# 61A Lecture 35

Monday, November 26

A **distributed computing application** consists of multiple programs running on multiple computers that together coordinate to perform some task.

A distributed computing application consists of multiple programs running on multiple computers that together coordinate to perform some task.

Computation is performed in parallel by many computers.

A distributed computing application consists of multiple programs running on multiple computers that together coordinate to perform some task.

- Computation is performed in parallel by many computers.
- Information can be restricted to certain computers.

A distributed computing application consists of multiple programs running on multiple computers that together coordinate to perform some task.

- Computation is performed in parallel by many computers.
- Information can be restricted to certain computers.
- Redundancy and geographic diversity improve reliability.

A distributed computing application consists of multiple programs running on multiple computers that together coordinate to perform some task.

- Computation is performed in parallel by many computers.
- Information can be restricted to certain computers.
- Redundancy and geographic diversity improve reliability.

Characteristics of distributed computing:

A **distributed computing application** consists of multiple programs running on multiple computers that together coordinate to perform some task.

- Computation is performed in parallel by many computers.
- Information can be restricted to certain computers.
- Redundancy and geographic diversity improve reliability.

#### Characteristics of distributed computing:

Computers are independent — they do not share memory.

A distributed computing application consists of multiple programs running on multiple computers that together coordinate to perform some task.

- Computation is performed in parallel by many computers.
- Information can be restricted to certain computers.
- Redundancy and geographic diversity improve reliability.

#### Characteristics of distributed computing:

- Computers are <u>independent</u> they do not share memory.
- Coordination is enabled by messages passed across a network.

A distributed computing application consists of multiple programs running on multiple computers that together coordinate to perform some task.

- Computation is performed in parallel by many computers.
- Information can be restricted to certain computers.
- Redundancy and geographic diversity improve reliability.

#### Characteristics of distributed computing:

- Computers are independent they do not share memory.
- Coordination is enabled by messages passed across a network.
- Individual programs have differentiating roles.

A distributed computing application consists of multiple programs running on multiple computers that together coordinate to perform some task.

- Computation is performed in parallel by many computers.
- Information can be restricted to certain computers.
- Redundancy and geographic diversity improve reliability.

#### Characteristics of distributed computing:

- Computers are independent they do not share memory.
- Coordination is enabled by messages passed across a network.
- Individual programs have differentiating roles.

Distributed computing for large-scale data processing:

A **distributed computing application** consists of multiple programs running on multiple computers that together coordinate to perform some task.

- Computation is performed in parallel by many computers.
- Information can be restricted to certain computers.
- Redundancy and geographic diversity improve reliability.

#### Characteristics of distributed computing:

- Computers are independent they do not share memory.
- Coordination is enabled by messages passed across a network.
- Individual programs have differentiating roles.

#### Distributed computing for large-scale data processing:

Databases respond to queries over a network.

A distributed computing application consists of multiple programs running on multiple computers that together coordinate to perform some task.

- Computation is performed in parallel by many computers.
- Information can be restricted to certain computers.
- Redundancy and geographic diversity improve reliability.

#### Characteristics of distributed computing:

- Computers are independent they do not share memory.
- Coordination is enabled by messages passed across a network.
- Individual programs have differentiating roles.

#### Distributed computing for large-scale data processing:

- Databases respond to queries over a network.
- Data sets can be spread across multiple machines (Wednesday).

Computers communicate via messages: sequences of bytes transmitted over a network.

Computers communicate via messages: sequences of bytes transmitted over a network.

Computers communicate via messages: sequences of bytes transmitted over a network.

Messages can serve many purposes:

Send data to another computer

Computers communicate via messages: sequences of bytes transmitted over a network.

- Send data to another computer
- Request data from another computer

Computers communicate via messages: sequences of bytes transmitted over a network.

- Send data to another computer
- Request data from another computer
- Instruct a program to call a function on some arguments.

Computers communicate via messages: sequences of bytes transmitted over a network.

- Send data to another computer
- Request data from another computer
- Instruct a program to call a function on some arguments.
- Transfer a program to be executed by another computer.

Computers communicate via messages: sequences of bytes transmitted over a network.

Messages can serve many purposes:

- Send data to another computer
- Request data from another computer
- Instruct a program to call a function on some arguments.
- Transfer a program to be executed by another computer.

Messages conform to a *message protocol* adopted by both the sender to encode the message & the receiver to interpret it.

Computers communicate via messages: sequences of bytes transmitted over a network.

Messages can serve many purposes:

- Send data to another computer
- Request data from another computer
- Instruct a program to call a function on some arguments.
- Transfer a program to be executed by another computer.

Messages conform to a *message protocol* adopted by both the sender to encode the message & the receiver to interpret it.

For example, bits at fixed positions may have fixed meanings.

Computers communicate via messages: sequences of bytes transmitted over a network.

Messages can serve many purposes:

- Send data to another computer
- Request data from another computer
- Instruct a program to call a function on some arguments.
- Transfer a program to be executed by another computer.

Messages conform to a *message protocol* adopted by both the sender to encode the message & the receiver to interpret it.

- For example, bits at fixed positions may have fixed meanings.
- Components of a message may be separated by delimiters.

Computers communicate via messages: sequences of bytes transmitted over a network.

Messages can serve many purposes:

- Send data to another computer
- Request data from another computer
- Instruct a program to call a function on some arguments.
- Transfer a program to be executed by another computer.

Messages conform to a *message protocol* adopted by both the sender to encode the message & the receiver to interpret it.

- For example, bits at fixed positions may have fixed meanings.
- Components of a message may be separated by delimiters.
- Protocols are designed to be implemented by many different programming languages on a variety of platforms.

The Internet Protocol (IP) specifies how to transfer *packets* of data among different networks.

Networks are inherently unreliable at any point.

- Networks are inherently unreliable at any point.
- The structure of a network is dynamic.

- Networks are inherently unreliable at any point.
- The structure of a network is dynamic.
- No system exists to monitor or track communications.

The Internet Protocol (IP) specifies how to transfer *packets* of data among different networks.

- Networks are inherently unreliable at any point.
- The structure of a network is dynamic.
- No system exists to monitor or track communications.

#### **IPv4 Header Format**

Offsets	Octet	0									1								2									3									
Octet	Bit	0	1	2	3	4	5	6	7	8	9	10	0 11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31				
0	0	Version IHL DSCP											ΕC	N	N Total Length																						
4	32	Identification													Flags Fragment Offset																						
8	64			Tim	ne 7	Γο L	ive	)					Pro	tocc	/			Header Checksum																			
12	96														S	our	ce I	P A	ddr	ess																	
16	128														Des	stina	itioi	n IP	Ad	dres	ss																
20	160														C	ptio	ns	(if II	HL >	> 5)																	

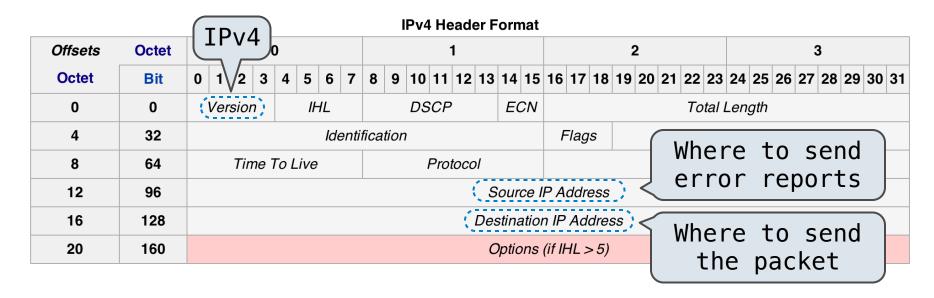
- Networks are inherently unreliable at any point.
- The structure of a network is dynamic.
- No system exists to monitor or track communications.

