Introduction to Computer Networks

Performance Analysis of Computer Networks

https://pages.cs.wisc.edu/~mgliu/CS640/F22/

Ming Liu
mgliu@cs.wisc.edu

Today

Last lecture

- What hardware elements are used to build computer networks?
- What software components are needed to build computer networks?

Today

- How fast is the network?

Announcements

- Quiz1 next Tuesday

Q: How fast is the network?

Q: How fast is the network?

Q1: How fast is the underlying network fabric?
Q2: How fast is the application running in an exclusive network?
Q3: How fast is the application running in a shared network?

Q1: How fast is the underlying network fabric?

Q1: How fast is the underlying network fabric?

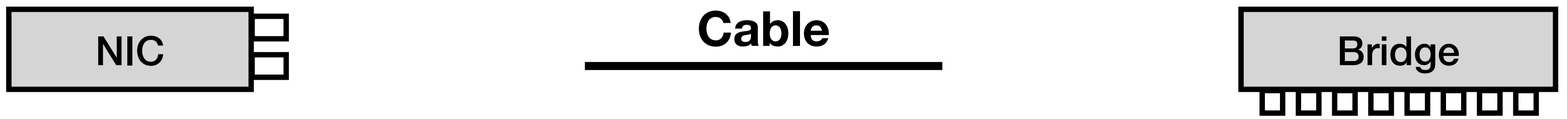
Metric 1: Bandwidth (throughput or data rate)

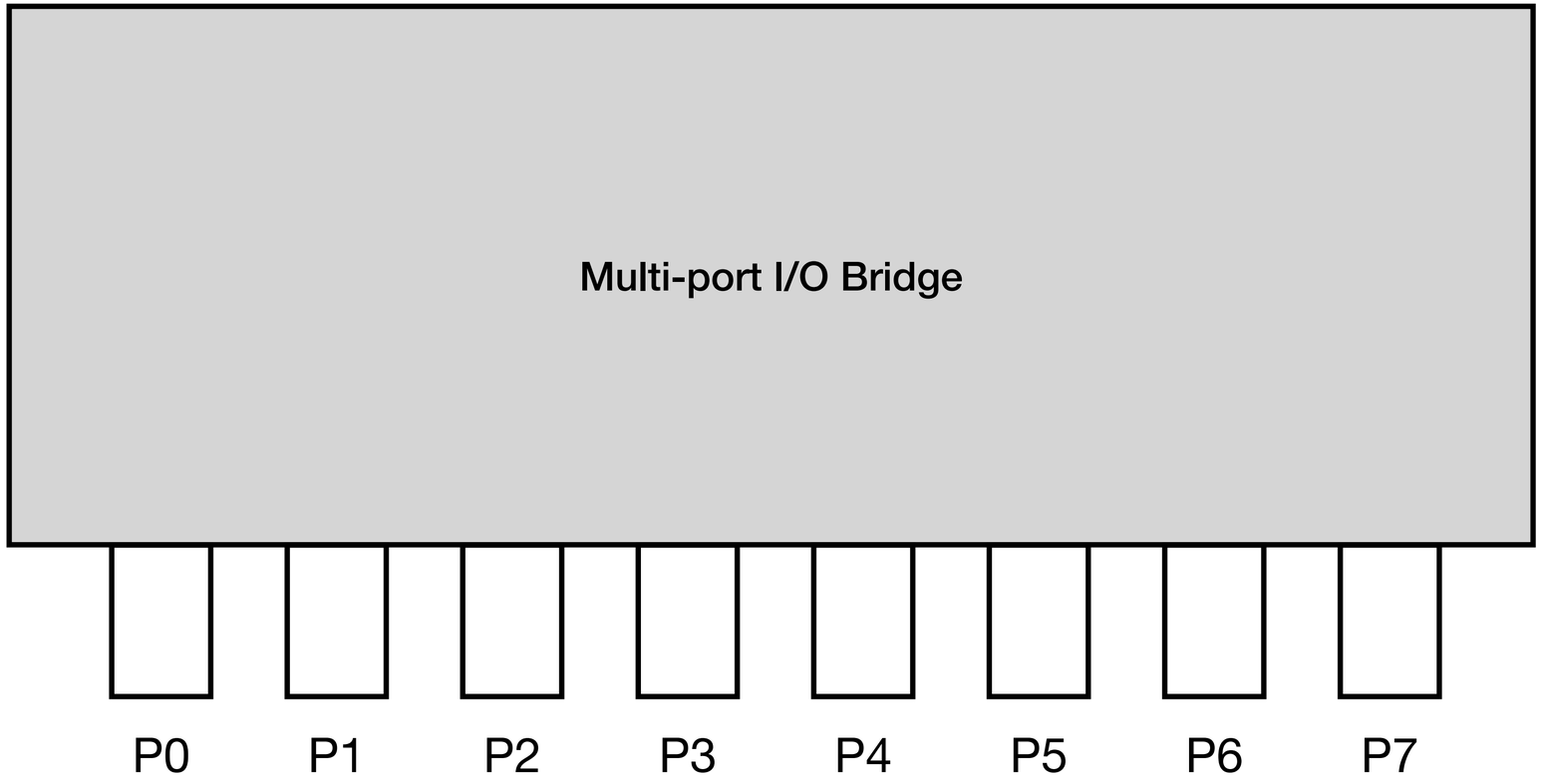
- The number of bits transmitted over one or several hardware elements in a certain period of time
- Unit: bits per second (bps), Kbps, Mbps, Gbps

