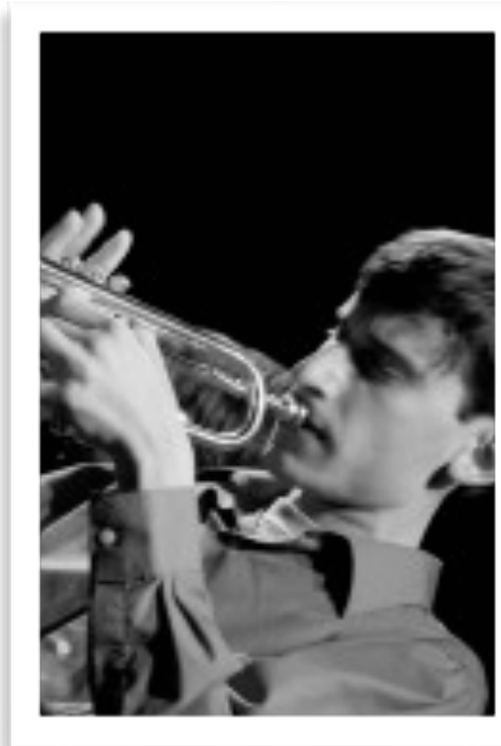
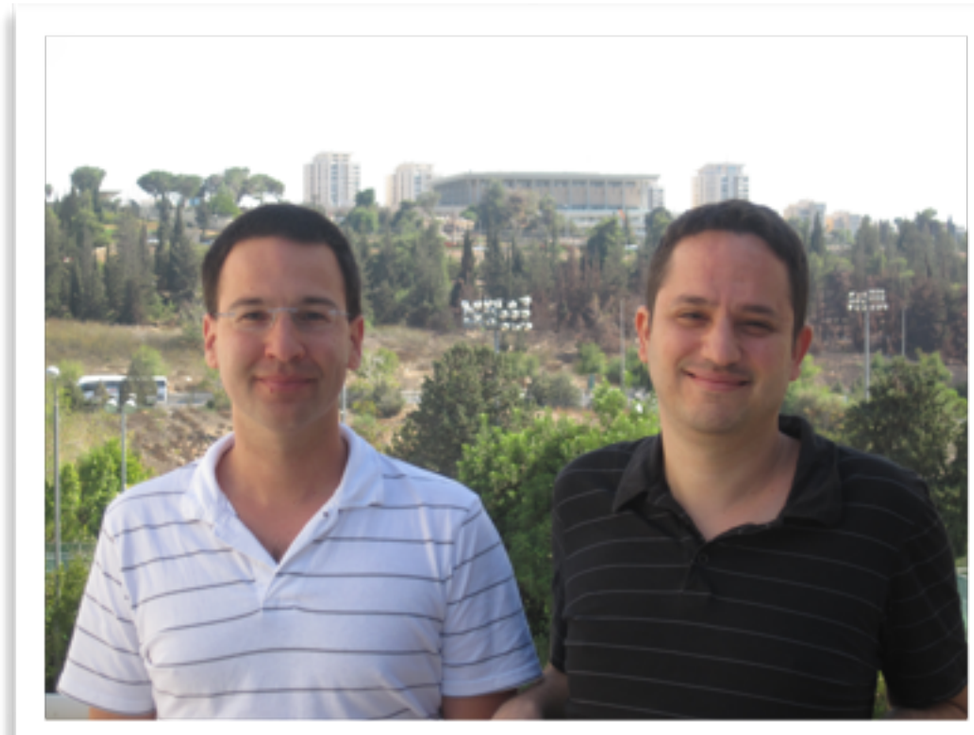
Performance-oriented Congestion Control

Mo Dong, Qingxi Li, Doron Zarchy*,
P. Brighten Godfrey and Michael Schapira*

University of Illinois at Urbana Champaign
*Hebrew University of Jerusalem





Qingxi Li

Brighten Godfrey

University of Illinois at Urbana-Champaign

Doron Zarchy and Michael Schapira

Hebrew University of Jerusalem


TCP
Est. 1988

High BDP	Wireless	Satellite	Inter-DC	Intra-DC
BIC H-TCP Compound CUBIC FAST TCP	Westwood Vegas Veno	Hybla STAR	Illinois SABUL	ICTCP DCTCP