		IPv4 Header Format																													
Offsets	Octet	1774	1								2									3											
Octet	Bit	0 1 2 3	7	8	9	10 1	1	12	13	14	15	16	17	18	19	20	21	22	23	24	1 2	25 2	26 2	27	28 2	29 3	30	31			
0	0	Version		DSCP ECN							Total									l Length											
4	32				lde	entii	ficat	tion	)						F	-lag	ıs	Fragment Offset													
8	64	Time 7	Γo L	.ive					P	rota	oco	1			Header Checksum																
12	96											S	our	ce I	e IP Address																
16	128										ı	Des	stina	atio	n IP	Ad	dre	ss													
20	160											C	ptic	ns	ns (if IHL > 5)																

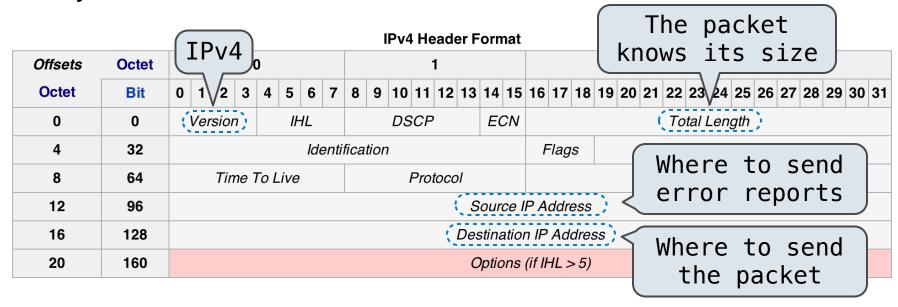
- Networks are inherently unreliable at any point.
- The structure of a network is dynamic.
- No system exists to monitor or track communications.

		TDVA						I	Pv4 H	ead	er F	ori	nat																	
Offsets	Octet	IPv4		1							2									3										
Octet	Bit	0 1 2 3	7	8	9	10 11	12	13	14	15	16	17	18	19	20	0 21	22	23	25 2	6 27	27 28 29 30				31					
0	0	Version	DSC	>		ECN							Total Length																	
4	32		Identification												Flags Fragment Offset															
8	64	Time	To L	ive					Pro	toc	ol			Header Checksum																
12	96										S	Soul	ce I	P A	ddr	ess														
16	128									- (	Des	stin	atio	n IF	Ad	dres	ss ,	<		Wh		ra		+ c	\ (		an.		7	
20	160		Options											Where to so the packe																
																					<u> </u>		1	Ja			. L			

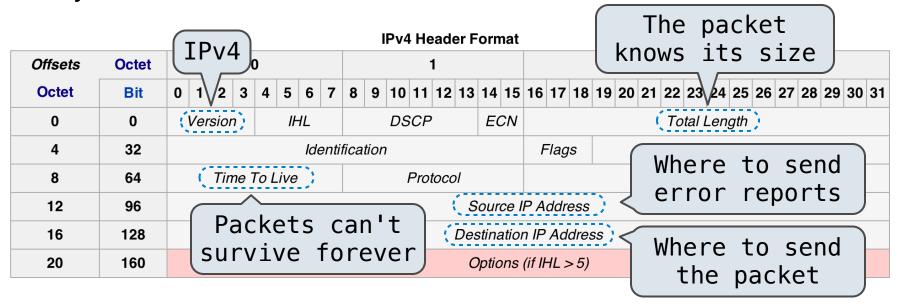
- Networks are inherently unreliable at any point.
- The structure of a network is dynamic.
- No system exists to monitor or track communications.



- Networks are inherently unreliable at any point.
- The structure of a network is dynamic.
- No system exists to monitor or track communications.

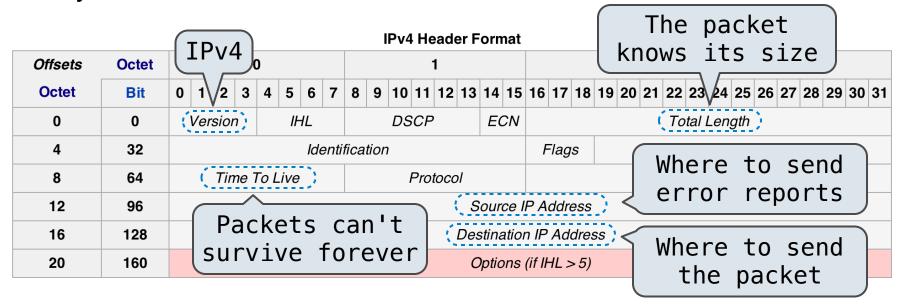


- Networks are inherently unreliable at any point.
- The structure of a network is dynamic.
- No system exists to monitor or track communications.



The Internet Protocol (IP) specifies how to transfer *packets* of data among different networks.

- Networks are inherently unreliable at any point.
- The structure of a network is dynamic.
- No system exists to monitor or track communications.



Packets are forwarded toward their destination using simple rules on a best-effort basis.

The design of the **Internet Protocol** (IP) imposes constraints:

The design of the Internet Protocol (IP) imposes constraints:

Packets are limited to 65,535 bytes each.

The design of the Internet Protocol (IP) imposes constraints:

- Packets are limited to 65,535 bytes each.
- Packets may arrive in a different order than they were sent.

The design of the Internet Protocol (IP) imposes constraints:

- Packets are limited to 65,535 bytes each.
- Packets may arrive in a different order than they were sent.
- Packets may be duplicated or lost.

The design of the Internet Protocol (IP) imposes constraints:

- Packets are limited to 65,535 bytes each.
- Packets may arrive in a different order than they were sent.
- Packets may be duplicated or lost.

The design of the Internet Protocol (IP) imposes constraints:

- Packets are limited to 65,535 bytes each.
- Packets may arrive in a different order than they were sent.
- Packets may be duplicated or lost.

The Transmission Control Protocol (TCP) improves reliability:

Ordered, reliable transmission of arbitrary byte streams.

The design of the Internet Protocol (IP) imposes constraints:

- Packets are limited to 65,535 bytes each.
- Packets may arrive in a different order than they were sent.
- Packets may be duplicated or lost.

- Ordered, reliable transmission of arbitrary byte streams.
- Implemented using the IP.

The design of the Internet Protocol (IP) imposes constraints:

- Packets are limited to 65,535 bytes each.
- Packets may arrive in a different order than they were sent.
- Packets may be duplicated or lost.

- Ordered, reliable transmission of arbitrary byte streams.
- Implemented using the IP.
- Correctly orders packets by including sequence numbers.

The design of the Internet Protocol (IP) imposes constraints:

- Packets are limited to 65,535 bytes each.
- Packets may arrive in a different order than they were sent.
- Packets may be duplicated or lost.

- Ordered, reliable transmission of arbitrary byte streams.
- Implemented using the IP.
- Correctly orders packets by including sequence numbers.
- Removes duplicates; requests retransmission of lost packets.

The design of the Internet Protocol (IP) imposes constraints:

- Packets are limited to 65,535 bytes each.
- Packets may arrive in a different order than they were sent.
- Packets may be duplicated or lost.

The Transmission Control Protocol (TCP) improves reliability:

- Ordered, reliable transmission of arbitrary byte streams.
- Implemented using the IP.
- Correctly orders packets by including sequence numbers.
- Removes duplicates; requests retransmission of lost packets.

TCP connection initiates with a "handshake" procedure.

The design of the Internet Protocol (IP) imposes constraints:

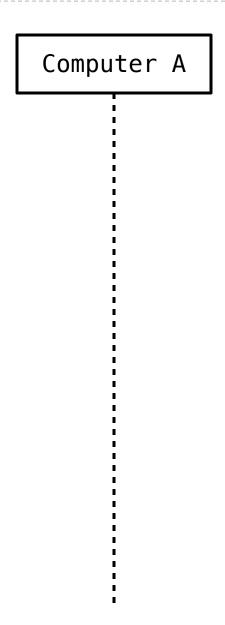
- Packets are limited to 65,535 bytes each.
- Packets may arrive in a different order than they were sent.
- Packets may be duplicated or lost.

