Congestion Control

Mark Allman
mallman@case.edu

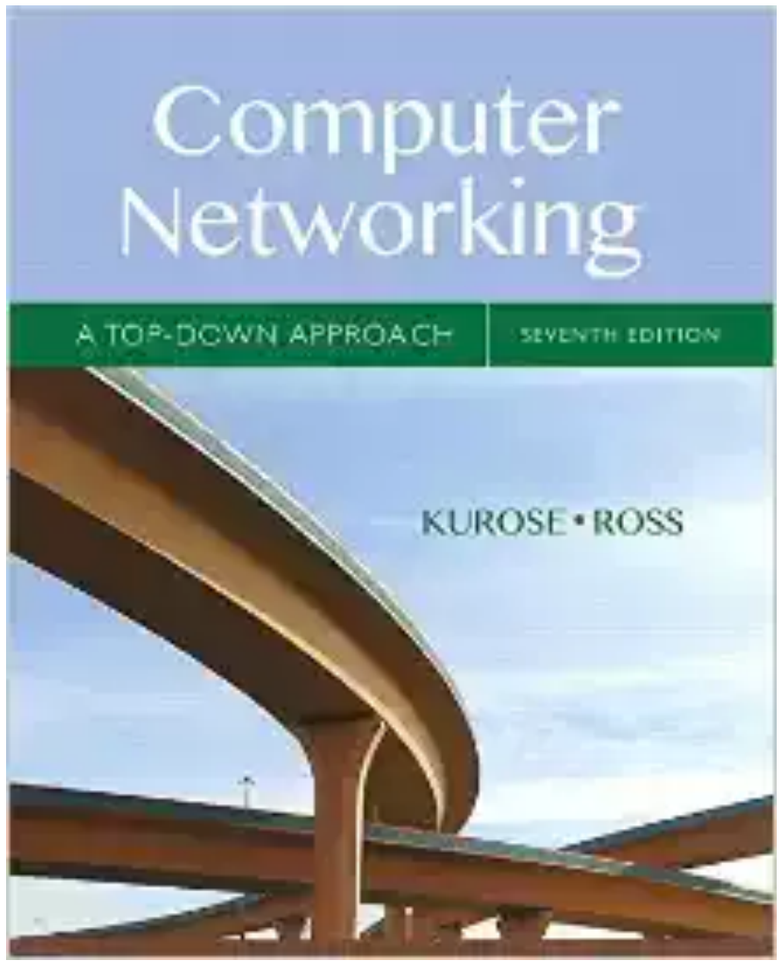
EECS 325/425
Fall 2018

"Well, we busted outa class, had to get away from those fools,
We learned more from a three minute record than we ever learned in school."

These slides are more-or-less directly from the slide set developed by Jim Kurose and Keith Ross for their book "Computer Networking: A Top Down Approach, 5th edition".

The slides have been lightly adapted for Mark Allman's EECS 325/425 Computer Networks class at Case Western Reserve University.

Reading Along …



- 3.6: Principles of Congestion Control

Congestion Control

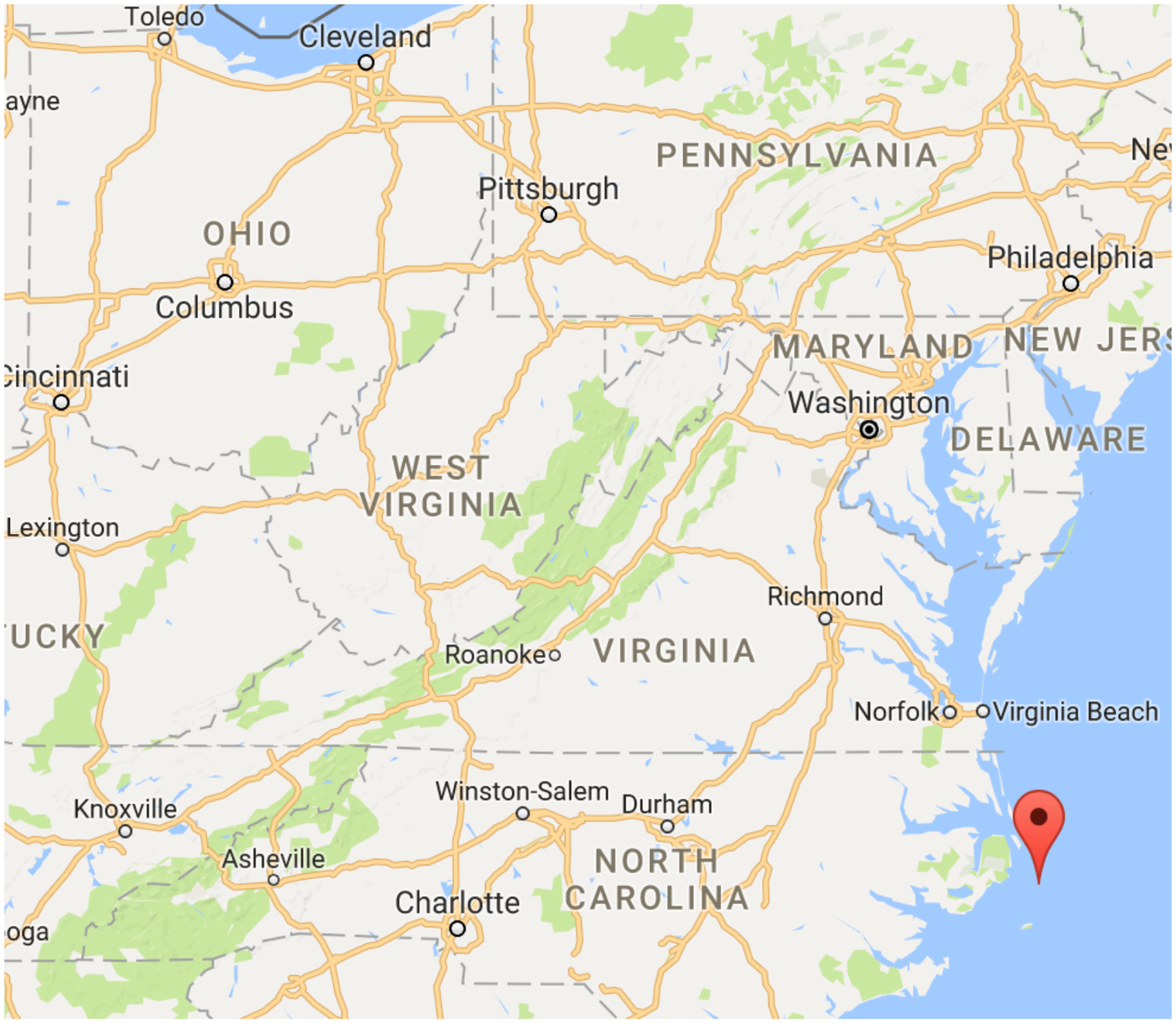
Congestion Control

- What is "congestion"?

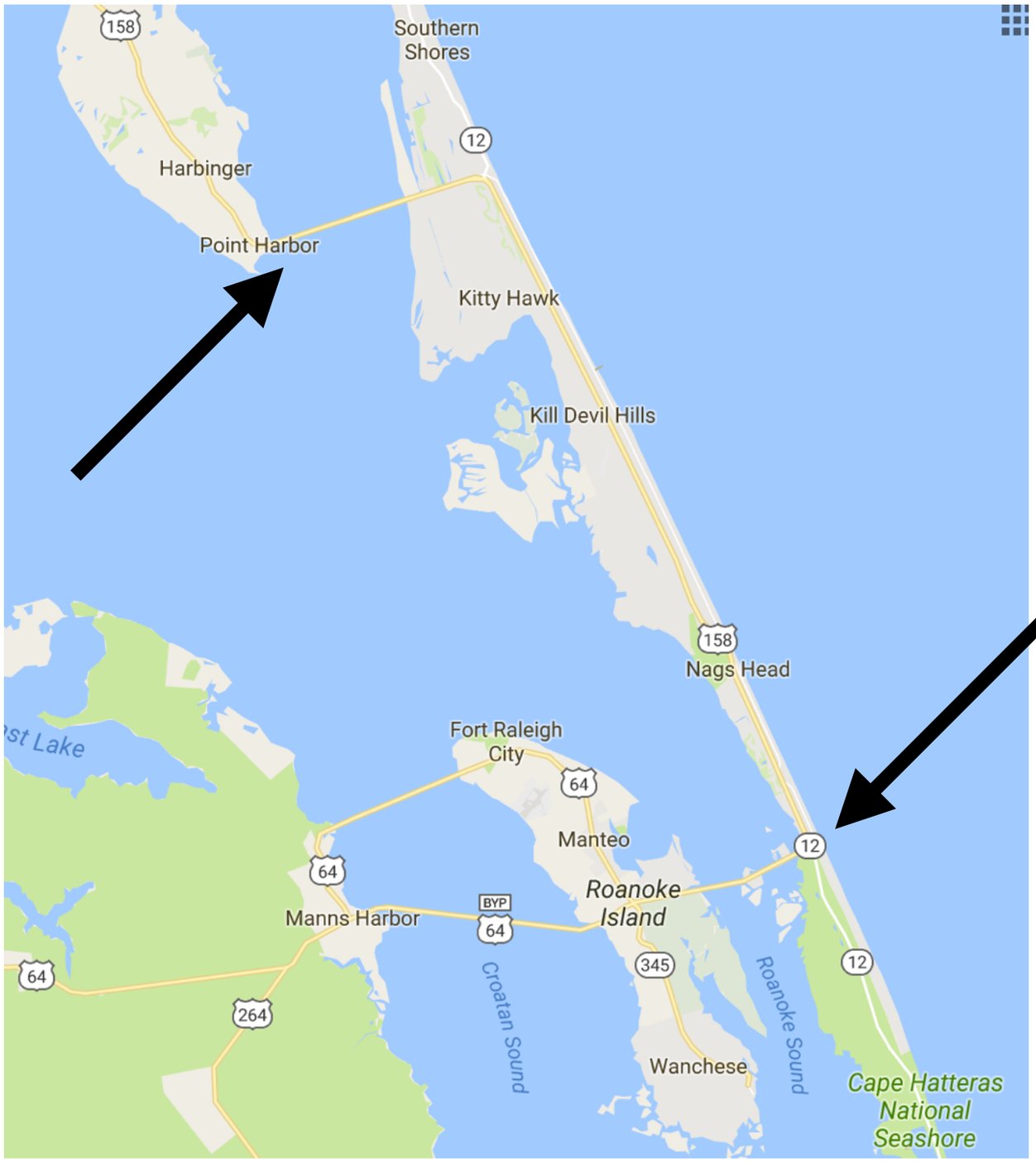
Congestion Control

- What is "congestion"?
- Why does it need "controlled"?

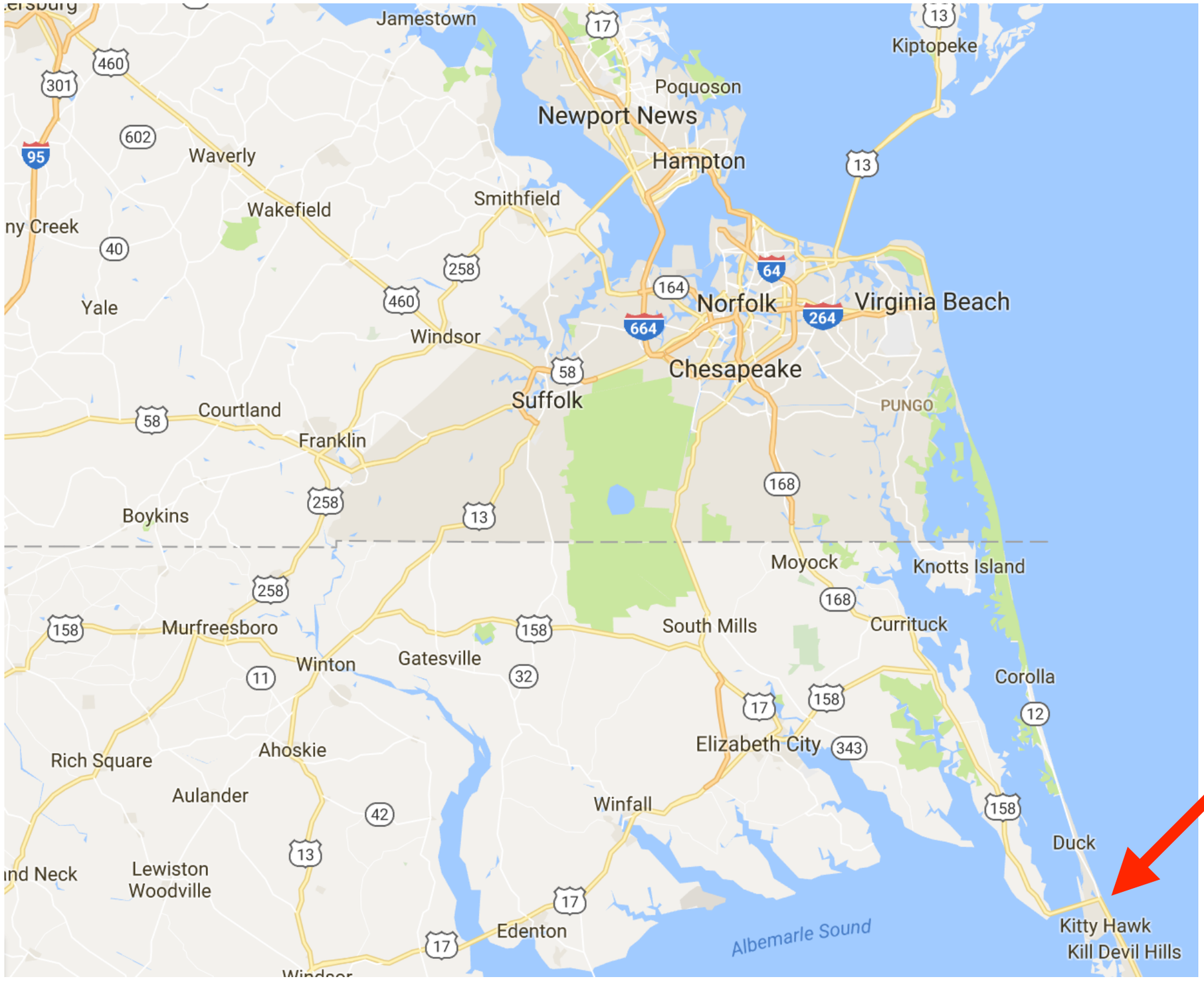
Congestion


Toledo
Cleveland
ayne
PENNSYLVANIA
Pittsburgh
OHIO
Philadelphia
Columbus
MARYLAND
NEW JERS
Cincinnati
Washington
DELAWARE
WEST
VIRGINIA
Lexington
Richmond
UCKY
Roanoke
VIRGINIA
Norfolk
Virginia Beach
Winston-Salem
Durham
Knoxville
Asheville
NORTH
CAROLINA
Charlotte
oga

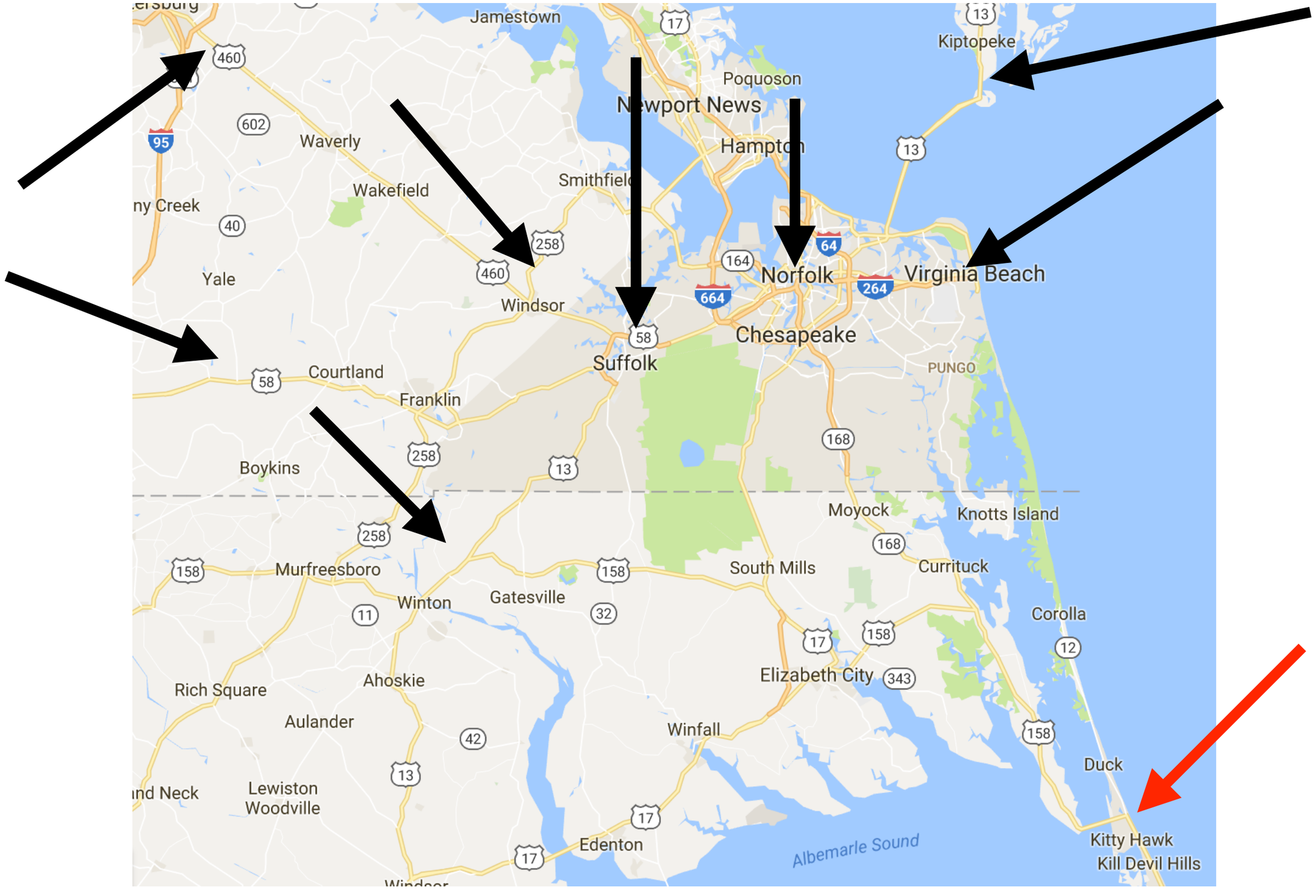
Congestion


158
Southern Shores
12
Harbinger
Point Harbor
Kitty Hawk
Kill Devil Hills
158
Nags Head
Fort Raleigh City
64
Manteo
12
64
Roanoke Island
BYP
Manns Harbor
64
345
12
64
264
Croatan Sound
Roanoke Sound
Wanchese
Cape Hatteras National Seashore

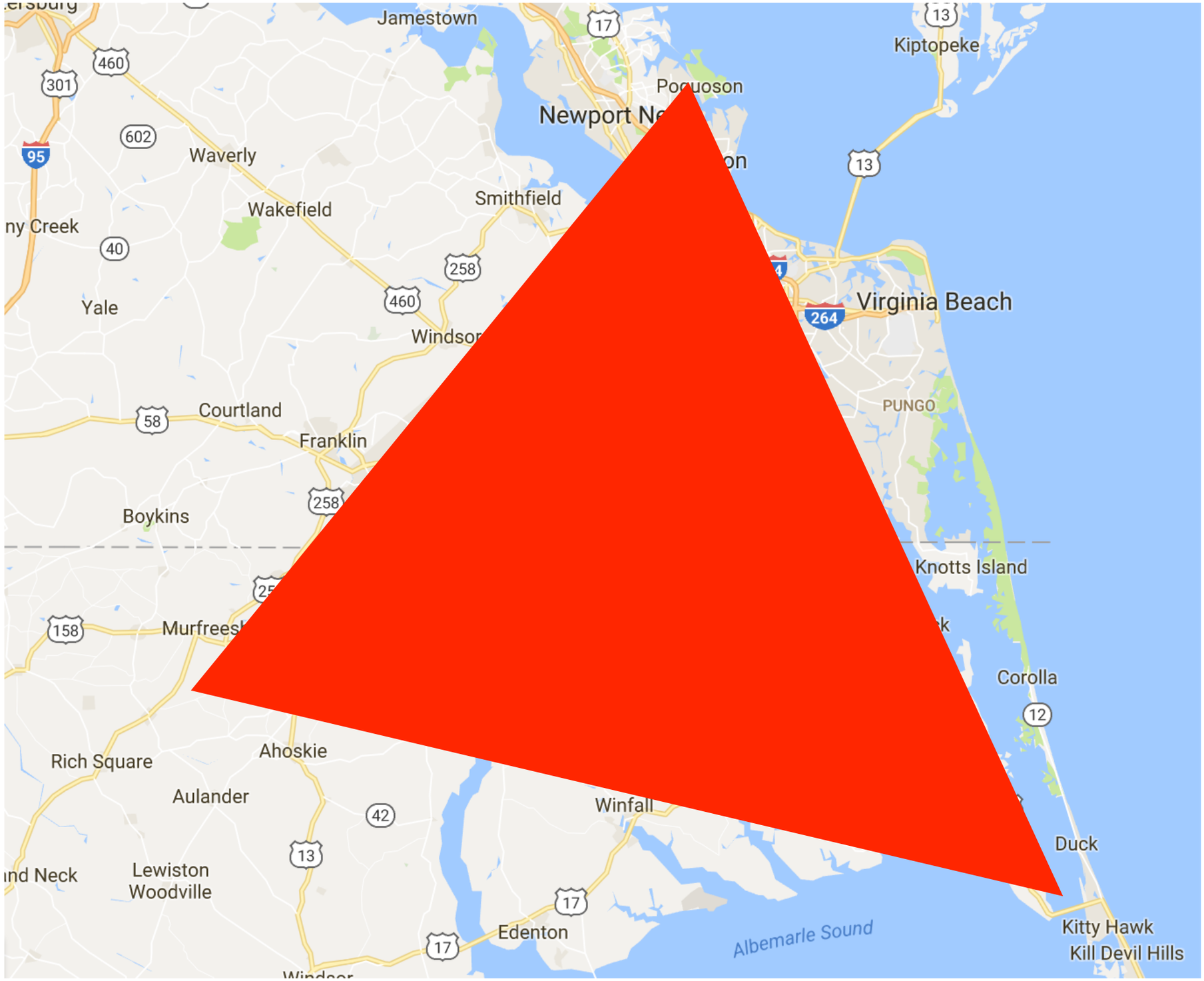
Congestion


Jamestown
17
13
Kiptopeke
460
301
Poquoson
Newport News
602
95
Waverly
Hampton
13
Smithfield
Wakefield
ny Creek
40
258
64
164
460
Norfolk
Virginia Beach
Yale
664
264
Windsor
58
Chesapeake
Suffolk
Courtland
PUNGO
58
Franklin
168
258
Boykins
13
Moyock
Knotts Island
168
258
158
Murfreesboro
158
South Mills
Currituck
Winton
Gatesville
11
32
Corolla
17
158
12
Rich Square
Ahoskie
Elizabeth City
343
Aulander
Winfall
42
158
Duck
13
Lewiston
Woodville
17
17
Edenton
Albemarle Sound
Kitty Hawk
Kill Devil Hills

Congestion

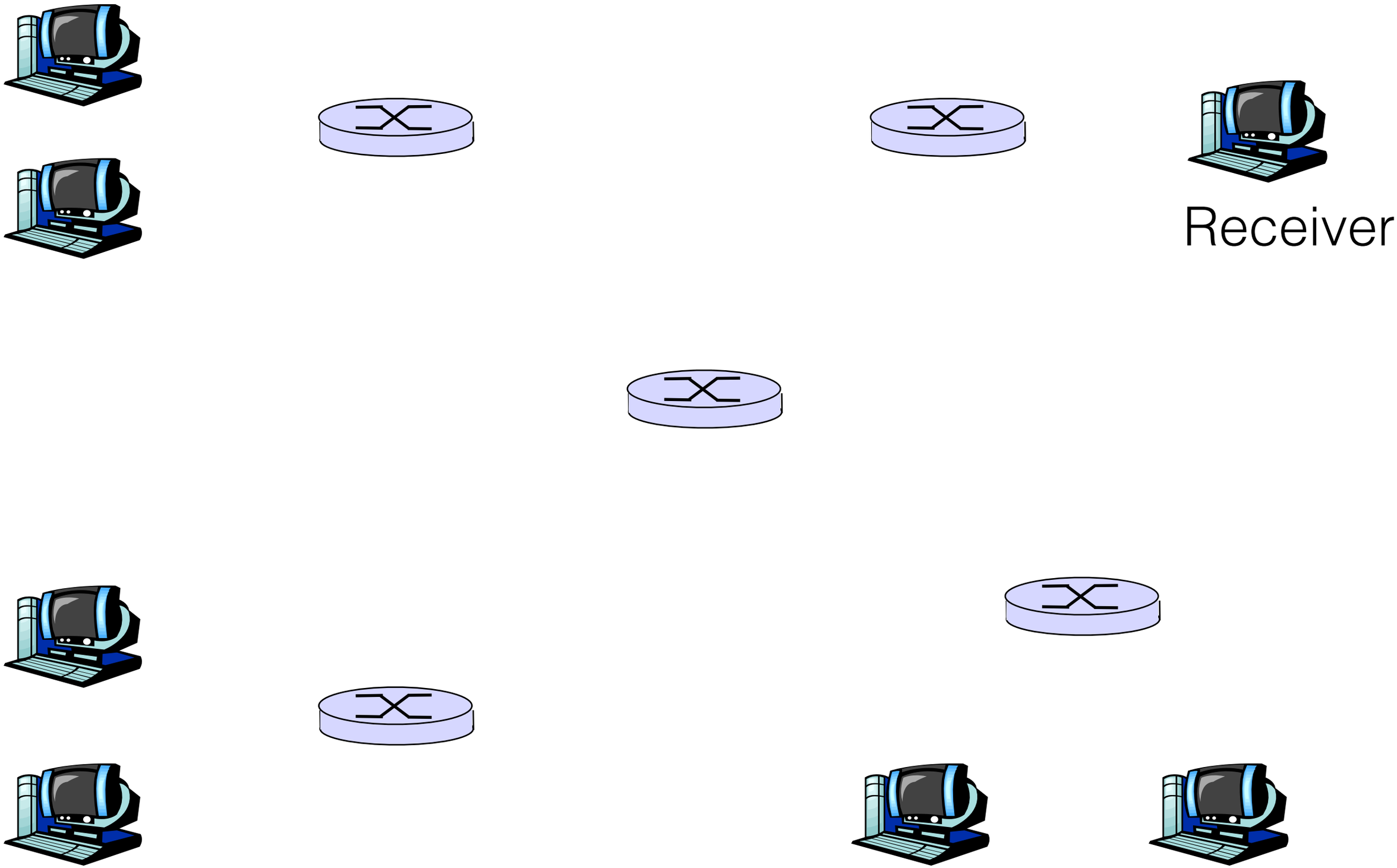

Jamestown
17
13
Kiptopeke
460
Poquoson
Newport News
602
95
Waverly
Hampton
13
Smithfield
Wakefield
ny Creek
40
258
64
164
460
Norfolk
Virginia Beach
Yale
664
264
Windsor
58
Chesapeake
Suffolk
Courtland
58
PUNGO
Franklin
168
258
13
Boykins
Moyock
Knotts Island
258
168
158
Murfreesboro
158
South Mills
Currituck
Winton
Gatesville
11
32
Corolla
17
158
12
Elizabeth City
343
Rich Square
Ahoskie
Aulander
Winfall
42
158
Duck
13
Lewiston
Woodville
17
Edenton
Albemarle Sound
Kitty Hawk
Kill Devil Hills
17