High BDP	Wireless	Satellite	Inter-DC	Intra-DC
BIC H-TCP Compound CUBIC FAST TCP	Westwood Vegas Veno	Hybla STAR	Illinois SABUL	ICTCP DCTCP

•••

High BDP	Wireless	Satellite	Inter-DC	Intra-DC
BIC H-TCP Compound CUBIC FAST TCP	Westwood Vegas Veno	Hybla STAR	Illinois SABUL	ICTCP DCTCP

...

Who can be not happy with TCP ?

C

CP

STAR

CP

are

~~can be~~ not happy with TCP ?

High BDP	Wireless	Satellite	Inter-DC	Intra-DC
BIC H-TCP Compound CUBIC FAST TCP	Westwood Vegas Veno	Hybla STAR	Illinois SABUL	ICTCP DCTCP

High BDP	Wireless	Satellite	Inter-DC	Intra-DC
BIC H-TCP Compound CUBIC FAST TCP	Westwood Vegas Veno	Hybla STAR	Illinois SABUL	ICTCP DCTCP

Point Solutions

High BDP	Wireless	Satellite	Inter-DC	Intra-DC
BIC H-TCP Compound CUBIC FAST TCP	Westwood Vegas Veno	Hybla STAR	Illinois SABUL	ICTCP DCTCP
10X	10X	17X	4X	

Unstable, RTT Unfair, Bufferbloat, Crash on Changing Networks, ...

Point Solutions

+

Performance

Far from Optimal

TCP fails to achieve consistent high performance

Why is it so hard?

Reno	1 pkt loss	cwnd/2
Scalable	ACK	cwnd+1
CUBIC	Time pass 1ms	cwnd+f(t,cwn,rtt)
FAST	RTT increase x%	Reduce cwnd to f(x)%
HTCP	100 ACK	cwnd+f(cwnd)/cwnd

	Event	Action
Reno	1 pkt loss	cwnd/2
Scalable	ACK	cwnd+1
CUBIC	Time pass 1ms	cwnd+f(t,cwn,rtt)
FAST	RTT increase x%	Reduce cwnd to f(x)%
HTCP	100 ACK	cwnd+f(cwnd)/cwnd

Hardwired Mapping

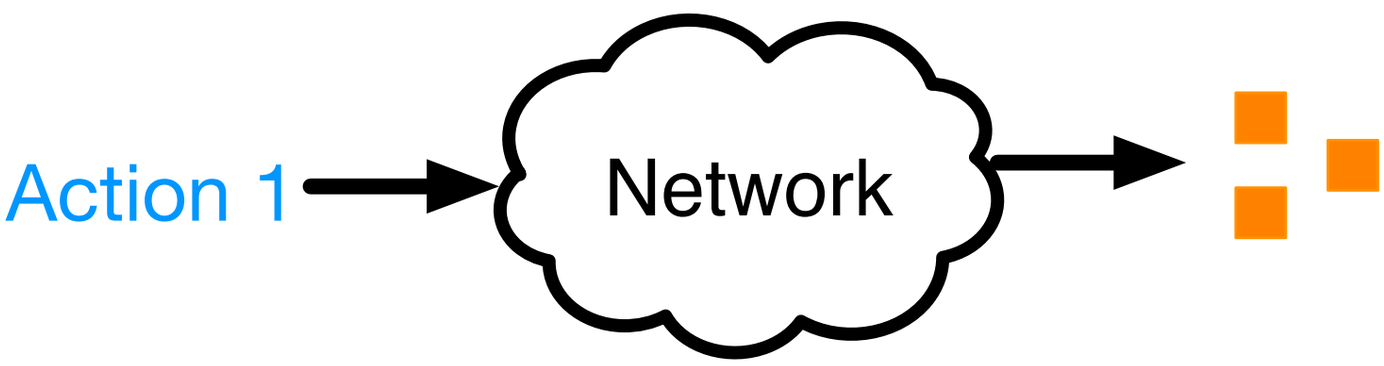
	Event	Action
Reno	1 pkt loss	cwnd/2
Scalable	ACK	cwnd+1
CUBIC	Time pass 1ms	cwnd+f(t,cwn,rtt)
FAST	RTT increase x%	Reduce cwnd to f(x)%
HTCP	100 ACK	cwnd+f(cwnd)/cwnd