The Transmission Control Protocol (TCP) improves reliability:

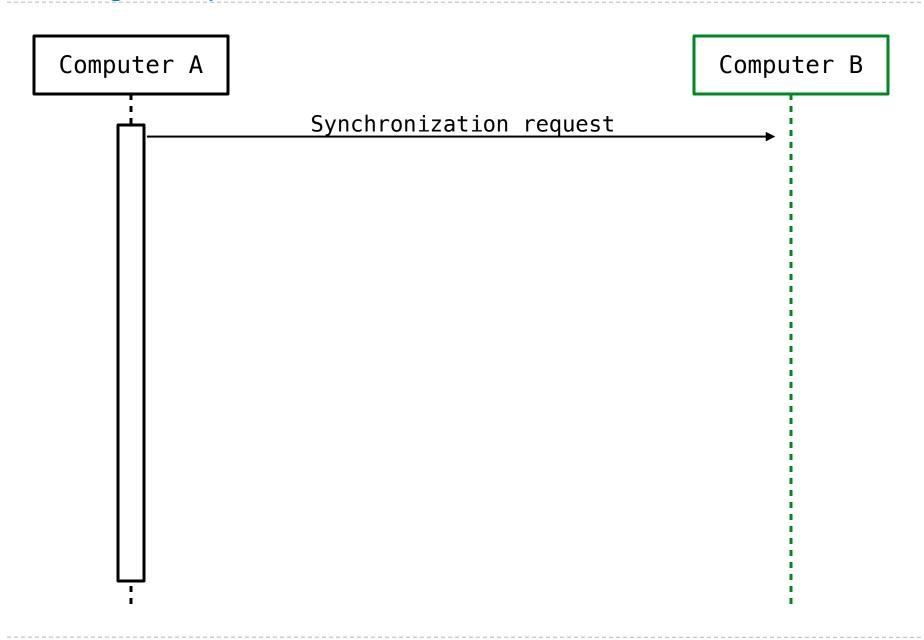
- Ordered, reliable transmission of arbitrary byte streams.
- Implemented using the IP.
- Correctly orders packets by including sequence numbers.
- Removes duplicates; requests retransmission of lost packets.

TCP connection initiates with a "handshake" procedure.

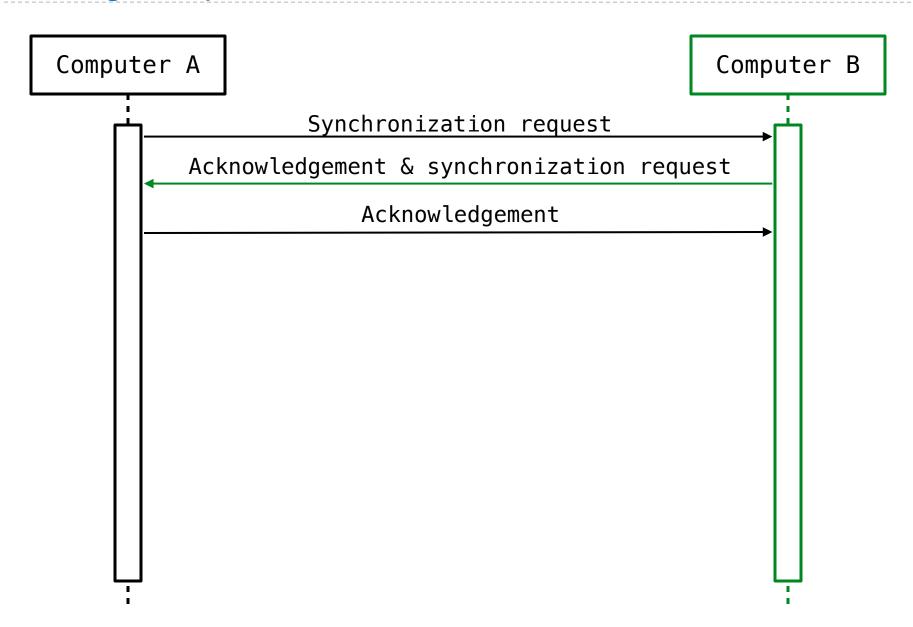
 What's the minimum number of messages needed to prove to both computers that two-way communication is possible?