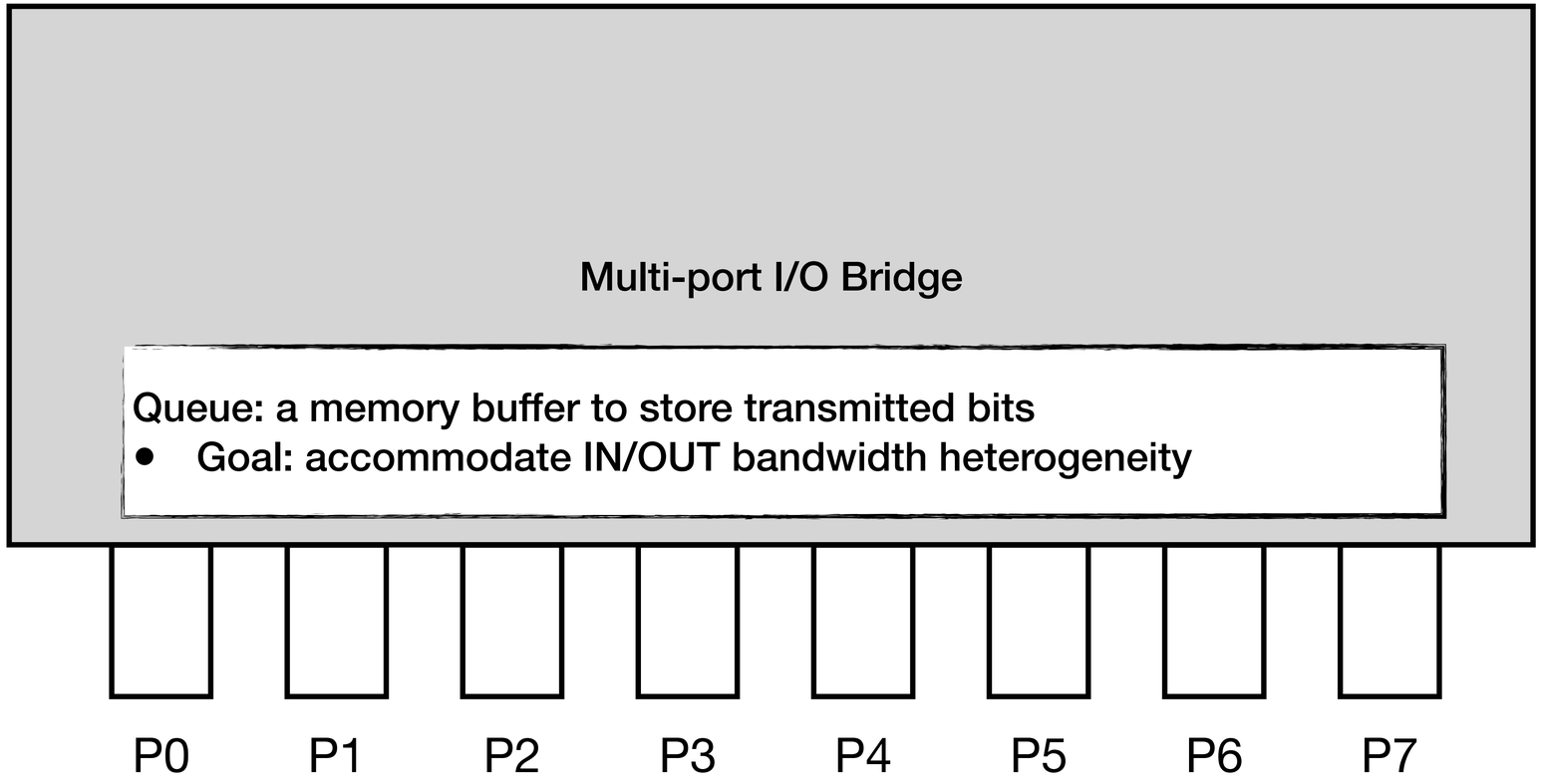
Q1: How fast is the underlying network fabric?

Metric 1: Bandwidth (throughput or data rate)

- The number of bits transmitted over one or several hardware elements in a certain period of time
- Unit: bits per second (bps), Kbps, Mbps, Gbps

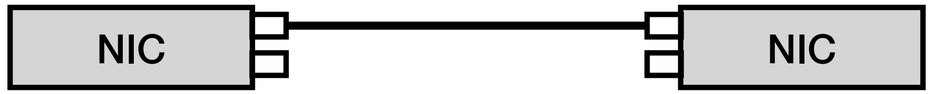



Multi-port I/O Bridge
P0
P1
P2
P3
P4
P5
P6
P7


Multi-port I/O Bridge
Queue: a memory buffer to store transmitted bits
• Goal: accommodate IN/OUT bandwidth heterogeneity
P0
P1
P2
P3
P4
P5
P6
P7