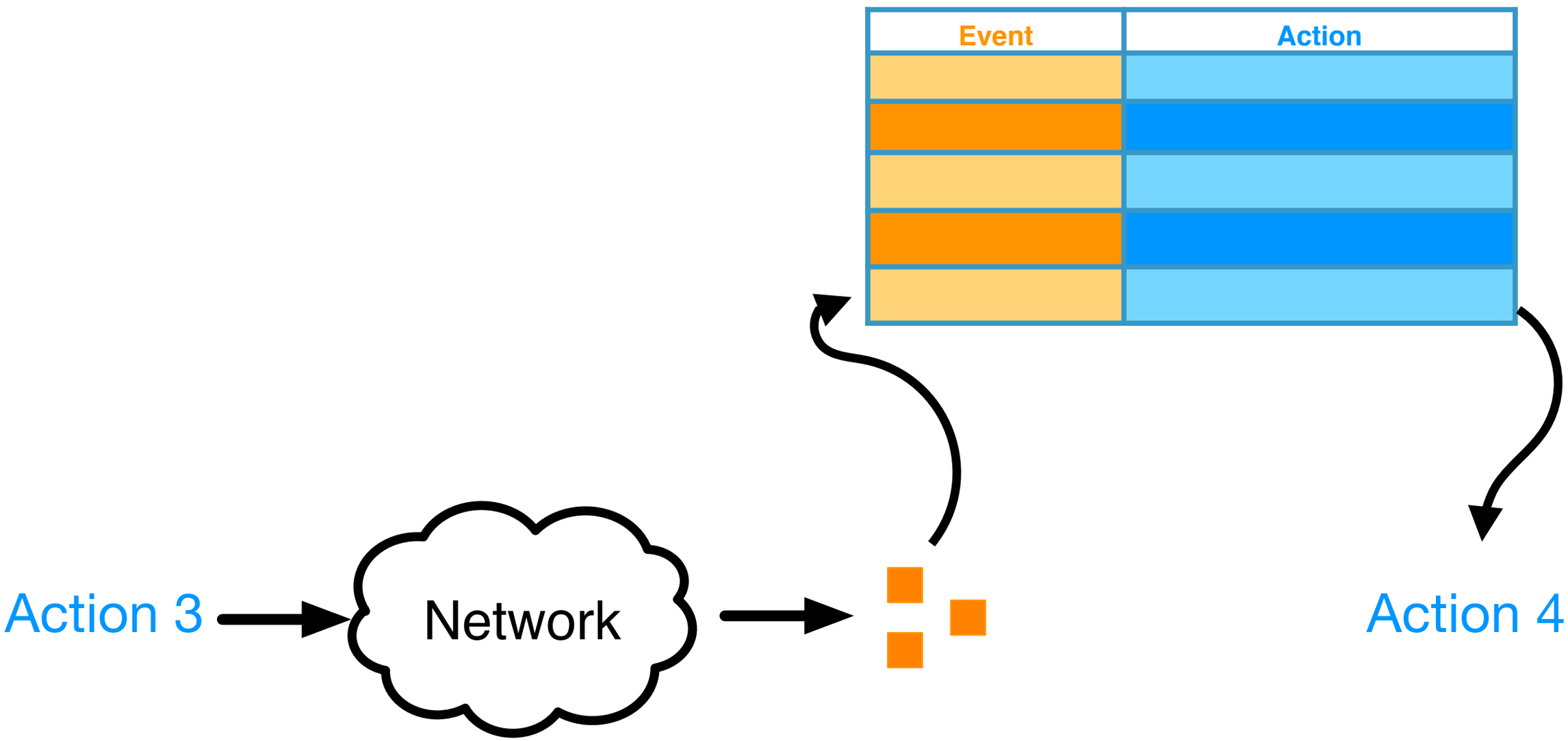
Hardwired Mapping