Congestion

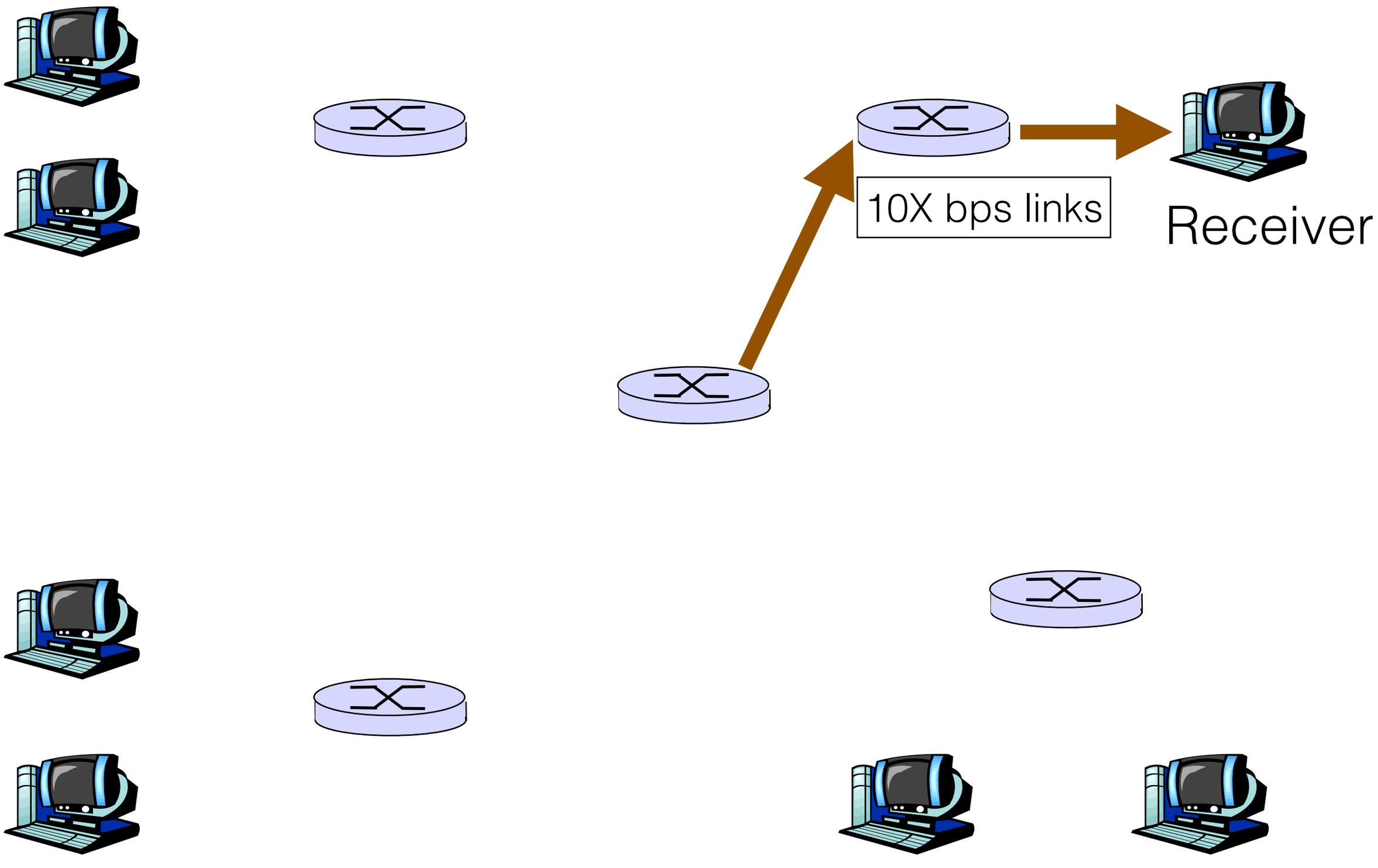

Jamestown
17
13
Kiptopeke
460
301
602
95
Waverly
Wakefield
Smithfield
13
ny Creek
40
258
460
Yale
Virginia Beach
264
58
Courtland
Franklin
PUNGO
Boykins
258
Knotts Island
158
Corolla
12
Rich Square
Ahoskie
Aulander
42
Winfall
13
Duck
Lewiston
Woodville
17
Edenton
17
Albemarle Sound
Kitty Hawk
Kill Devil Hills

Worth the Effort!