Bandwidth of the three building blocks

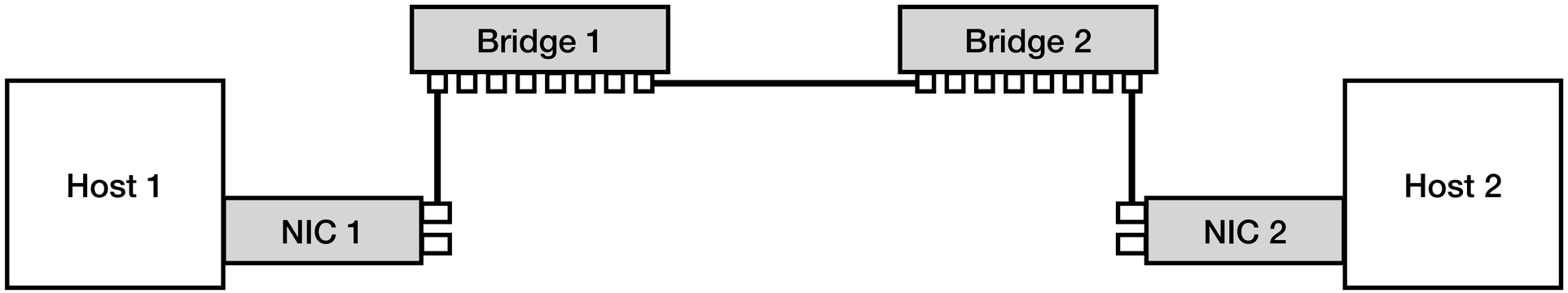
#1: NIC-Cable-NIC



#2: NIC-Cable-Bridge

#3: Bridge-Cable-Bridge

Bandwidth of this network system



Q1: How fast is the underlying network fabric?

Metric 2: Latency (delay)

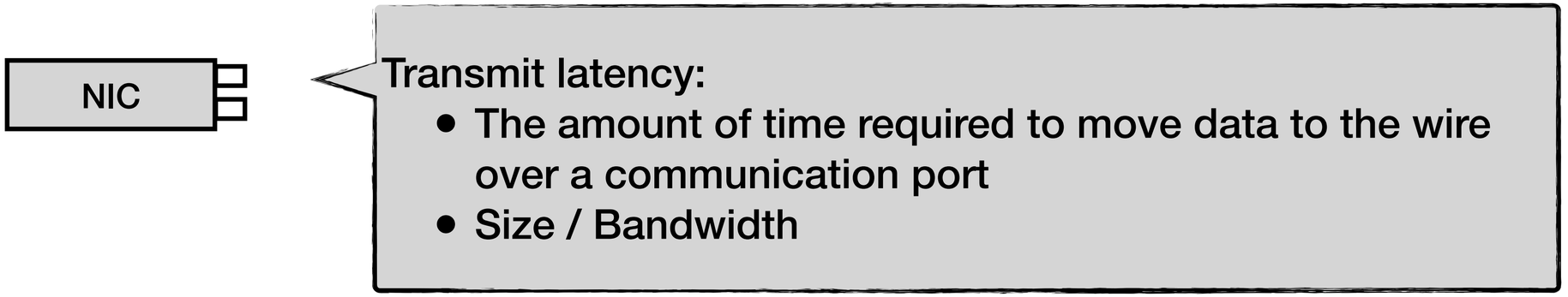
- The amount of time it takes to transfer **a fixed number of bits** over **one or several hardware elements**
- What is the source? What is the destination?
- Unit: second(s), millisecond(ms), microsecond(us), nanosecond(ns)

What is the latency of the NIC?

~~What is the latency of the NIC?~~

Transmit latency:

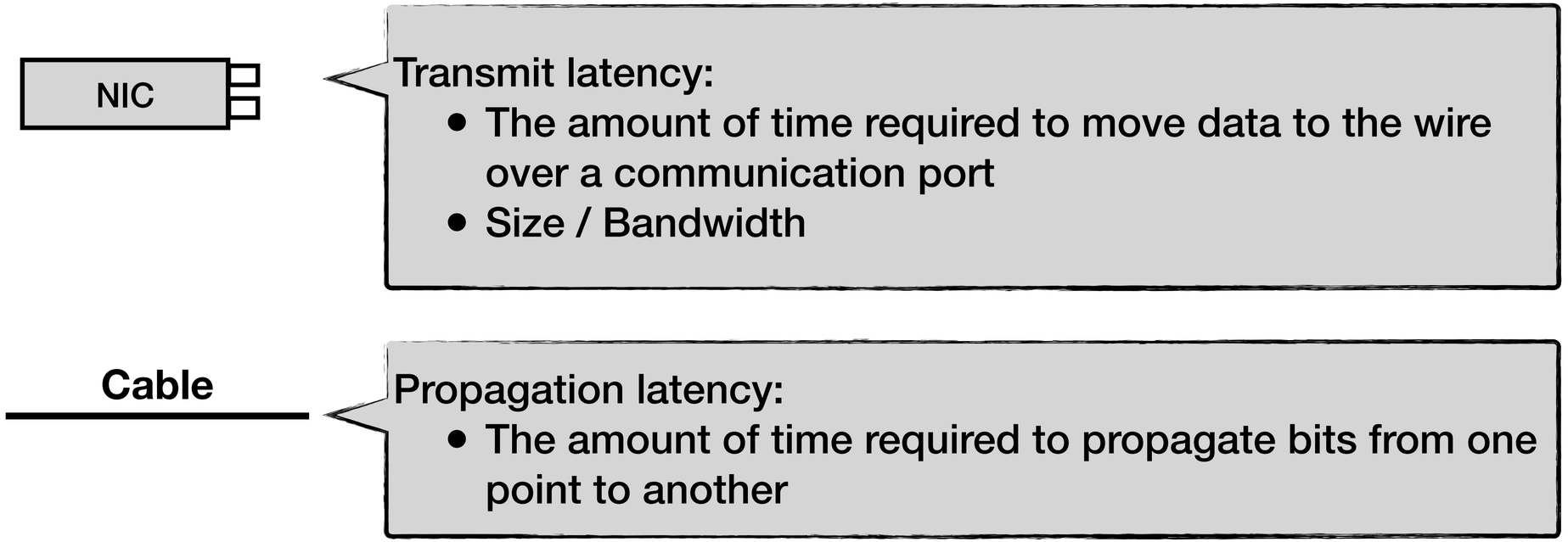
- The amount of time required to move data to the wire over a communication port