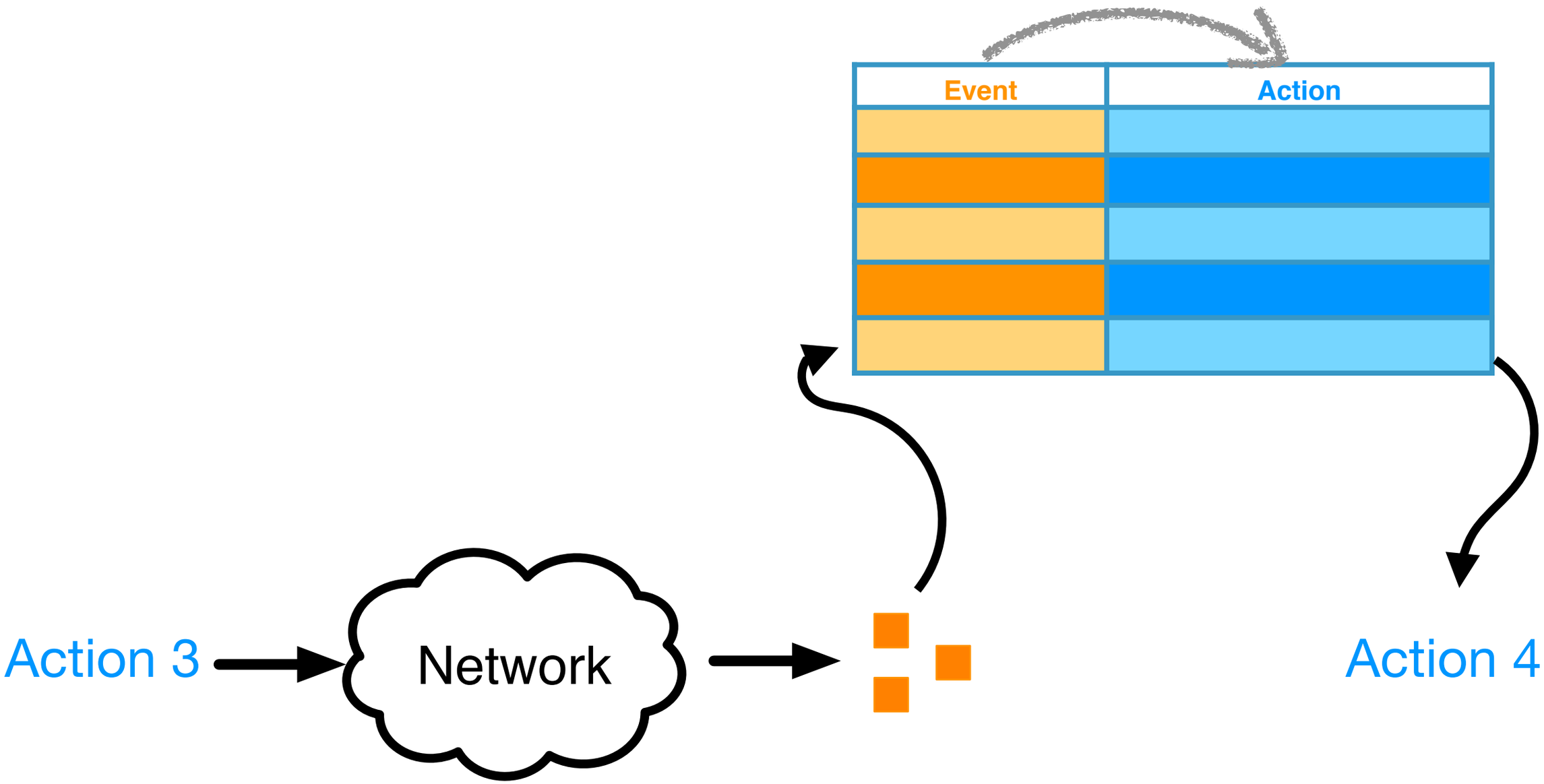
Event	Action