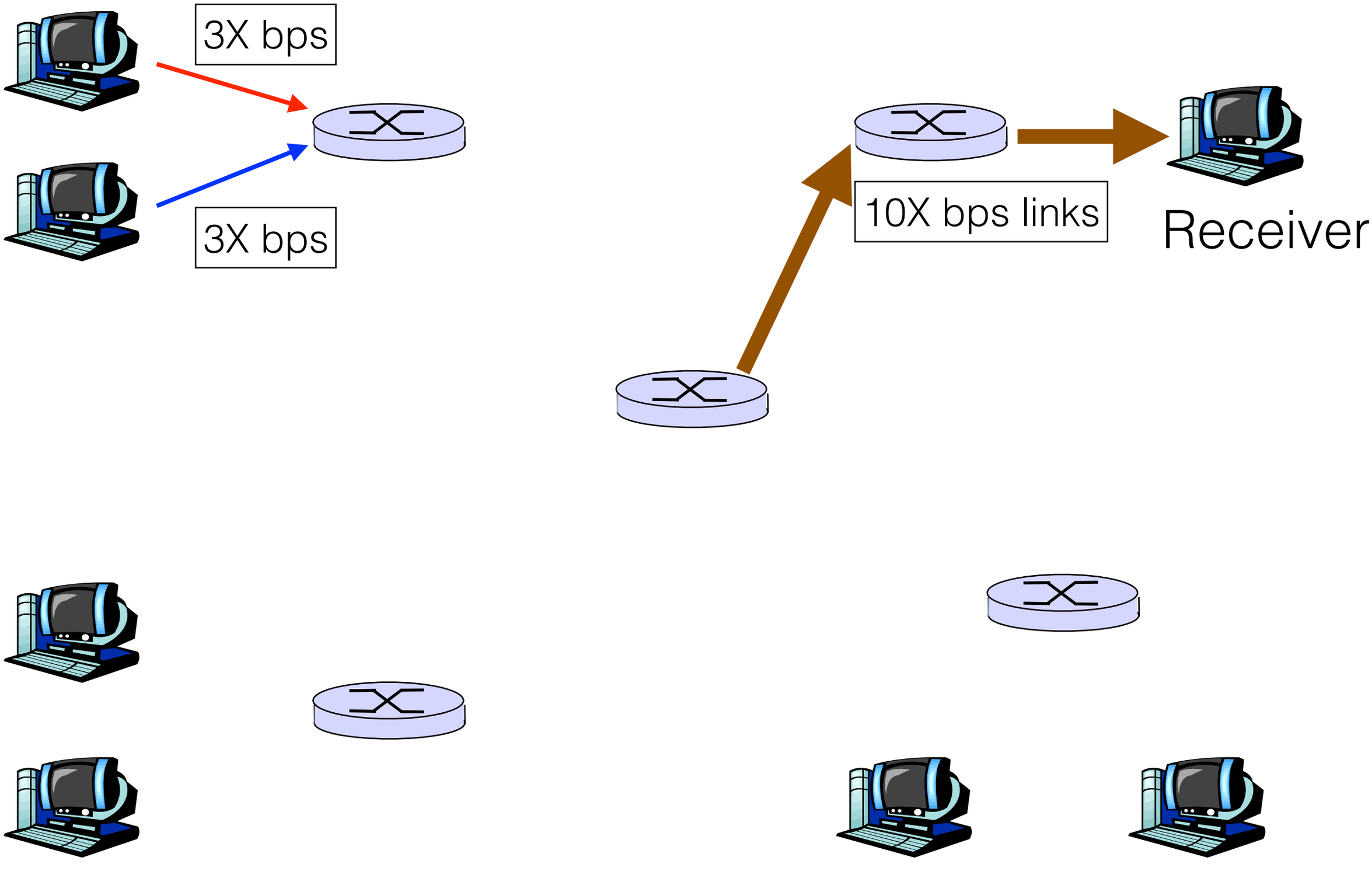
A Networking Example



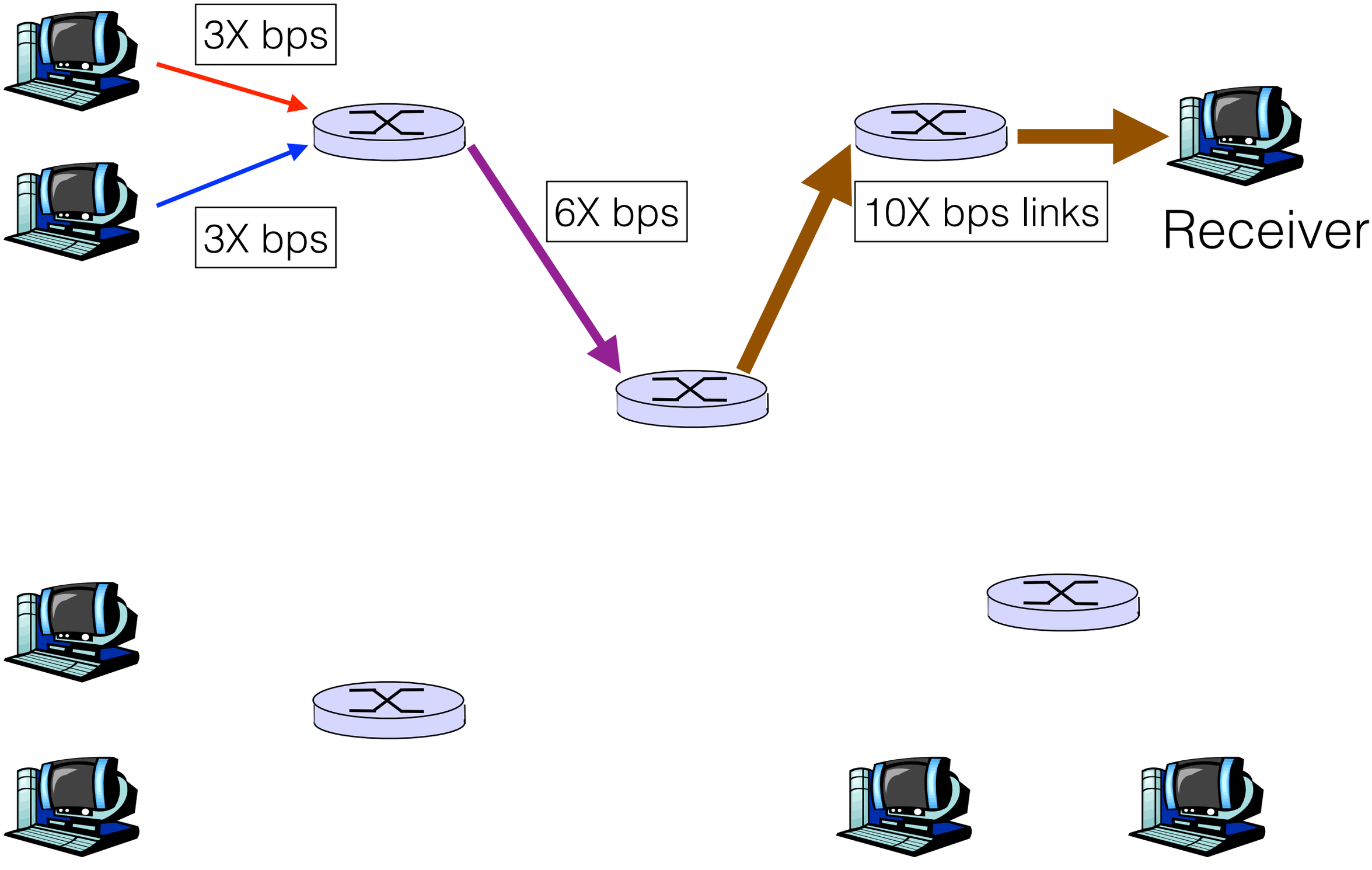
A Networking Example



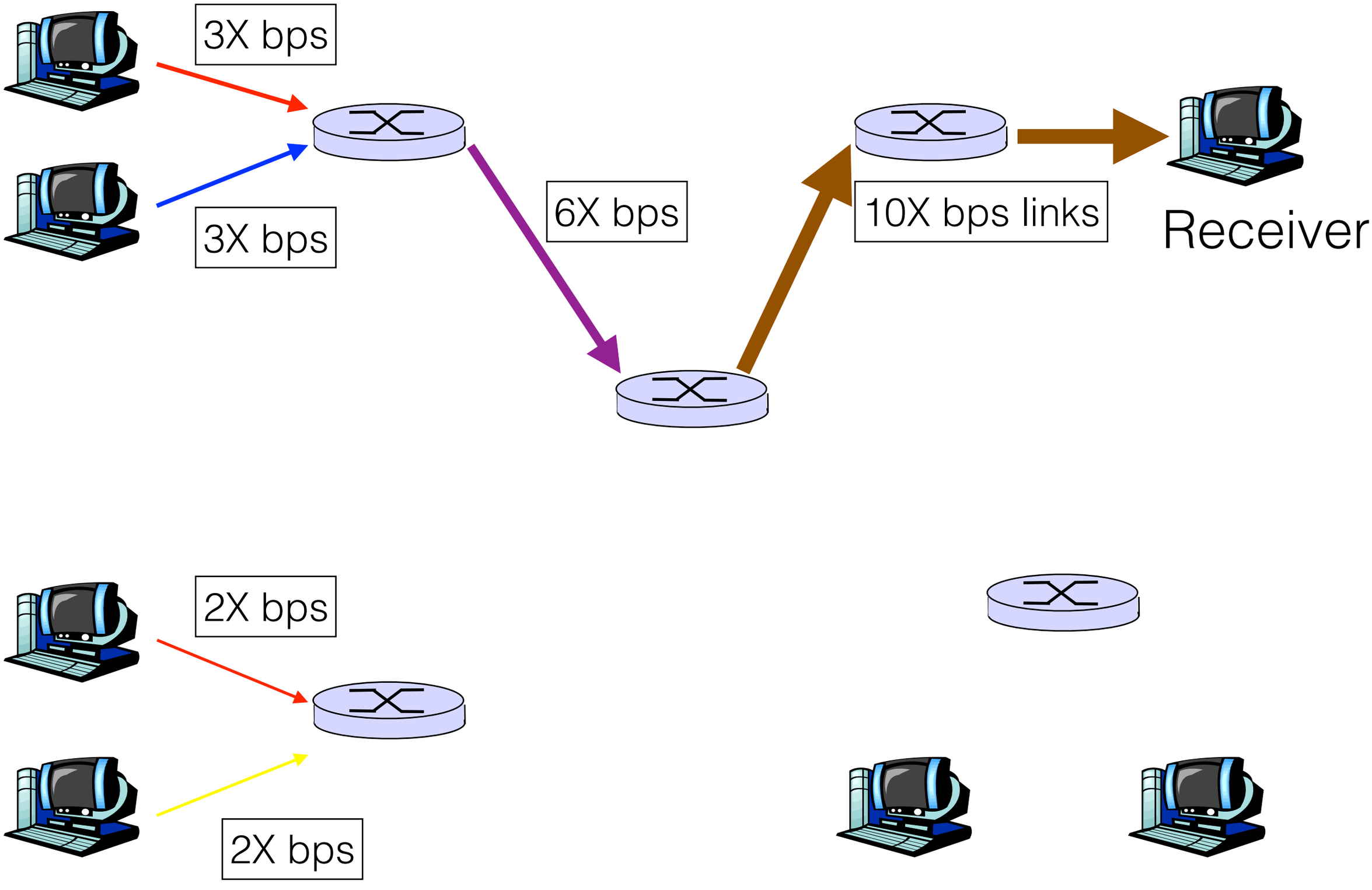
A Networking Example



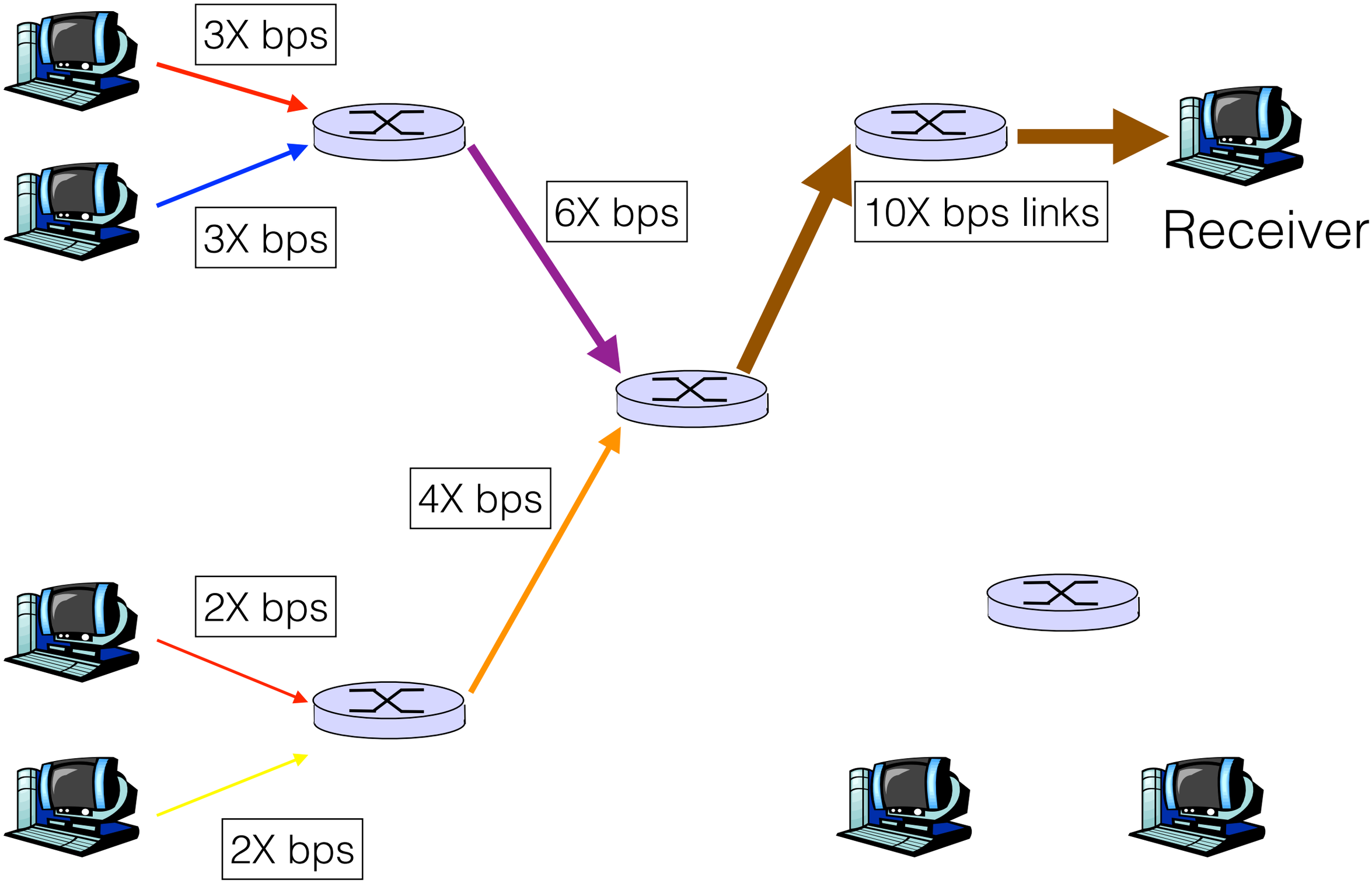
A Networking Example



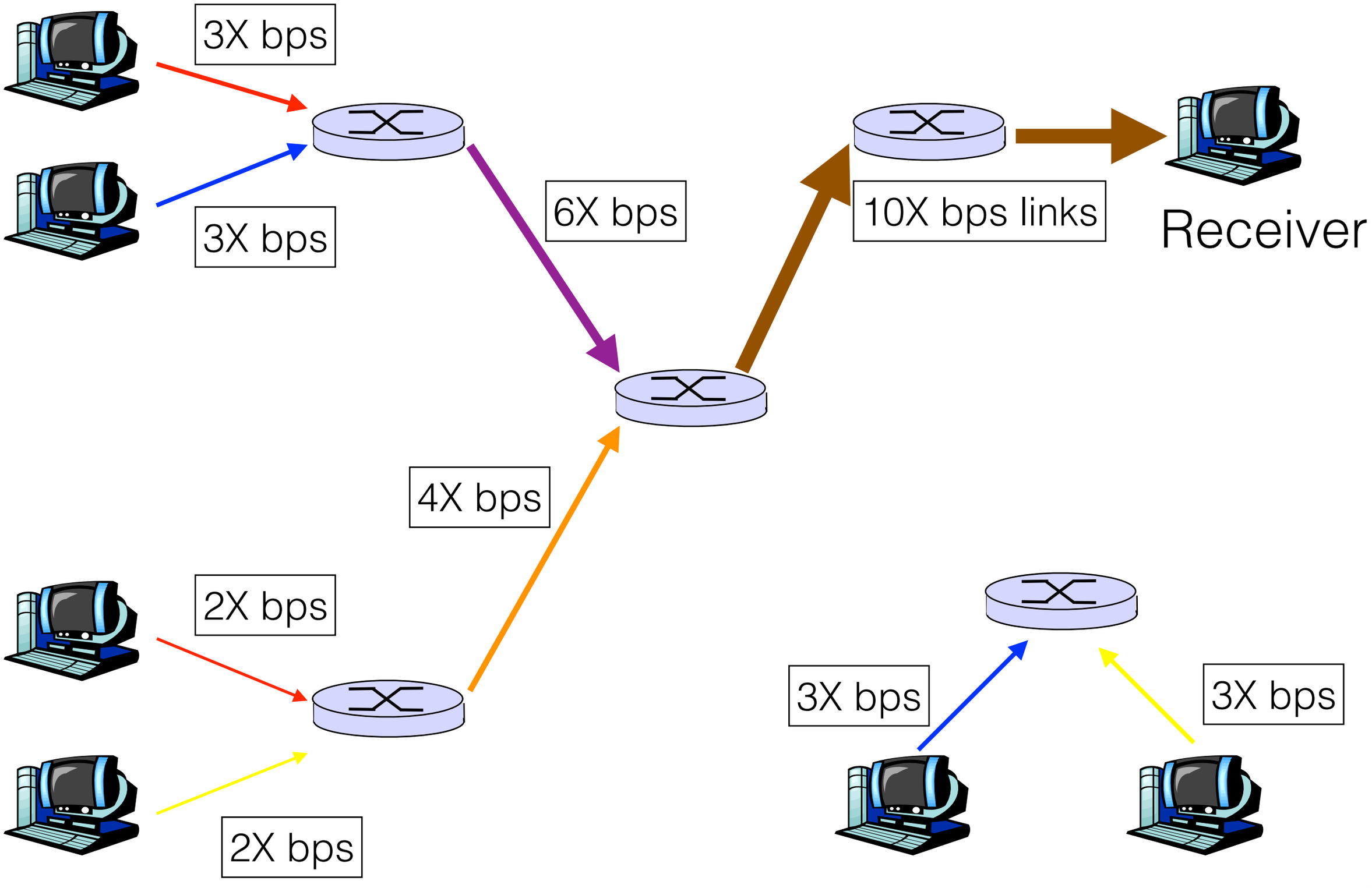
A Networking Example



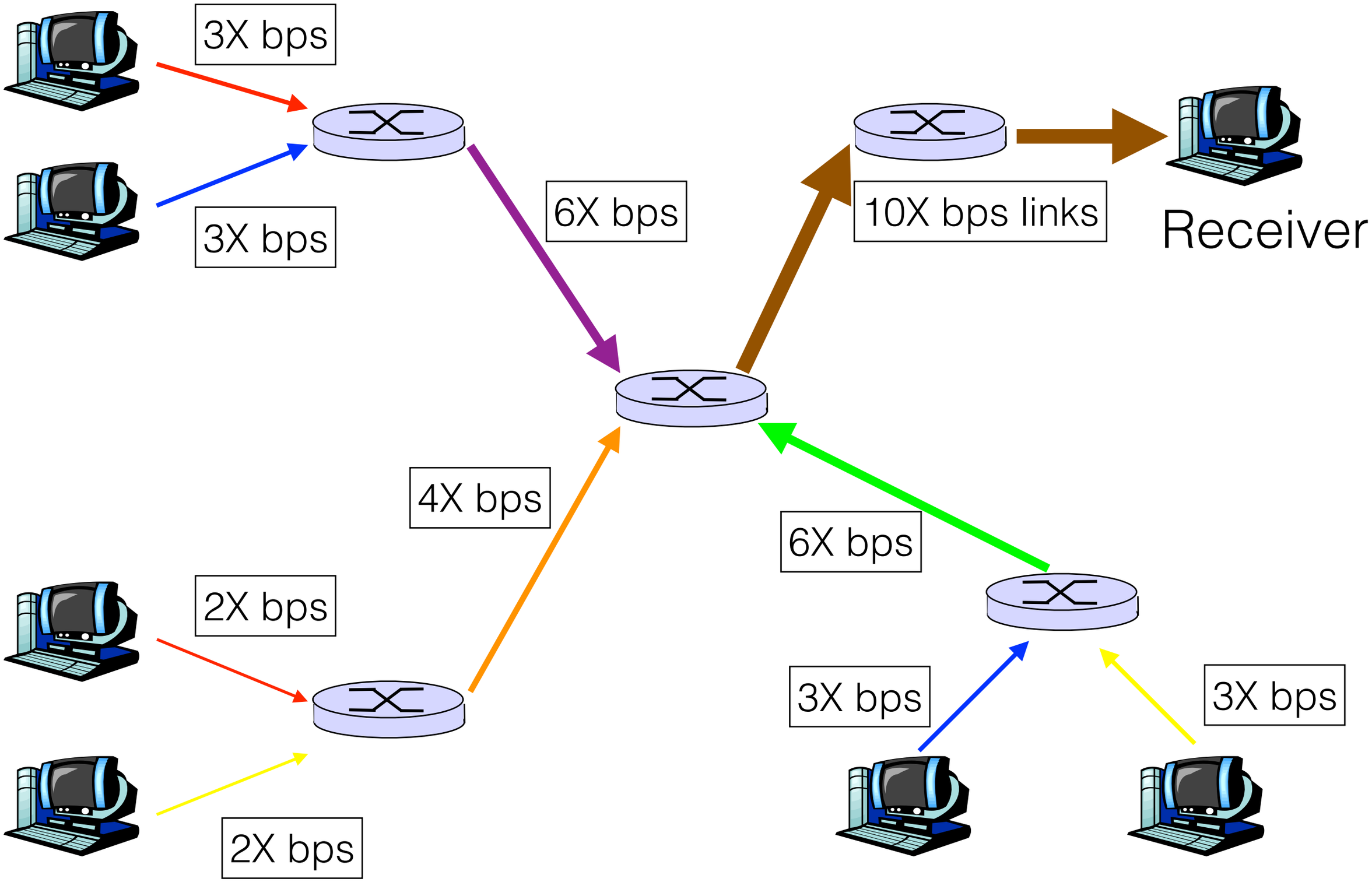
A Networking Example



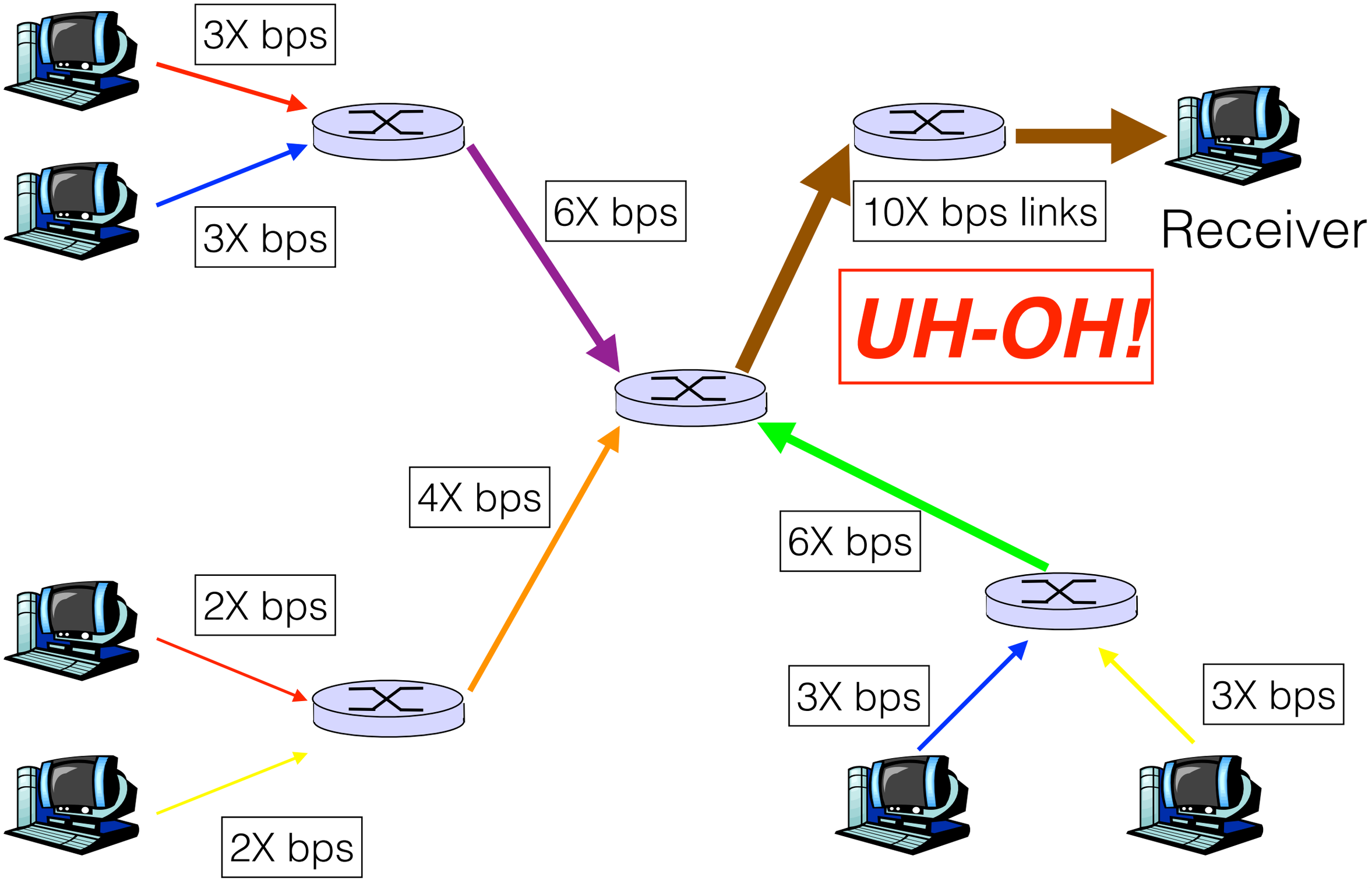
A Networking Example



A Networking Example



A Networking Example



Principles of Congestion Control

Principles of Congestion Control

Congestion:

- informally: "too many sources sending too much data too fast for network to handle"
- different from flow control!

Principles of Congestion Control

Congestion:

- informally: "too many sources sending too much data too fast for network to handle"
- different from flow control!
- manifestations:

Principles of Congestion Control

Congestion:

- informally: "too many sources sending too much data too fast for network to handle"
- different from flow control!
- manifestations:
  - lost packets (buffer overflow at routers)
  - long delays (queueing in router buffers)

Principles of Congestion Control

Congestion:

- informally: "too many sources sending too much data too fast for network to handle"
- different from flow control!
- manifestations:
  - lost packets (buffer overflow at routers)
  - long delays (queueing in router buffers)
- a top-10 problem!

Costs of Congestion

Costs of Congestion

- more work for a given sending rate
  - i.e., more retransmissions

Costs of Congestion

- more work for a given sending rate
  - i.e., more retransmissions
- retransmissions take network resources

Costs of Congestion

- more work for a given sending rate
  - i.e., more retransmissions
- retransmissions take network resources
- less timely data delivery

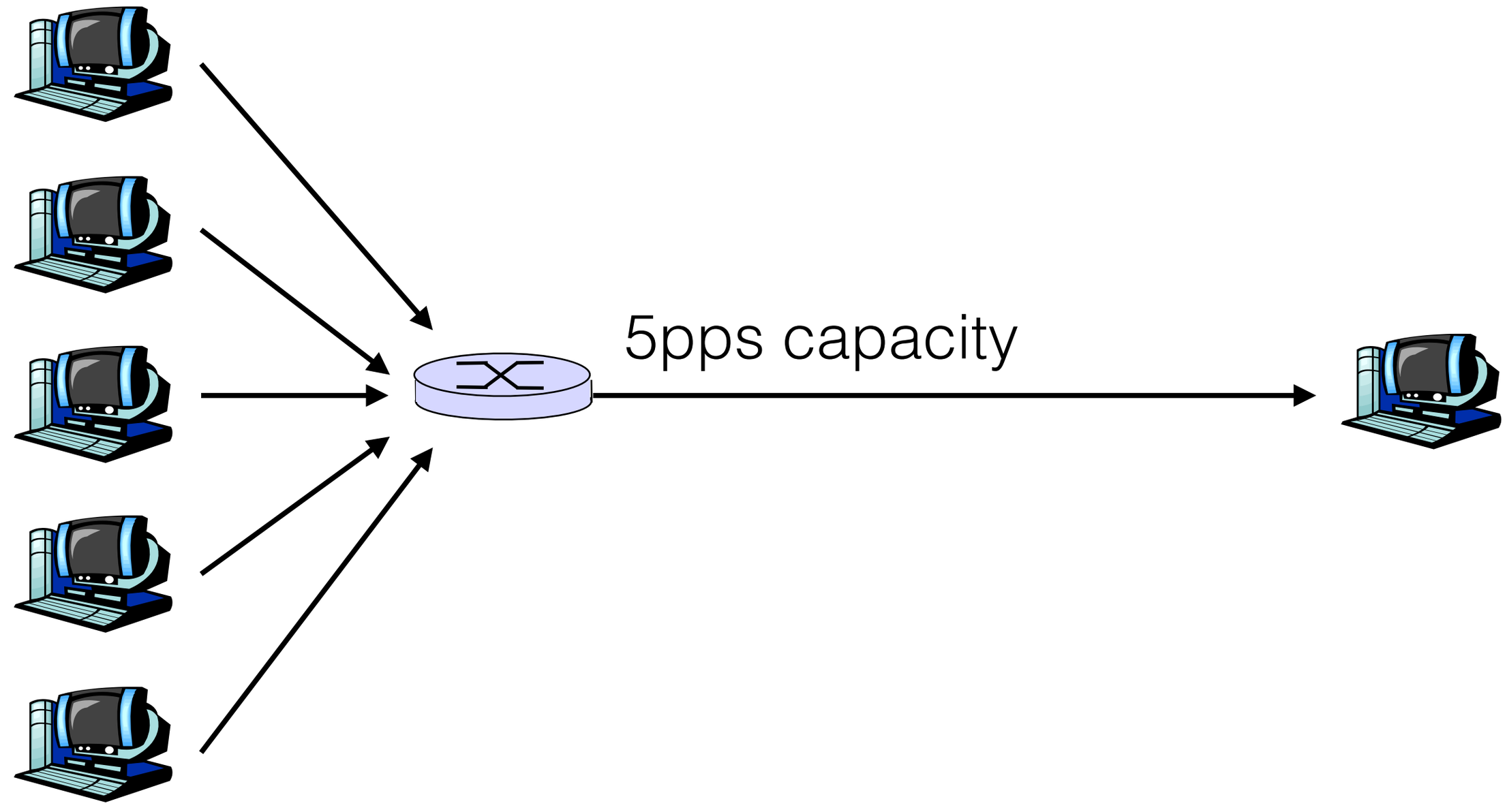
Costs of Congestion

- more work for a given sending rate
  - i.e., more retransmissions
- retransmissions take network resources
- less timely data delivery
- but, are these annoyance or fundamental issues?!

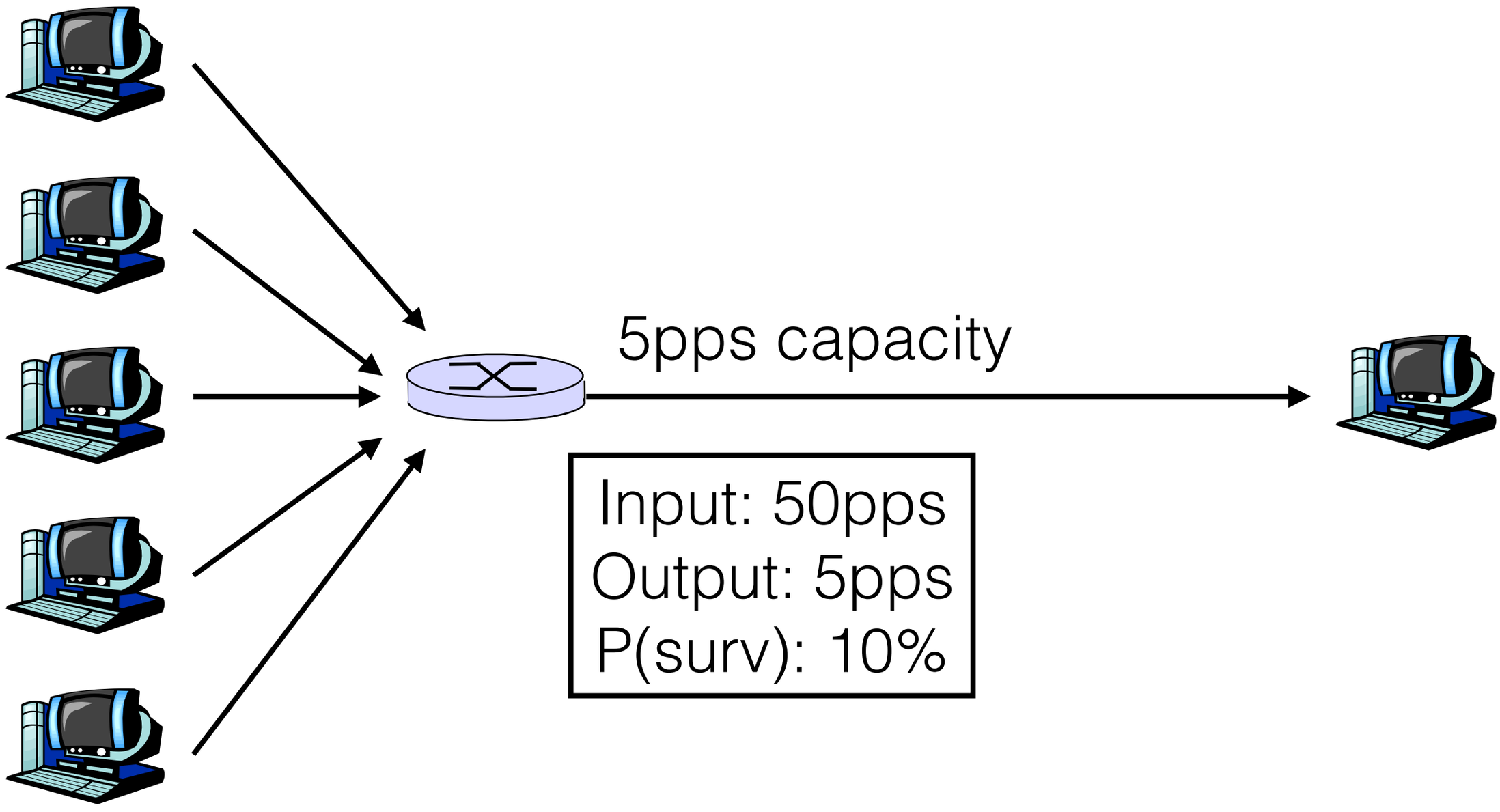
Costs of Congestion

- more work for a given sending rate
  - i.e., more retransmissions
- retransmissions take network resources
- less timely data delivery
- but, are these annoyance or fundamental issues?!
  - left unchecked congestion can lead to "collapse" whereby a network expends scarce resources on packets that will ultimately be dropped
    - i.e., accomplish little—if any—meaningful work