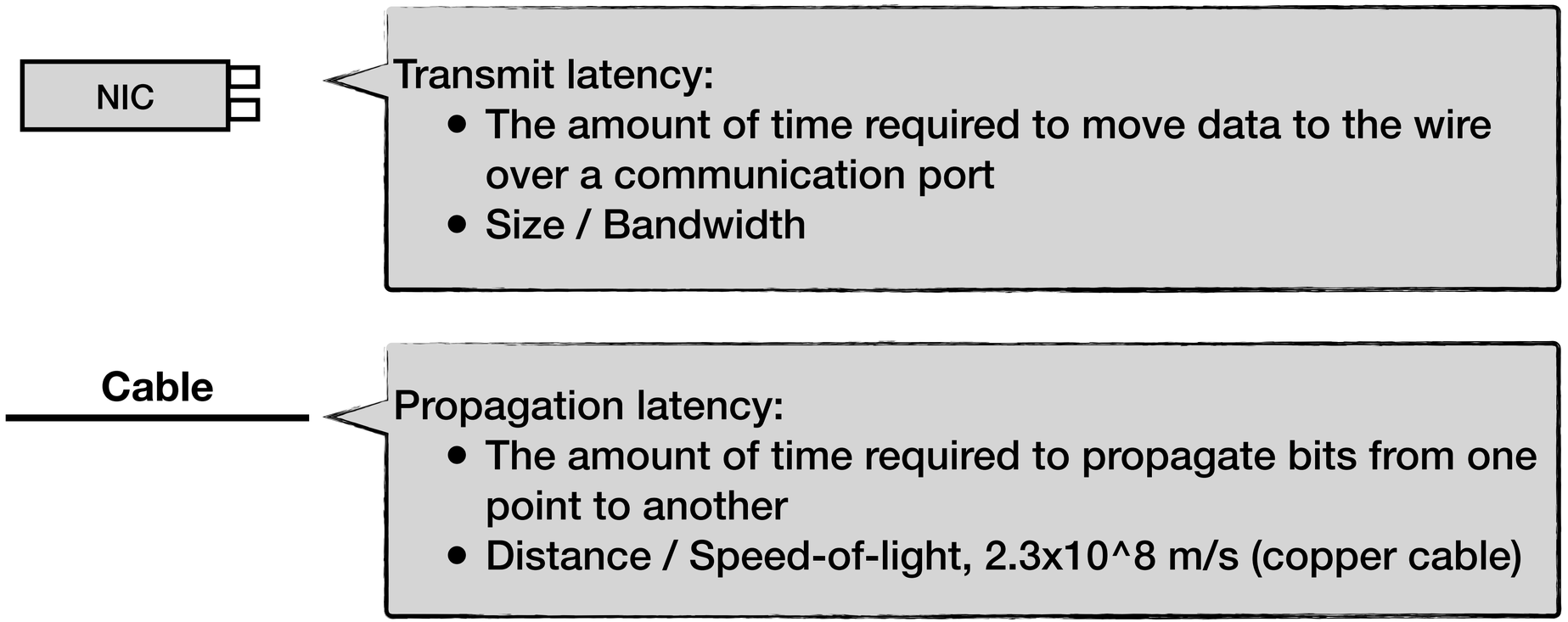

NIC
Transmit latency:
The amount of time required to move data to the wire over a communication port
Size / Bandwidth

Transmit latency:

- The amount of time required to move data to the wire over a communication port
- Size / Bandwidth

Cable


NIC
Transmit latency:
• The amount of time required to move data to the wire over a communication port
• Size / Bandwidth
Cable
Propagation latency:
• The amount of time required to propagate bits from one point to another


NIC
Transmit latency:
The amount of time required to move data to the wire over a communication port
Size / Bandwidth
Cable
Propagation latency:
The amount of time required to propagate bits from one point to another
Distance / Speed-of-light, 2.3x10^8 m/s (copper cable)

