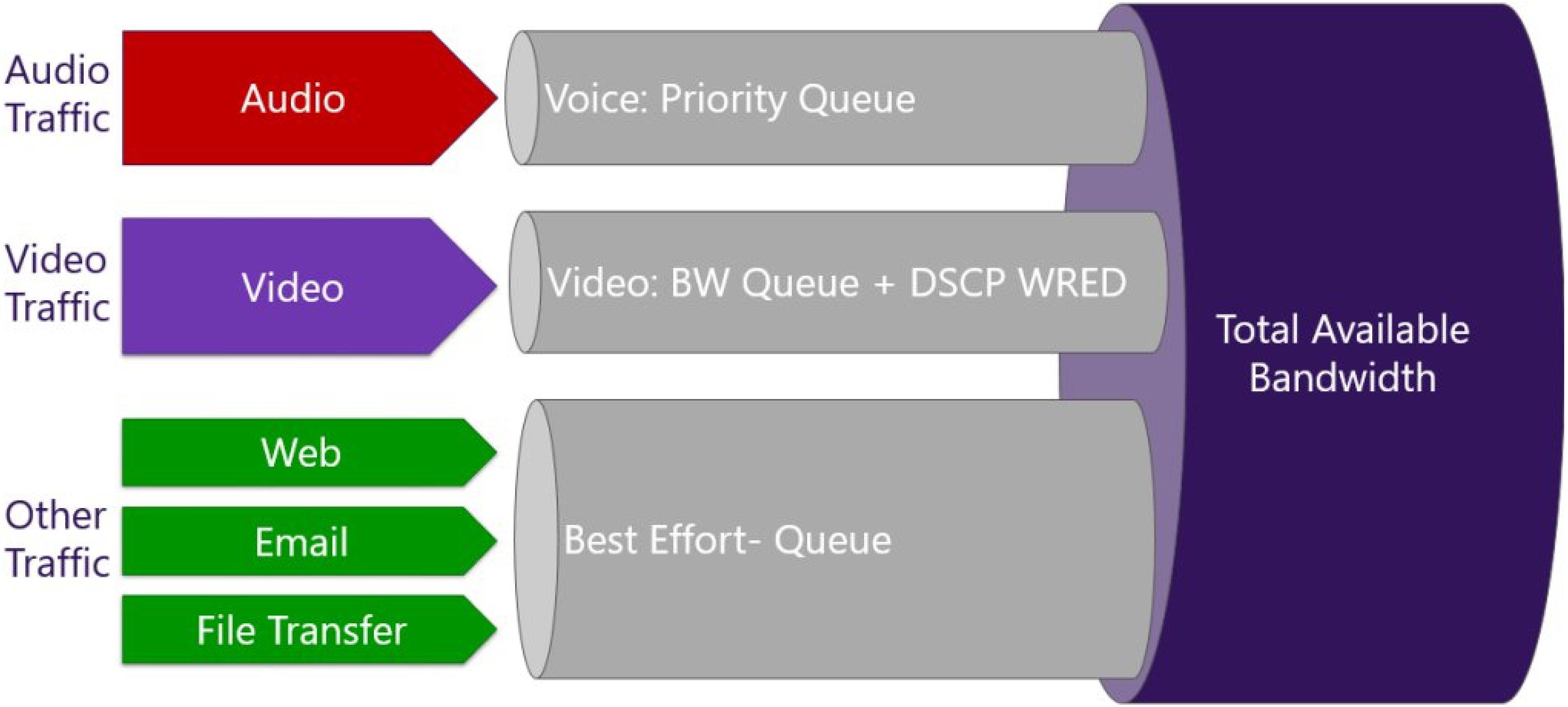
IPv4 – DiffServ: EF PHB – Low Latency Queue (LLQ)



What happens if LLQ is always busy?