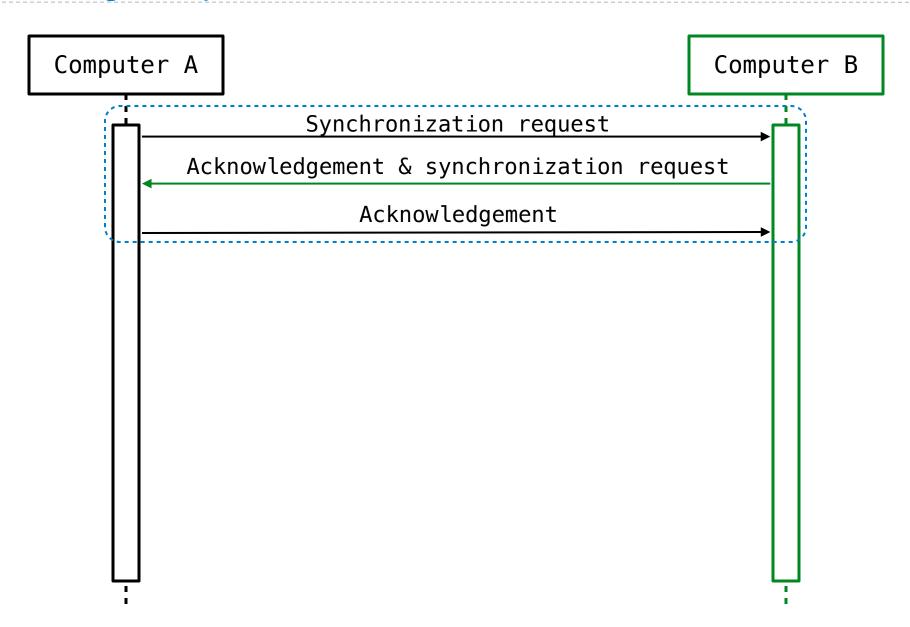


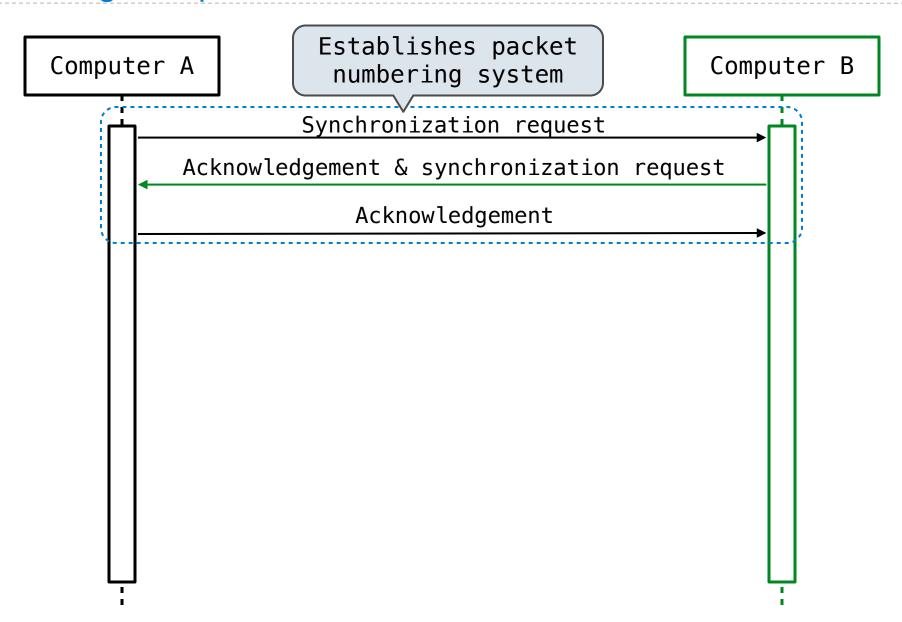


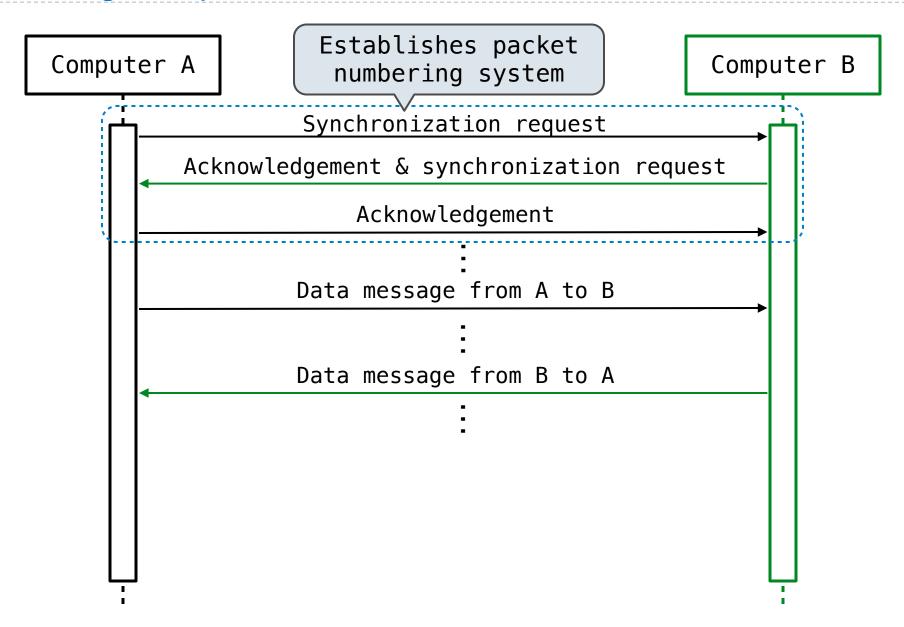


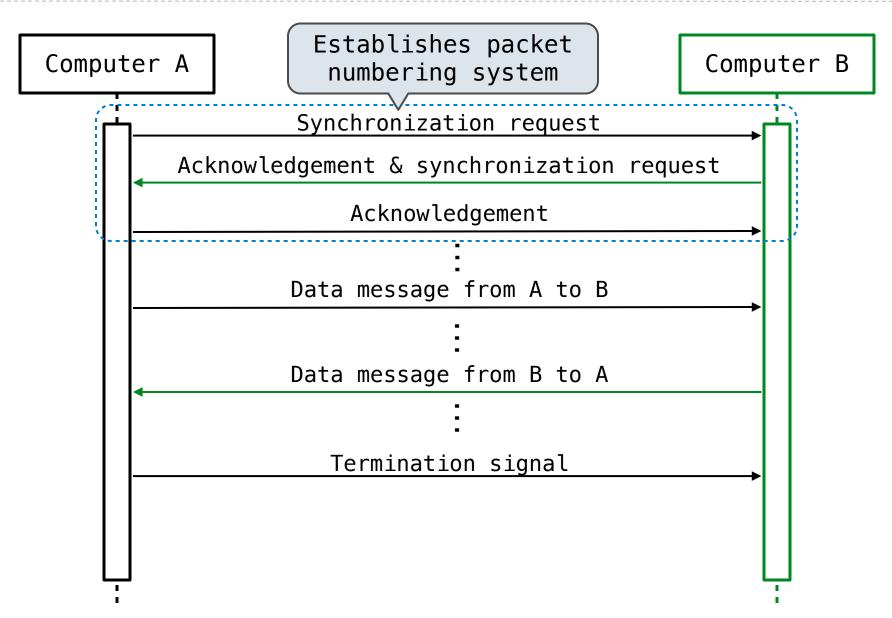


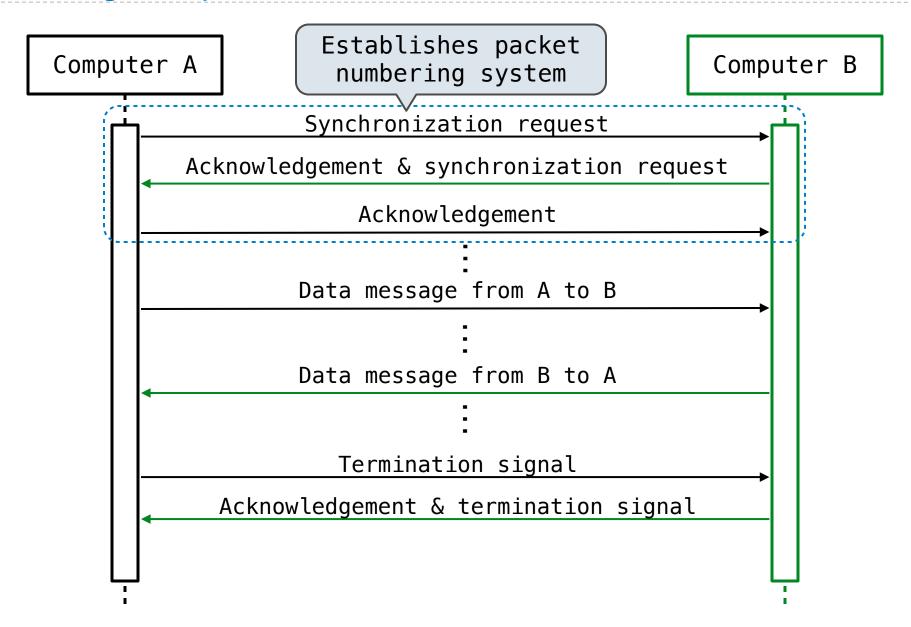


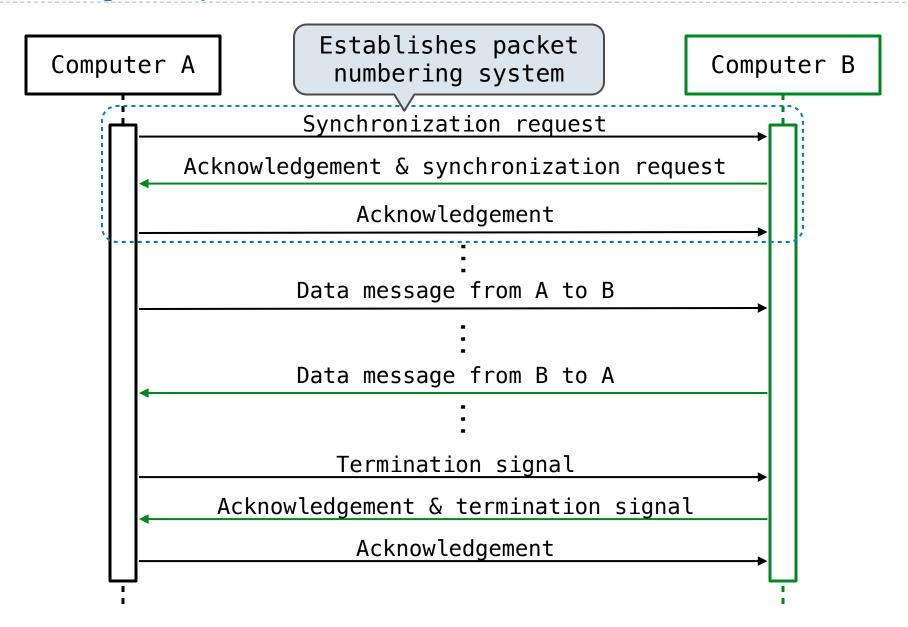


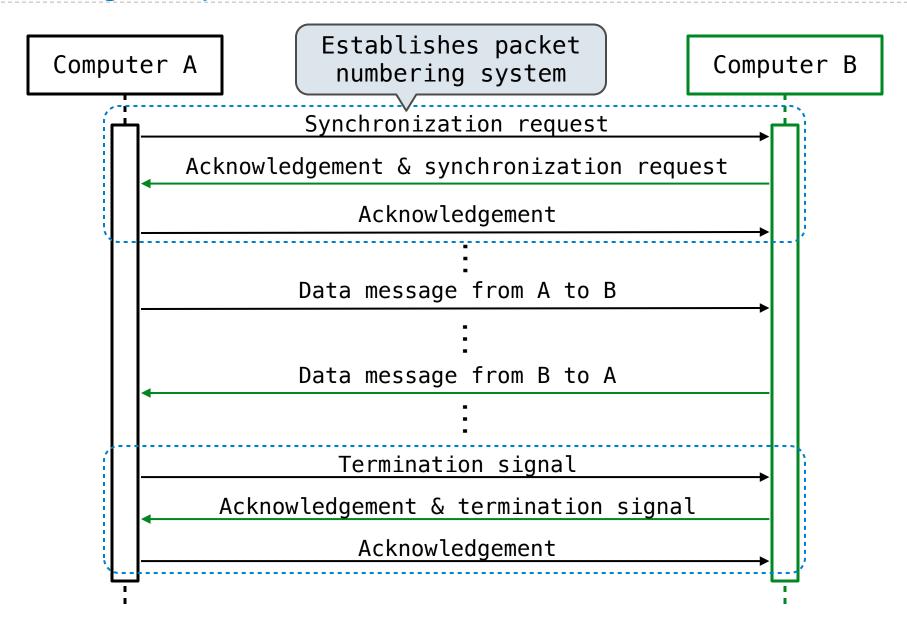






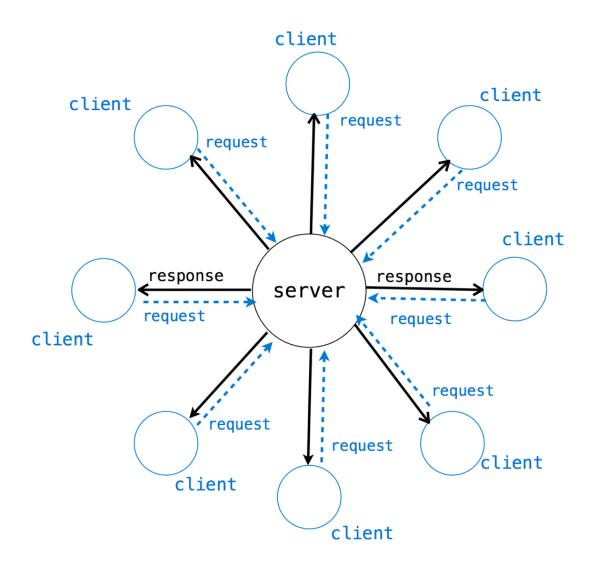






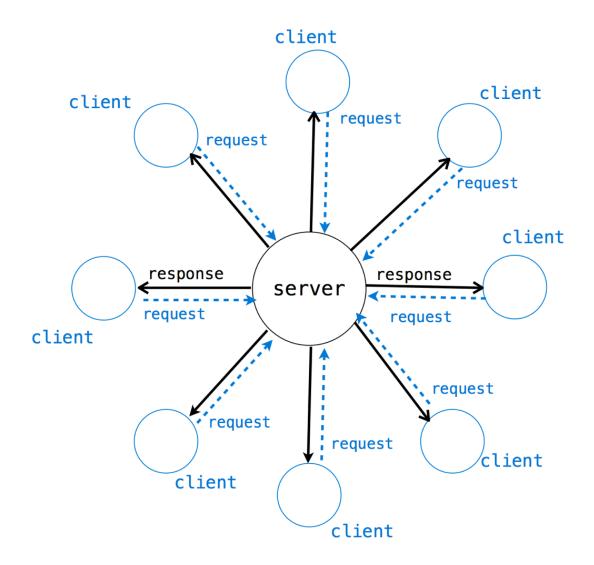
One server provides information to multiple clients through *request* and *response* messages.

One server provides information to multiple clients through *request* and *response* messages.



One server provides information to multiple clients through *request* and *response* messages.

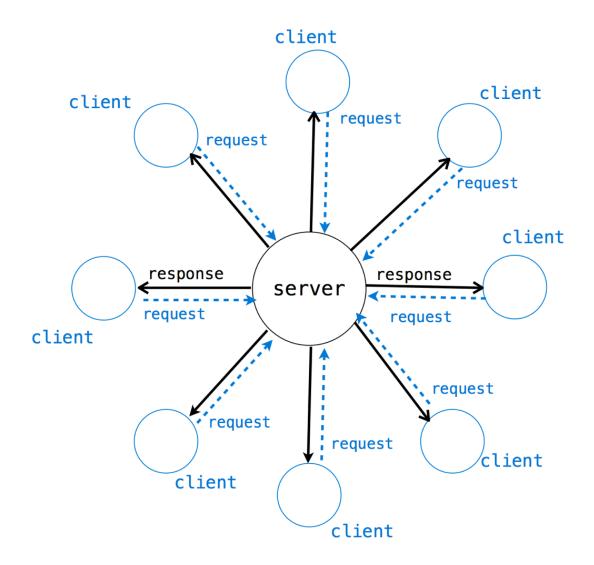
**Server role:** Respond to service requests with requested information.



One server provides information to multiple clients through *request* and *response* messages.

**Server role:** Respond to service requests with requested information.

Client role: Request information and make use of the response.

