QUALITY OF SERVICE

CLASS 1

INTRODUCTION

About the instructor

- Name: Guilherme IECKER RICARDO
- Office : F-402
- Email : guilherme.ricardo@irit.fr
- Web page : http://irit.fr/~Guilherme.Ricardo

COURSE PRESENTATION

Course Material

- No textbook !
- *Moodle* : N8EN18A - Qualité de Service
  - Course slides
  - TP and TD handouts
  - Additional resources (e.g., papers, software)

Course Organization

Class	Data	Time	Topic
CM1	24/01/2024	16h15 - 18h00	Networking, traffic characteristics and fairness
CM2	08/02/2024	08h00 - 10h00	Classification, Marking, and Congestion avoidance
CM3	13/02/2024	14h00 - 15h45	Policing, Shaping, and Scheduling
CM4	14/02/2024	14h00 - 15h45	Traffic Engineering: MPLS and RSVP
TD	16/02/2024	R: 08h00 - 09h45 A: 16h15 - 18h00	-
TP1	R: 01/03/2024 A: 01/03/2024	14h00 - 15h45 16h15 - 18h00	TCP Flows and Congestion Control
CM5	05/03/2024	10h15 - 12h00	QoS in modern mobile networks
TP2	R: 06/03/2024 A: 08/03/2024	14h00 - 15h45 08h00 - 09h45	Packet Processing and Queuing
Éxamen	22/03/2024	10h15 - 12h00	-

Before we start…

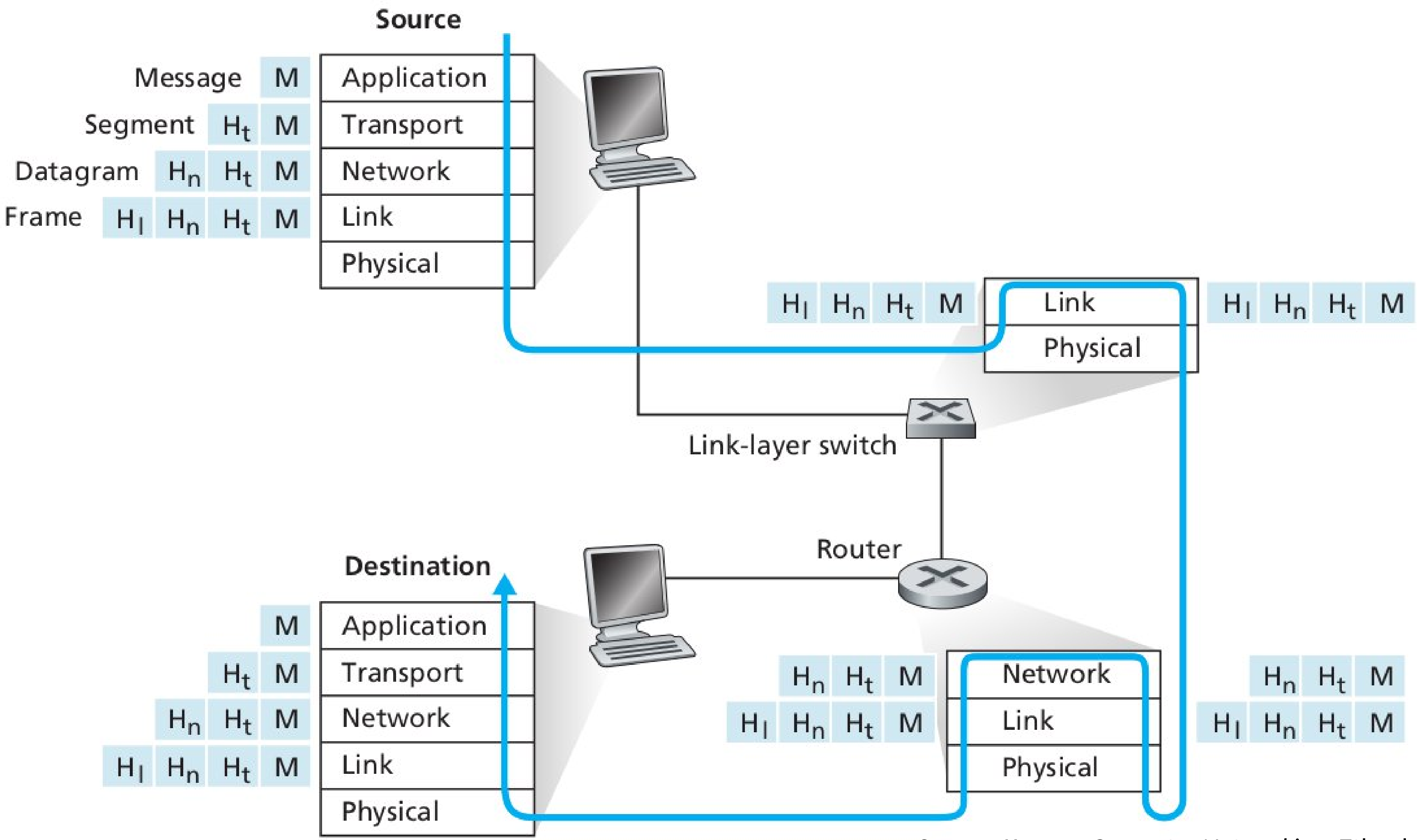
<u>**Objectives**</u>:

- Review of fundamental network operations and performance indicators
- Different traffic characteristics
- Definition of QoS

RECAP OF L3 NETWORKING

The Internet Networking Model

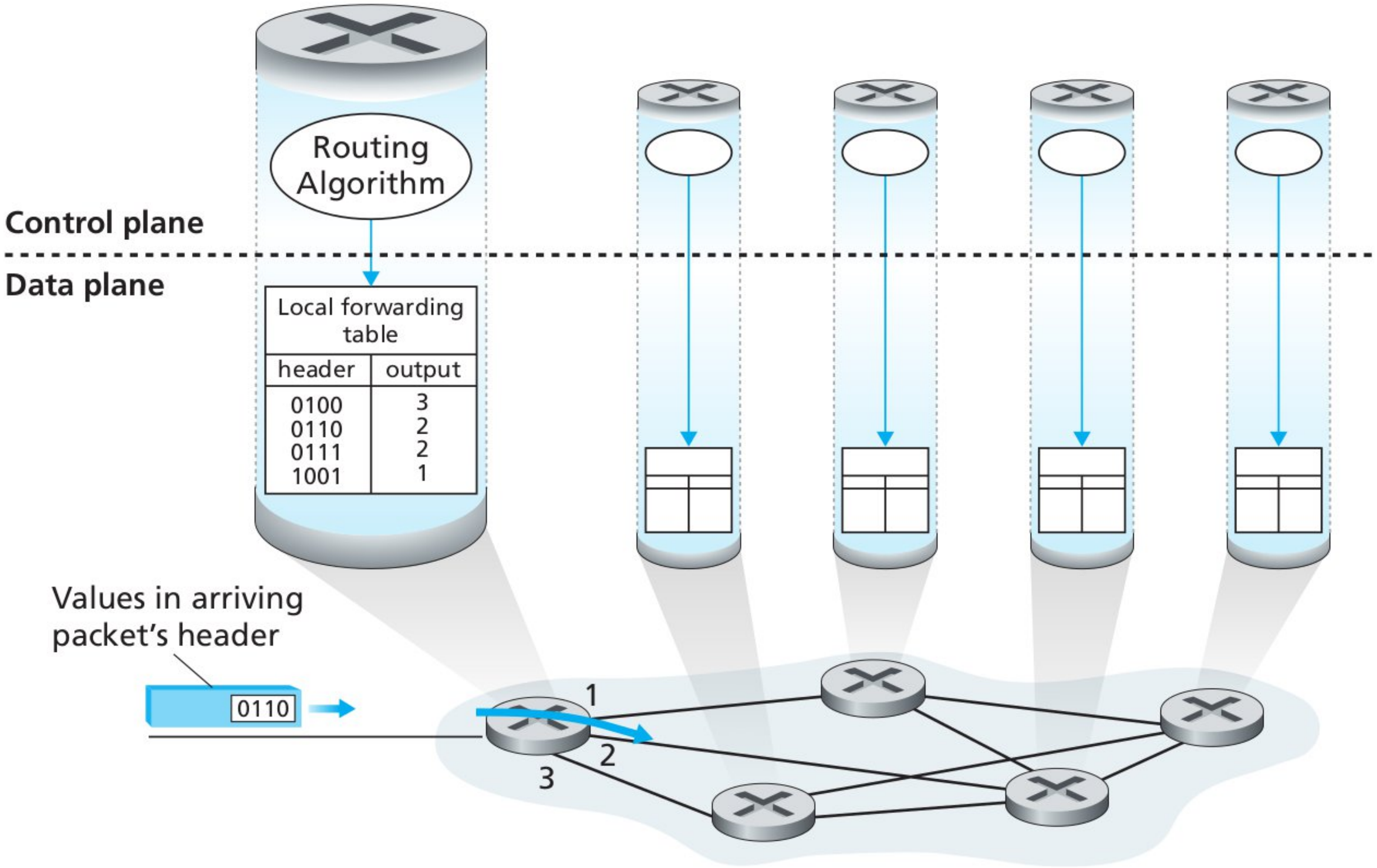
Layered Abstraction of End-to-End Communication



Source: Kurose, *Computer Networking*. 7th ed.

Network Layer

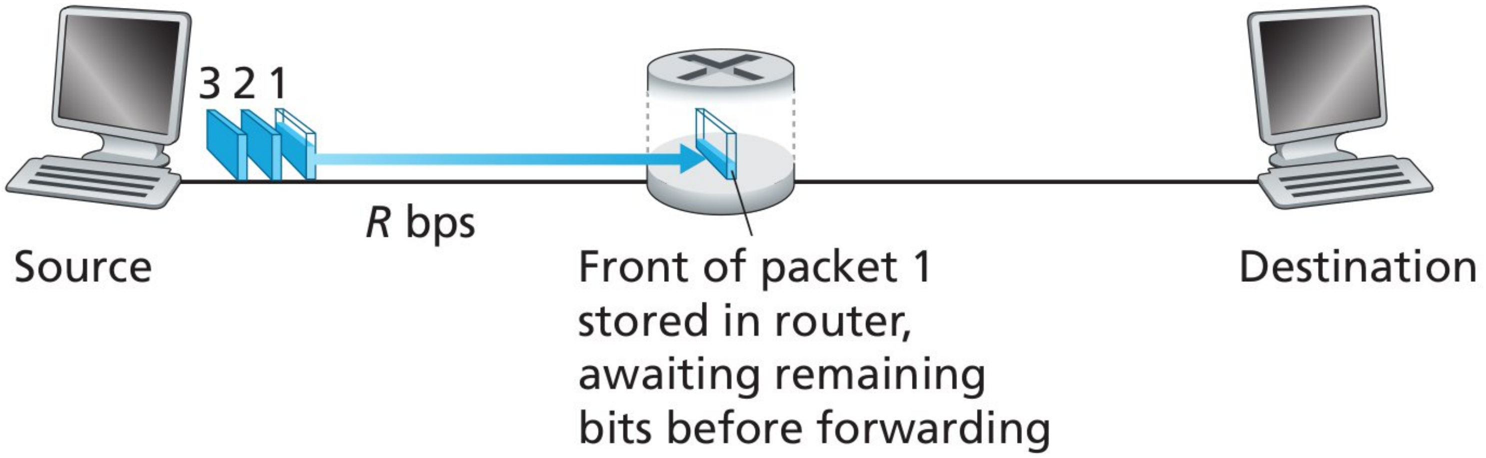
Routing vs. Forwarding



Source: Kurose, *Computer Networking*. 7th ed.