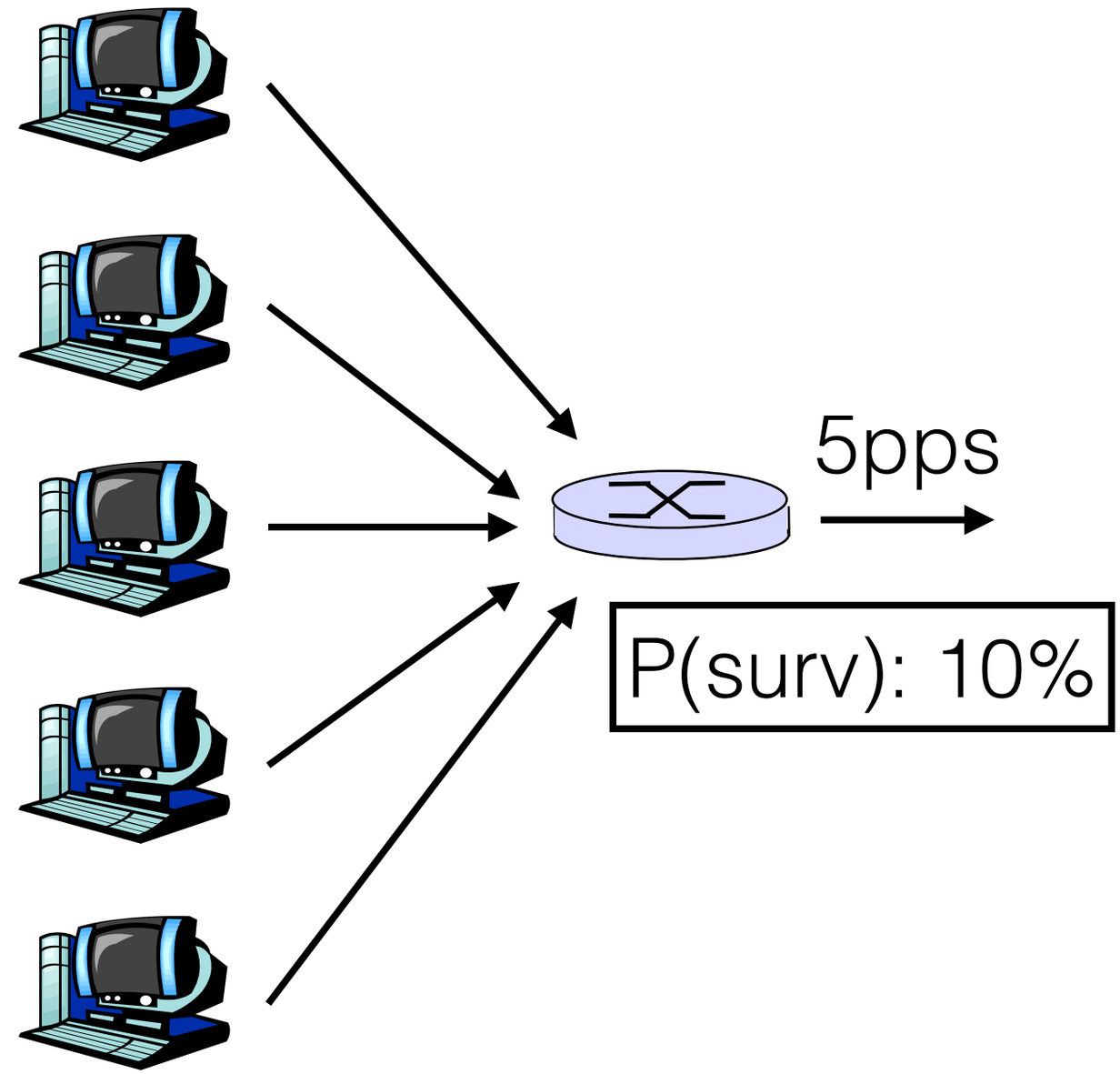
Congestion: Single Bottleneck



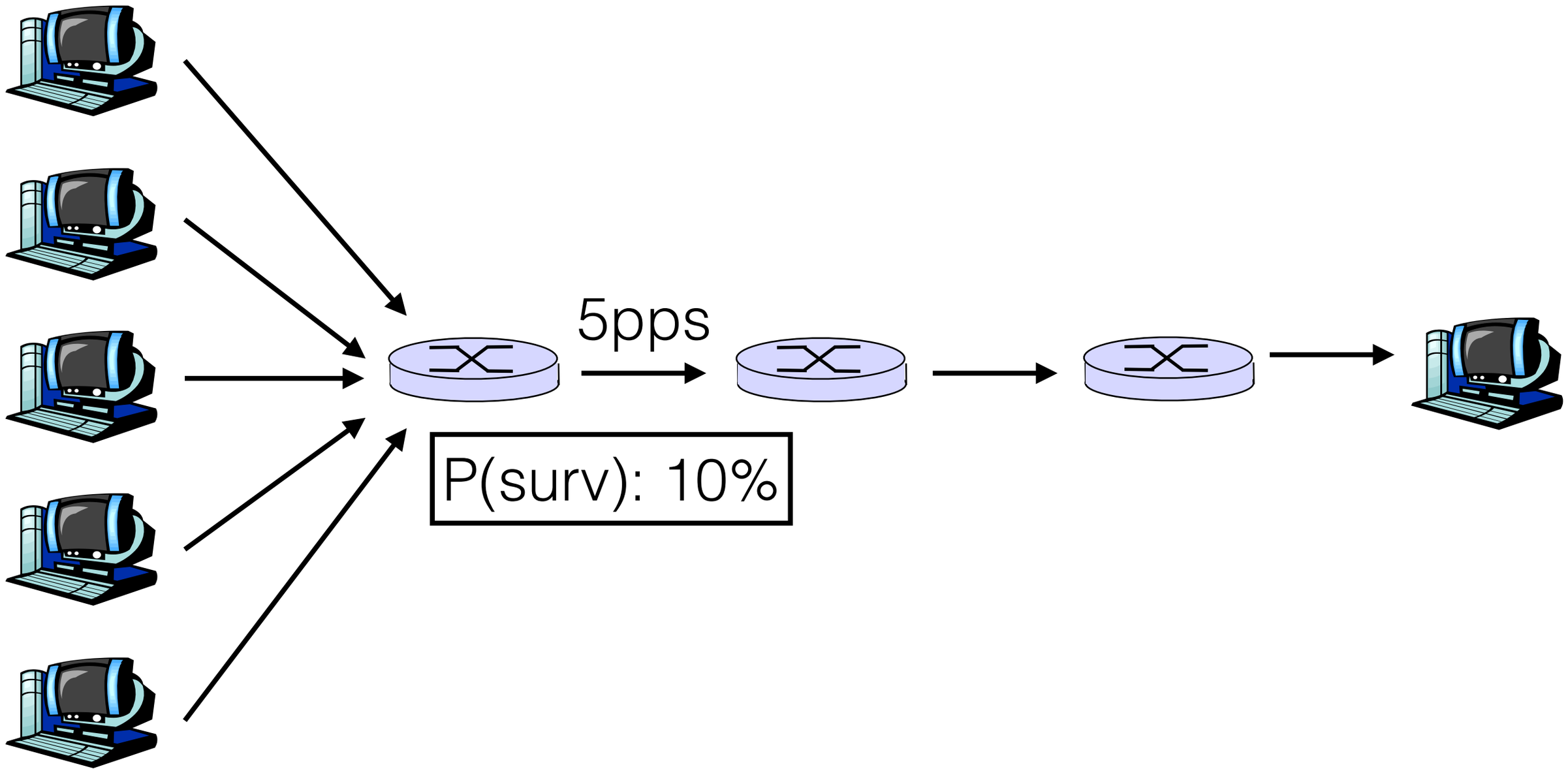
Congestion: Single Bottleneck



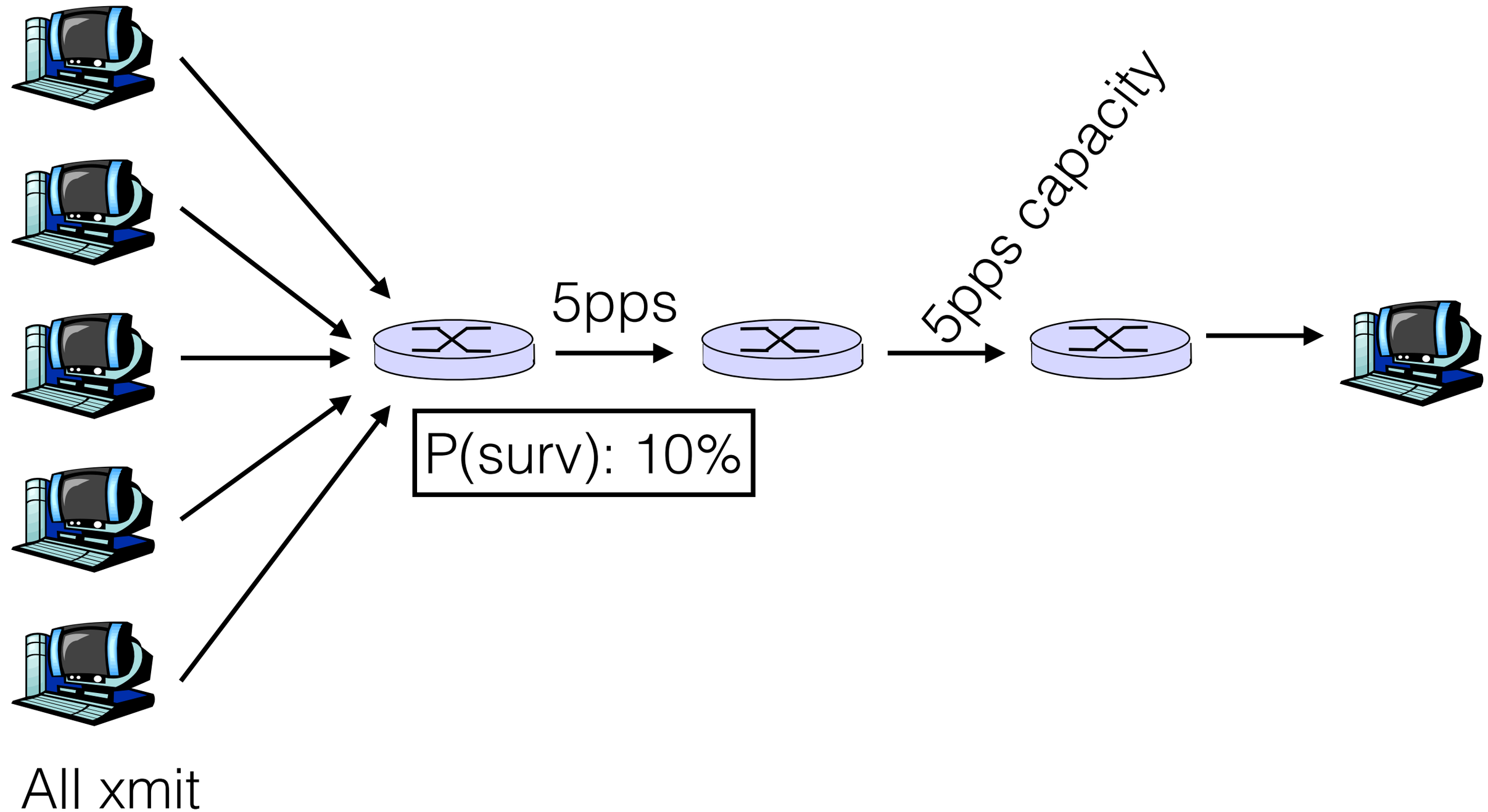
Congestion: Multiple Bottlenecks



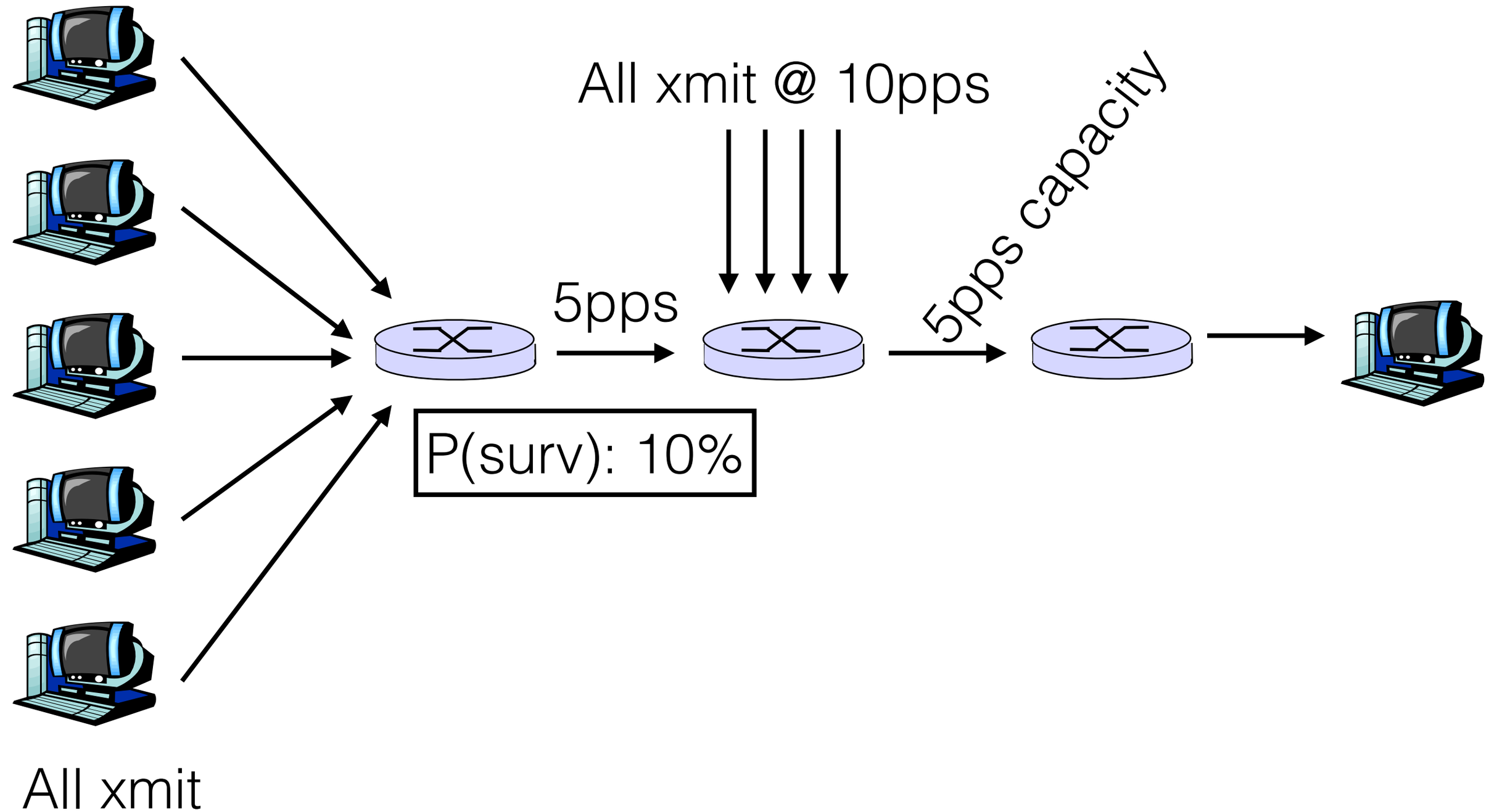
Congestion: Multiple Bottlenecks



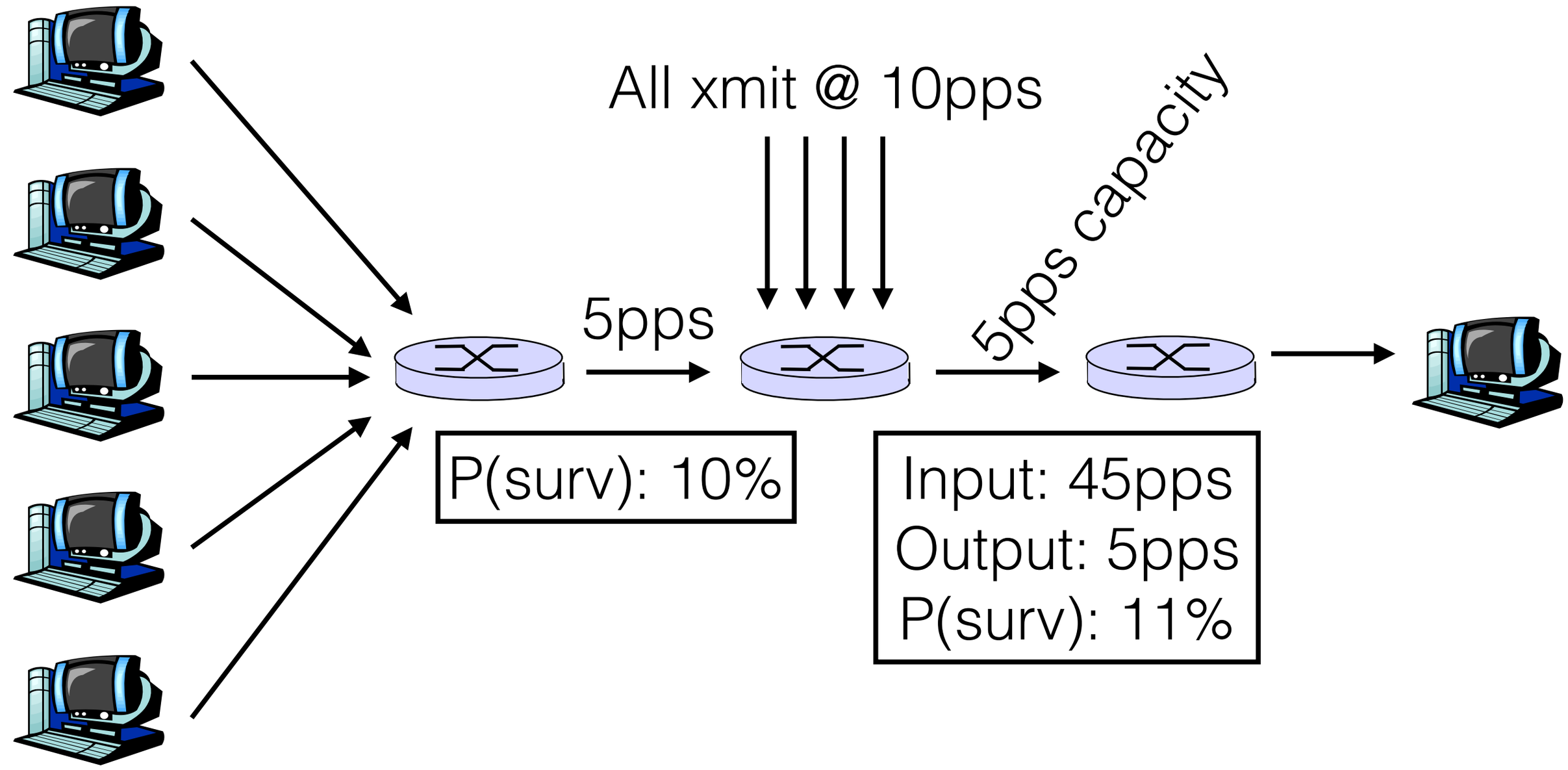
Congestion: Multiple Bottlenecks



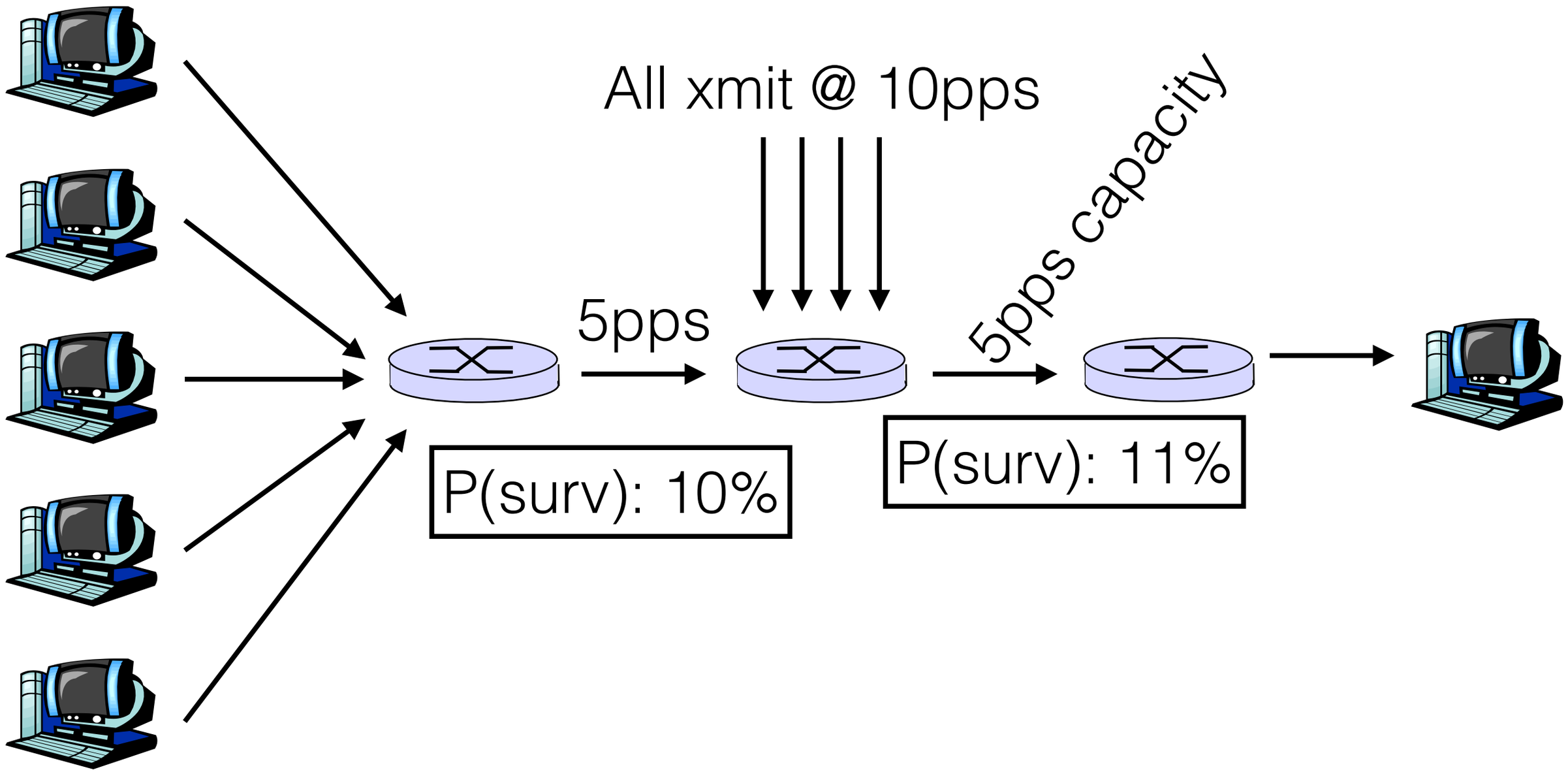
Congestion: Multiple Bottlenecks



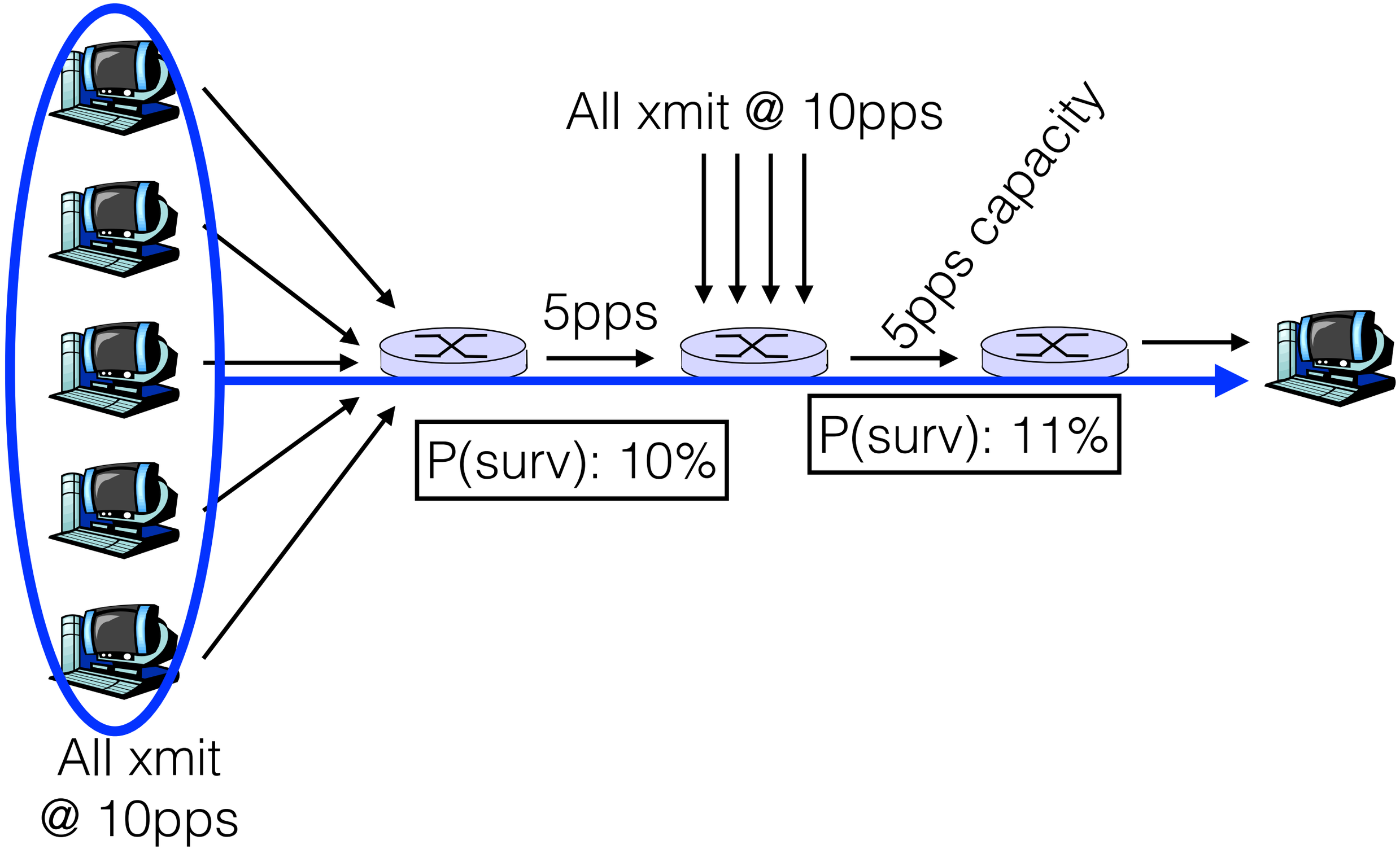
Congestion: Multiple Bottlenecks



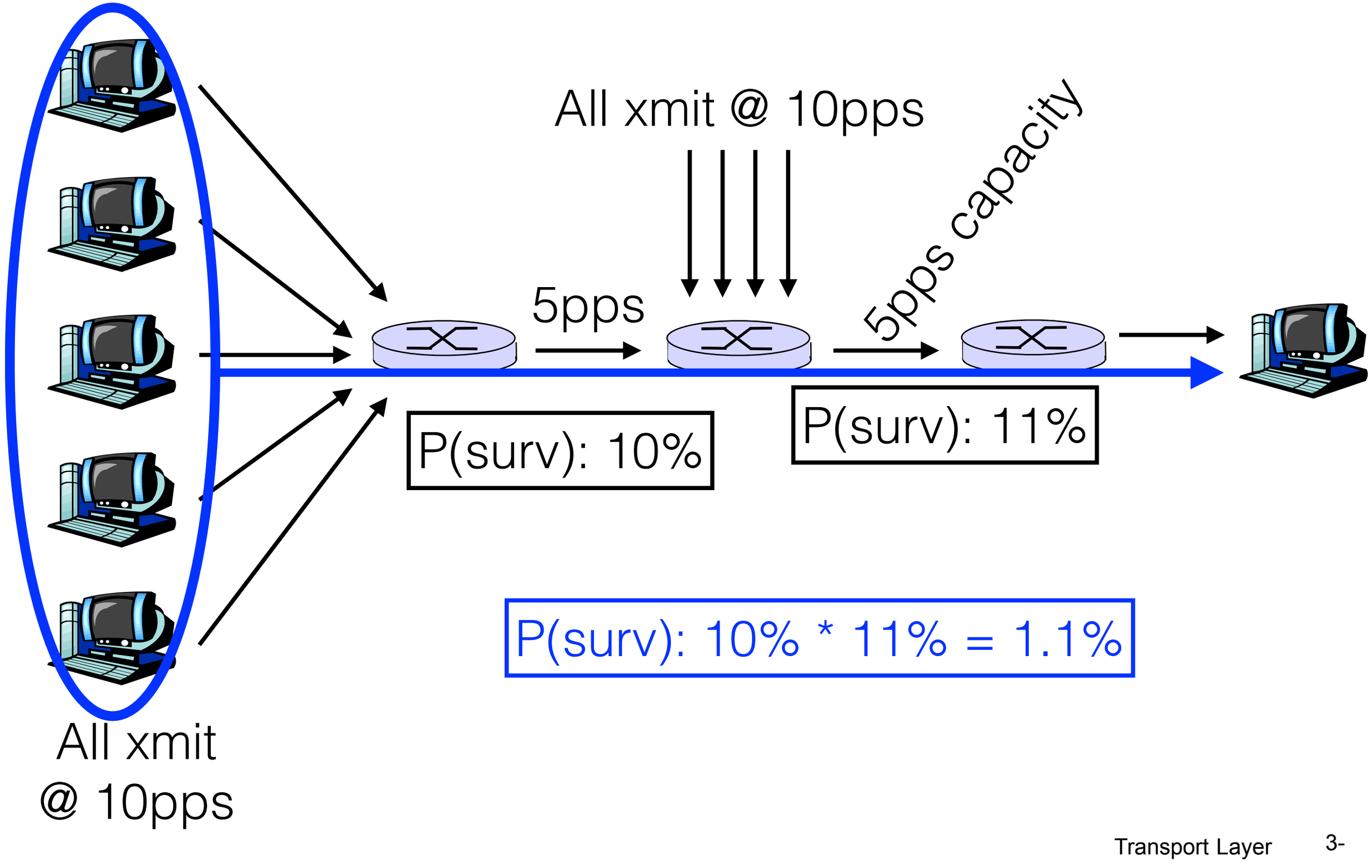
Congestion: Multiple Bottlenecks



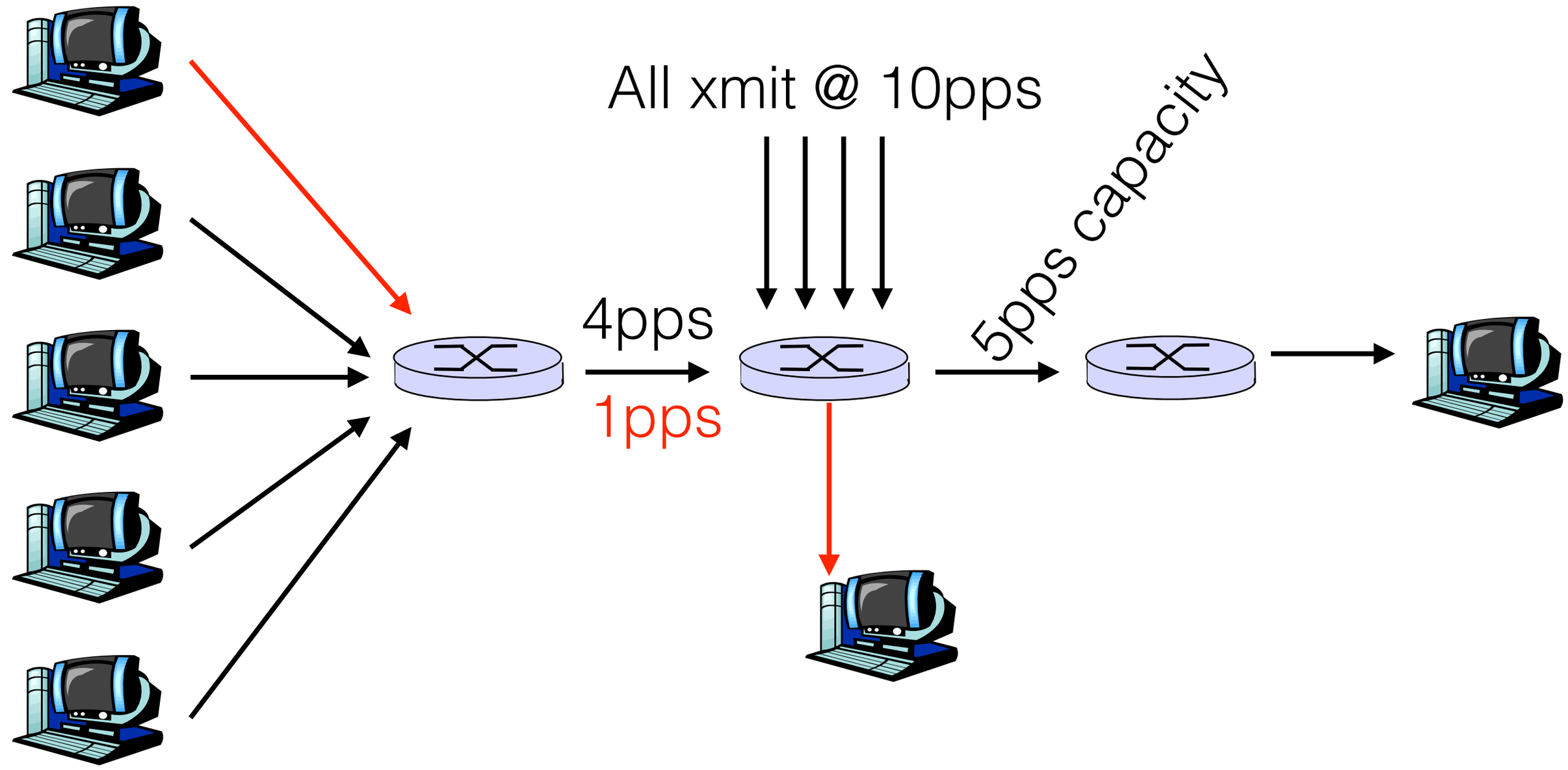
Congestion: Multiple Bottlenecks



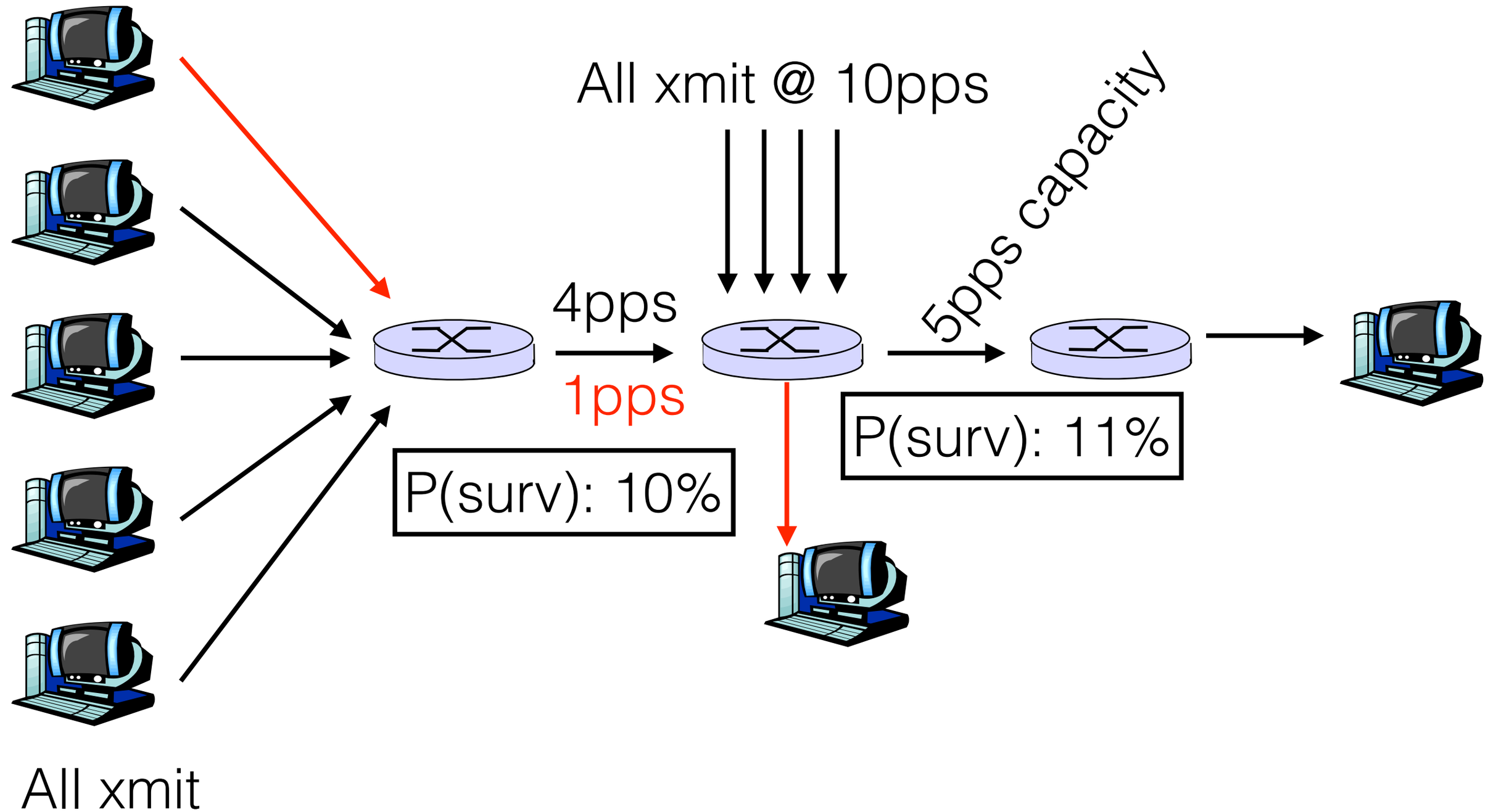
Congestion: Multiple Bottlenecks



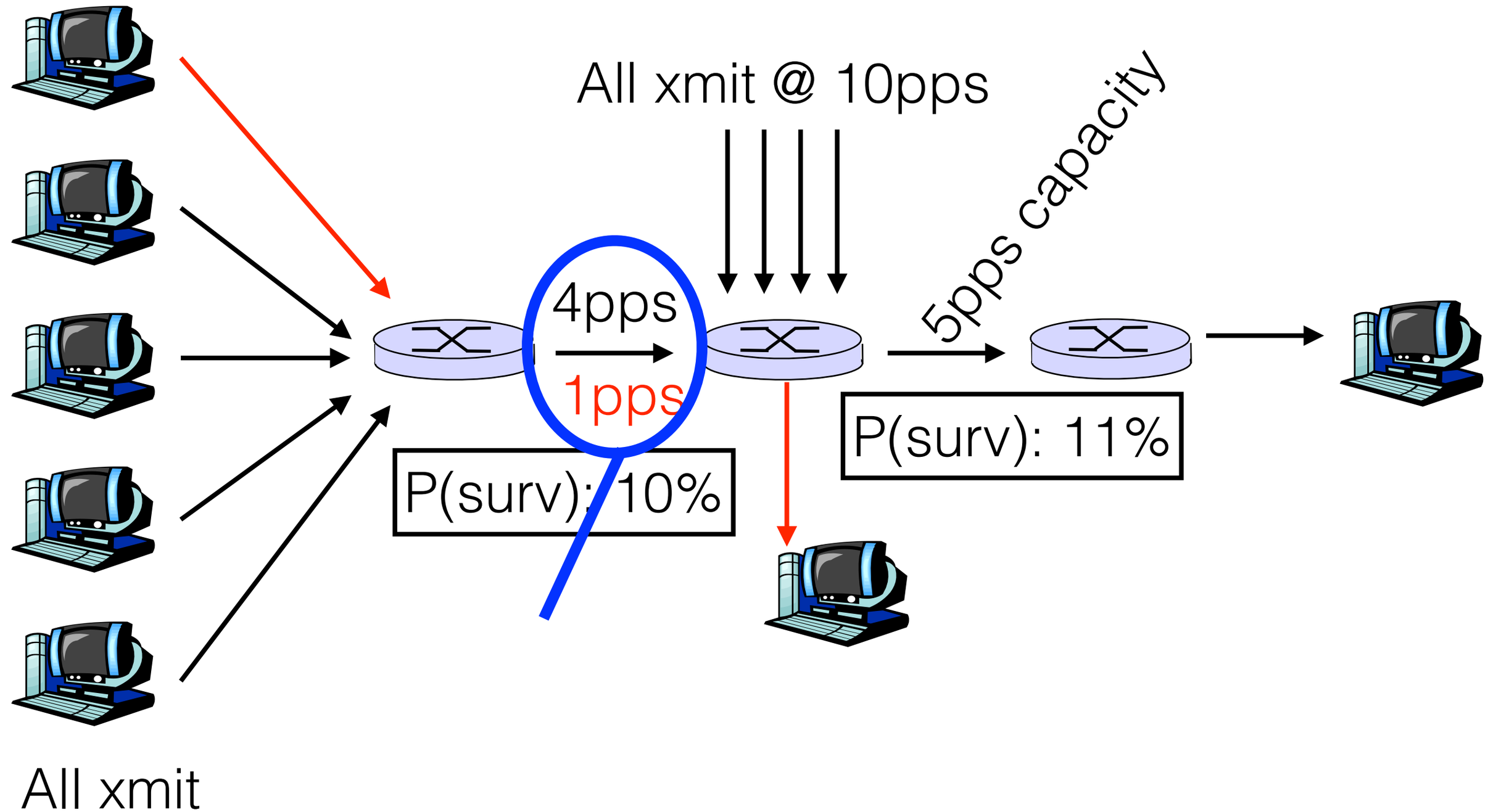
Congestion: Multiple Bottlenecks



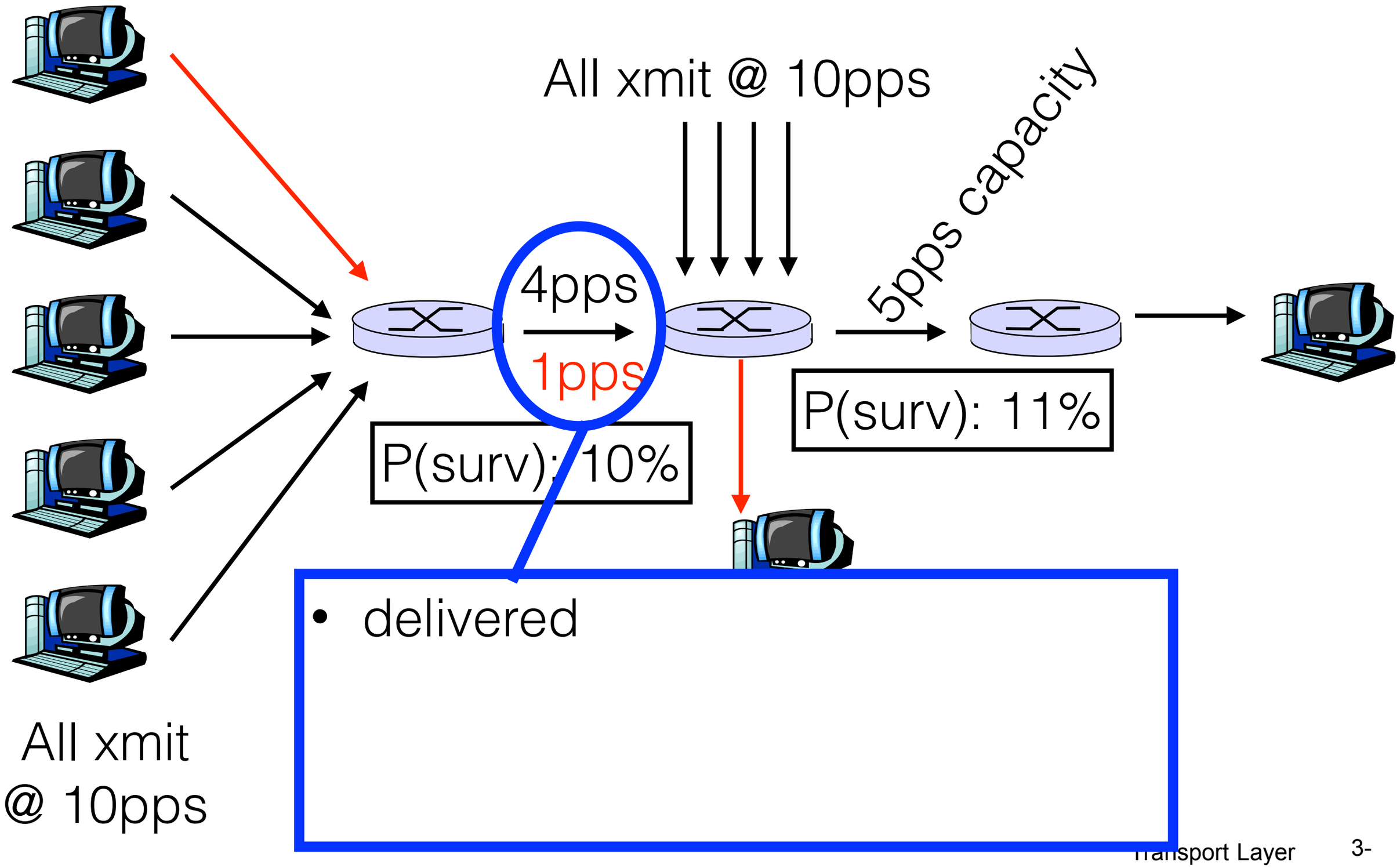
Congestion: Multiple Bottlenecks



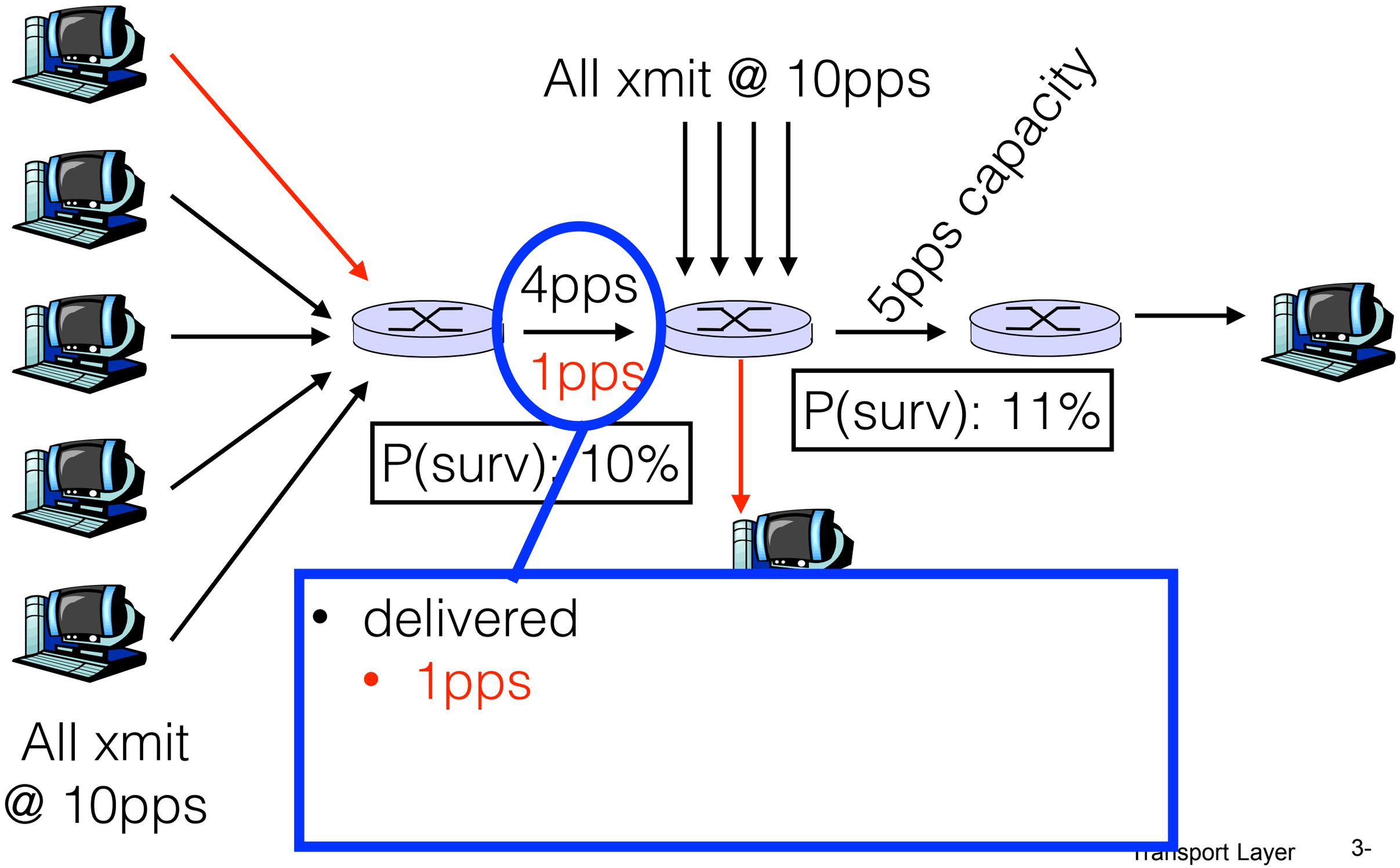
Congestion: Multiple Bottlenecks



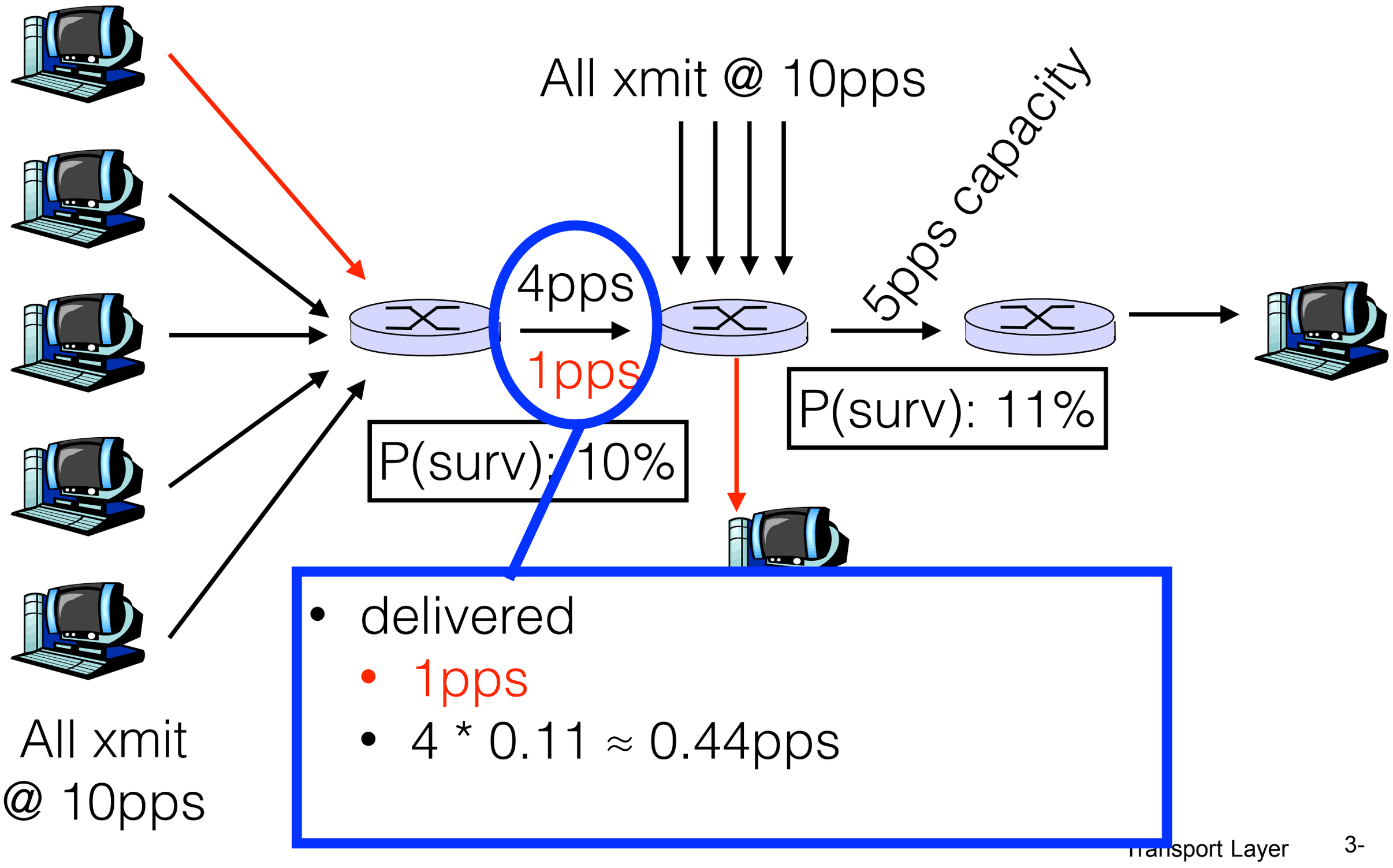
Congestion: Multiple Bottlenecks



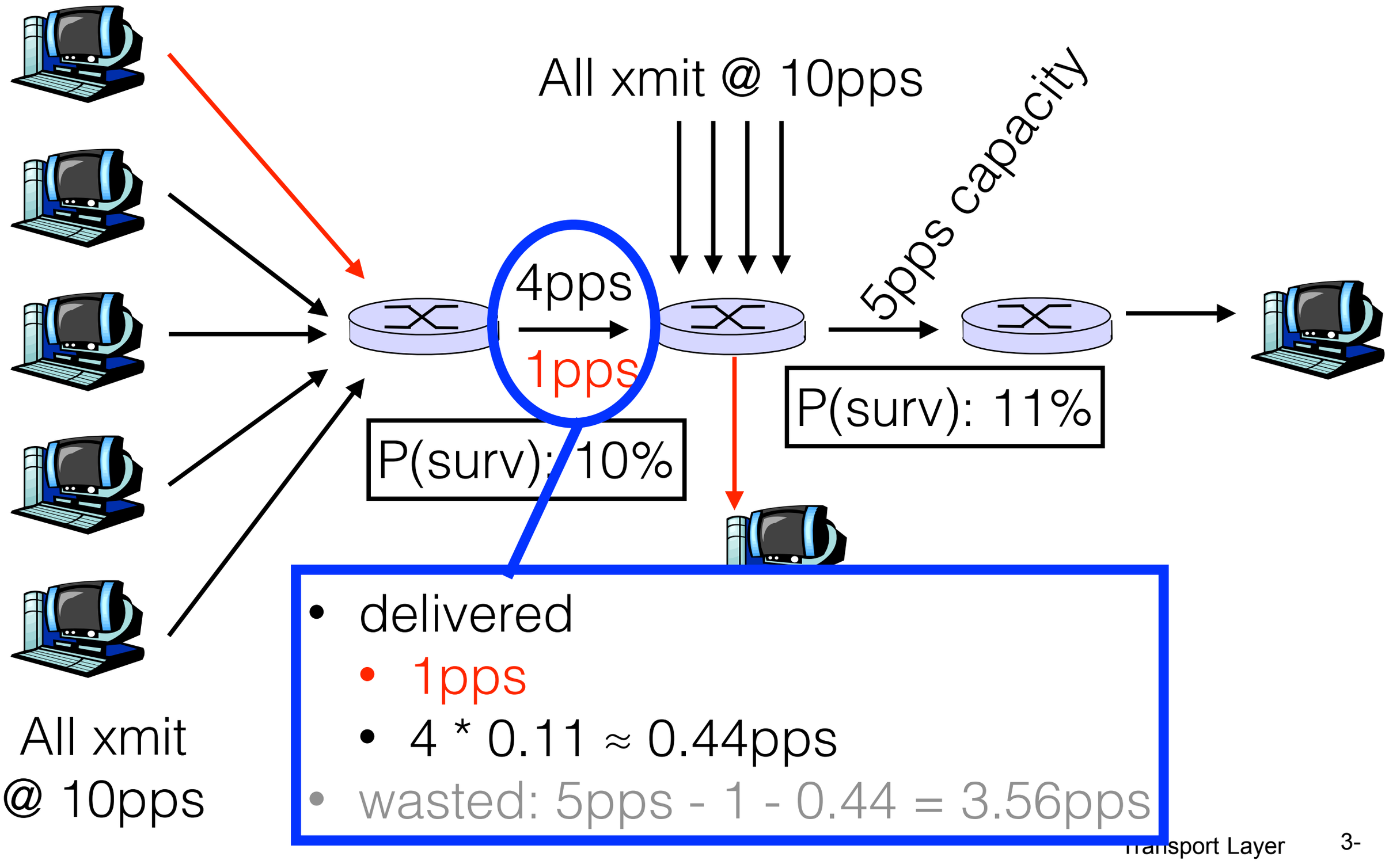
Congestion: Multiple Bottlenecks



Congestion: Multiple Bottlenecks

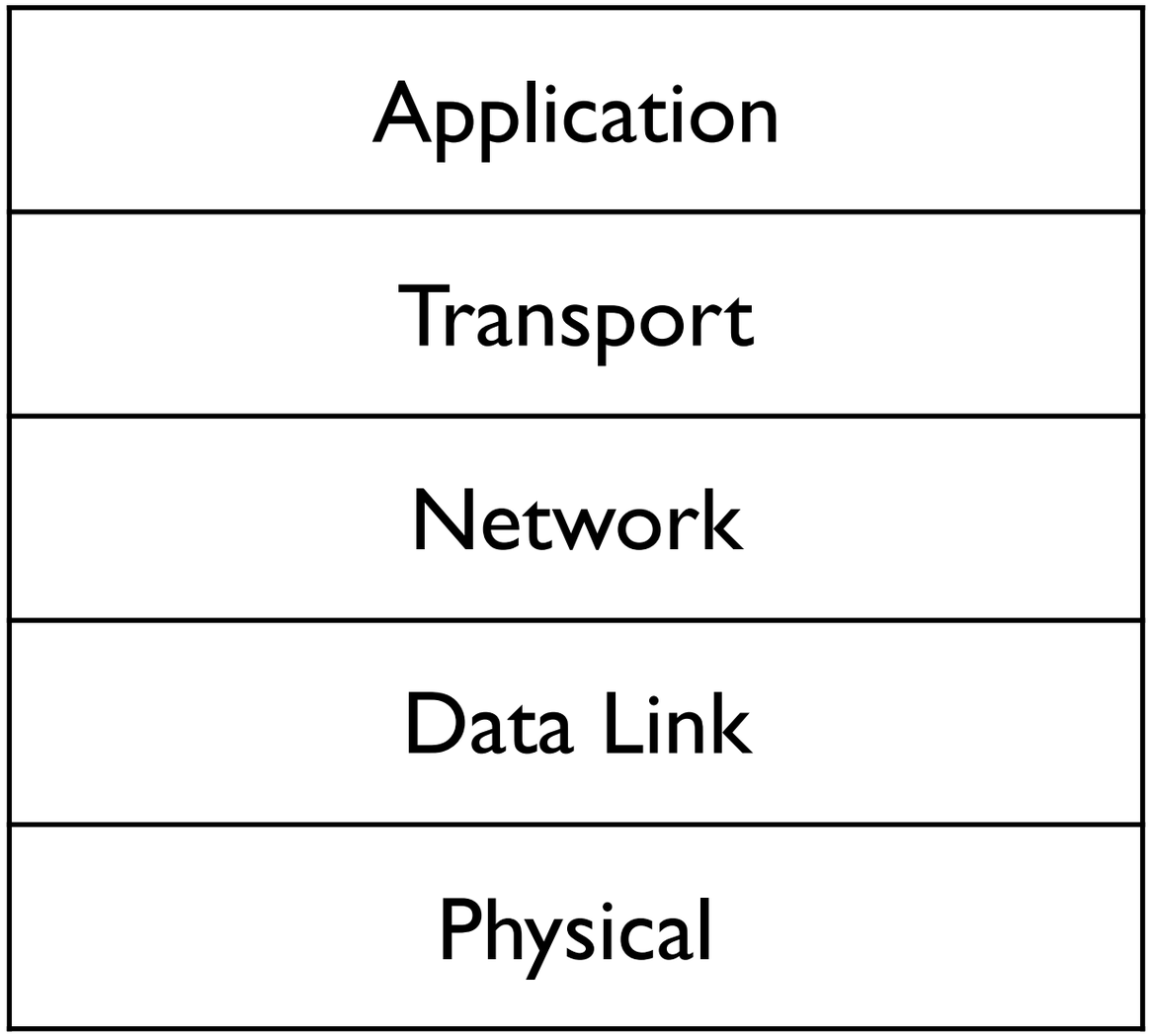


Congestion: Multiple Bottlenecks



Approaches towards congestion control

Approaches towards congestion control



Approaches towards congestion control

Approaches towards congestion control

- Fundamental problem …

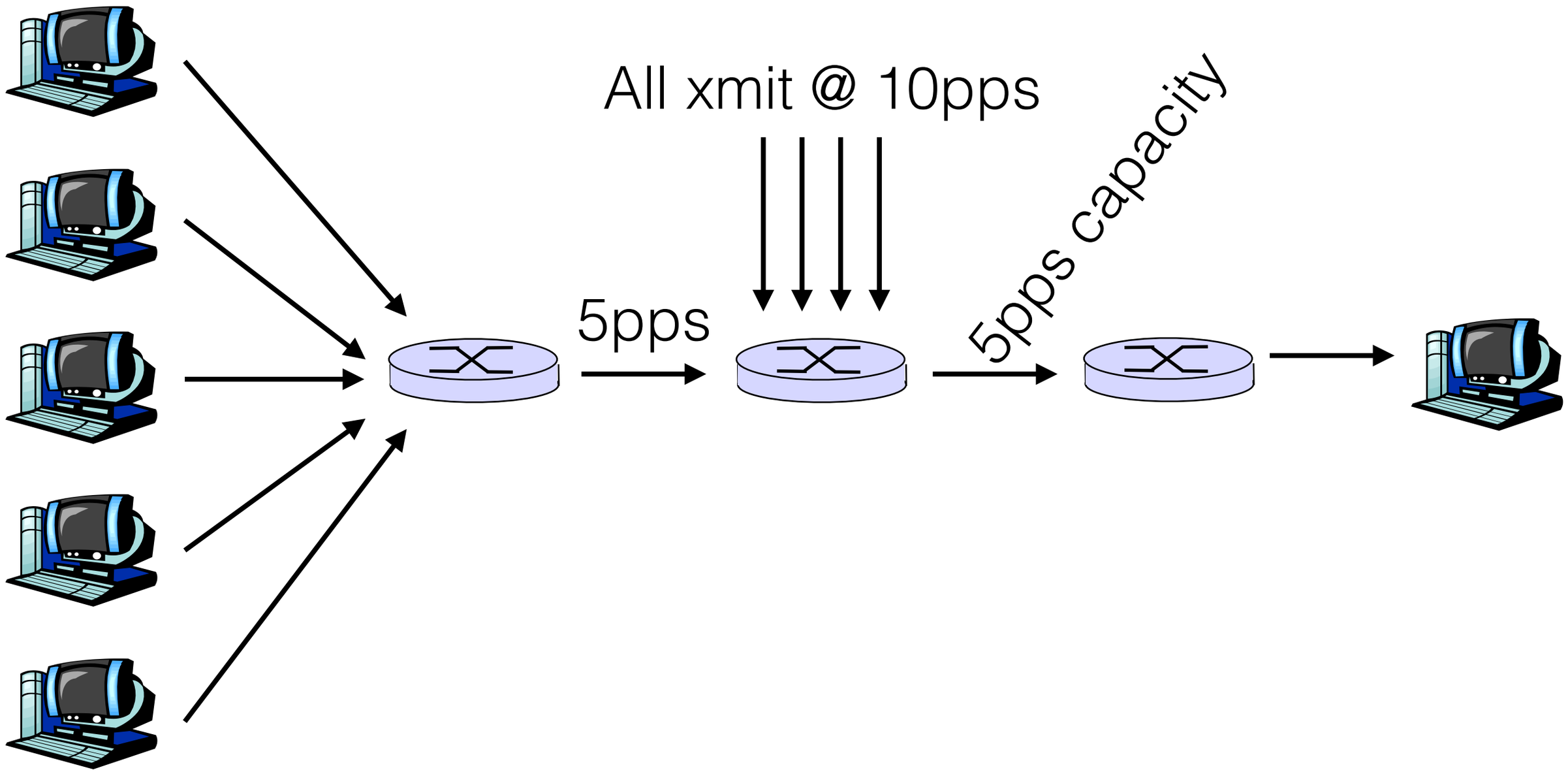
Approaches towards congestion control

- Fundamental problem …
  - congestion happens at a location that has no control over the imposed load

Approaches towards congestion control

- Fundamental problem …
  - congestion happens at a location that has no control over the imposed load
  - the devices that impose load have no explicit understanding of congestion

Congestion: Multiple Bottlenecks



Approaches towards congestion control

Approaches towards congestion control

- Fundamental problem …
  - congestion happens at a location that has no control over the imposed load
  - the devices that impose load have no explicit understanding of congestion

Approaches towards congestion control

- Fundamental problem …
  - congestion happens at a location that has no control over the imposed load
  - the devices that impose load have no explicit understanding of congestion
  - rather, congestion is generally caused by aggregate behavior from many hosts
    - so, it's hard to blame one entity
    - and, it's hard for one entity to mitigate

Approaches towards congestion control

- Fundamental problem …