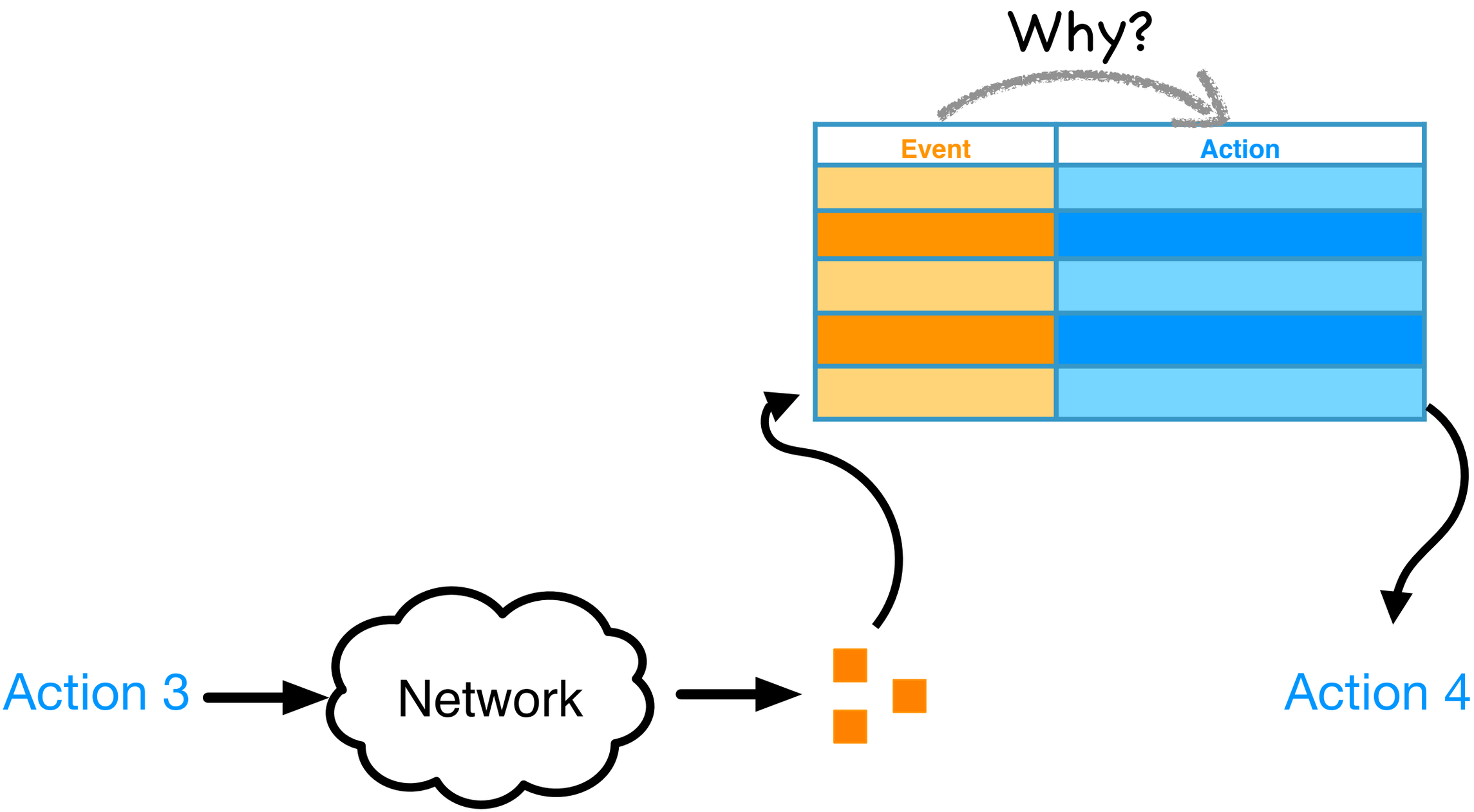
Event	Action