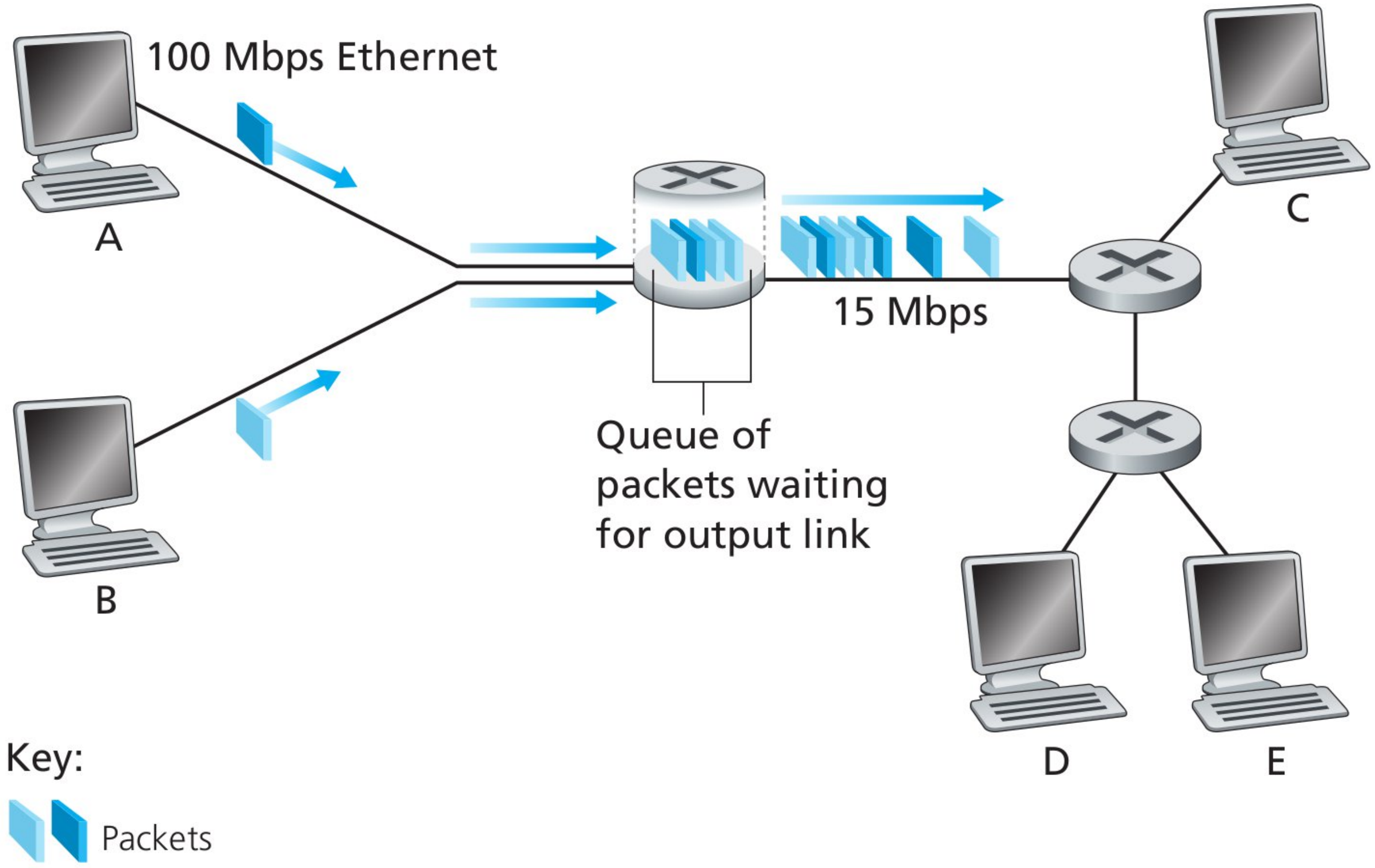
Packet Switching

Store-and-Forward Transmission



Source: Kurose, *Computer Networking*. 7th ed.

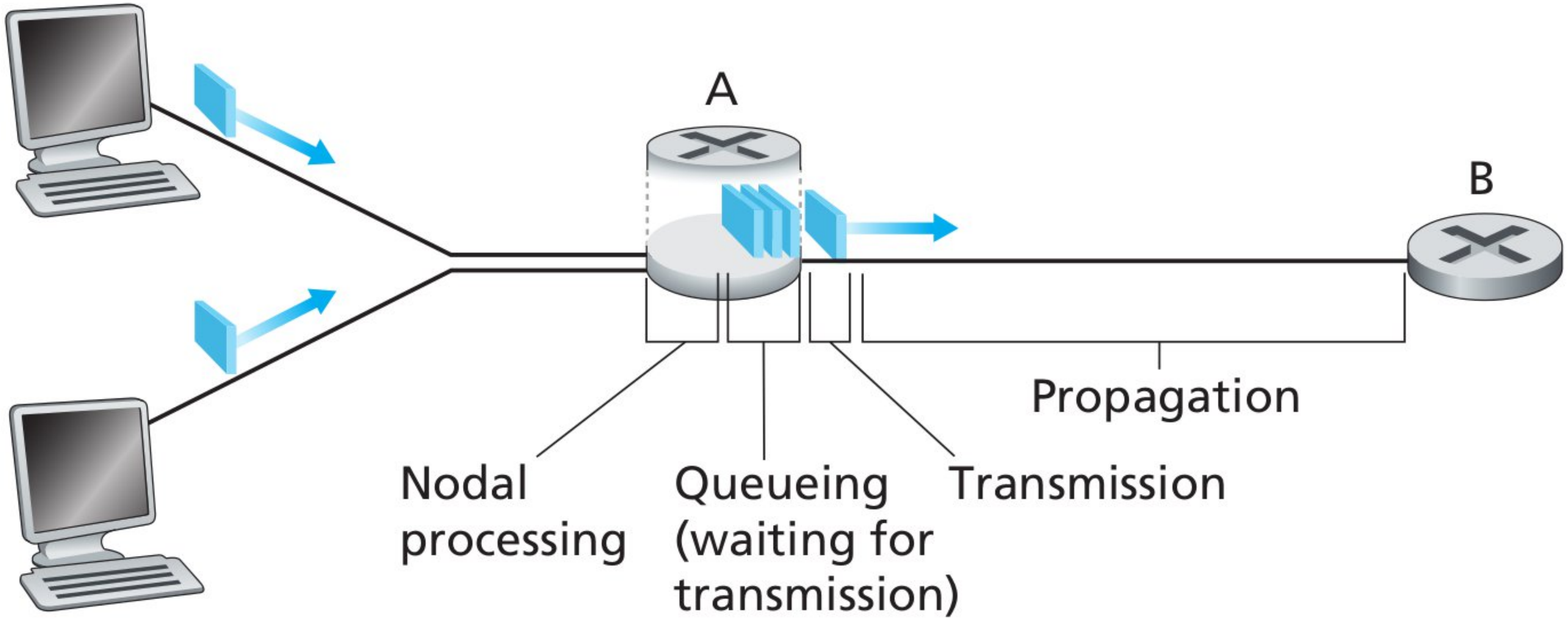
Packet Switching
Queuing



Source: Kurose, *Computer Networking*. 7th ed.

Packet Switching

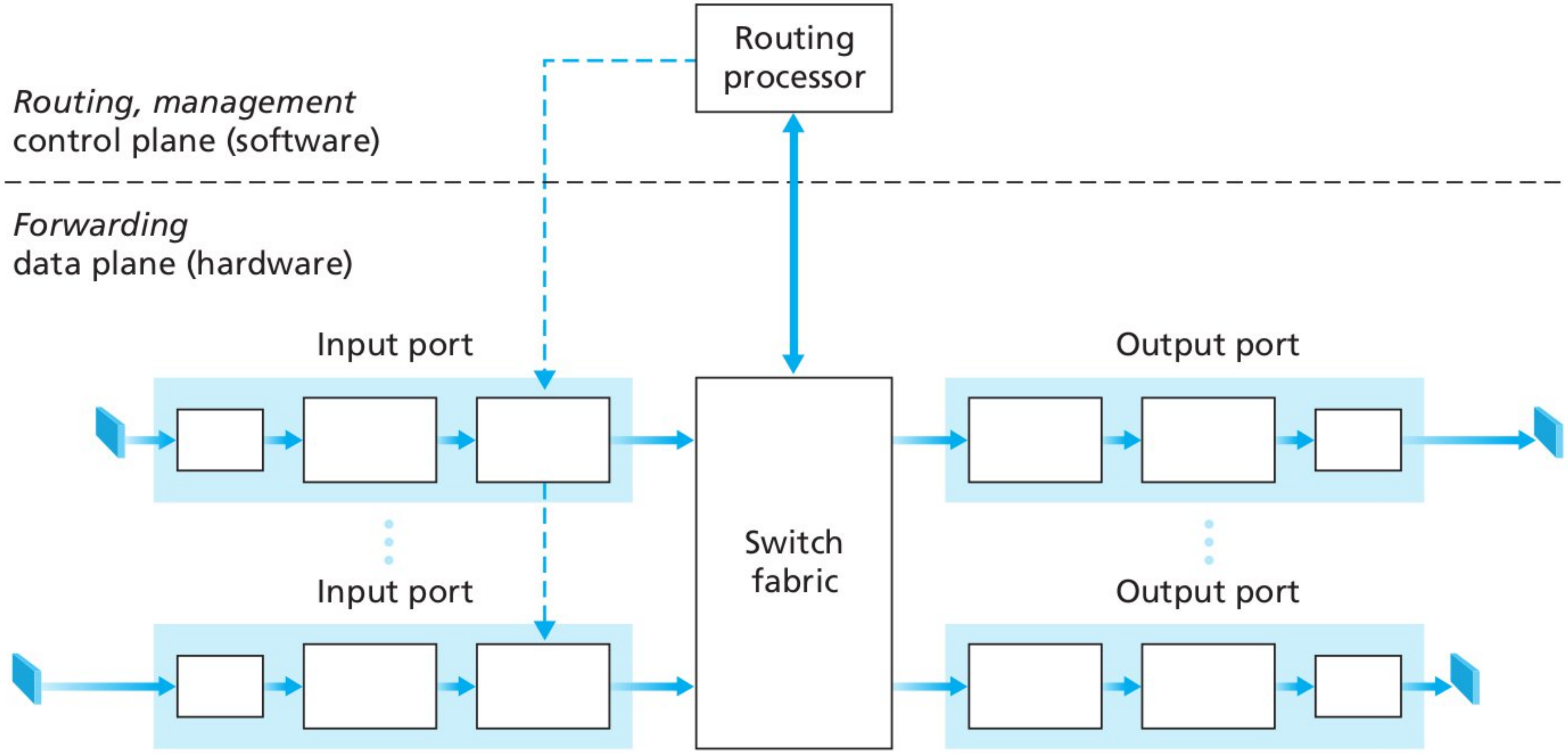
Transmission Stages



Source: Kurose, *Computer Networking*. 7th ed.