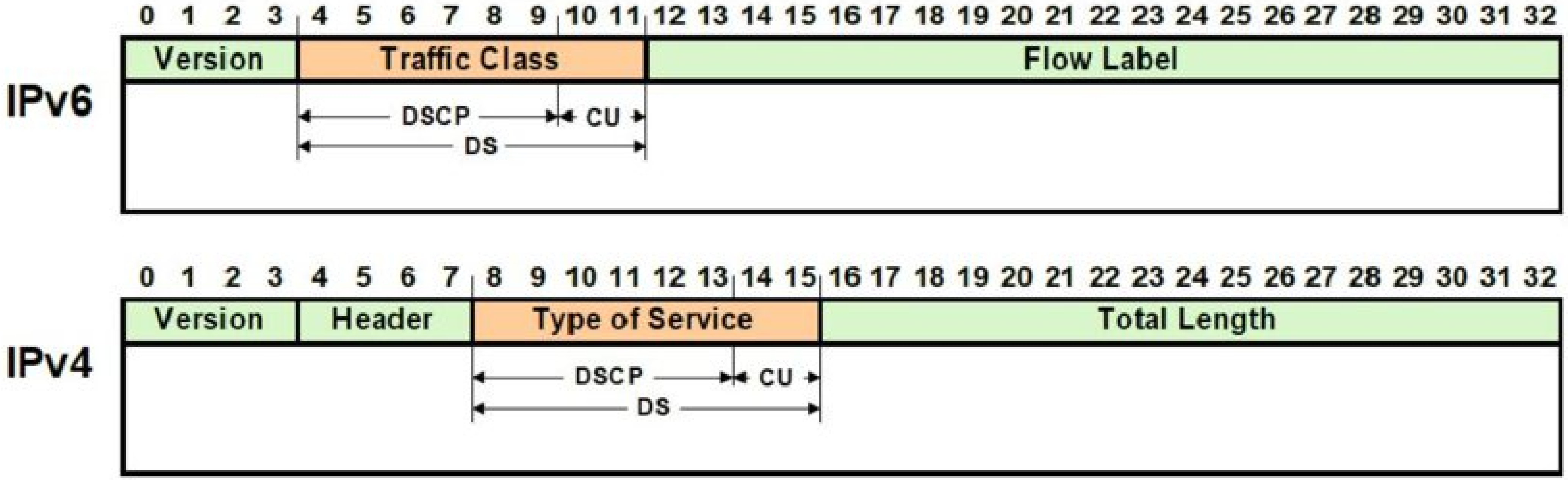
Marking

IPv4 – Hybrid DiffServ Networks



Marking

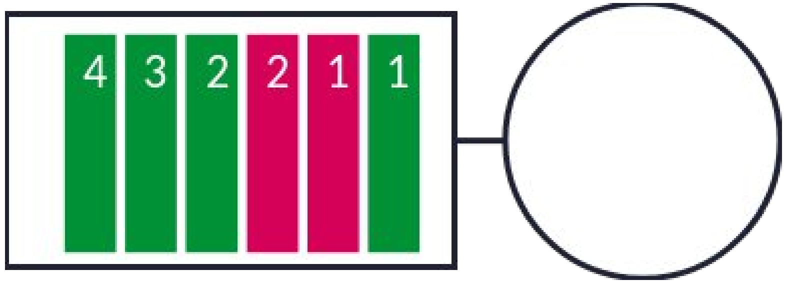
IPv6 – DiffServ: Traffic Class



DS – *Differentiated Service* , DSCP – *Differentiated Service Code Point*, CU – *Currently Unused*