  - congestion happens at a location that has no control over the imposed load
  - the devices that impose load have no explicit understanding of congestion
  - rather, congestion is generally caused by aggregate behavior from many hosts
    - so, it's hard to blame one entity
    - and, it's hard for one entity to mitigate

- "tragedy of the commons" problem

Approaches towards congestion control

Approaches towards congestion control

Two broad approaches towards congestion control:

end-end congestion control:

- no explicit feedback from network
- congestion inferred from end-system observed loss, delay
- approach taken by TCP

Approaches towards congestion control

Two broad approaches towards congestion control:

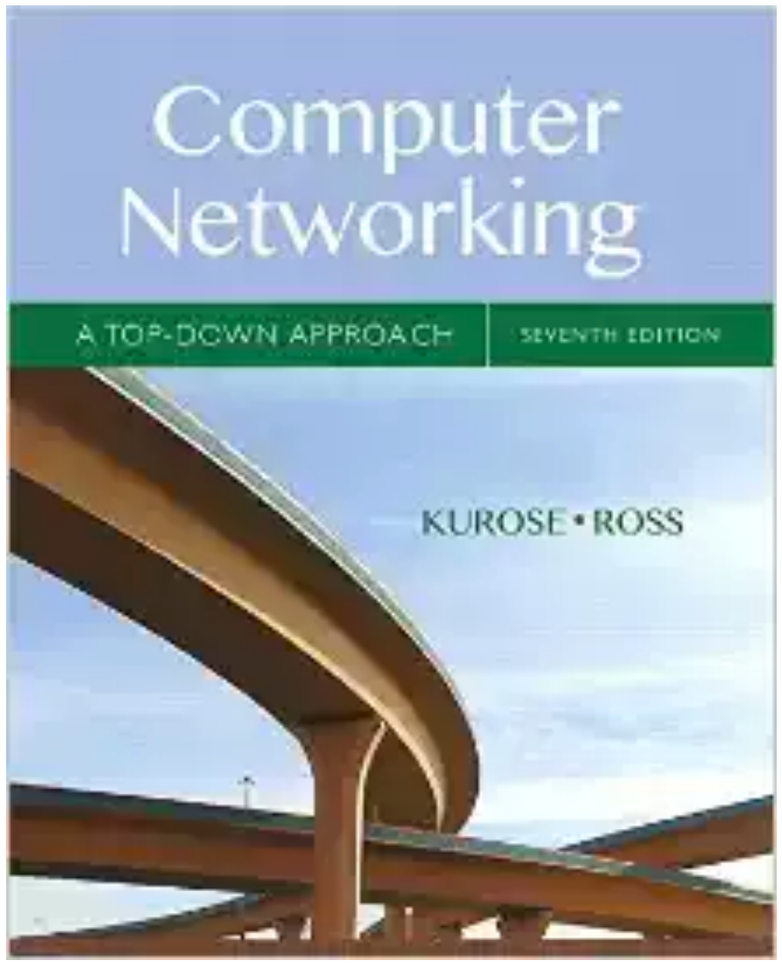
end-end congestion control:

- no explicit feedback from network
- congestion inferred from end-system observed loss, delay
- approach taken by TCP

network-assisted congestion control:

- routers provide feedback to end systems
  - single bit indicating congestion (SNA, DECbit, TCP/IP ECN, ATM)
  - explicit rate sender should send at

Reading Along …



- 3.7: TCP Congestion Control

TCP Congestion Control

TCP Congestion Control

- RFCs 3390, 3517, 3782, 5681, etc.

TCP Congestion Control

- RFCs 3390, 3517, 3782, 5681, etc.
- Sender defines another window:
  - the "congestion window" or cwnd
  - limits amount of unACKed data

TCP Congestion Control

- RFCs 3390, 3517, 3782, 5681, etc.
- Sender defines another window:
  - the "congestion window" or cwnd
  - limits amount of unACKed data
  - strictly limited by the flow control window

TCP Congestion Control

- RFCs 3390, 3517, 3782, 5681, etc.
- Sender defines another window:
  - the "congestion window" or cwnd
  - limits amount of unACKed data
  - strictly limited by the flow control window
  - dynamic
  - adjusts based on the estimated state of the network

TCP Congestion Control

- RFCs 3390, 3517, 3782, 5681, etc.
- Sender defines another window:
  - the "congestion window" or cwnd
  - limits amount of unACKed data
  - strictly limited by the flow control window
  - dynamic
  - adjusts based on the estimated state of the network
  - cwnd is sender state and not transmitted to the receiver

TCP Congestion Control

TCP Congestion Control

- Assume loss means the network is congested
  - not always a good assumption

TCP Congestion Control

- Additive increase, multiplicative decrease (AIMD)
  - when no congestion detected, increase the rate (cwnd) linearly
  - when congestion detected, decrease the rate (cwnd) by half

TCP Congestion Control: details

- sender limits transmission:

LastByteSent-LastByteAcked

≤ cwnd

- roughly,

$$\text{rate} = \frac{\text{cwnd}}{\text{RTT}} \text{ Bytes/sec}$$

- **cwnd** is dynamic, function of perceived network congestion

How does sender perceive congestion?

- loss event = timeout or 3 duplicate acks

Getting Started

- OK, so, send at most cwnd bytes

- But, how do we get started?

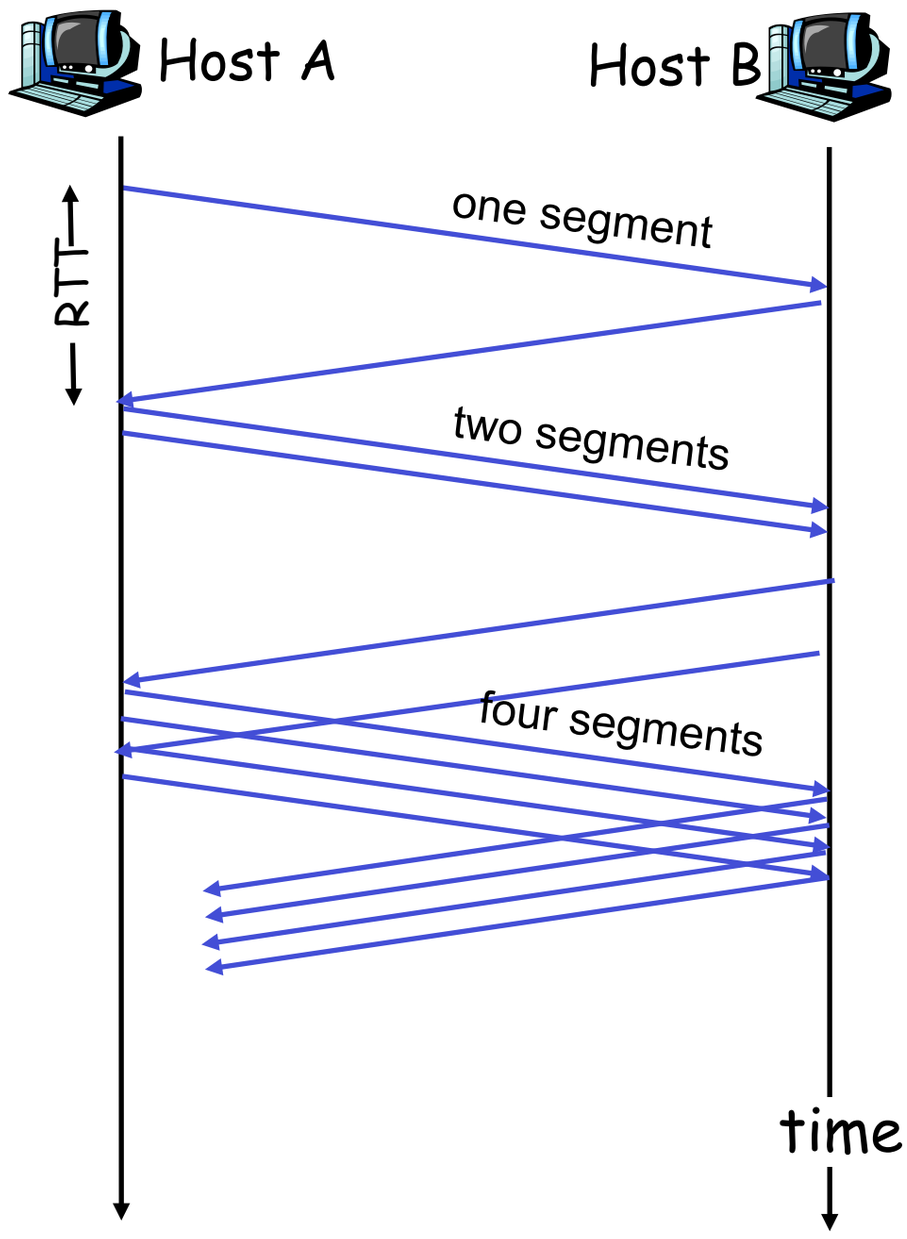
TCP Slow Start

TCP Slow Start

- when connection begins, increase rate exponentially until first loss event:
  - initially **cwnd** = 1 MSS
  - double **cwnd** every RTT
  - done by incrementing **cwnd** for every ACK received
- summary: initial rate is slow but ramps up exponentially fast

TCP Slow Start

- when connection begins, increase rate exponentially until first loss event:
  - initially **cwnd** = 1 MSS
  - double **cwnd** every RTT
  - done by incrementing **cwnd** for every ACK received
- summary: initial rate is slow but ramps up exponentially fast



Evolving the Initial cwnd

Evolving the Initial cwnd

- Late 80s: 1 segment (RFC 2001)

Evolving the Initial cwnd

- Late 80s: 1 segment (RFC 2001)
  - why?

Evolving the Initial cwnd

- Late 80s: 1 segment (RFC 2001)
  - why?

- Late 90s: 2-4 segments (RFCs 2414 & 3390)
  - depends on segment size

Evolving the Initial cwnd

- Late 80s: 1 segment (RFC 2001)
  - why?

- Late 90s: 2-4 segments (RFCs 2414 & 3390)
  - depends on segment size
  - why?

Evolving the Initial cwnd

- Late 80s: 1 segment (RFC 2001)
  - why?

- Late 90s: 2-4 segments (RFCs 2414 & 3390)
  - depends on segment size
  - why?

- 2013: 10 segments (RFC 6928)

Evolving the Initial cwnd

- Late 80s: 1 segment (RFC 2001)
  - why?

- Late 90s: 2-4 segments (RFCs 2414 & 3390)
  - depends on segment size
  - why?

- 2013: 10 segments (RFC 6928)
  - why?

Inferring congestion

Inferring congestion

- after 3 dup ACKs:
  - **cwnd** is cut in half
  - window then grows linearly

Inferring congestion

- after 3 dup ACKs:
  - **cwnd** is cut in half
  - window then grows linearly
- but after timeout event:
  - **cwnd** instead set to 1 MSS;
  - window then grows exponentially
  - to a threshold, then grows linearly

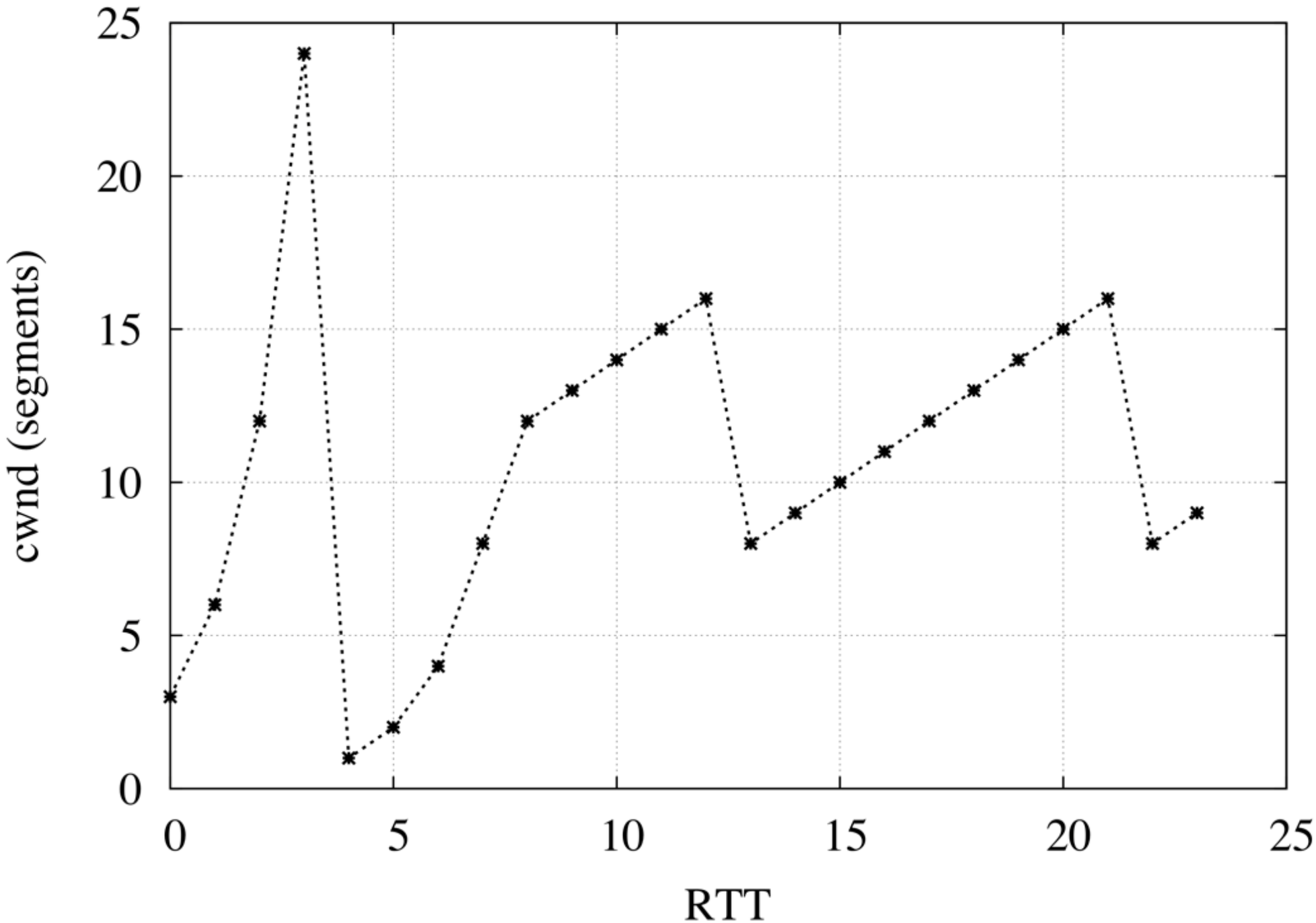
Inferring congestion

- after 3 dup ACKs:
  - **cwnd** is cut in half
  - window then grows linearly
- but after timeout event:
  - **cwnd** instead set to 1 MSS;
  - window then grows exponentially
  - to a threshold, then grows linearly

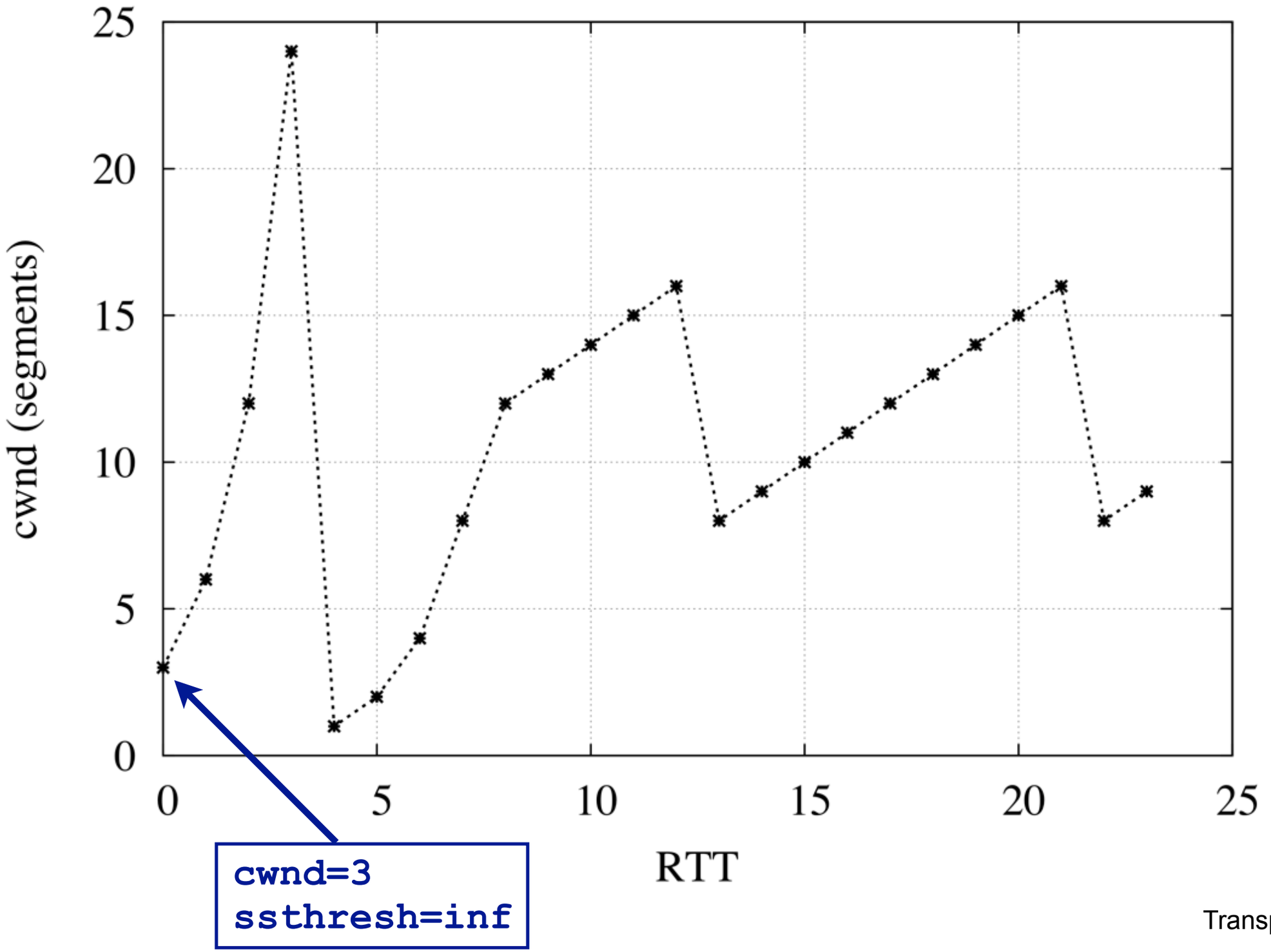
Philosophy:

- 3 dup ACKs indicates network capable of delivering some segments
- timeout indicates a "more alarming" congestion scenario