Packet Switching

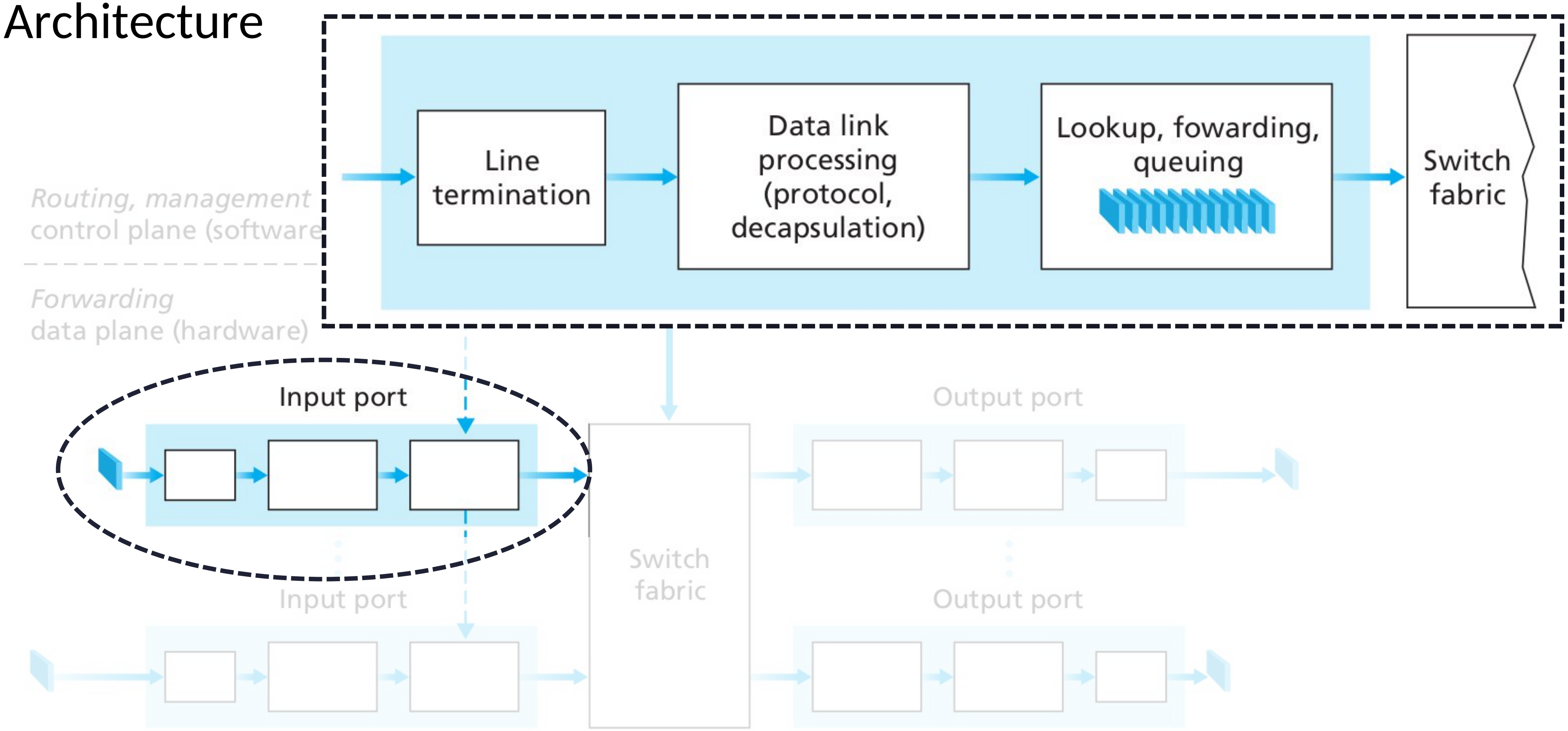
Router Architecture



Source: Kurose, *Computer Networking*. 7th ed.

Packet Switching

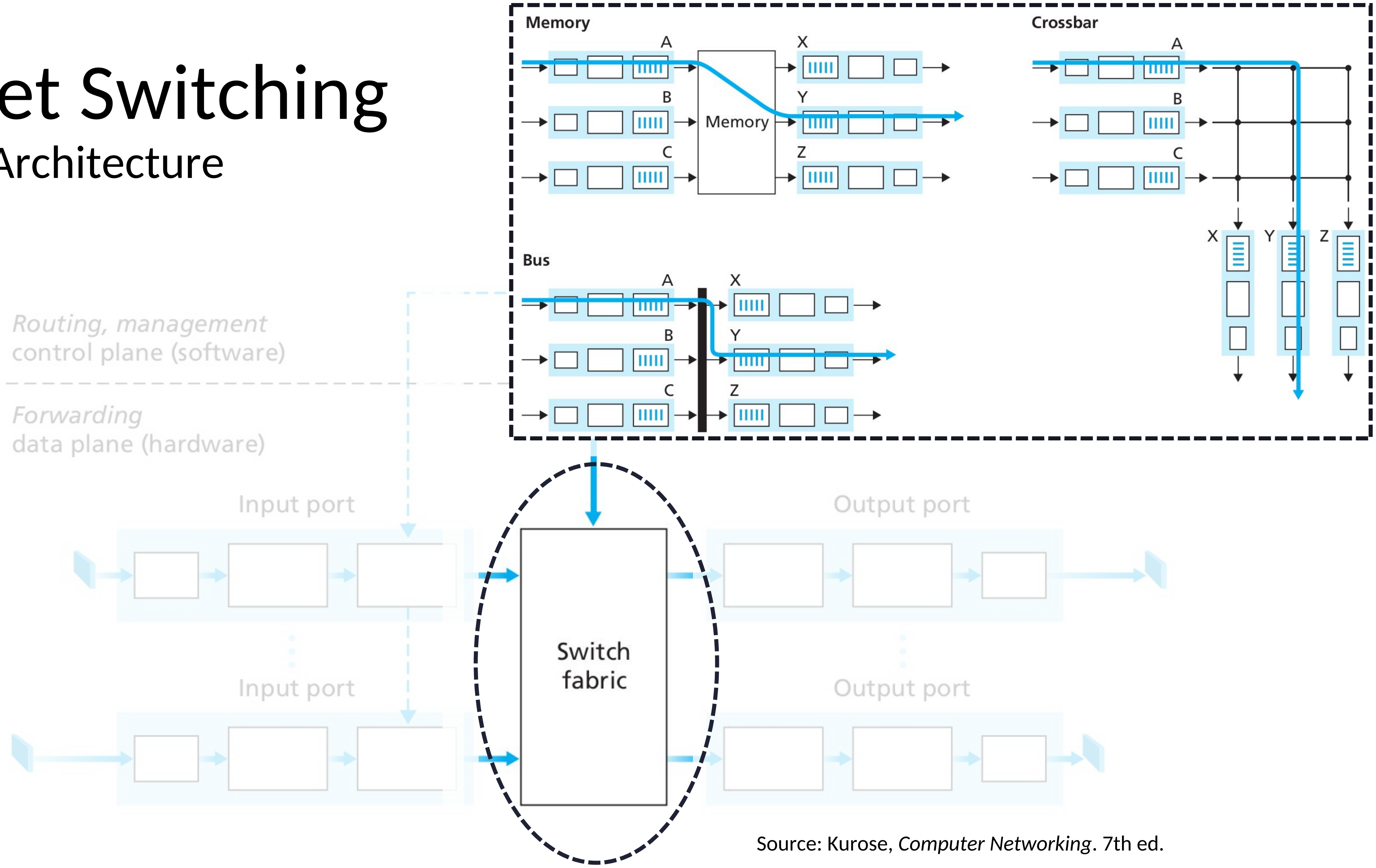
Router Architecture



Source: Kurose, *Computer Networking*. 7th ed.

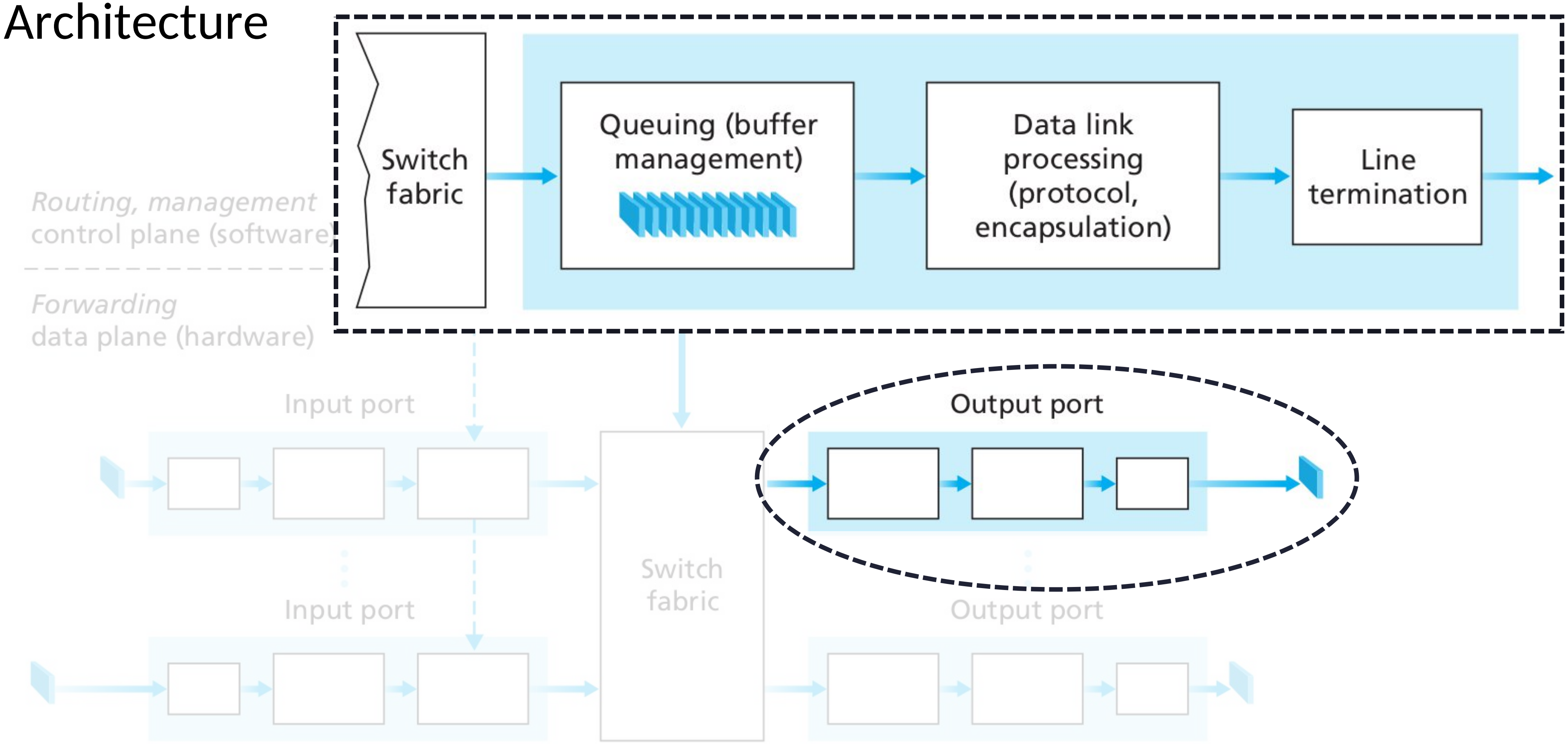
Packet Switching

Router Architecture



Source: Kurose, *Computer Networking*. 7th ed.