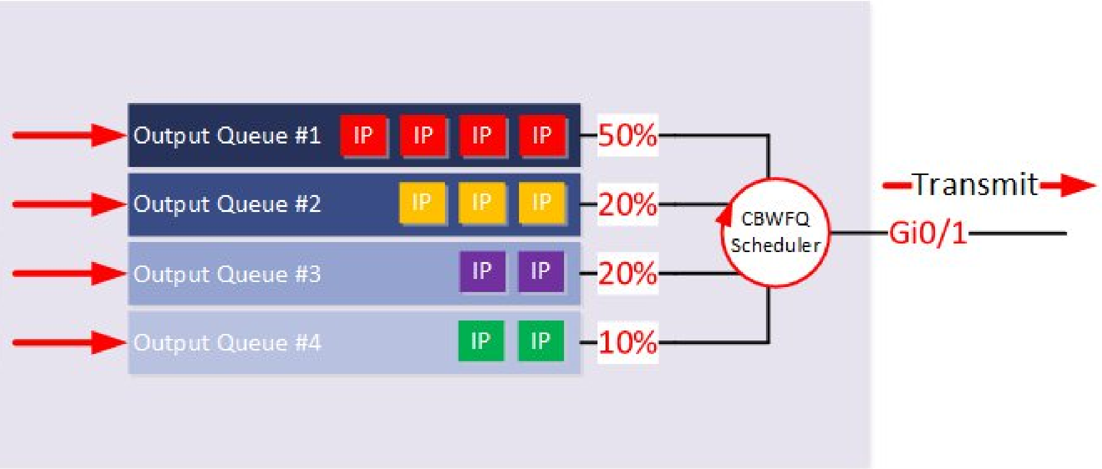
Marking

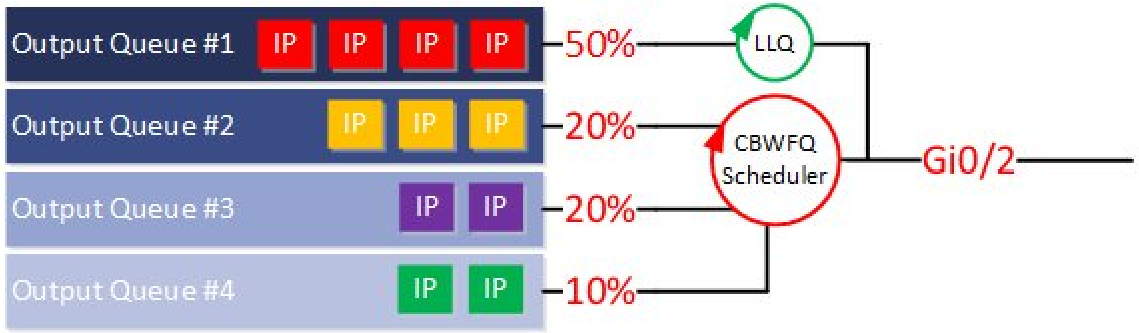
The different types of queues



How to select packets to drop when queue is full?

How to organize the queues so they have their promised performance?