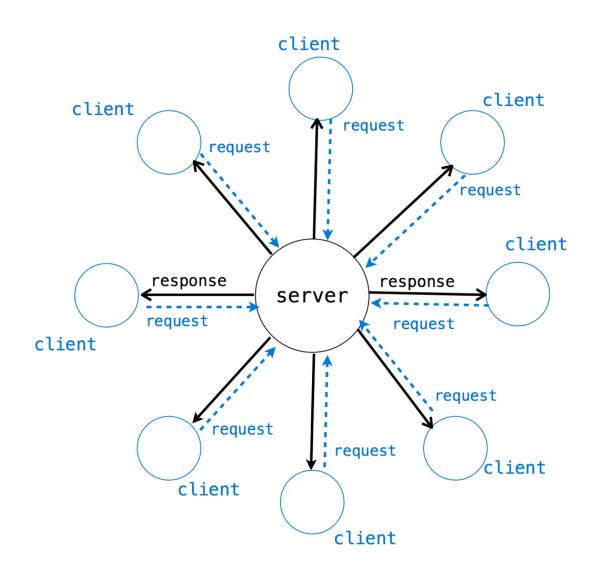


One server provides information to multiple clients through *request* and *response* messages.

**Server role:** Respond to service requests with requested information.

Client role: Request information and make use of the response.

Abstraction: The client knows what service a server provides but not how it is provided.



The client is a web browser (e.g., Firefox):

The client is a web browser (e.g., Firefox):

• Request content from a location on behalf of the user.

The client is a web browser (e.g., Firefox):

- Request content from a location on behalf of the user.
- Display the content to the user.

The client is a web browser (e.g., Firefox):

- Request content from a location on behalf of the user.
- Display the content to the user.

The **server** is a web server (e.g., <a href="https://www.nytimes.com">www.nytimes.com</a>)

The client is a web browser (e.g., Firefox):

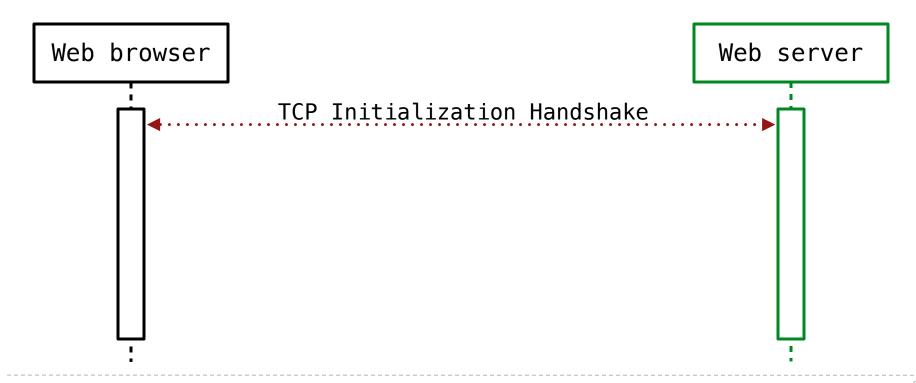
- Request content from a location on behalf of the user.
- Display the content to the user.