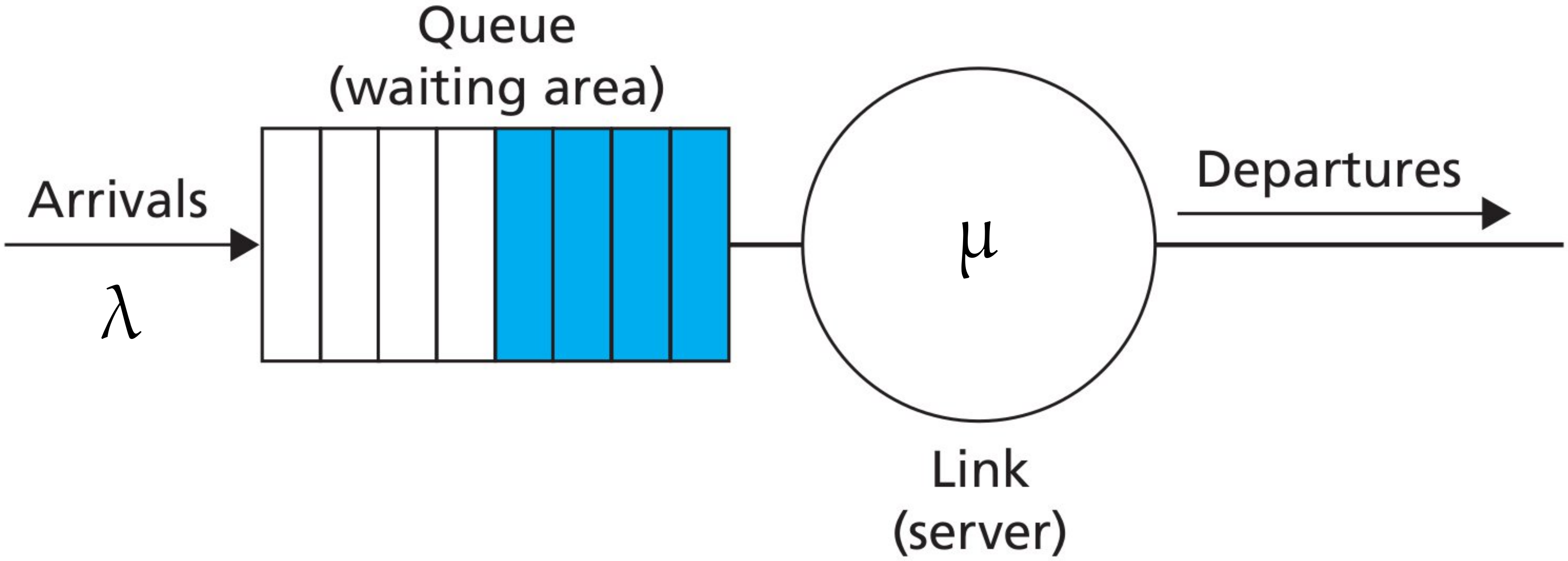
Packet Switching
Router Architecture



Source: Kurose, *Computer Networking*. 7th ed.

Managing Output Queues

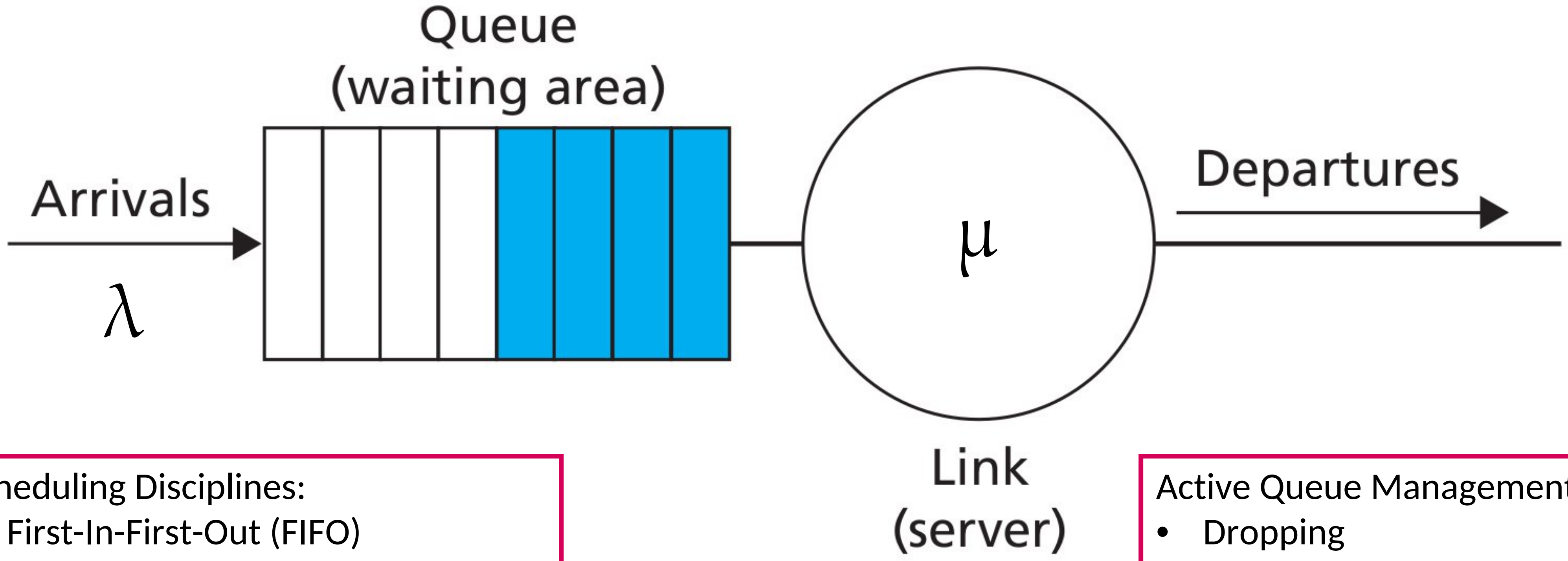
Queuing Theory Representation



Managing Output Queues

Parameters and Scheduling

Parameters:
- Service rate (output link capacity)
- Number of queues
- Queue(s) size
- Arrival rate



Scheduling Disciplines:
- First-In-First-Out (FIFO)
- Round Robin
- Weighted Fair Queuing (WFQ)
- etc.

Active Queue Management (AQM)
- Dropping
- Marking
- Random Early Detection (RED)
- etc.

Network Service Models

General Network Services

Guaranteed Delivery	• Every generated packet is delivered to the destination host, eventually
Delay-bounded Guaranteed Delivery	• Every generated packet is delivered to the destination host • The packet arrives within a given delay tolerance
In-Order Delivery	• Packets arrive at the destination in the order that they were sent
Minimal-Rate Delivery	• Every generated packet is delivered to the destination host, if it is sent at a rate larger than a given bound
Secure Delivery	• Datagram’s payload is encrypted at the source and only decrypted at the destination host, enforcing confidentiality

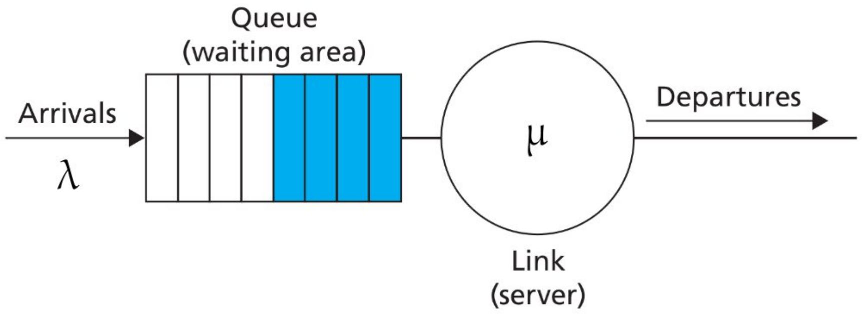
Network Service Models

Example: Internet's Service Model

- ~~Guaranteed Delivery~~
- ~~Delay-bounded Guaranteed Delivery~~
- ~~In-Order Delivery~~
- ~~Minimal-Rate Delivery~~
- ~~Secure Delivery~~

Internet's service model is *Best Effort*, aka., no service at all!

? How to manage output queues in order to implement **Best Effort**?



Network Service Models

Example: Internet's Service Model

Guaranteed Delivery

Delay-bounde
Gu nteed De

In-Ord elive

M mal-Rate Deli y

Secure Delivery

Internet's service model is

all!

Is *Best Effort* a good network service model?
Does it work well?
Does it work the same way for everyone?

We are about to discuss performance (or quality) of a service model. But first, let's define a few performance indicators in networking.

eues in order

NETWORK TRAFFIC CHARACTERISTICS

Characteristics of Network Traffic

Throughput	Delay	Jitter	Loss

Characteristics of Network Traffic

Throughput	Delay	Jitter	Loss

<ins>**Definition**</ins>: Link's capacity, commonly (and **wrongly**) referred to as bandwidth, is the maximum data rate achievable in a given link. Simply put, it is the link's speed. Capacity is measured in *bits per second* (bps).

Example of capacity model is the Shannon-Hartley theorem:

$$R = B \cdot log_2\left(1 + \frac{S}{N}\right)$$

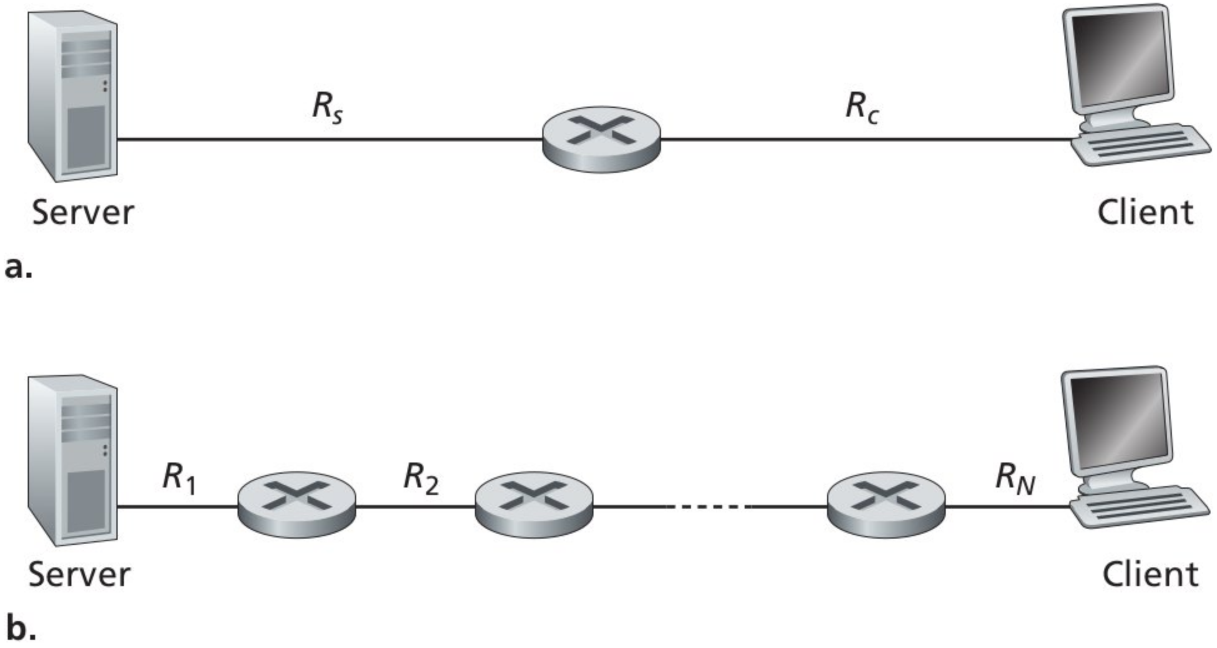
R: Capacity [bps]
B: Bandwidth [Hz]
S: Transmission power [W]
N: Noise and Interference [W]

? What is the difference between bandwidth and capacity?