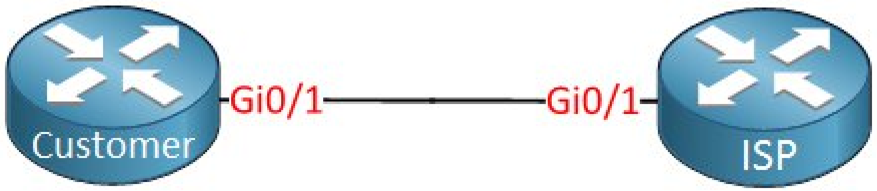


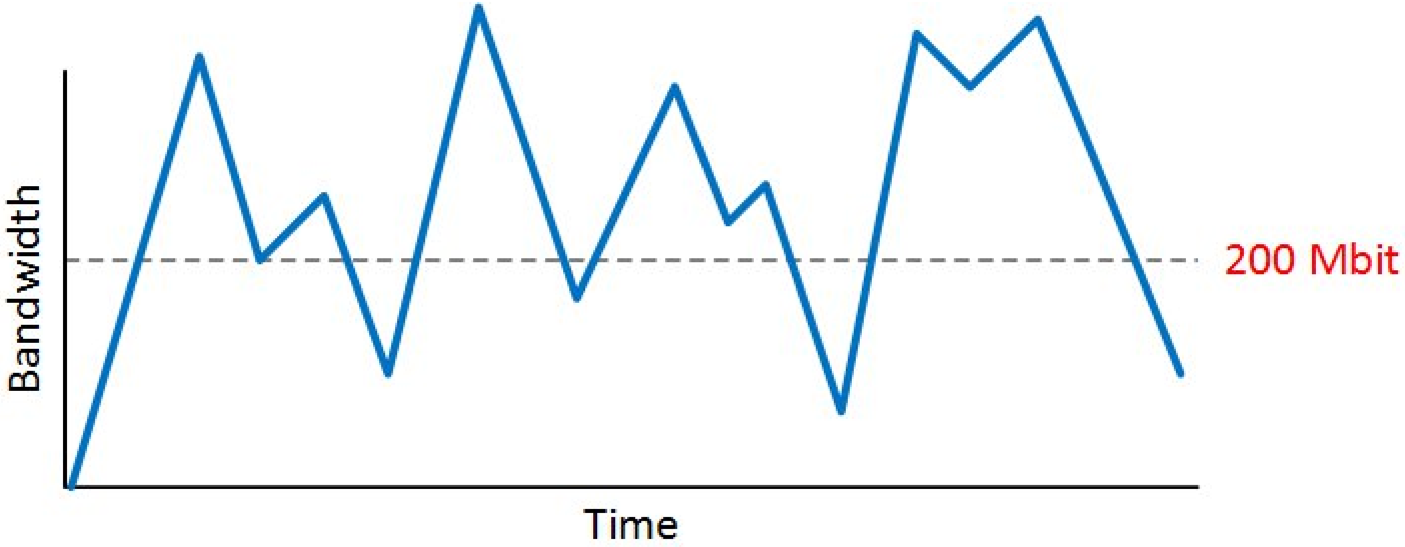


How to organize the queues so that non-priority queues have a chance to transmit?

POLICING

Use Case: Committed Information Rate (CIR)



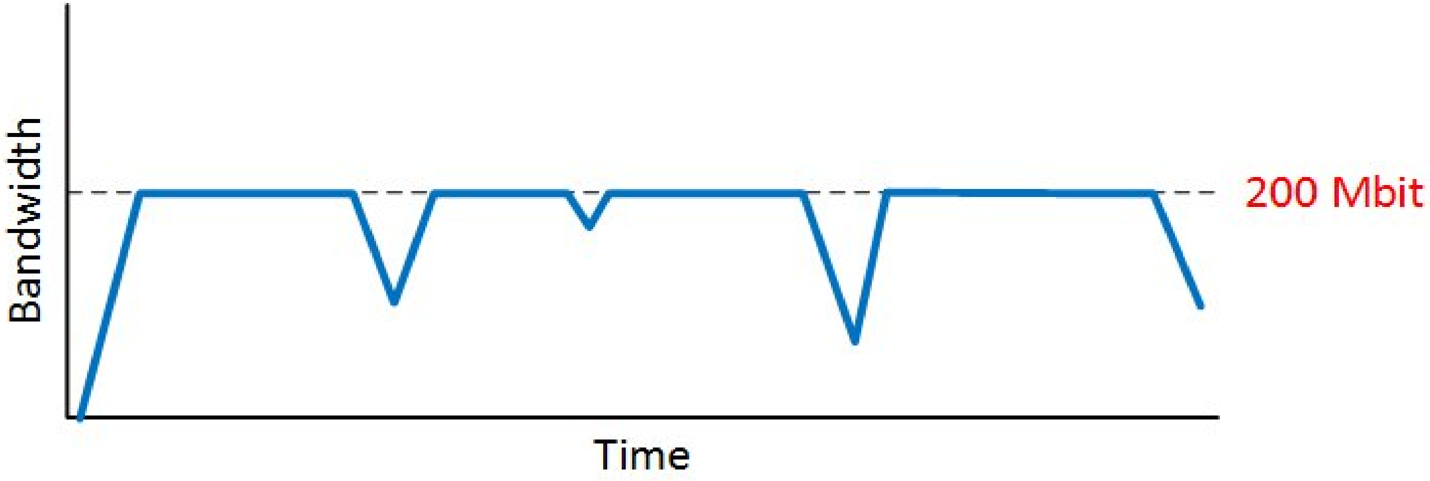


How can we guarantee that the user does not get more throughput than it is actually paying for?