The **server** is a web server (e.g., <a href="https://www.nytimes.com">www.nytimes.com</a>)

The client is a web browser (e.g., Firefox):

- Request content from a location on behalf of the user.
- Display the content to the user.

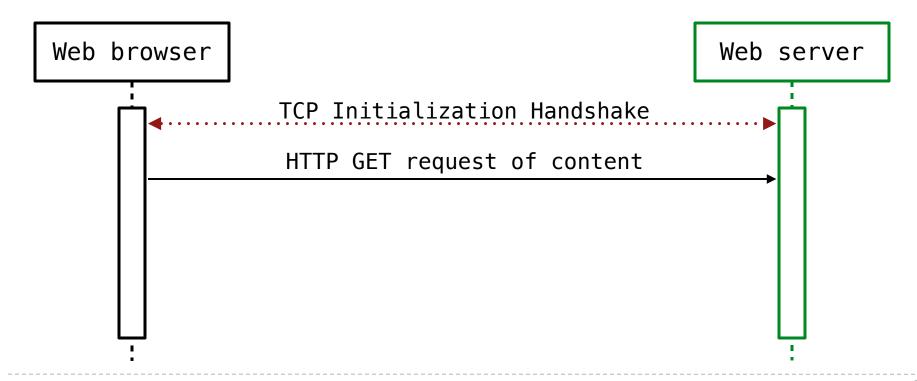
The **server** is a web server (e.g., <a href="https://www.nytimes.com">www.nytimes.com</a>)



The client is a web browser (e.g., Firefox):

- Request content from a location on behalf of the user.
- Display the content to the user.

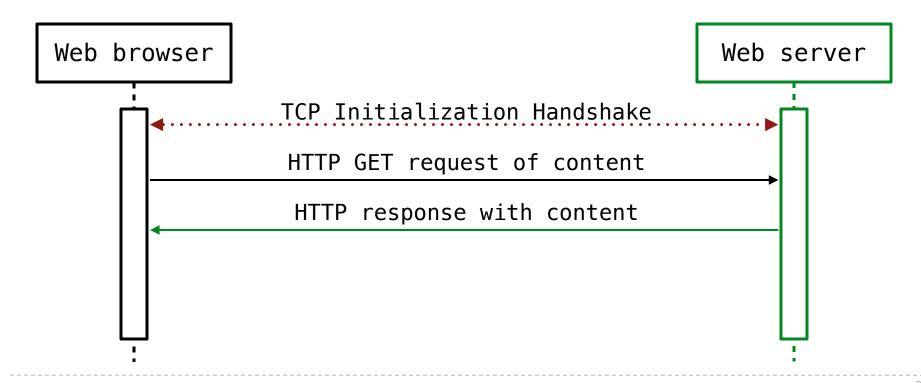
The **server** is a web server (e.g., <a href="https://www.nytimes.com">www.nytimes.com</a>)



The client is a web browser (e.g., Firefox):

- Request content from a location on behalf of the user.
- Display the content to the user.

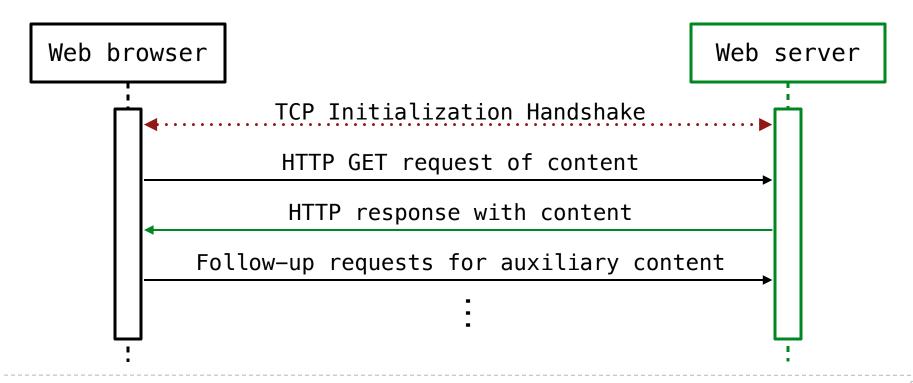
The **server** is a web server (e.g., <a href="https://www.nytimes.com">www.nytimes.com</a>)



The client is a web browser (e.g., Firefox):

- Request content from a location on behalf of the user.
- Display the content to the user.

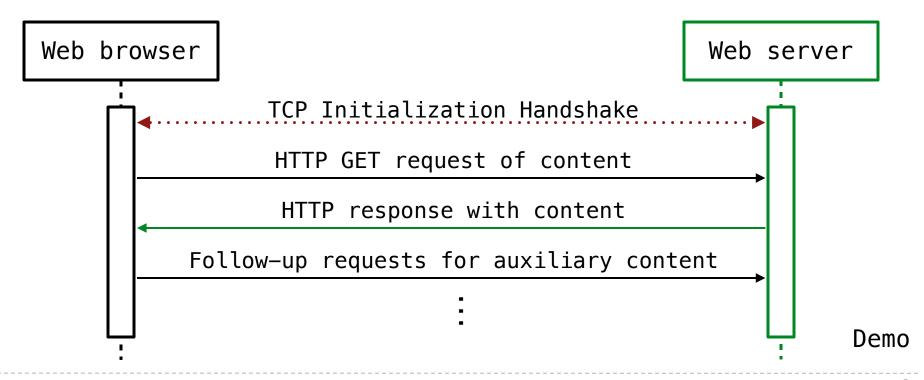
The **server** is a web server (e.g., <a href="https://www.nytimes.com">www.nytimes.com</a>)



The client is a web browser (e.g., Firefox):

- Request content from a location on behalf of the user.
- Display the content to the user.

The **server** is a web server (e.g., <a href="https://www.nytimes.com">www.nytimes.com</a>)



The Hypertext Transfer Protocol (HTTP) is a protocol designed to implement a Client/Server architecture.

The Hypertext Transfer Protocol (HTTP) is a protocol designed to implement a Client/Server architecture.



The Hypertext Transfer Protocol (HTTP) is a protocol designed to implement a Client/Server architecture.

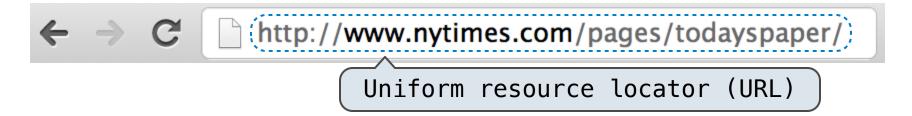


The Hypertext Transfer Protocol (HTTP) is a protocol designed to implement a Client/Server architecture.



Browser issues a GET request to <a href="www.nytimes.com">www.nytimes.com</a> for the content (resource) at location "pages/todayspaper".

The Hypertext Transfer Protocol (HTTP) is a protocol designed to implement a Client/Server architecture.



Browser issues a GET request to <a href="www.nytimes.com">www.nytimes.com</a> for the content (resource) at location "pages/todayspaper".

Server response contains more than just the resource itself:

The Hypertext Transfer Protocol (HTTP) is a protocol designed to implement a Client/Server architecture.



Browser issues a GET request to <a href="www.nytimes.com">www.nytimes.com</a> for the content (resource) at location "pages/todayspaper".

Server response contains more than just the resource itself:

Status code, e.g. 200 OK, 404 Not Found, 403 Forbidden, etc.

The Hypertext Transfer Protocol (HTTP) is a protocol designed to implement a Client/Server architecture.



Browser issues a GET request to <a href="www.nytimes.com">www.nytimes.com</a> for the content (resource) at location "pages/todayspaper".

Server response contains more than just the resource itself:

- Status code, e.g. 200 OK, 404 Not Found, 403 Forbidden, etc.
- Date of response; type of server responding

The Hypertext Transfer Protocol (HTTP) is a protocol designed to implement a Client/Server architecture.