Characteristics of Network Traffic

Throughput | Delay | Jitter | Loss

<u>**Definition**</u>: End-to-end throughput, is the speed, in bits/s, at which a client host receives data from the server. It is bounded by the link with the smallest capacity in the communication path, which is called a bottleneck link.



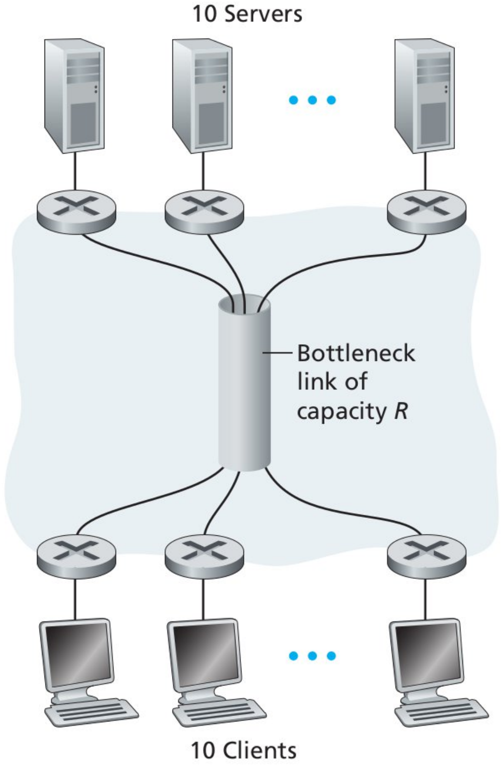
$$T = min(R_s, R_c)$$

$$T = \min_{i=1,\ldots,N} \{R_i\}$$

Characteristics of Network Traffic

Throughput	Delay	Jitter	Loss

Bottlenecks can also be formed when multiple data flows simultaneously share the link.



What is the throughput of each server-client communication in this case?

Answer:

- If R is sufficiently large, R >> Rc and R >> Rs, then:

$$T=\min(R_c, R_s).$$

- Otherwise, considering that R is equally split between the 10 data flows, then:

$$T=\min(R_c, R_s, R/10)$$

Characteristics of Network Traffic

Throughput | **Delay** | **Jitter** | **Loss**

Definition: The nodal delay is the time for a packet to be transmitted from one network node to the next one. It the sum of the delay component at every communication stage, i.e.,

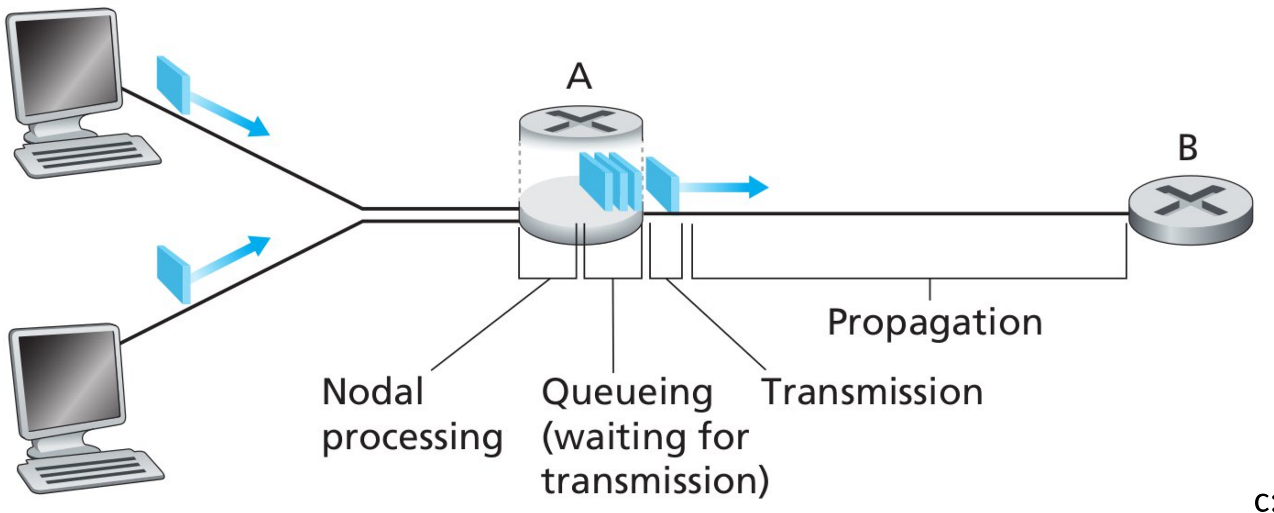
$$d_i^{Nodal} = d_i^{Proc} + d_{i,i+1}^{Queue} + d_{i,i+1}^{Trans} + d_{i,i+1}^{Prop}$$

Processing Delay

d_i^{Proc} (constant)

Queuing Delay

$d_{i,j}^{Queue}$ (later)



Transmission Delay

$$d_{i,j}^{Trans} = \frac{L}{R_{i,j}}$$

L: Packet size [bits/packet]

Propagation Delay

$$d_{i,j}^{Prop} = \frac{dist_{i,j}}{c}$$

c: Speed of light $3.0 \cdot 10^8$m/s

Characteristics of Network Traffic

Throughput	**Delay**	Jitter	Loss

Special case: The **queuing delay** is sensitive to the experienced **traffic intensity** $\rho_{i,j}$ going through link (i,j).

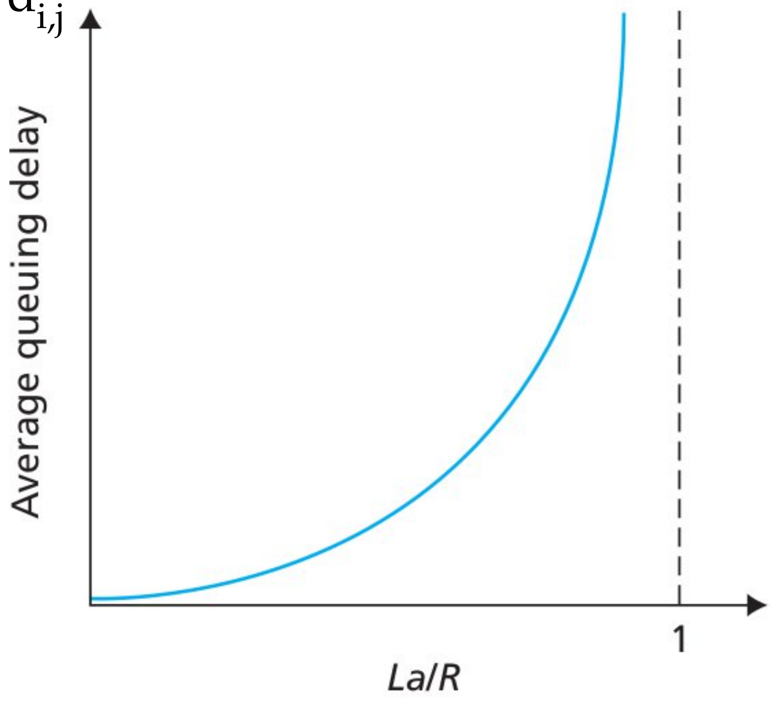
$$\rho_{i,j} = \frac{L \cdot a}{R_{i,j}}$$

L: Packet size [bits/packet]
a: Average rate [packets/s]
$R_{i,j}$: Link (i,j) transmission rate [bps]

Traffic engineering design rule

$$0 \leq \rho_{i,j} \leq 1$$

$d_{i,j}^{Queue}$



Characteristics of Network Traffic

Throughput	**Delay**	Jitter	Loss

<u>Definition</u>: The end-to-end delay (or one-way delay) is the time, in *seconds* (s), for a packet to get from the source to a destination host. It is given by the sum of the nodal delay at all <u>transmitting</u> nodes in the data flow $F_{s \to c}$ from server s to client c:

$$d^{E2E} = \sum_{i \in F_{s \to c}} d_i^{Nodal} = \sum_{i \in F_{s \to c}} d_i^{Proc} + d_{i,i+1}^{Queue} + d_{i,i+1}^{Trans} + d_{i,i+1}^{Prop}$$

The time it takes to get from a source to the destination and back is called the round-trip delay, commonly referred to as RTT (round-trip time).

$$RTT = \sum_{i \in F_{s \to c}} d_i^{Nodal} + \sum_{j \in F_{c \to s}} d_j^{Nodal}$$

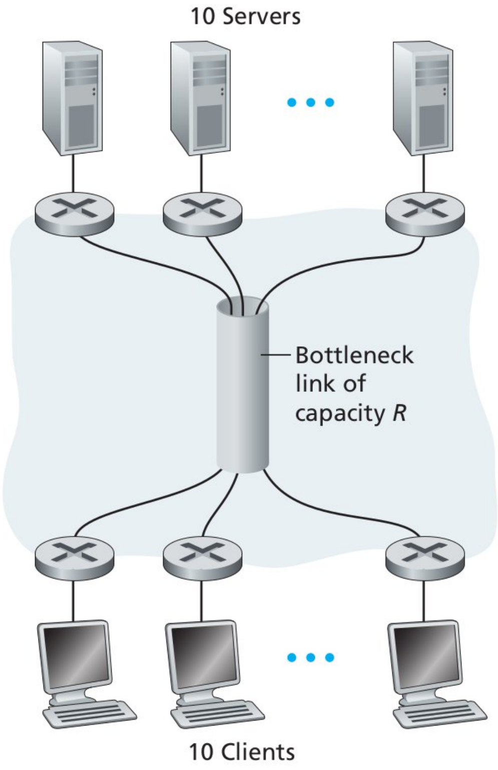
Characteristics of Network Traffic

Throughput	**Delay**	Jitter	Loss

Consider the network of the figure.

We observed that the amount of traffic sent by the servers doubled and it caused a significant increase in the end-to-end delay for the packets transiting from servers to clients.