Policing

Definition: Policing is a QoS mechanism used to limit throughput of a given traffic flow by performing one of the following actions to arriving packets:

- Allow packet to pass
- Drop the packet
- Re-mark the packet with different priority



Policing

- In Policing, packets may be categorized in terms of conformity to the traffic contract, i.e.,
  - Conforming: OK rate
  - Exceeding: using the excess burst capacity (more about it later)
  - Violating: higher rate than allowed
- Categories are optional and must be configured
- Example of actions are:
  - Conforming – pass
  - Exceeding – lower priority [optional]
  - Violating – drop

Policing

Single-Rate, Two-Color Policer (Single Bucket)

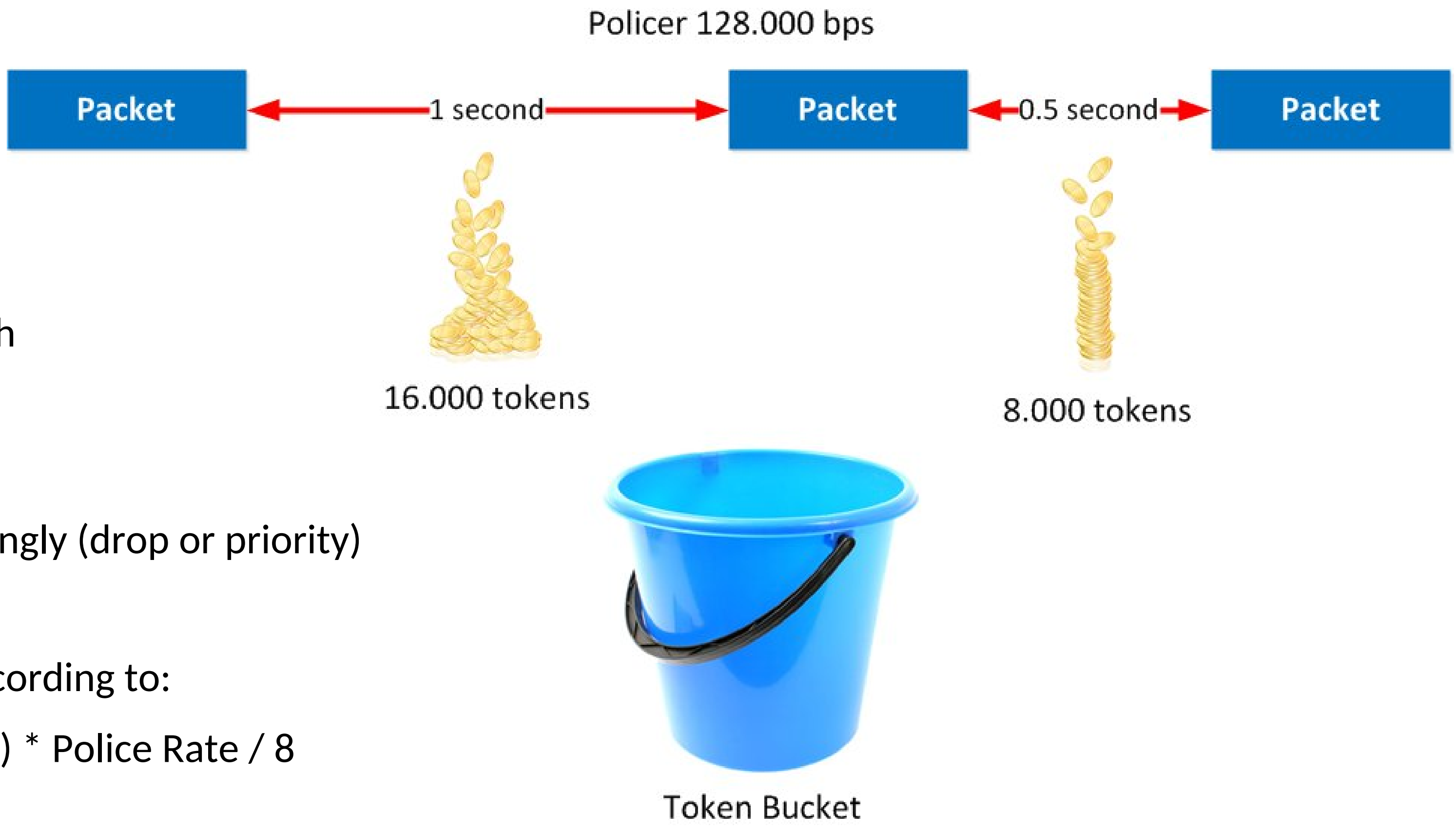
When a new packet arrives:

```
If packet size (Bytes) <= Token budget, then:
    Packet is conforming
    Tokens are consumed and packet goes through
Else:
    Packet is exceeding
    Tokens are kept and packet is handled accordingly (drop or priority)
```