Transmit latency:

- The amount of time required to move data to the wire over a communication port
- Size / Bandwidth

Propagation latency:

- The amount of time required to propagate bits from one point to another
- Distance / Speed-of-light, 2.3x10^8 m/s (copper cable)

Transmit latency:

- The amount of time required to move data to the wire over a communication port
- Size / Bandwidth

Propagation latency:

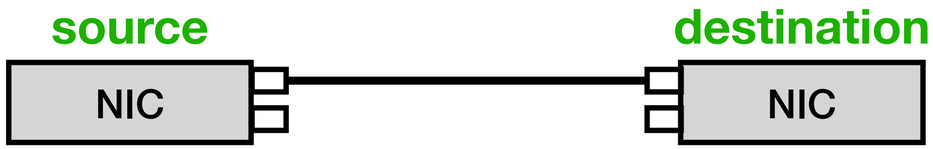
- The amount of time required to propagate bits from one point to another
- Distance / Speed-of-light, 2.3x10^8 m/s (copper cable)

Queuing latency:

- The amount of time required to stay in the memory of a multi-port I/O bridge

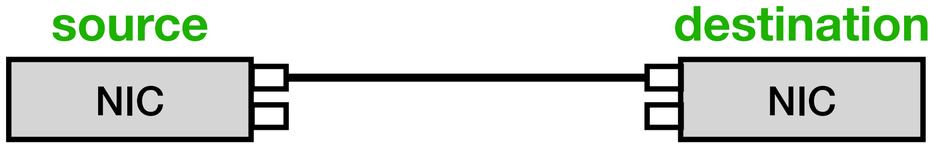
Latency of the three building blocks

#1: NIC-Cable-NIC



Latency of the three building blocks

#1: NIC-Cable-NIC

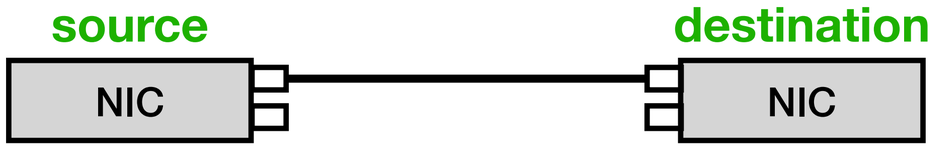


Processing latency

- The amount of time to process the transmitted data

Latency of the three building blocks

#1: NIC-Cable-NIC



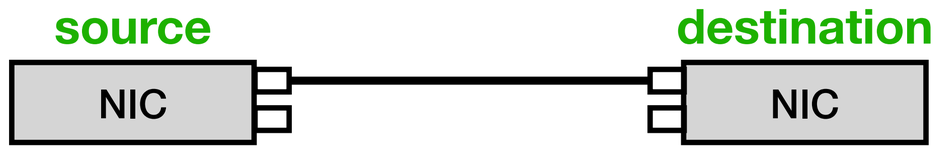
Processing latency

- The amount of time to process the transmitted data

Latency = Transmit latency + Propagation latency + Queuing Latency + Processing Latency
= Size / Bandwidth + Distance / Speed-of-Light + Queuing Latency + Processing Latency