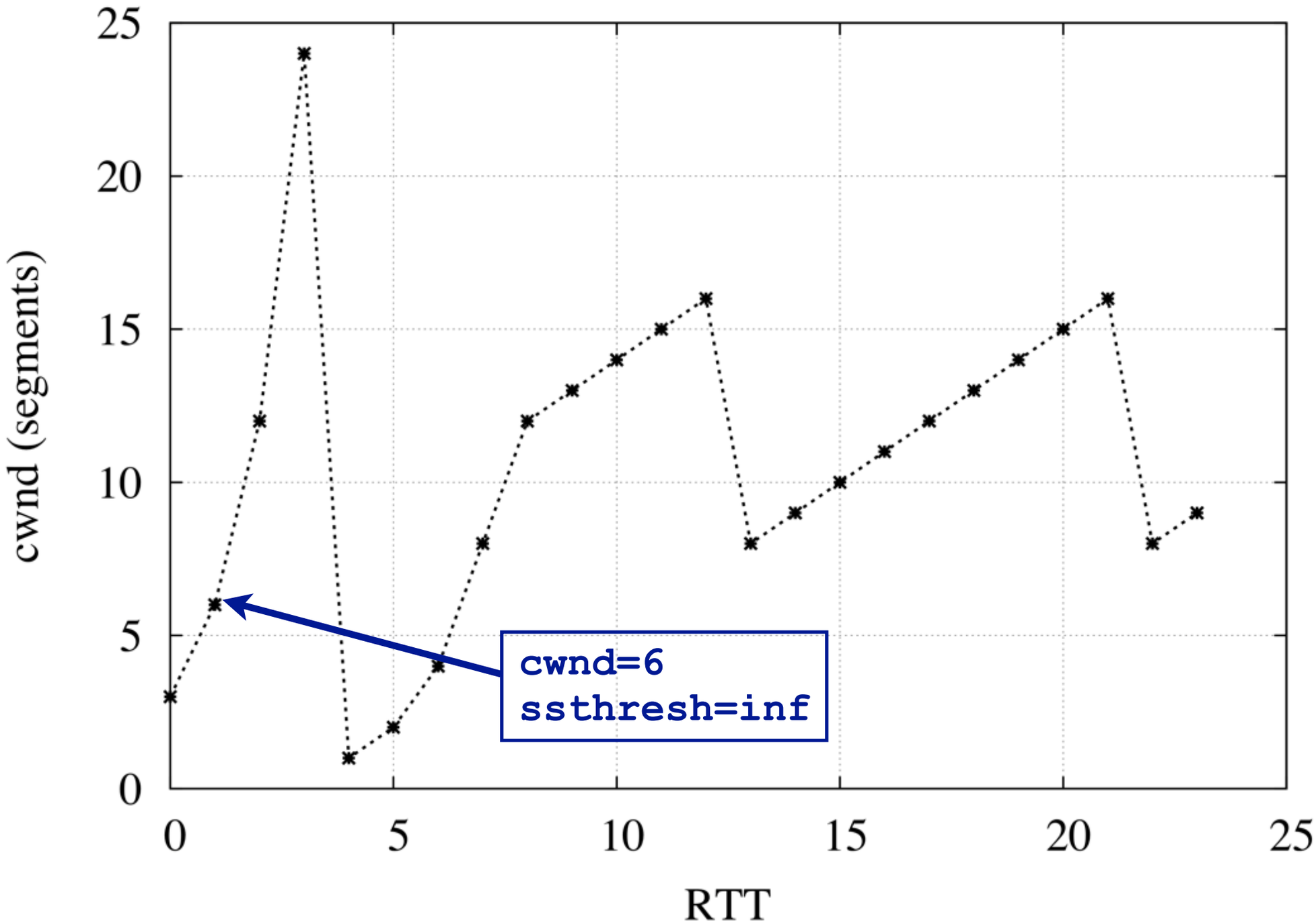
Congestion Window Evolution



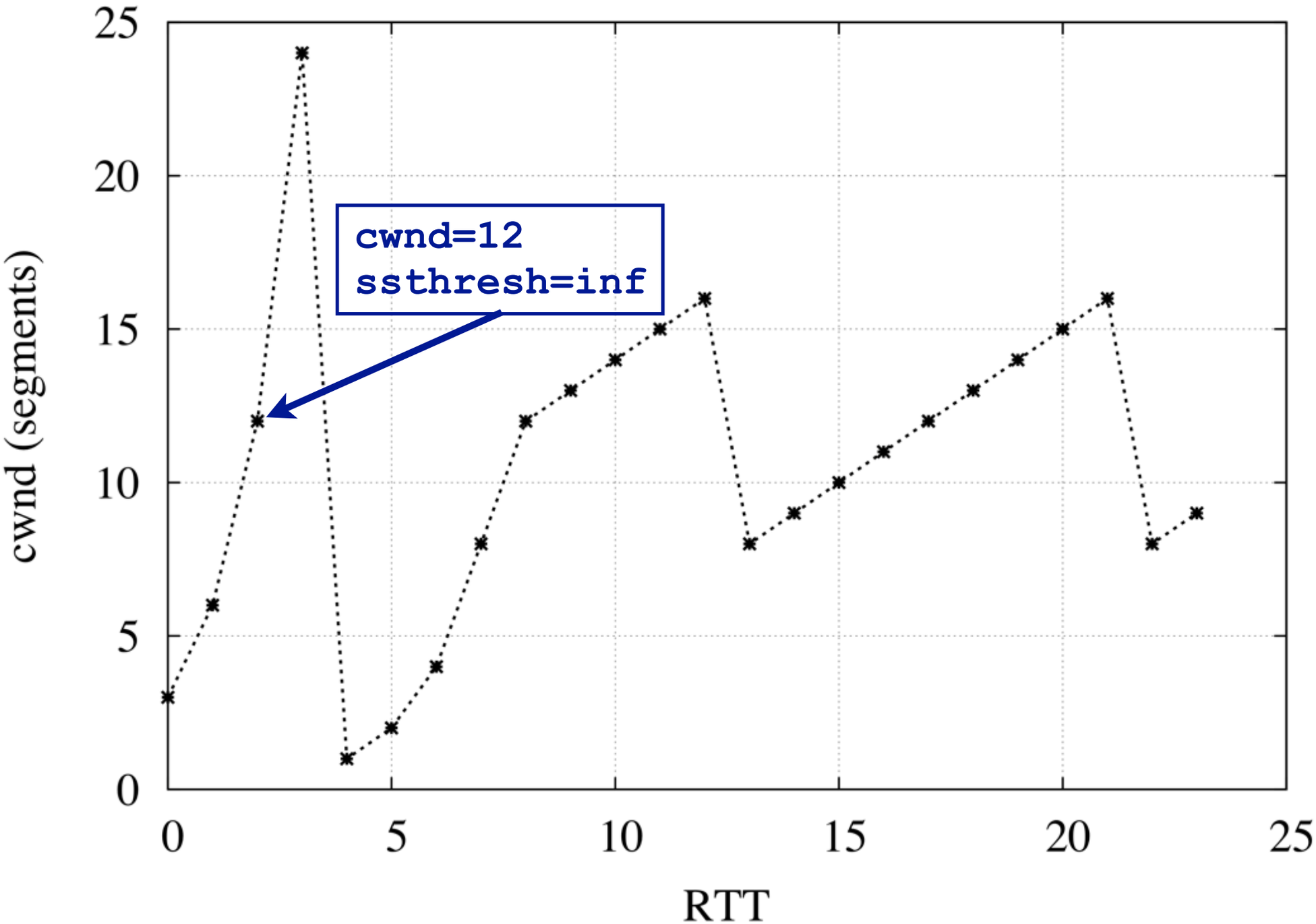
Congestion Window Evolution



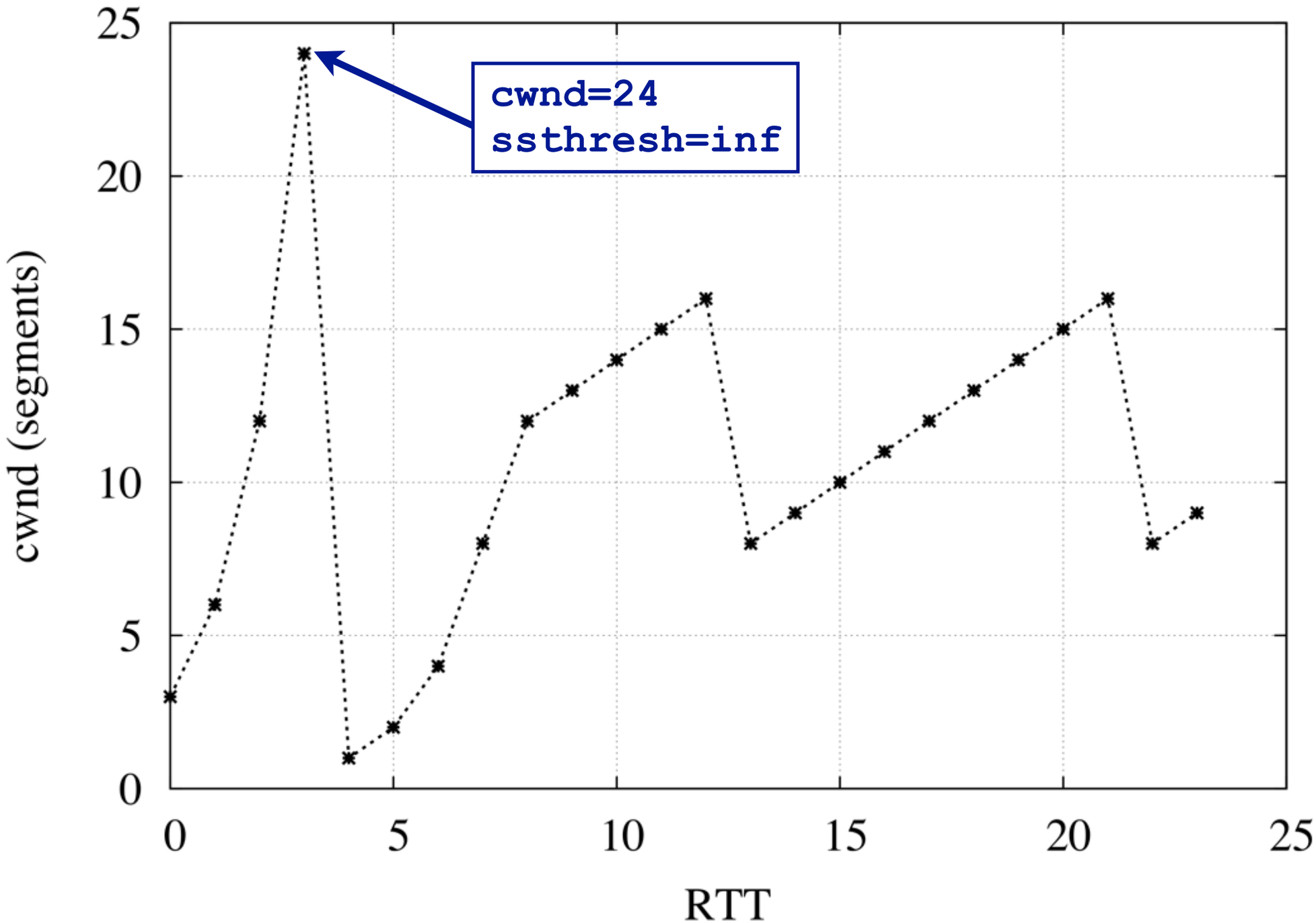
Congestion Window Evolution



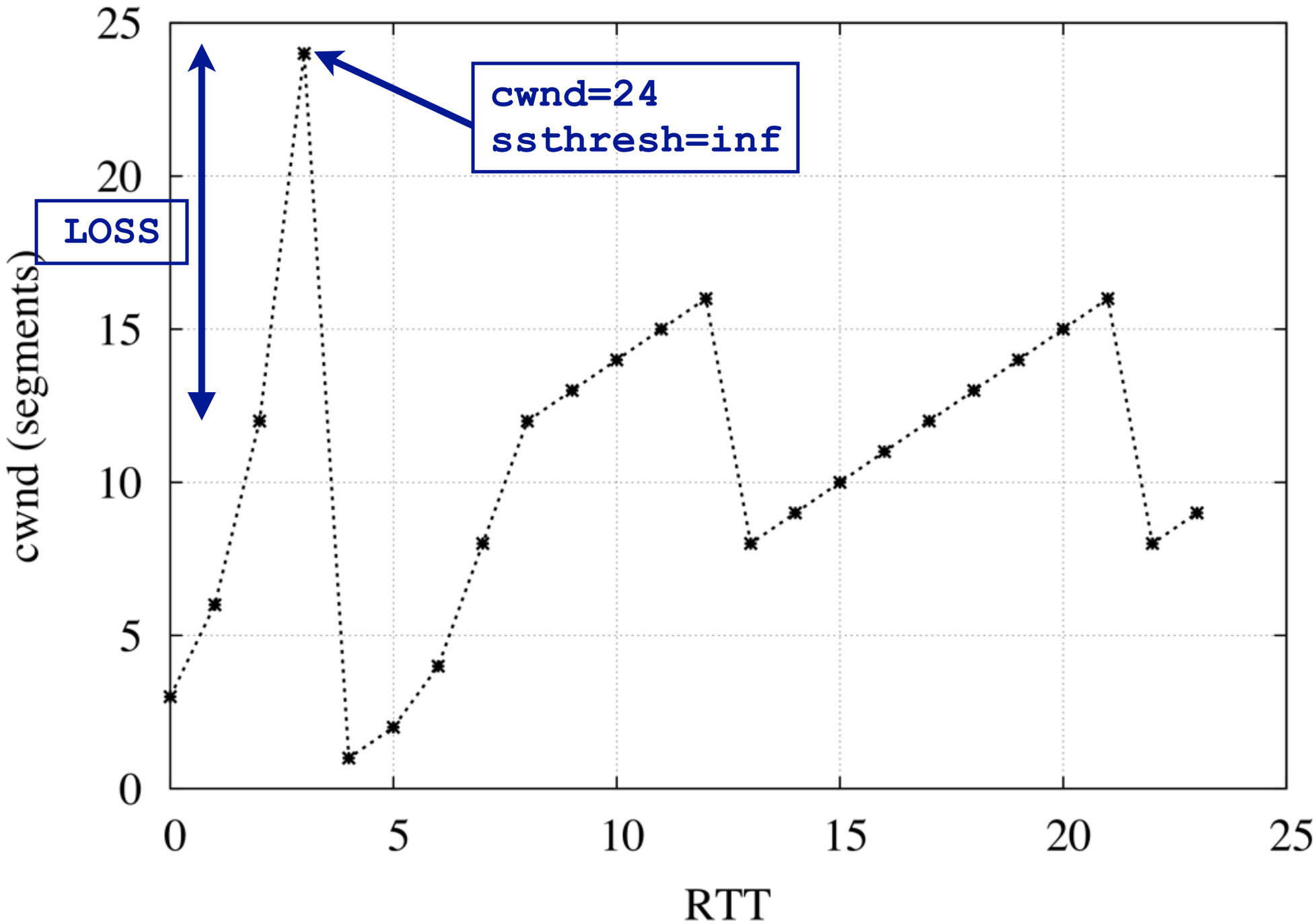
Congestion Window Evolution



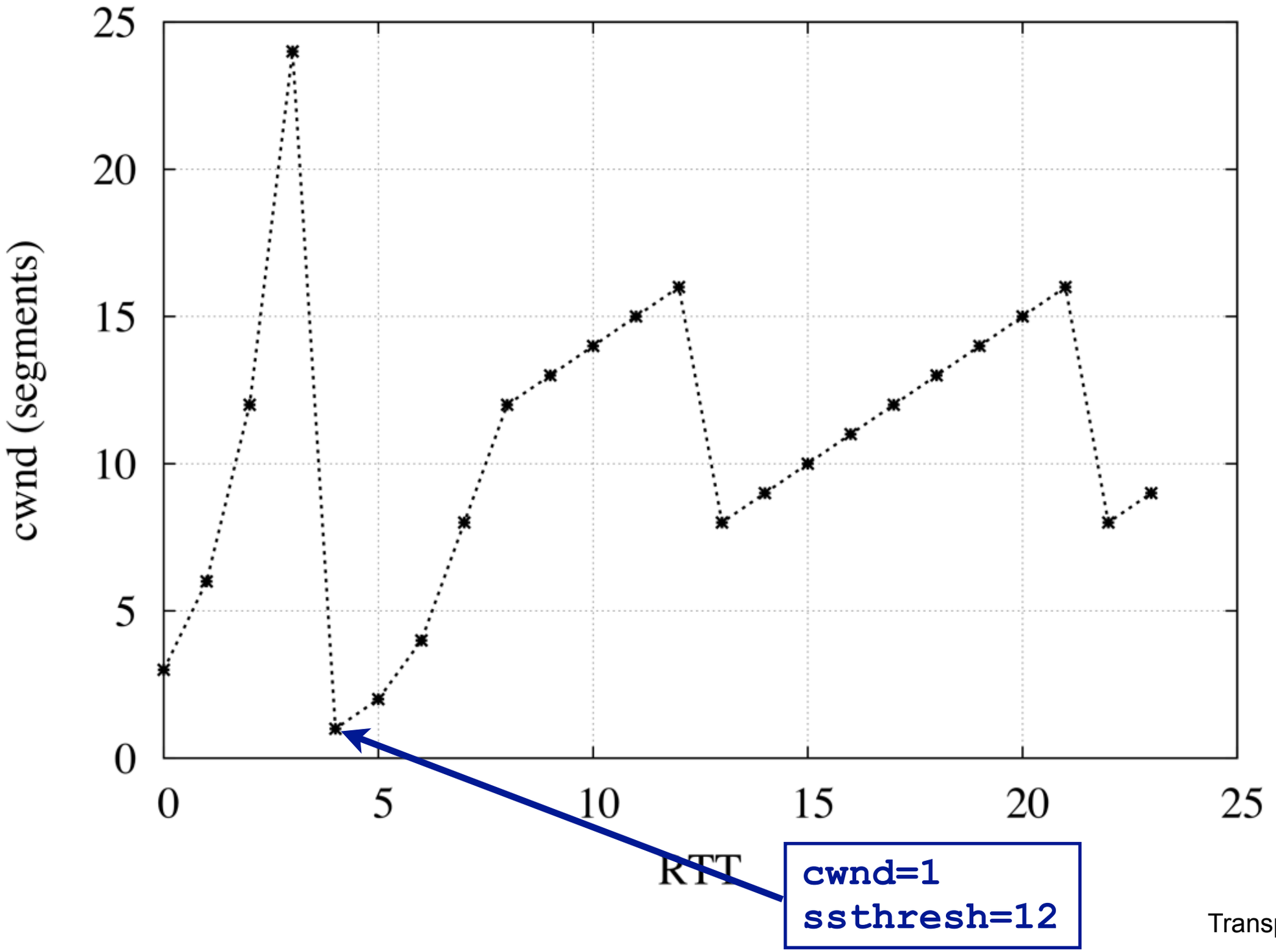
Congestion Window Evolution



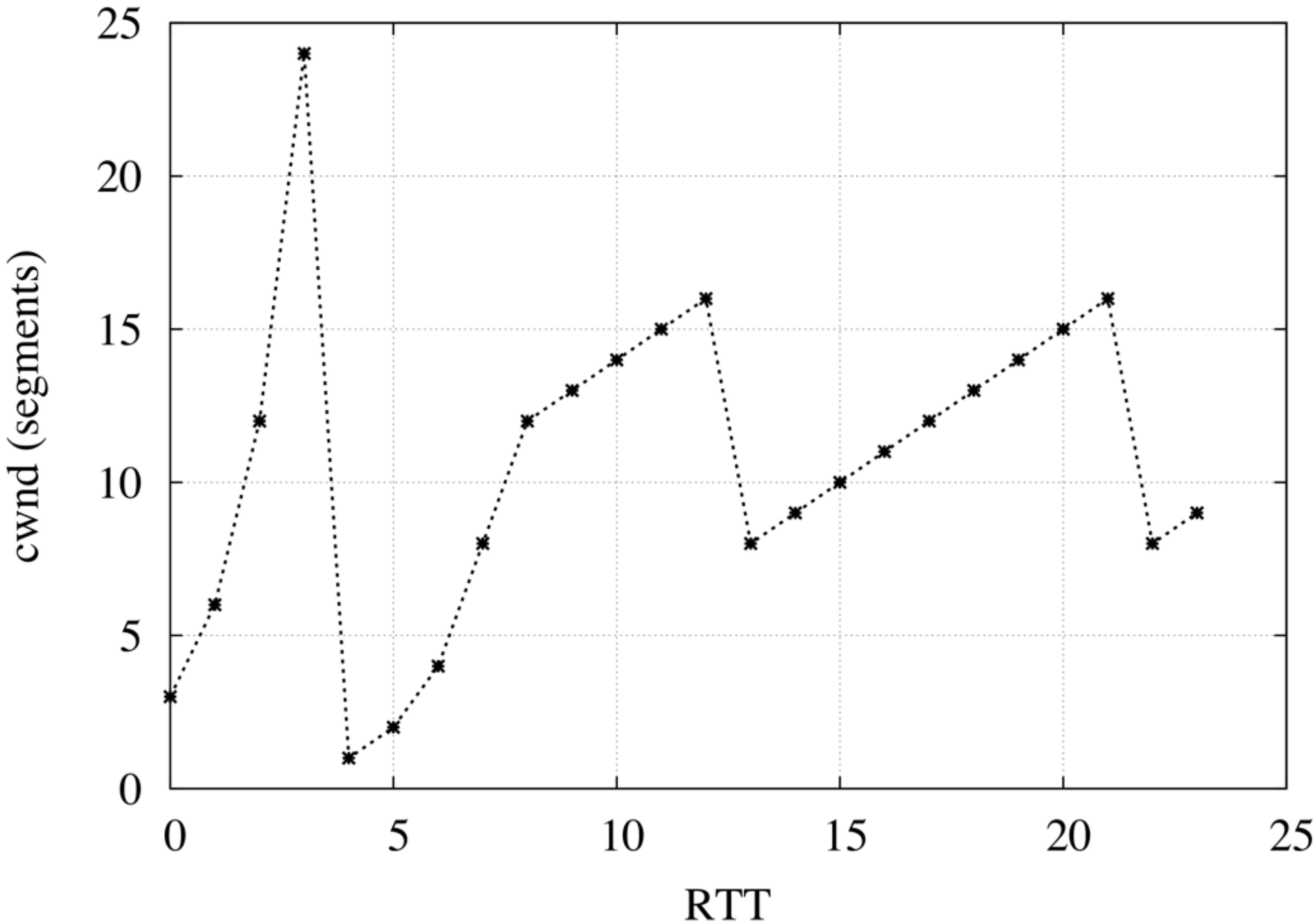
Congestion Window Evolution



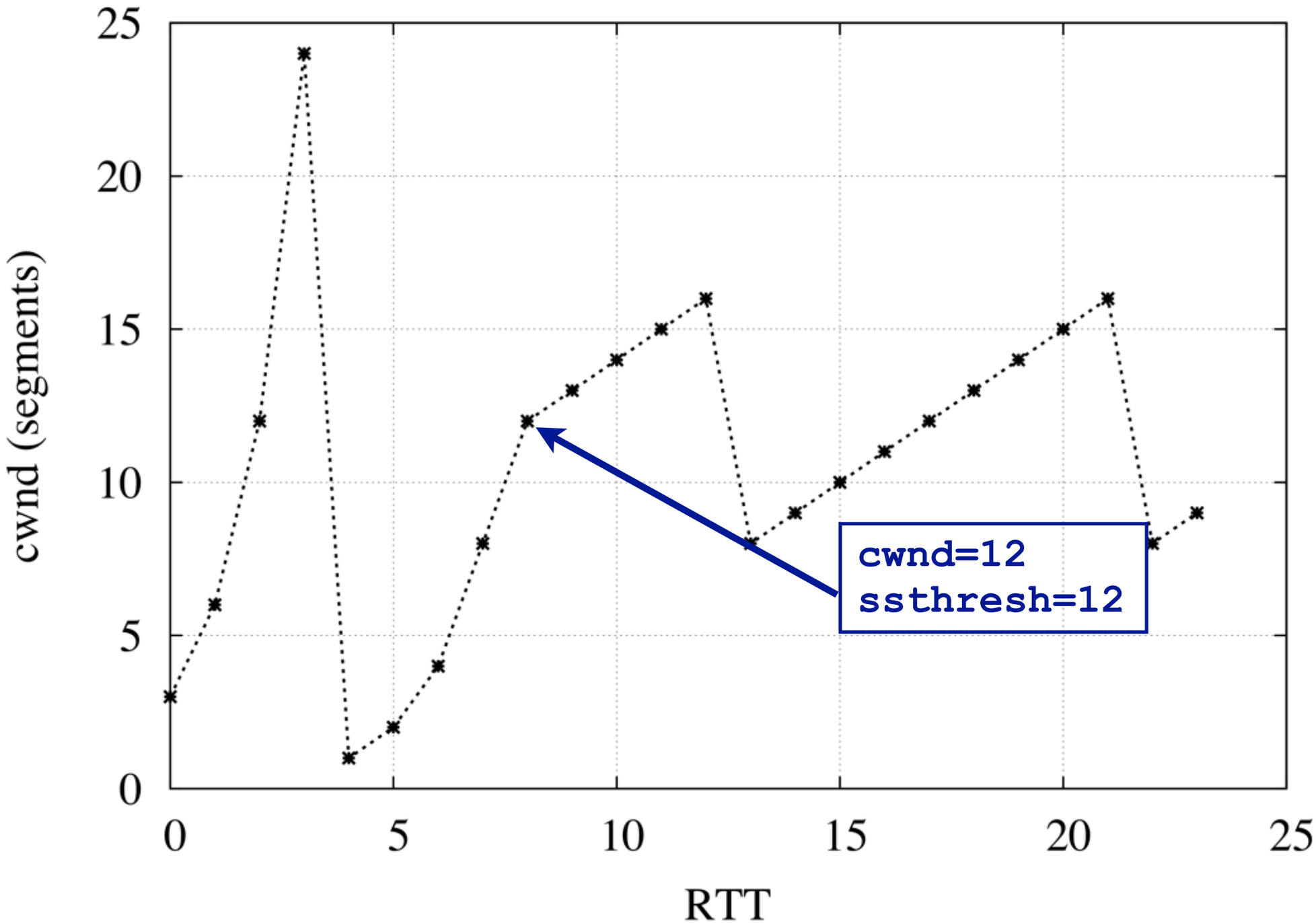
Congestion Window Evolution