Tokens are **replenished** into the token bucket according to:

(Packet arrival time - Previous packet arrival time) * Police Rate / 8

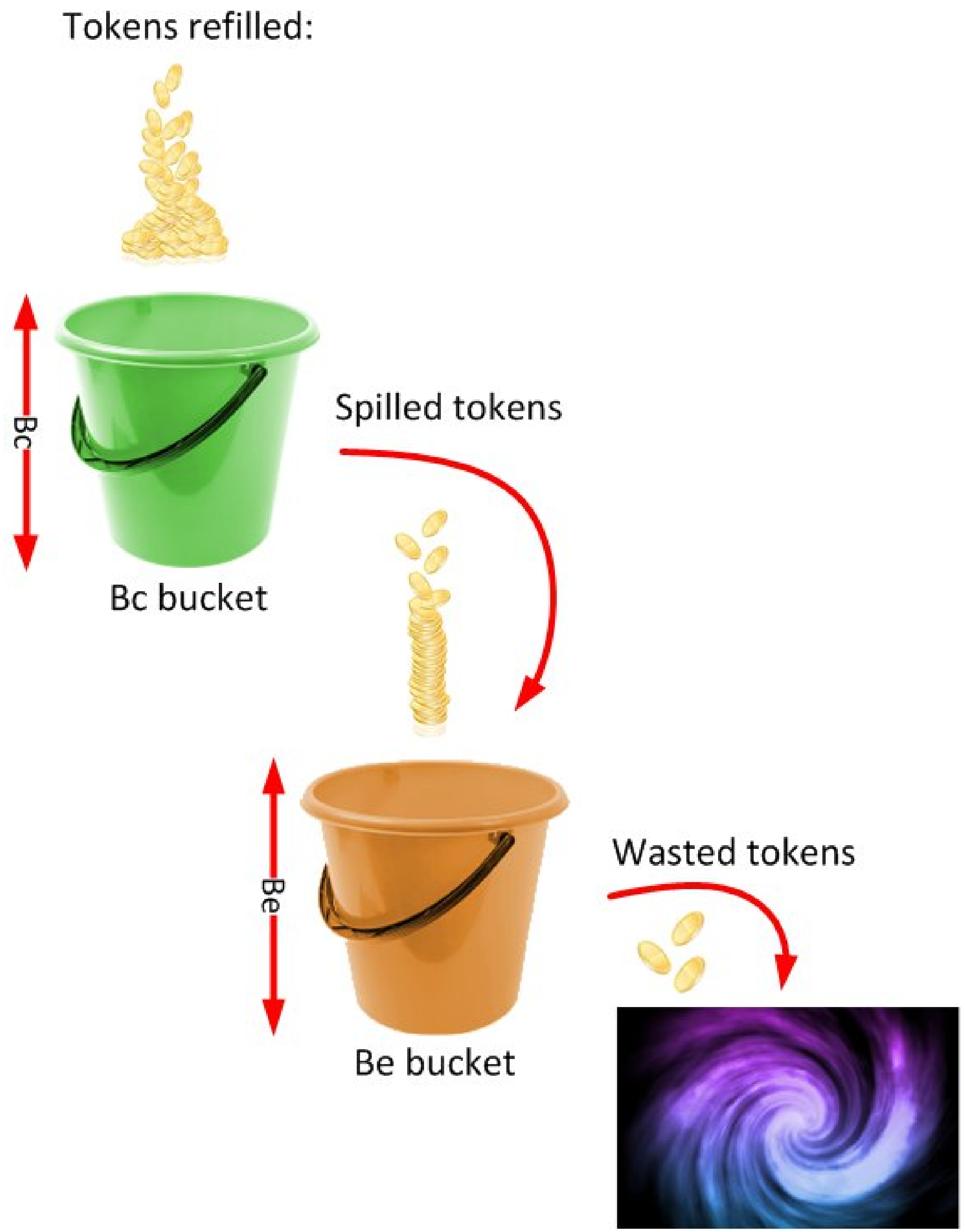
Replenished tokens are "**spilled**" if bucket is full