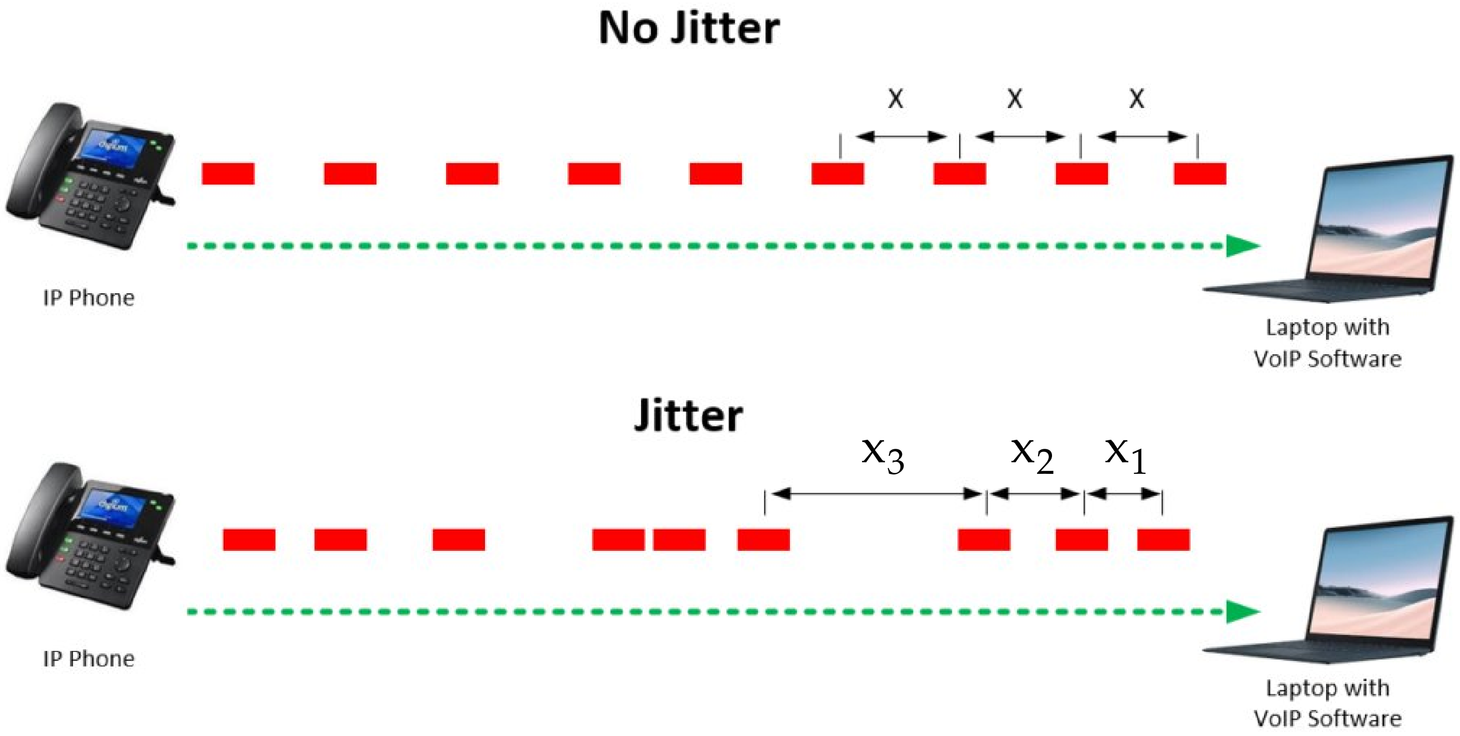
– What delay component (processing, queuing, transmission, or propagation) is most likely to be the cause of the end-to-end delay increase and why?
– Can you think of a way to reduce the end-to-end delay in this case?



Characteristics of Network Traffic

Throughput	Delay	**Jitter**	Loss

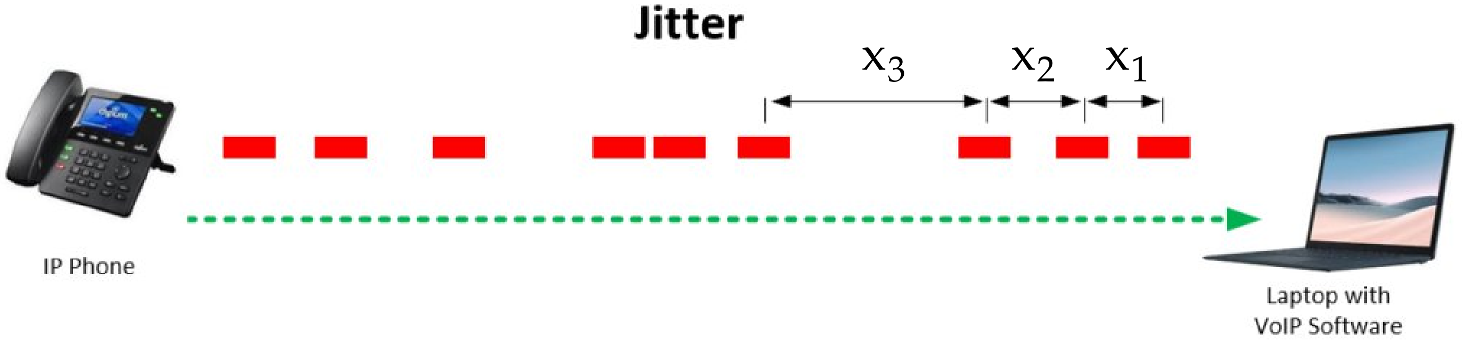
Definition: Instantaneous jitter is the variability of one-way delay between 2 consecutive packets in a data flow (or stream of packets), given in *seconds* (s).



Characteristics of Network Traffic

Throughput	Delay	**Jitter**	Loss

Example:



Data flow of N=3 packets.
E2E delay for packet i=1 is $d^{E2E}(1) = 20\text{ms}.$

For packet i=2, $d^{E2E}(2) = 30\text{ms}.$ $\quad$ $\text{Jitter}(2) = d^{E2E}(2) - d^{E2E}(1) = 30 - 20 = +10\text{ms}$ (Dispersion)

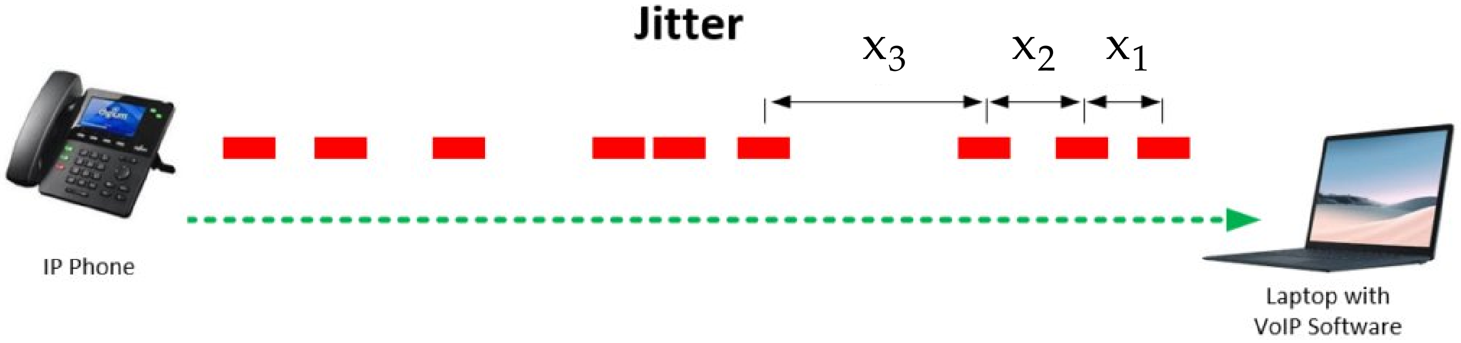
For packet i=3, $d^{E2E}(3) = 10\text{ms}.$ $\quad$ $\text{Jitter}(3) = d^{E2E}(3) - d^{E2E}(2) = 10 - 30 = -20\text{ms}$ (Clumping)

Characteristics of Network Traffic

Throughput	Delay	**Jitter**	Loss

Definition: Average jitter is the variance of one-way delay in a stream of packets in *seconds* (s).

Let X be a random variable capturing the inter-arrival time between two consecutive packets of the same flow. Consider that its expected value $E[X]$ and its variance $Var[X]$ are known.



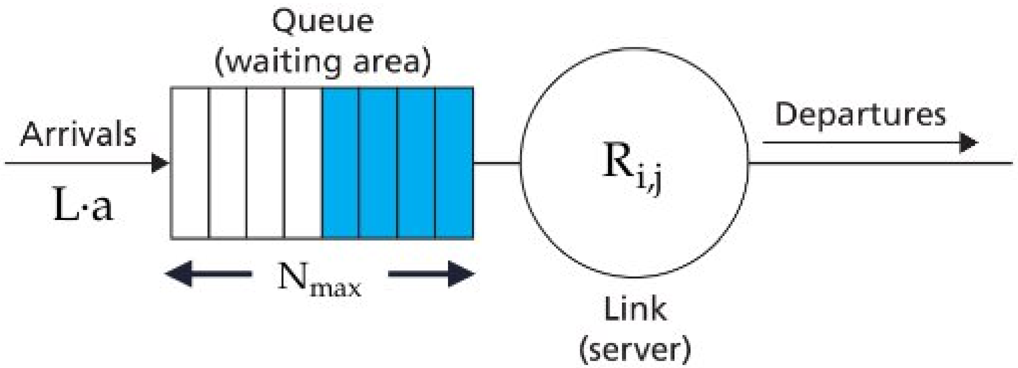
In practice, if X={x_1,..., x_N} is the sample space containing the inter-arrival time of N packets, Jitter is the **sample variance** S_X^2, i.e.,

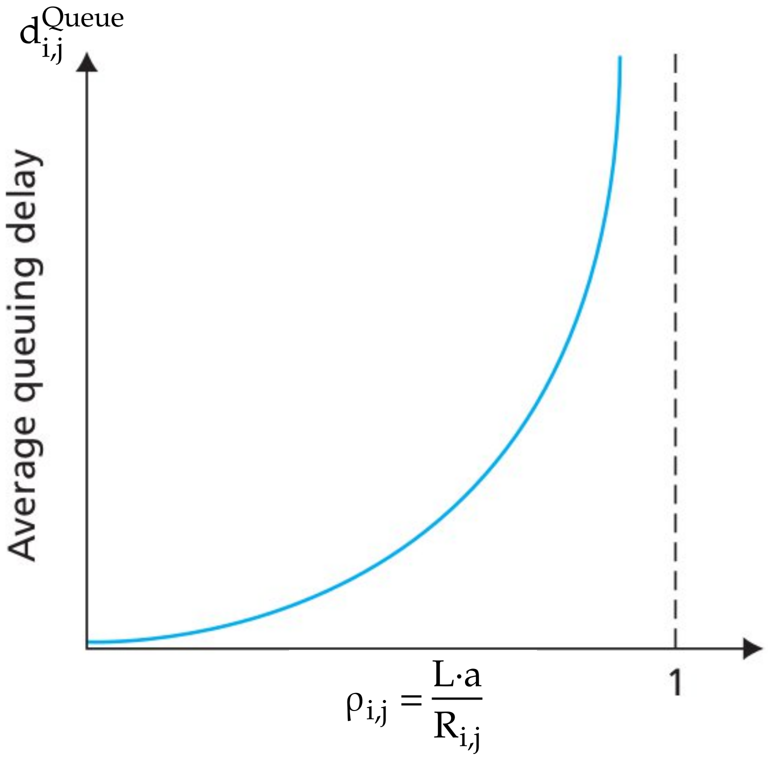
$$S_X^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2, \quad \text{where } \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Characteristics of Network Traffic

Throughput	Delay	Jitter	**Loss**

Definition: The *percentage* of packets that are dropped due to full queues is called packet loss.