Latency of the three building blocks

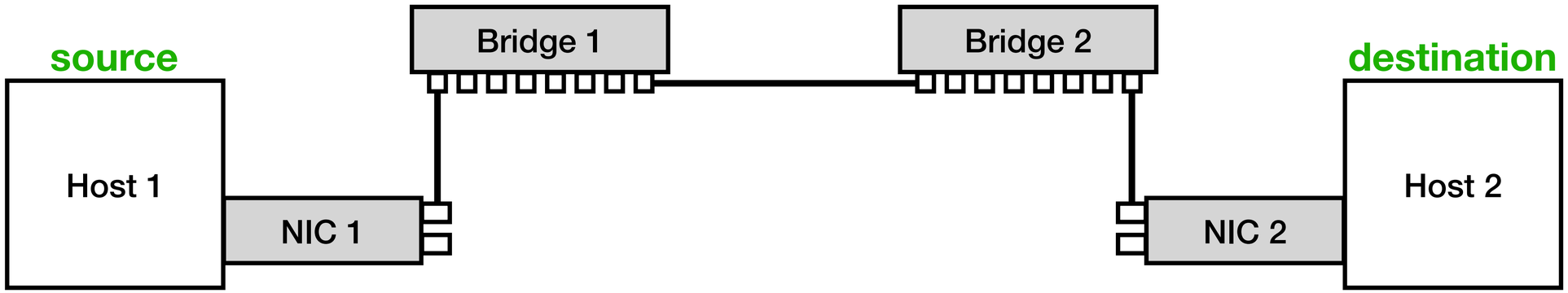
#1: NIC-Cable-NIC



#2: NIC-Cable-Bridge

#3: Bridge-Cable-Bridge

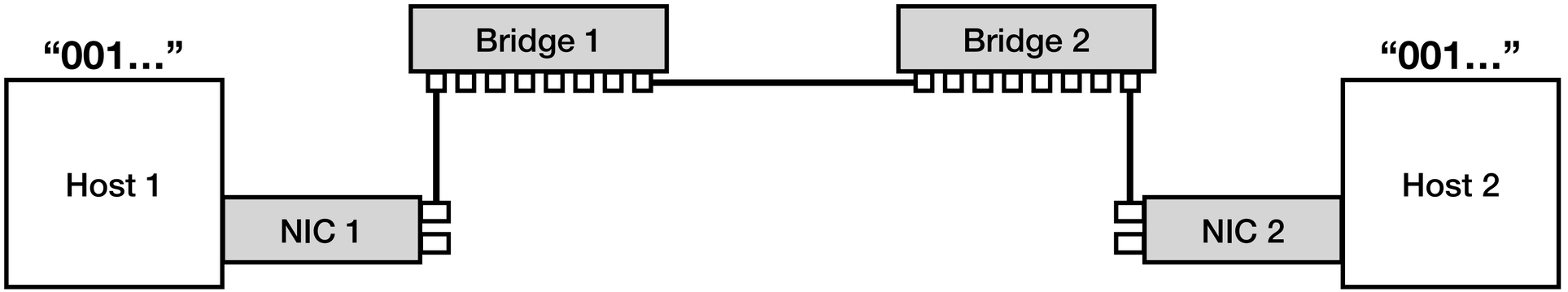
Latency of this network system



Q2: How fast is the application running in an exclusive network?

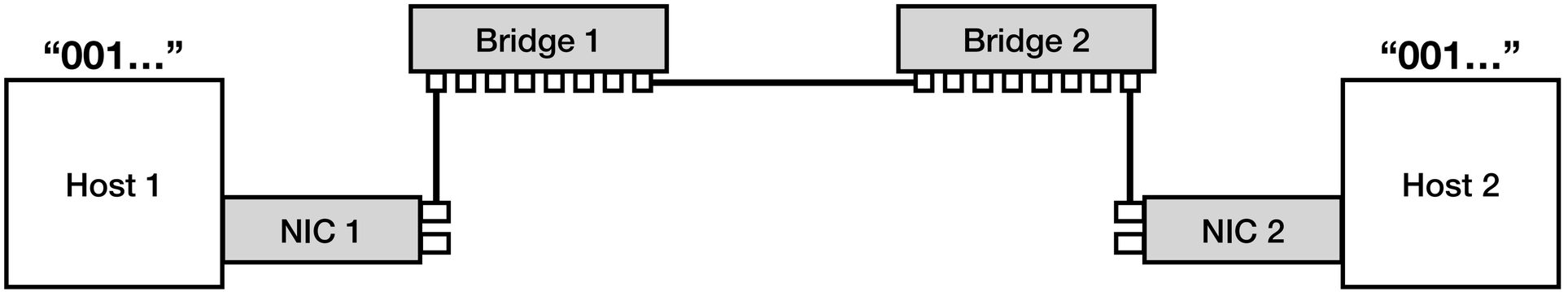
Q2: How fast is the application running in an exclusive network?

From the network perspective, an application can be viewed as a sequence of data transfers.



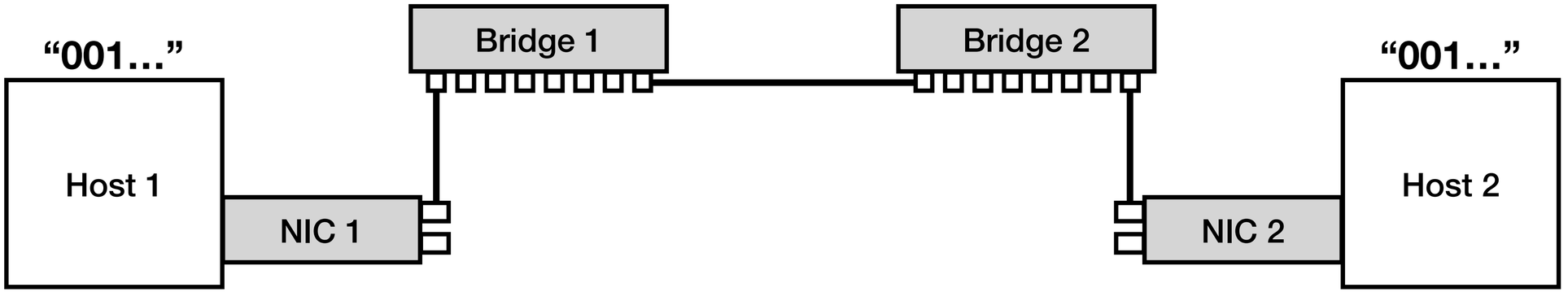
What is the latency and bandwidth of an application sending 1KB of data over the above network?
Suppose:

- The bandwidth of NIC1, NIC2, Bridge1, and Bridge2 is 100Mbps
- There is no processing latency and queueing latency
- The length of three cables is 230km (speed-of-light = 2.3x10^8 m/s)



What is the latency and bandwidth of an application sending 1GB of data over the above network?
Suppose:

- The bandwidth of NIC1, NIC2, Bridge1, and Bridge2 is 100Mbps
- There is no processing latency and queueing latency
- The length of three cables is 230km (speed-of-light = 2.3x10^8 m/s)



What is the latency and bandwidth of an application sending 1KB of data over the above network?
Suppose:

- The bandwidth of NIC1, NIC2, Bridge1, and Bridge2 is 100Mbps
- The queueing latency of Bridge 1 and Bridge 2 is 0.5ms
- The length of three cables is 230km (speed-of-light = 2.3x10^8 m/s)

Q2: How fast is the application running in an exclusive network?