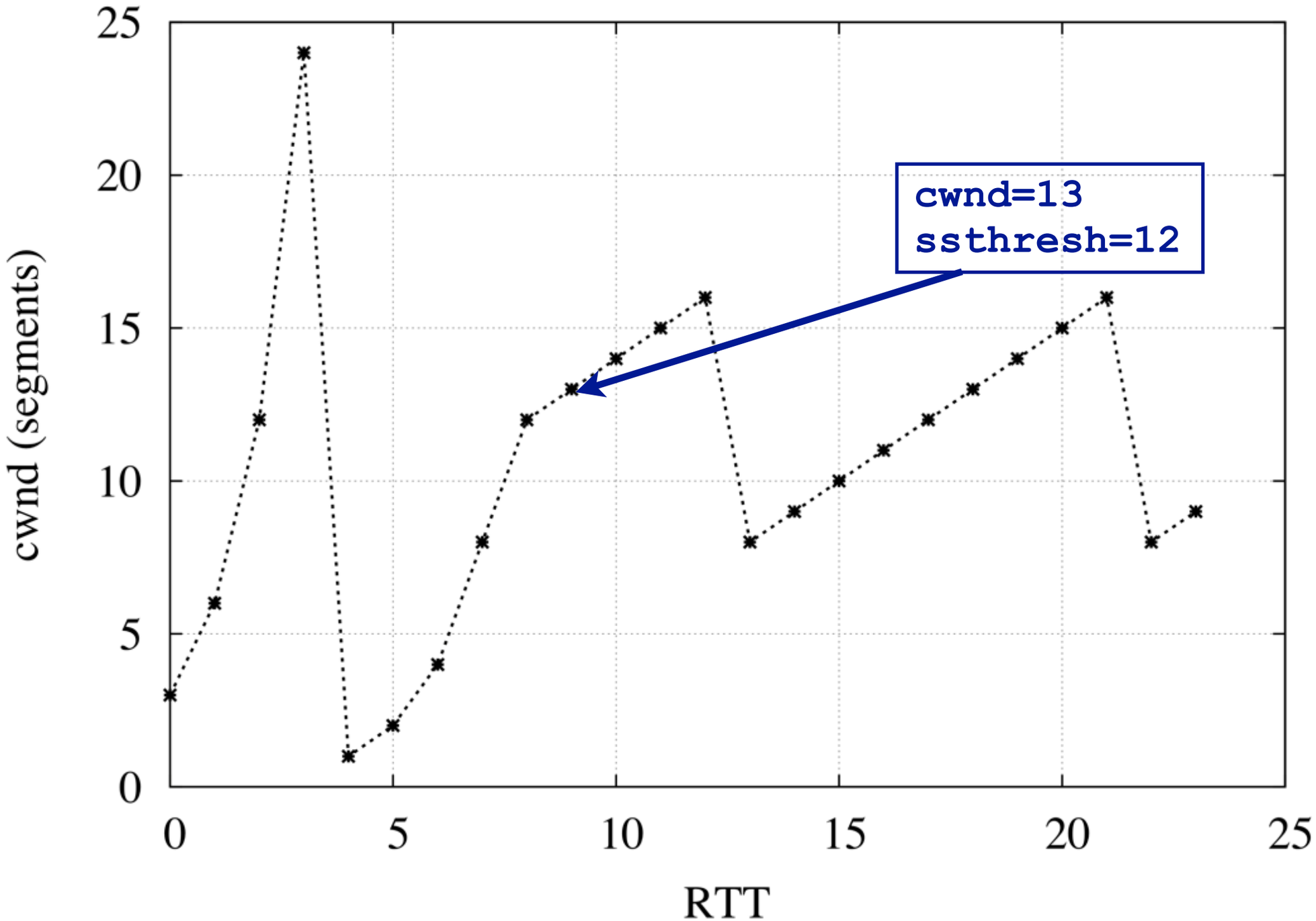
Congestion Window Evolution



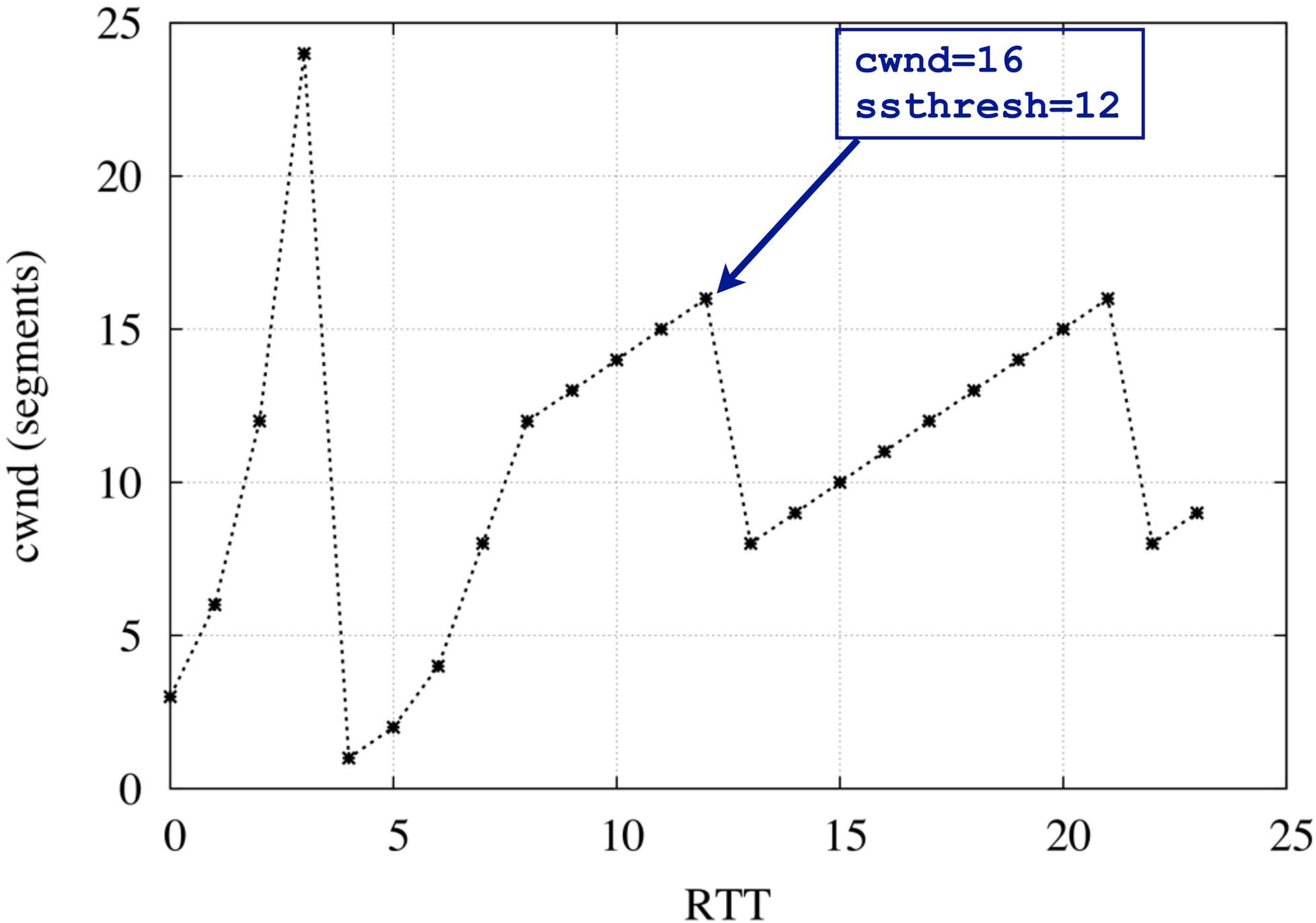
Congestion Window Evolution



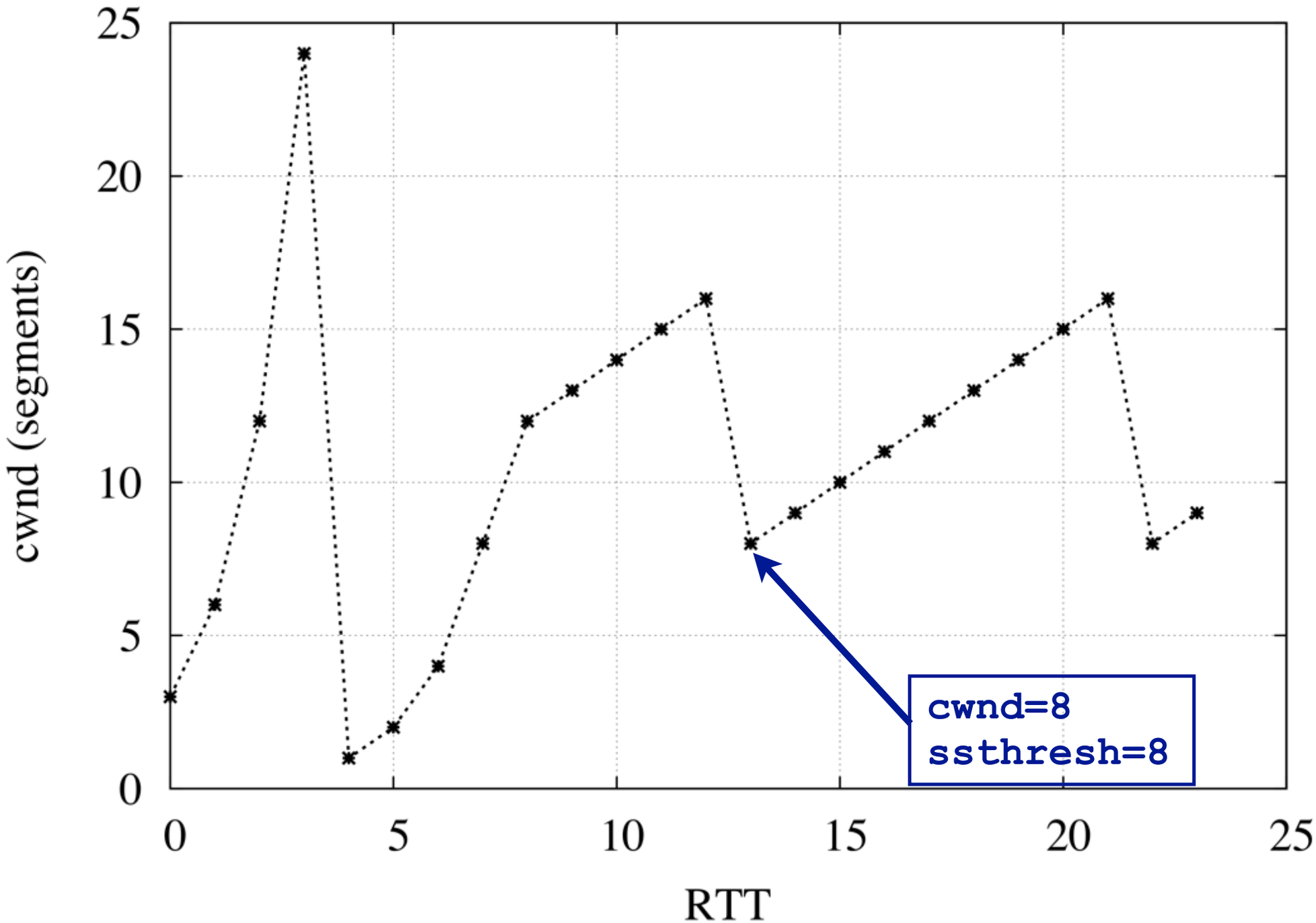
Congestion Window Evolution



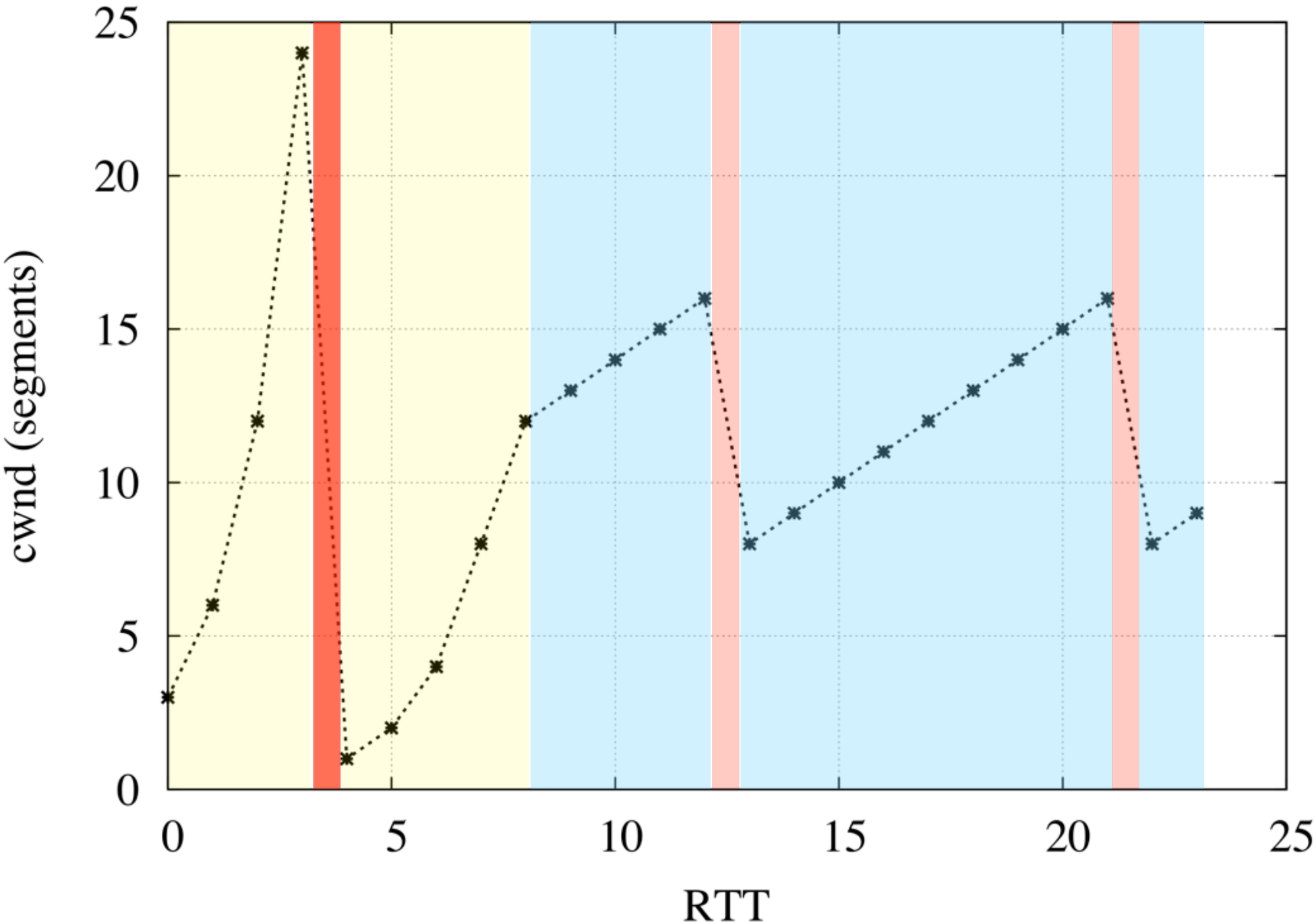
Congestion Window Evolution



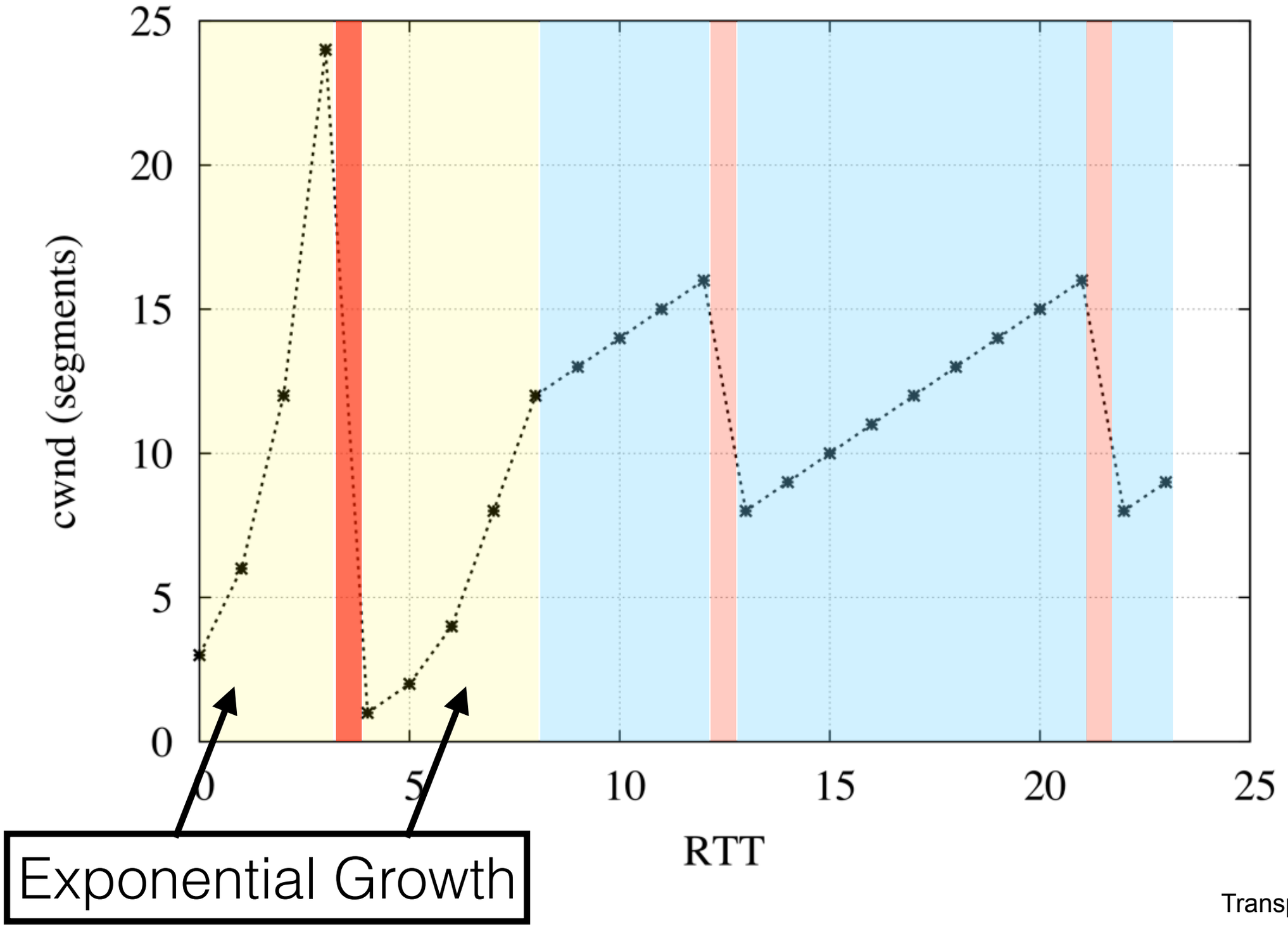
Congestion Window Evolution



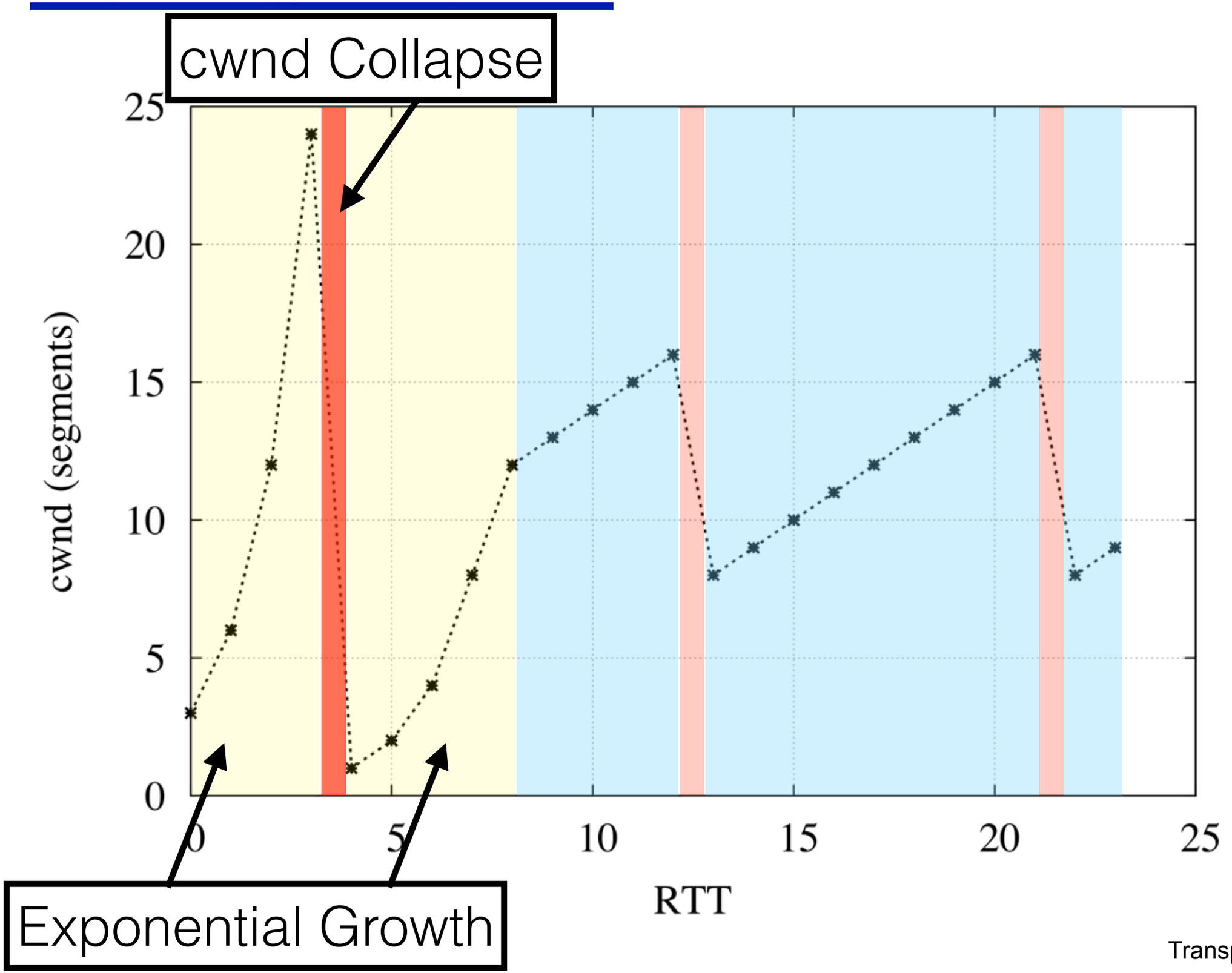
cwnd Evolution



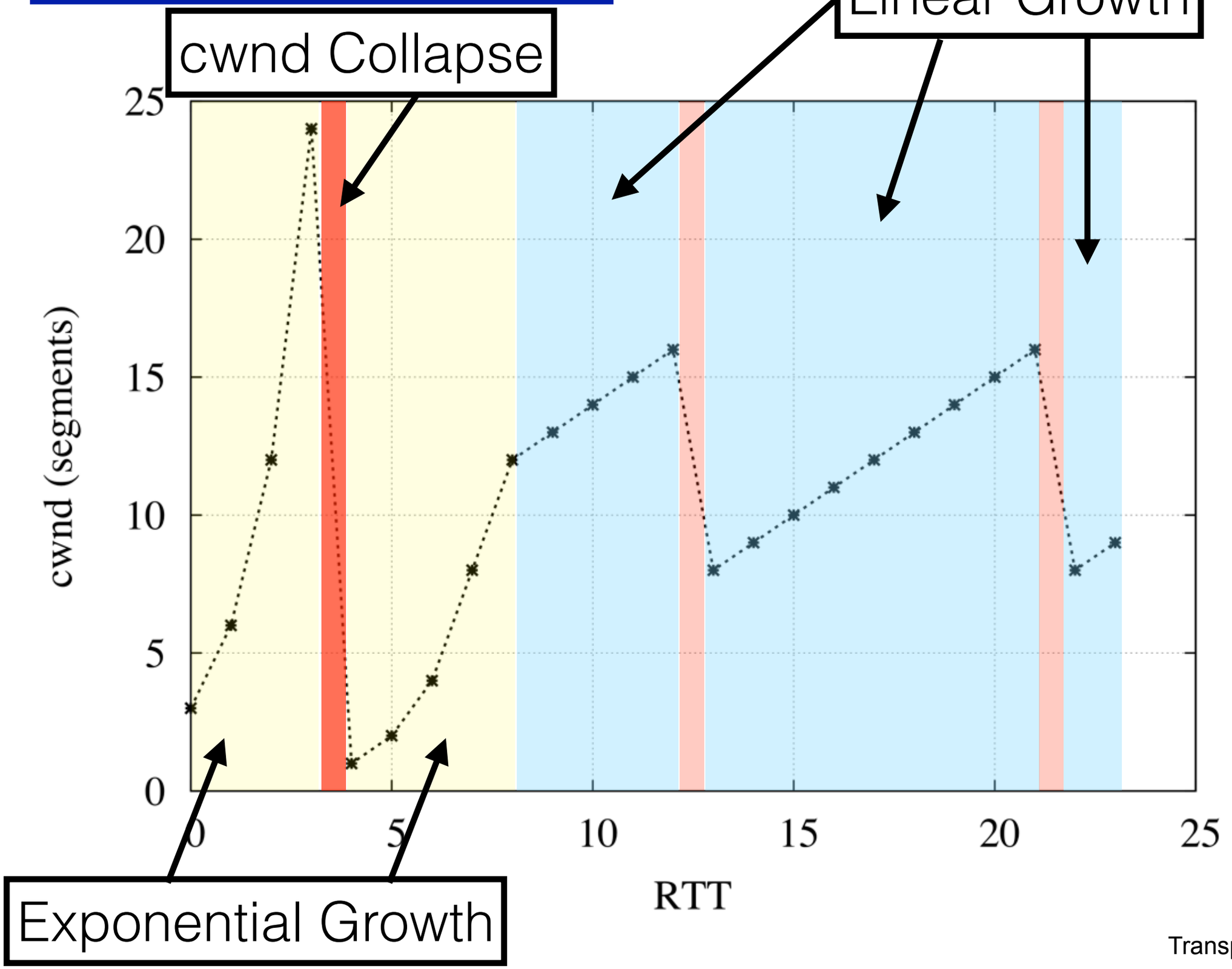
cwnd Evolution



cwnd Evolution



cwnd Evolution



cwnd Evolution