Policing

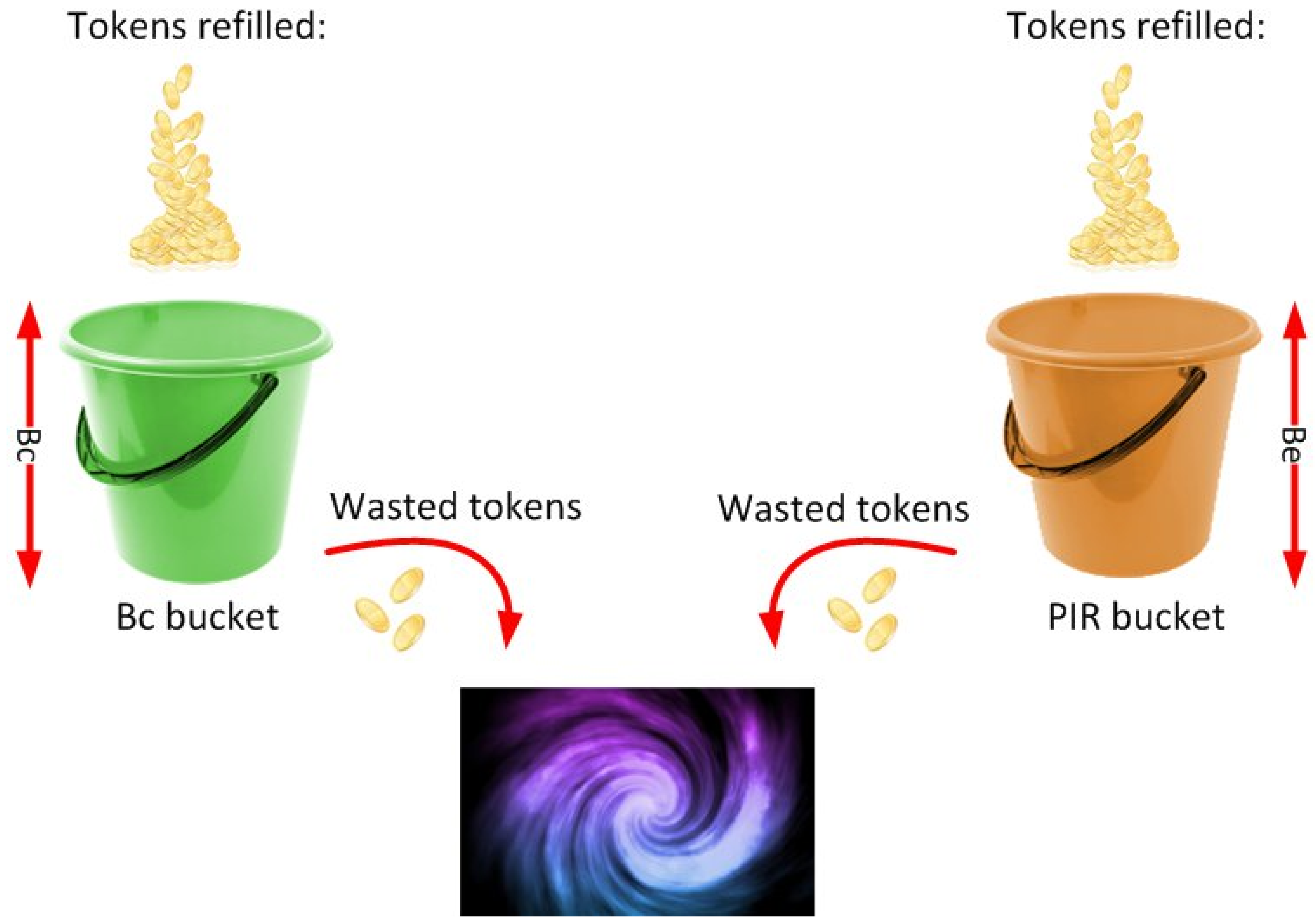
Single-Rate, Three-Color Policer (Two Bucket)