Browser issues a GET request to <a href="www.nytimes.com">www.nytimes.com</a> for the content (resource) at location "pages/todayspaper".

Server response contains more than just the resource itself:

- Status code, e.g. 200 OK, 404 Not Found, 403 Forbidden, etc.
- Date of response; type of server responding
- Last-modified time of the resource

The Hypertext Transfer Protocol (HTTP) is a protocol designed to implement a Client/Server architecture.

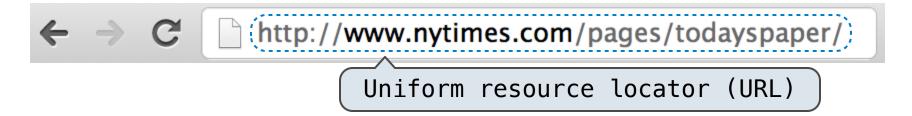


Browser issues a GET request to <a href="www.nytimes.com">www.nytimes.com</a> for the content (resource) at location "pages/todayspaper".

Server response contains more than just the resource itself:

- Status code, e.g. 200 OK, 404 Not Found, 403 Forbidden, etc.
- Date of response; type of server responding
- Last-modified time of the resource
- Type of content and length of content

The Hypertext Transfer Protocol (HTTP) is a protocol designed to implement a Client/Server architecture.



Browser issues a GET request to <a href="www.nytimes.com">www.nytimes.com</a> for the content (resource) at location "pages/todayspaper".

Server response contains more than just the resource itself:

- Status code, e.g. 200 OK, 404 Not Found, 403 Forbidden, etc.
- Date of response; type of server responding
- Last-modified time of the resource
- Type of content and length of content

Demo

**Benefits:** 

#### **Benefits:**

Creates a separation of concerns among components.

#### **Benefits:**

- Creates a separation of concerns among components.
- Enforces an abstraction barrier between client and server.

#### **Benefits:**

- Creates a separation of concerns among components.
- Enforces an abstraction barrier between client and server.
- A centralized server can reuse computation across clients.

#### **Benefits:**

- Creates a separation of concerns among components.
- Enforces an abstraction barrier between client and server.
- A centralized server can reuse computation across clients.

#### Liabilities:

#### **Benefits:**

- Creates a separation of concerns among components.
- Enforces an abstraction barrier between client and server.
- A centralized server can reuse computation across clients.

#### Liabilities:

A single point of failure: the server.

#### **Benefits:**

- Creates a separation of concerns among components.
- Enforces an abstraction barrier between client and server.
- A centralized server can reuse computation across clients.

#### Liabilities:

- A single point of failure: the server.
- Computing resources become scarce with increasing demand.

#### **Benefits:**

- Creates a separation of concerns among components.
- Enforces an abstraction barrier between client and server.
- A centralized server can reuse computation across clients.

#### Liabilities:

- A single point of failure: the server.
- Computing resources become scarce with increasing demand.

#### Common use cases:

#### **Benefits:**

- Creates a separation of concerns among components.
- Enforces an abstraction barrier between client and server.
- A centralized server can reuse computation across clients.

#### Liabilities:

- A single point of failure: the server.
- Computing resources become scarce with increasing demand.

#### Common use cases:

Databases — The database serves responses to query requests.

#### **Benefits:**

- Creates a separation of concerns among components.
- Enforces an abstraction barrier between client and server.
- A centralized server can reuse computation across clients.

#### Liabilities:

- A single point of failure: the server.
- Computing resources become scarce with increasing demand.

#### Common use cases:

- Databases The database serves responses to query requests.
- Open Graphics Library (OpenGL) A graphics processing unit (GPU) serves images to a central processing unit (CPU).

#### **Benefits:**

- Creates a separation of concerns among components.
- Enforces an abstraction barrier between client and server.
- A centralized server can reuse computation across clients.

#### Liabilities:

- A single point of failure: the server.
- Computing resources become scarce with increasing demand.

#### Common use cases:

- Databases The database serves responses to query requests.
- Open Graphics Library (OpenGL) A graphics processing unit (GPU) serves images to a central processing unit (CPU).
- File and resource transfer: HTTP, FTP, email, etc.

All participants in a distributed application contribute computational resources: processing, storage, and network.

All participants in a distributed application contribute computational resources: processing, storage, and network.

Messages are relayed through a network of participants.

All participants in a distributed application contribute computational resources: processing, storage, and network.

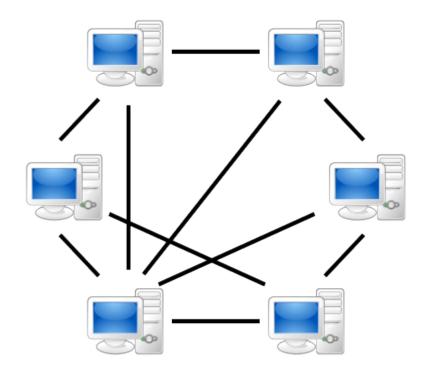
Messages are relayed through a network of participants.

Each participant has only partial knowledge of the network.

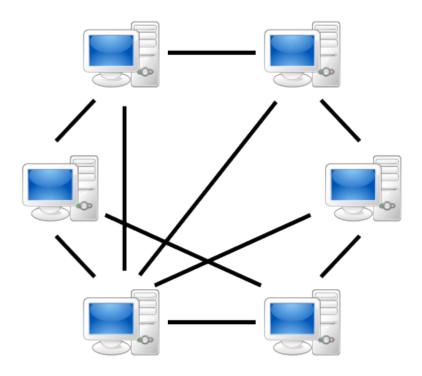
All participants in a distributed application contribute computational resources: processing, storage, and network.

Messages are relayed through a network of participants.

Each participant has only partial knowledge of the network.

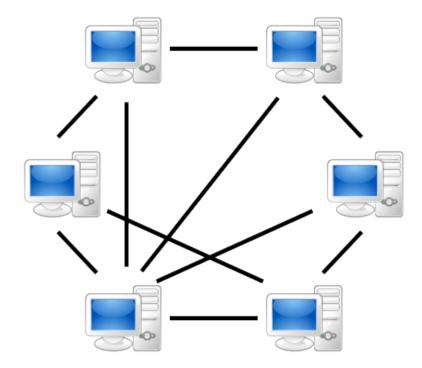


# **Network Structure Concerns**

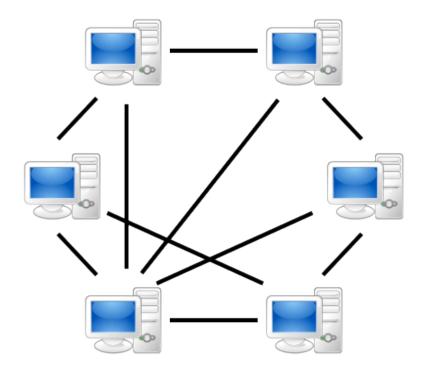


### **Network Structure Concerns**

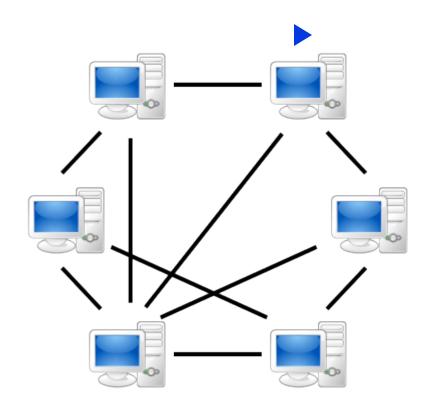
Some data transfers on the Internet are faster than others.



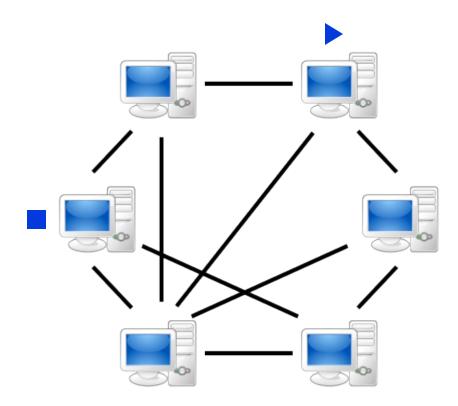
Some data transfers on the Internet are faster than others.