Q2: How fast is the application running in an exclusive network?

A: One-time communication performance is deterministic. However, overall performance depends on the communication pattern.

More discussion

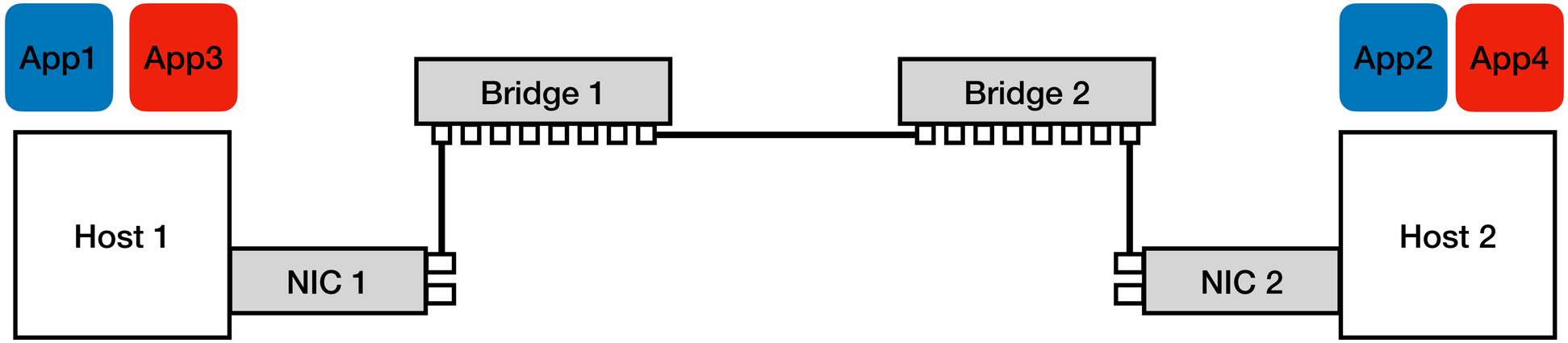
Round trip time (RTT)

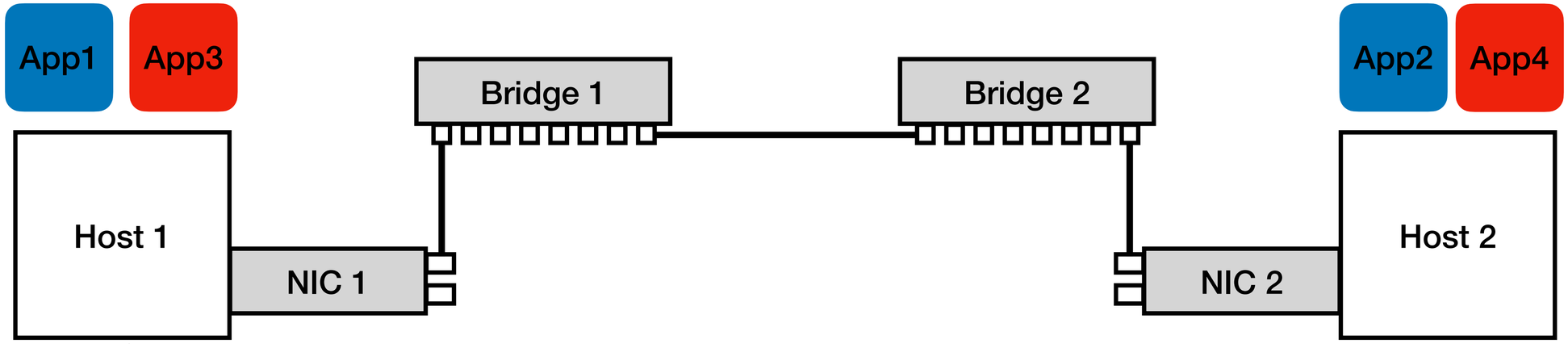
- Latency of data A (sender -> receiver) + Latency of data B (receiver -> sender)

Application performance is determined by

- Data size
- Raw hardware bandwidth
- Physical distance
- Computations of hardware elements