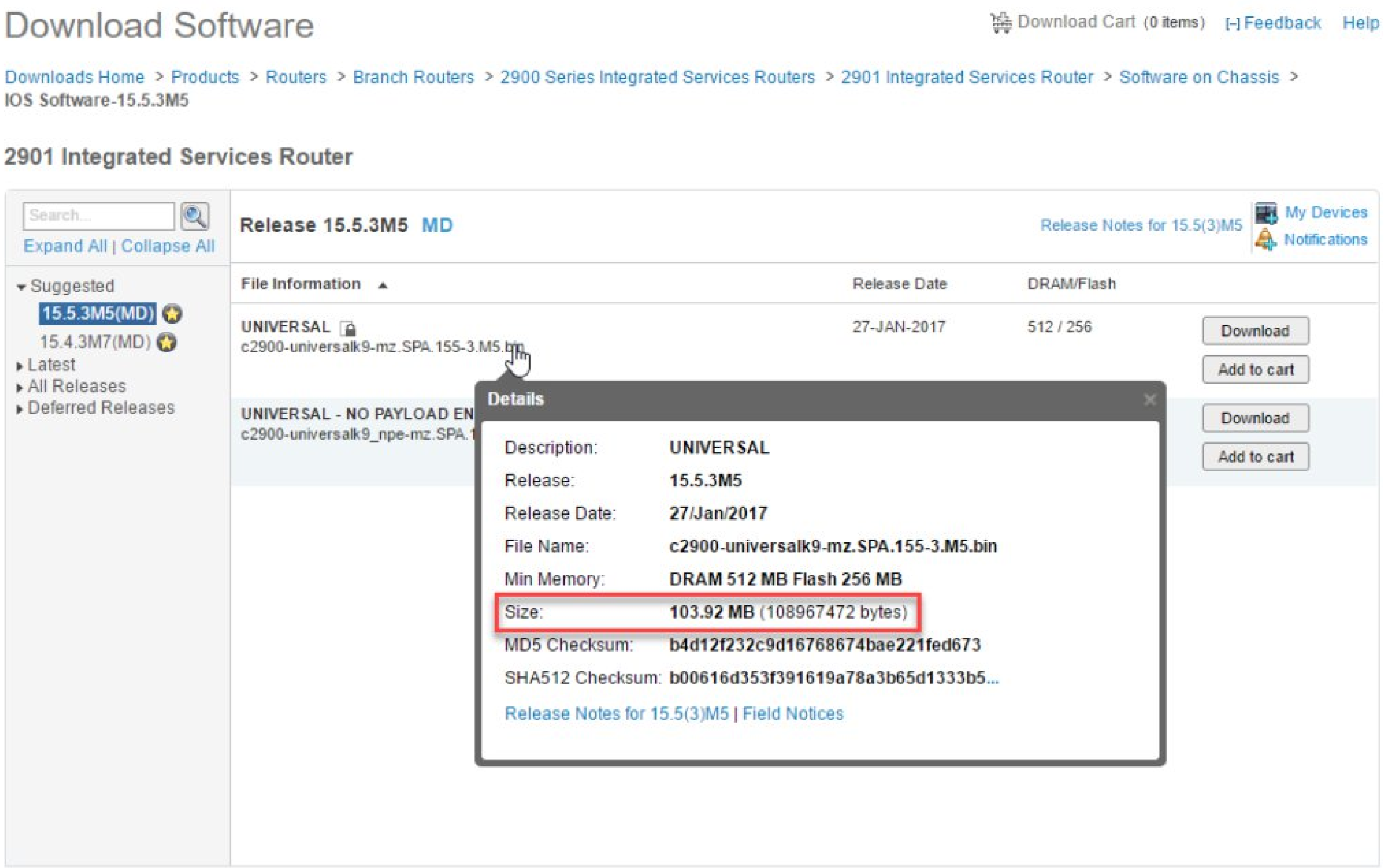
The expected number of packets in the queue is given by Little's Law:

$E[N] = a \cdot d_{i,j}^{Queue}$ [# of packets in the queue]

Packet drop condition:

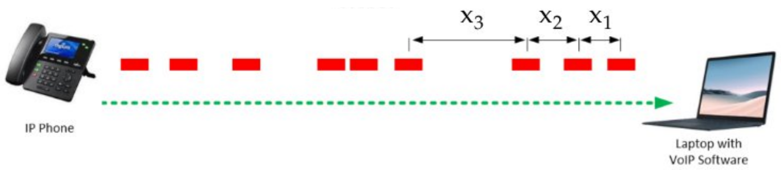
$$a \cdot d_{i,j}^{Queue} > N_{max}$$

Traffic types
Batch Application



File size: 103.92MB = 103.92x10^6Bytes
Datagram size: 1.5x10^3 Bytes
Datagram payload: 1.46x10^3 Bytes

Nb. of datagrams: 71,179 datagrams

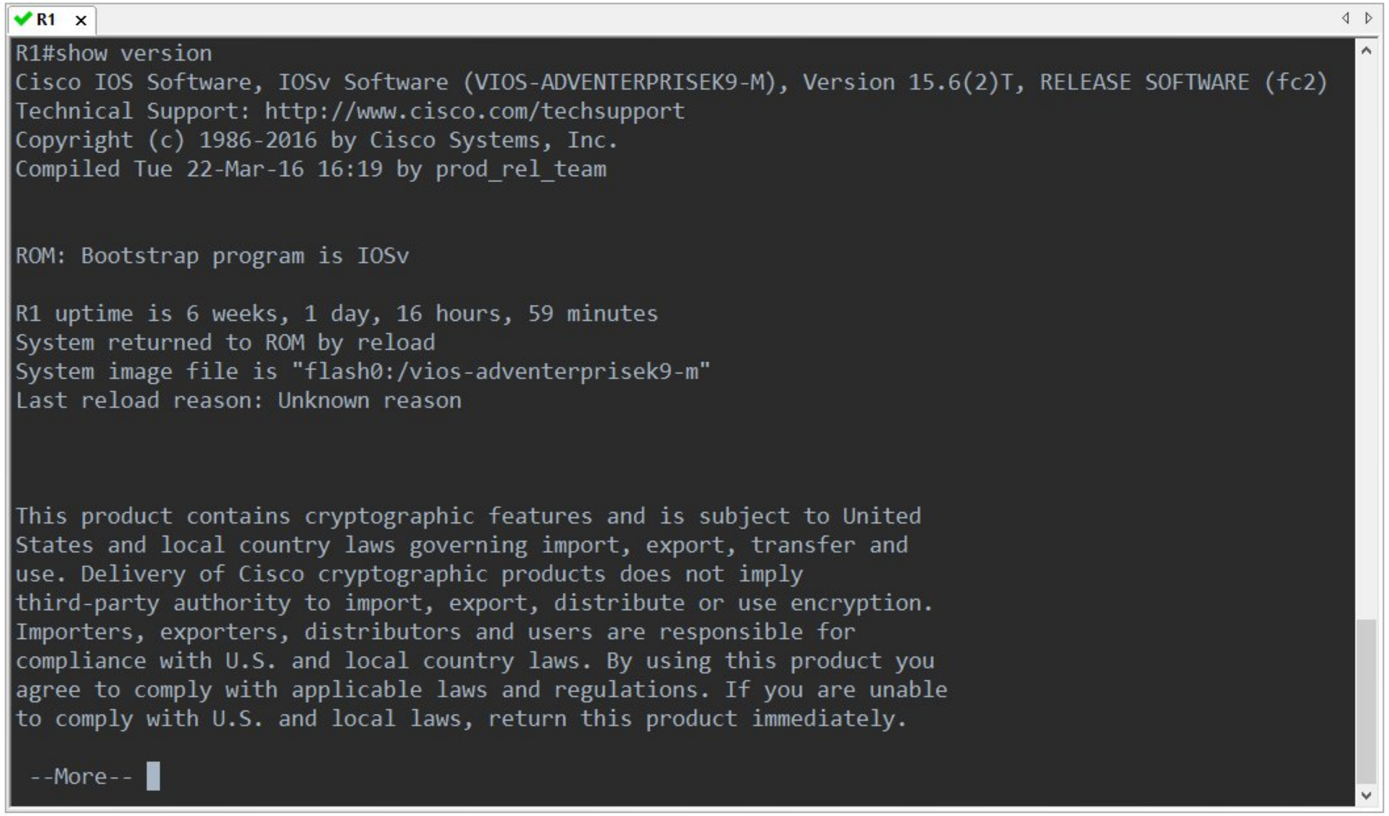


What characteristics are important?

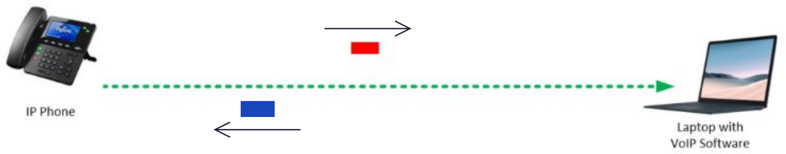
- Throughput
- Delay
- Jitter
- Loss

Traffic types

Interactive Application



Simple Instruction-Response Application: SSH

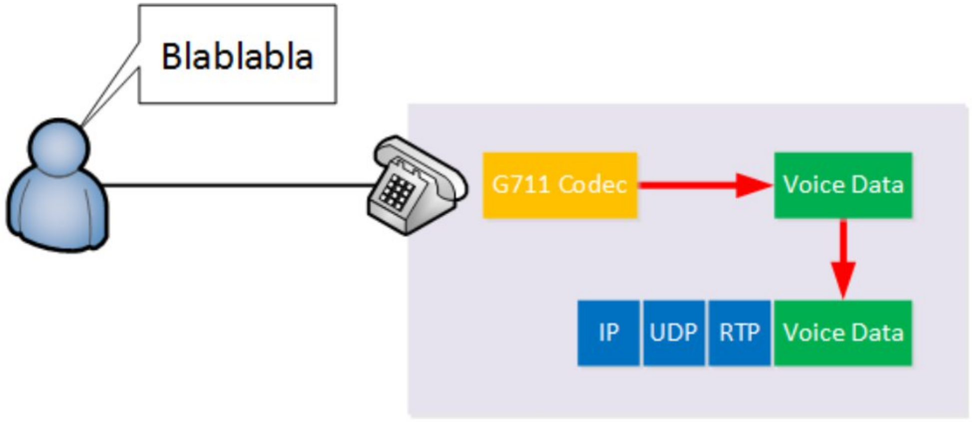


What characteristics are important?

- Throughput
- Delay
- Jitter
- Loss

Traffic types

Voice Application



Steps:

- Analog-Digital converter Codec
- Generate digital voice data
- Encapsulate data with IP/UDP/RTP headers

A few statistics:

- 1 s of voice data = 50 datagrams
- 24 - 80 kbps of required throughput
- Analog sound is digitized every 20ms

What characteristics are important?

- Throughput
- Delay
- Jitter
- Loss

How does this analysis change for voice and video applications?