Bc: Committed Burst
Be: Excess Burst



Policing

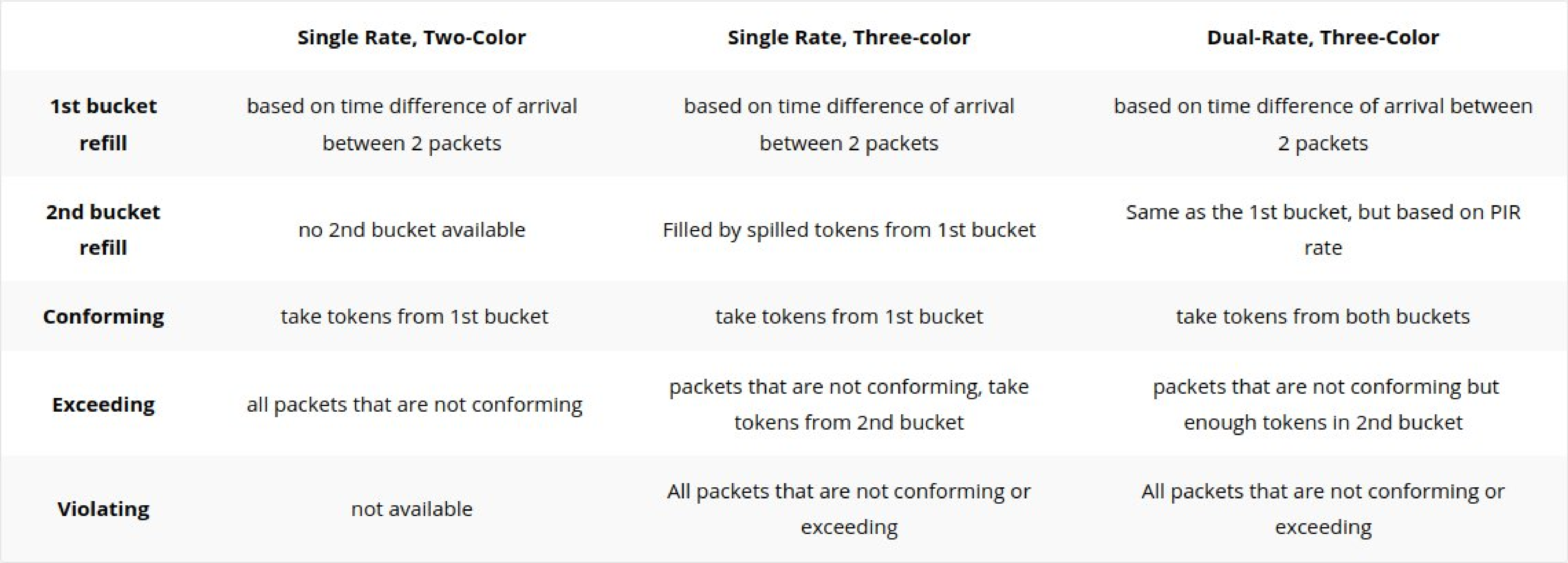
Double-Rate, Three-Color Policer (Two Bucket)



CIR: Committed Information Rate
PIR: Peak Information Rate

Policing

Summary



	Single Rate, Two-Color	Single Rate, Three-color	Dual-Rate, Three-Color
1st bucket refill	based on time difference of arrival between 2 packets	based on time difference of arrival between 2 packets	based on time difference of arrival between 2 packets
2nd bucket refill	no 2nd bucket available	Filled by spilled tokens from 1st bucket	Same as the 1st bucket, but based on PIR rate
Conforming	take tokens from 1st bucket	take tokens from 1st bucket	take tokens from both buckets
Exceeding	all packets that are not conforming	packets that are not conforming, take tokens from 2nd bucket	packets that are not conforming but enough tokens in 2nd bucket
Violating	not available	All packets that are not conforming or exceeding	All packets that are not conforming or exceeding

Wrapping up

What did you learn today?

- Classification
- Marking
  - IP Precedence
  - DiffServ (AF, EF)
- Policing
  - Single-Rate, Two-Color
  - Single-Rate, Three-Color
  - Double-Rate, Three-Color