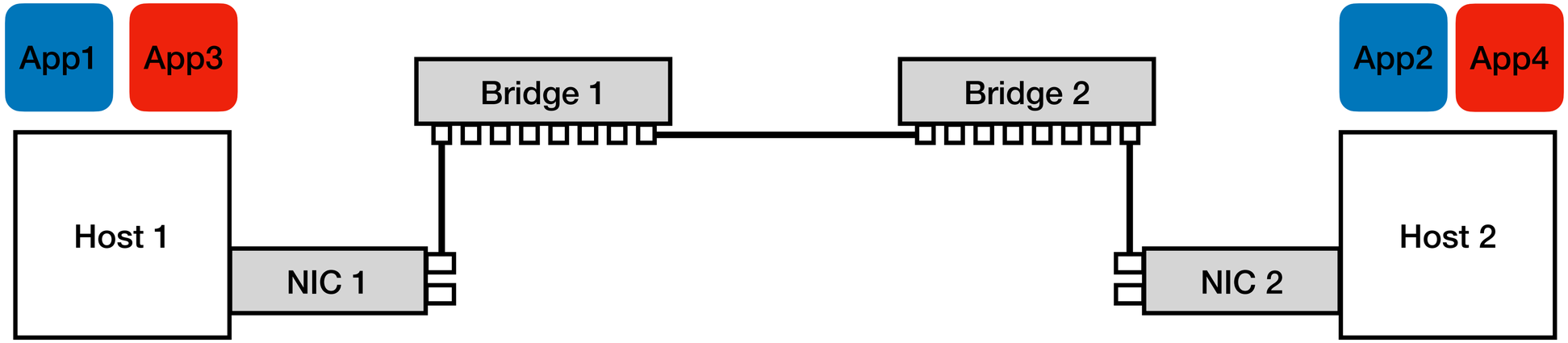
Q3: How fast is the application running in a shared network?


App1
App3
Bridge 1
Bridge 2
App2
App4
Host 1
NIC 1
NIC 2
Host 2



Time-division multiplexing (TDM)

- Allocate time slots to each application in a round-robin order
- Applications send data in its allocated time slots



Time-division multiplexing (TDM)

- Allocate time slots to each application in a round-robin order
- Applications send data in its allocated time slots

Network packet

- A formatted unit of data with a size boundary transmitted over the network
- Consist of control information (called header) and user data (called payload)

Q3: How fast is the application running in a shared network?

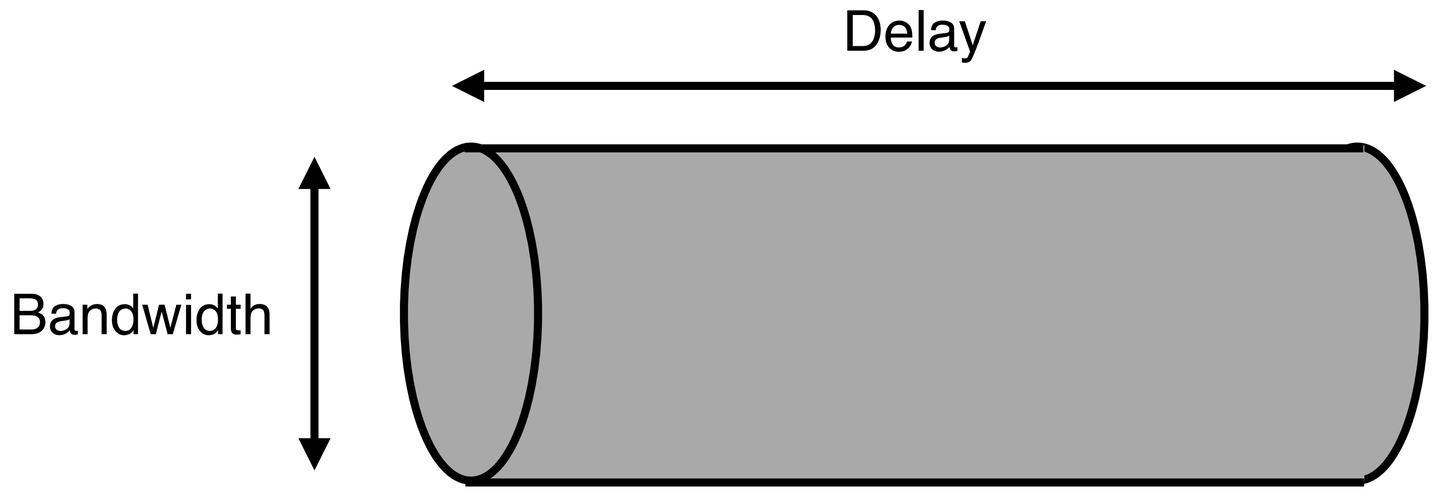
A: It depends on the internal network support.

- **Best case: fairness**
- **Worst case: anything**

Bandwidth-Delay Product (BDP)

The volume of a pipe = Delay (D) * Bandwidth (C)

- The number of bits have left the sender and are yet to reach the receiver



Understanding BDP

Network utilization

- The total number of bits can be packed into the link at any given time
- If the application keeps fewer outstanding, it is under-utilized

Transmission coordination with the receiver

- BDP = C * RTT
- The sender can send C * RTT of data to a receiver before hearing from the receiver about the first bit of data it sent

Terminology

1. Host
2. NIC
3. Multi-port I/O bridge
4. Protocol
5. RTT
6. Packet
7. Header
8. Payload
9. BDP

Principle

1. Layering

Technique

Summary

Today's takeaways

#1: Latency and bandwidth are two major metrics to evaluate a computer network

#2: Latency includes four components: transmit, propagation, queueing, processing

#3: Application performance under a network depends on its communication pattern, the underlying hardware fabric, and the sharing condition

Next lecture

- Bits on the cable; Encoding
- Lab1 overview