QUALITY OF SERVICE

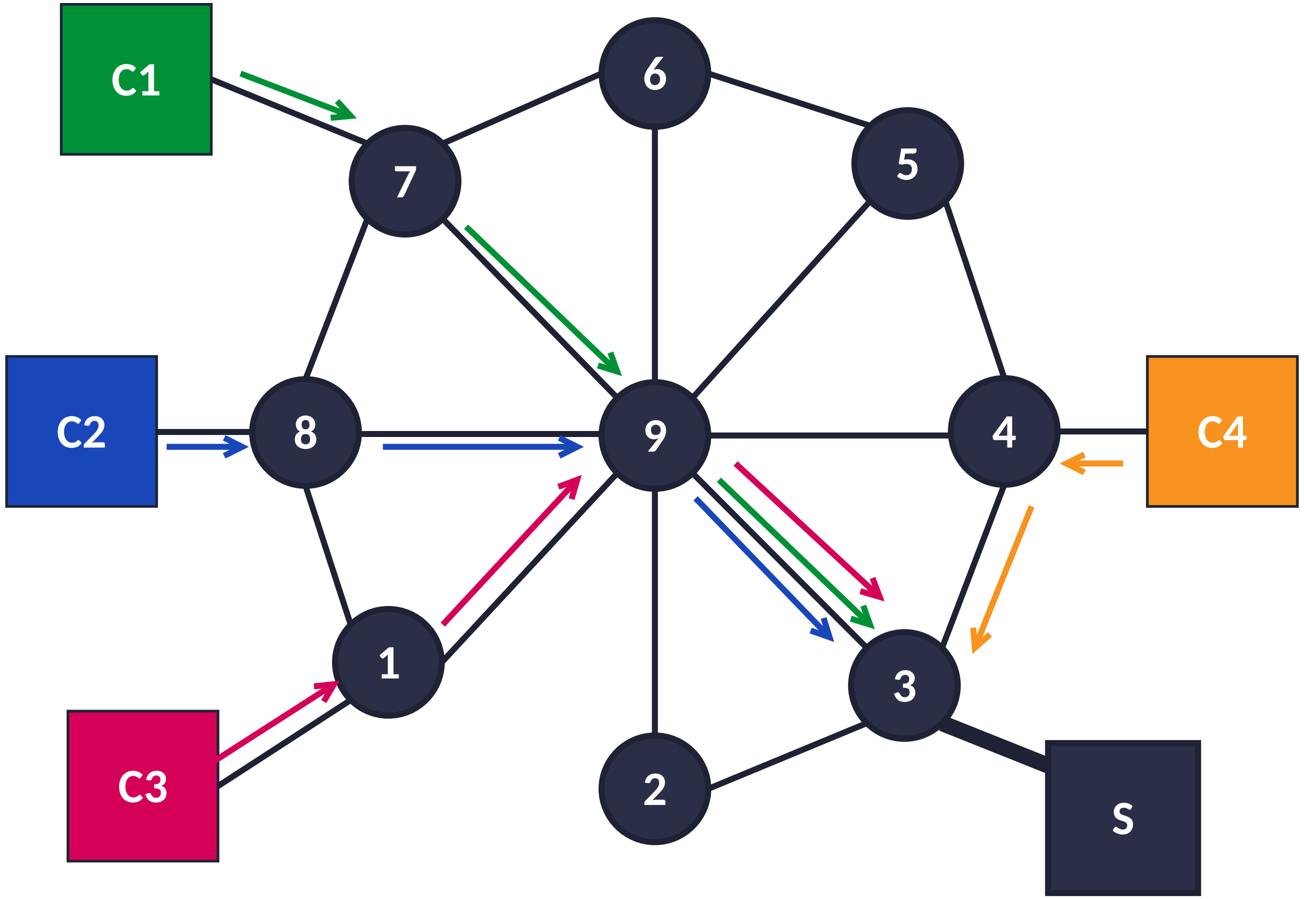
Fairness

- Fuzzy concept in networking. It may have multiple definitions, e.g.,
  - All clients’ traffics are processed under the conditions (equal rules)
  - All clients’ traffics are promised the same characteristics (equal experience)
  - All clients’ traffics are preferentially treated (for example, according to economic or geographical attributes)
- As network engineers, how to provide fair service to users?

Fairness

Example: Shortest Path Routing

Applying the same rule, might provide different performance levels to different users..



Quality of Service

<u>Definition</u>: Quality of Service (QoS) is the description of performance requirements for different traffic types. It also refers to the set of

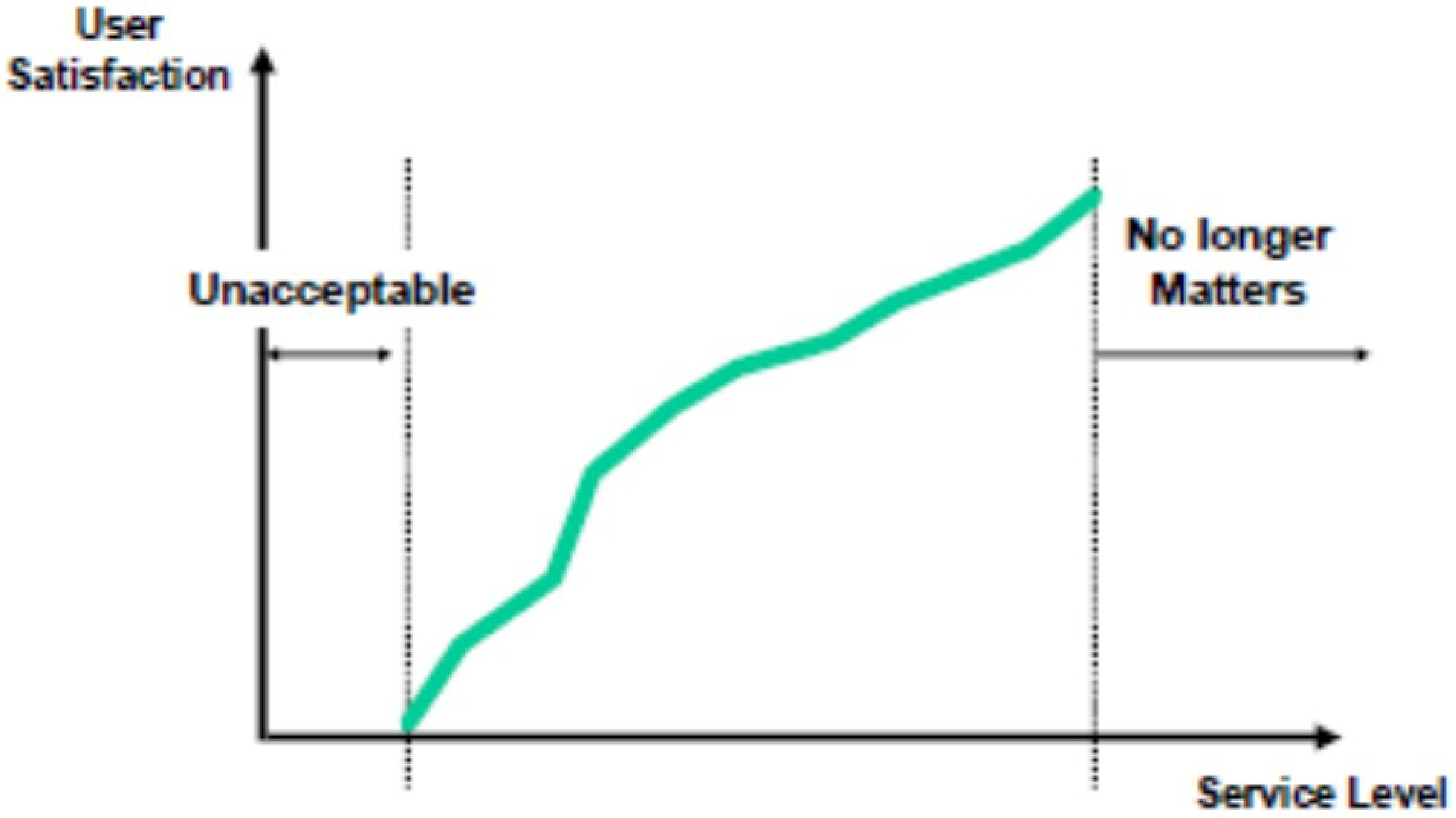
- Architectures,
- Mechanisms, and
- Protocols

that are used to meet these requirements.

“Treating users differently to provide fairness
by meeting their requirements”

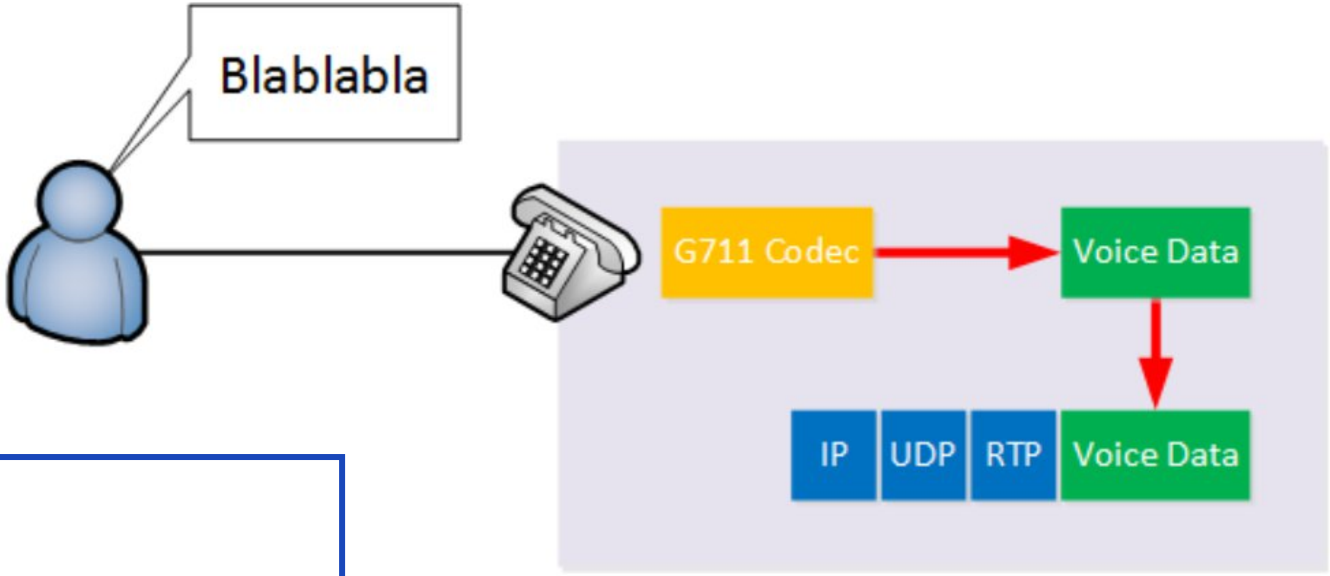
Quality of Service

Performance vs. Satisfaction



Quality of Service

Example of requirements – Voice traffic



Requirements:

- E2E delay < 150ms
- Jitter < 30ms
- Loss < 1%

Wrapping up

What did you learn today?

- Detailed review on L2 networking and service models
- Introduction to the 4 traffic characteristics important for QoS
- Definition of QoS