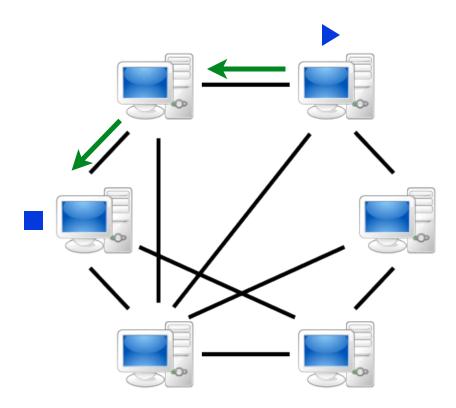
Some data transfers on the Internet are faster than others.



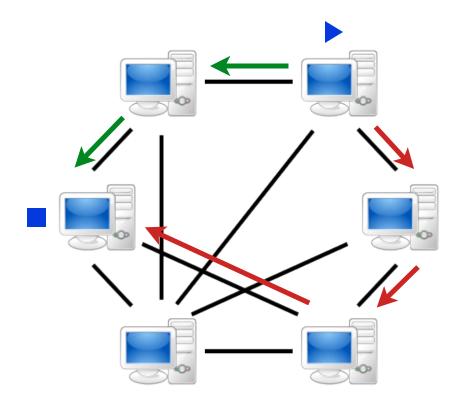
Some data transfers on the Internet are faster than others.



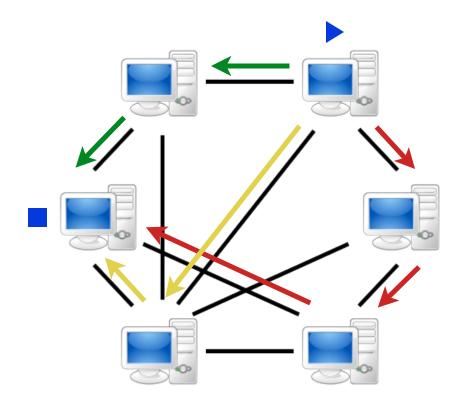
Some data transfers on the Internet are faster than others.



Some data transfers on the Internet are faster than others.



Some data transfers on the Internet are faster than others.



Skype is a Voice Over IP (VOIP) system that uses a hybrid peer-to-peer architecture.

Skype is a Voice Over IP (VOIP) system that uses a hybrid peer-to-peer architecture.

Login & contacts are handled via a centralized server.

Skype is a Voice Over IP (VOIP) system that uses a hybrid peer—to—peer architecture.

Login & contacts are handled via a centralized server.

Conversations between two computers that cannot send messages to each other directly are relayed through *supernodes*.

Skype is a Voice Over IP (VOIP) system that uses a hybrid peer-to-peer architecture.

Login & contacts are handled via a centralized server.

Conversations between two computers that cannot send messages to each other directly are relayed through *supernodes*.

Skype is a Voice Over IP (VOIP) system that uses a hybrid peer-to-peer architecture.

Login & contacts are handled via a centralized server.

Conversations between two computers that cannot send messages to each other directly are relayed through *supernodes*.

Any Skype client with its own IP address may be a supernode.

Client A

Skype is a Voice Over IP (VOIP) system that uses a hybrid peer-to-peer architecture.

Login & contacts are handled via a centralized server.

Conversations between two computers that cannot send messages to each other directly are relayed through *supernodes*.

Any Skype client with its own IP address may be a supernode.

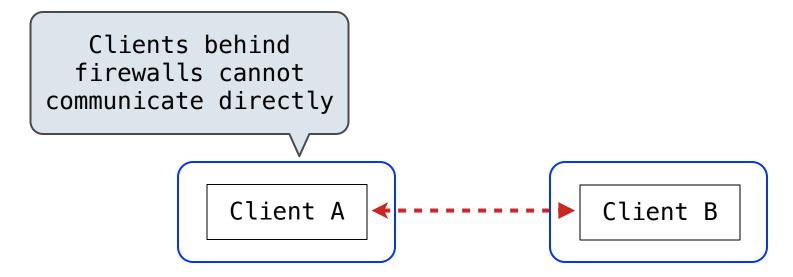
Client A

Client B

Skype is a Voice Over IP (VOIP) system that uses a hybrid peer—to—peer architecture.

Login & contacts are handled via a centralized server.

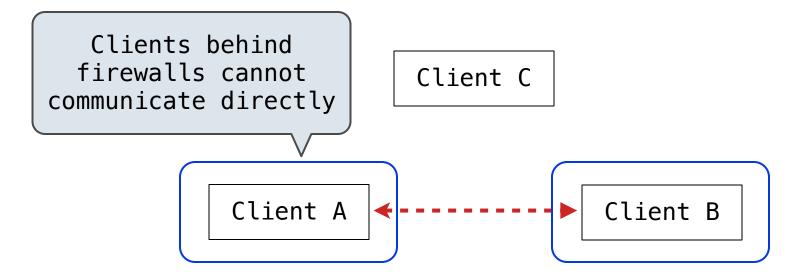
Conversations between two computers that cannot send messages to each other directly are relayed through *supernodes*.



Skype is a Voice Over IP (VOIP) system that uses a hybrid peer—to—peer architecture.

Login & contacts are handled via a centralized server.

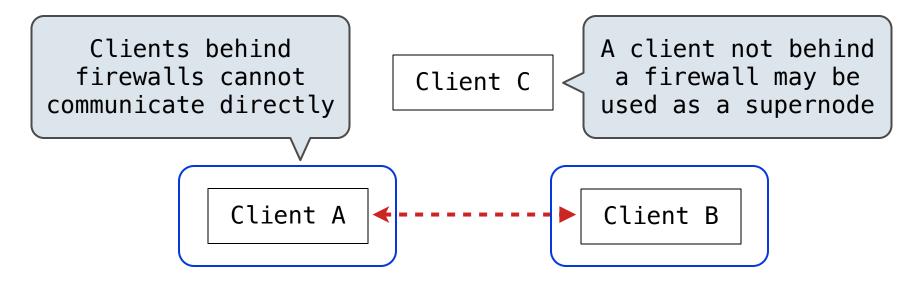
Conversations between two computers that cannot send messages to each other directly are relayed through *supernodes*.



Skype is a Voice Over IP (VOIP) system that uses a hybrid peer—to—peer architecture.

Login & contacts are handled via a centralized server.

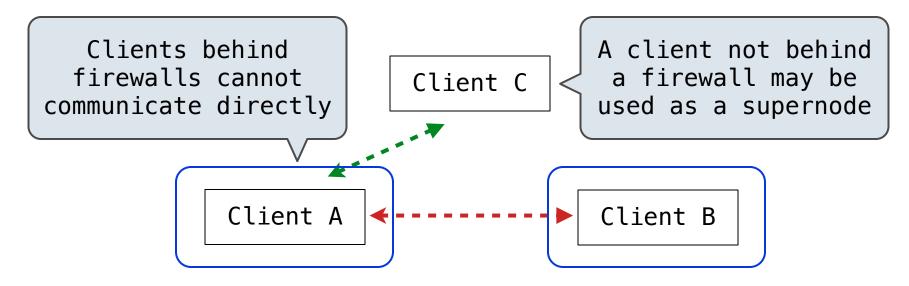
Conversations between two computers that cannot send messages to each other directly are relayed through *supernodes*.



Skype is a Voice Over IP (VOIP) system that uses a hybrid peer—to—peer architecture.

Login & contacts are handled via a centralized server.

Conversations between two computers that cannot send messages to each other directly are relayed through *supernodes*.



Skype is a Voice Over IP (VOIP) system that uses a hybrid peer—to—peer architecture.

Login & contacts are handled via a centralized server.

Conversations between two computers that cannot send messages to each other directly are relayed through *supernodes*.