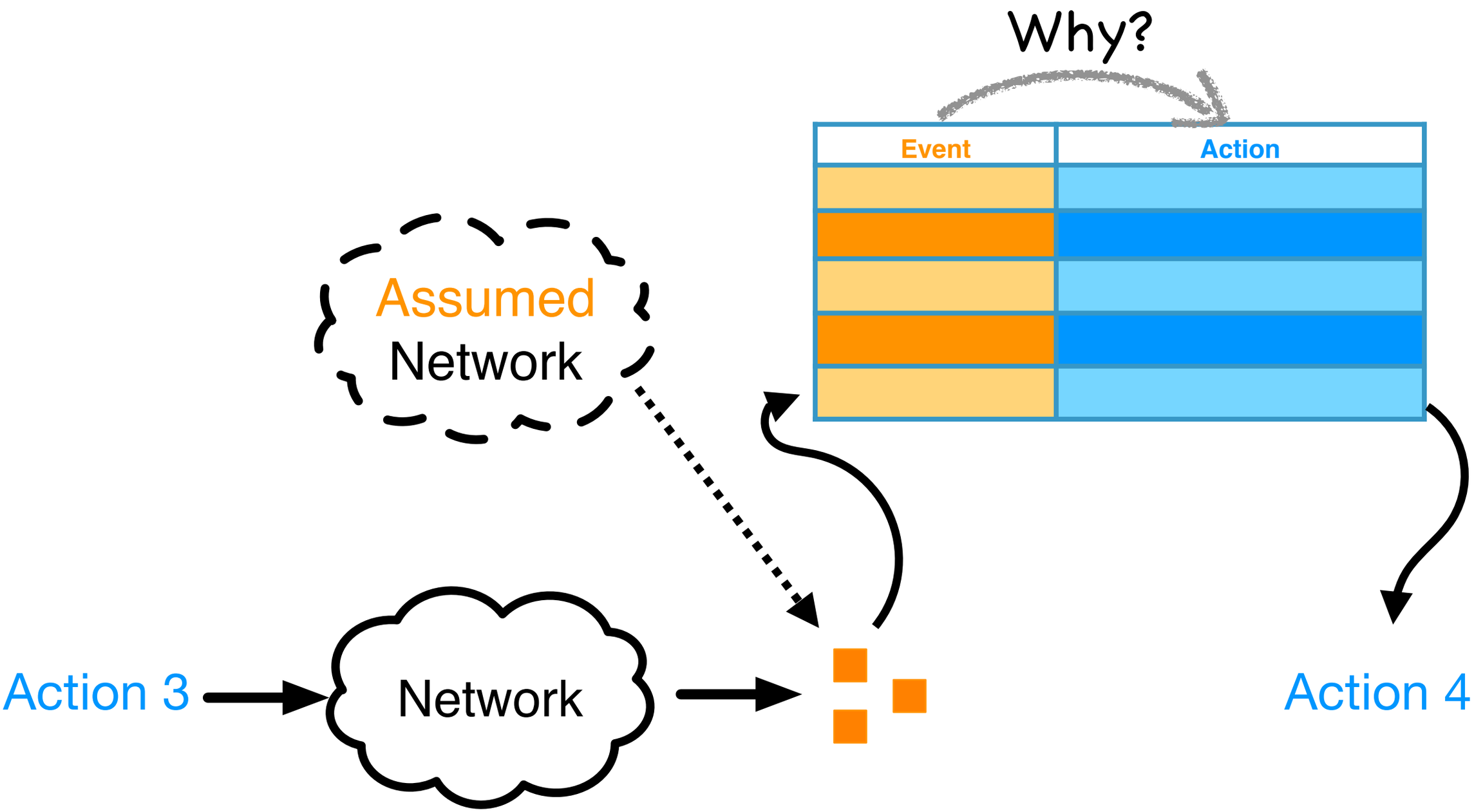
Event	Action

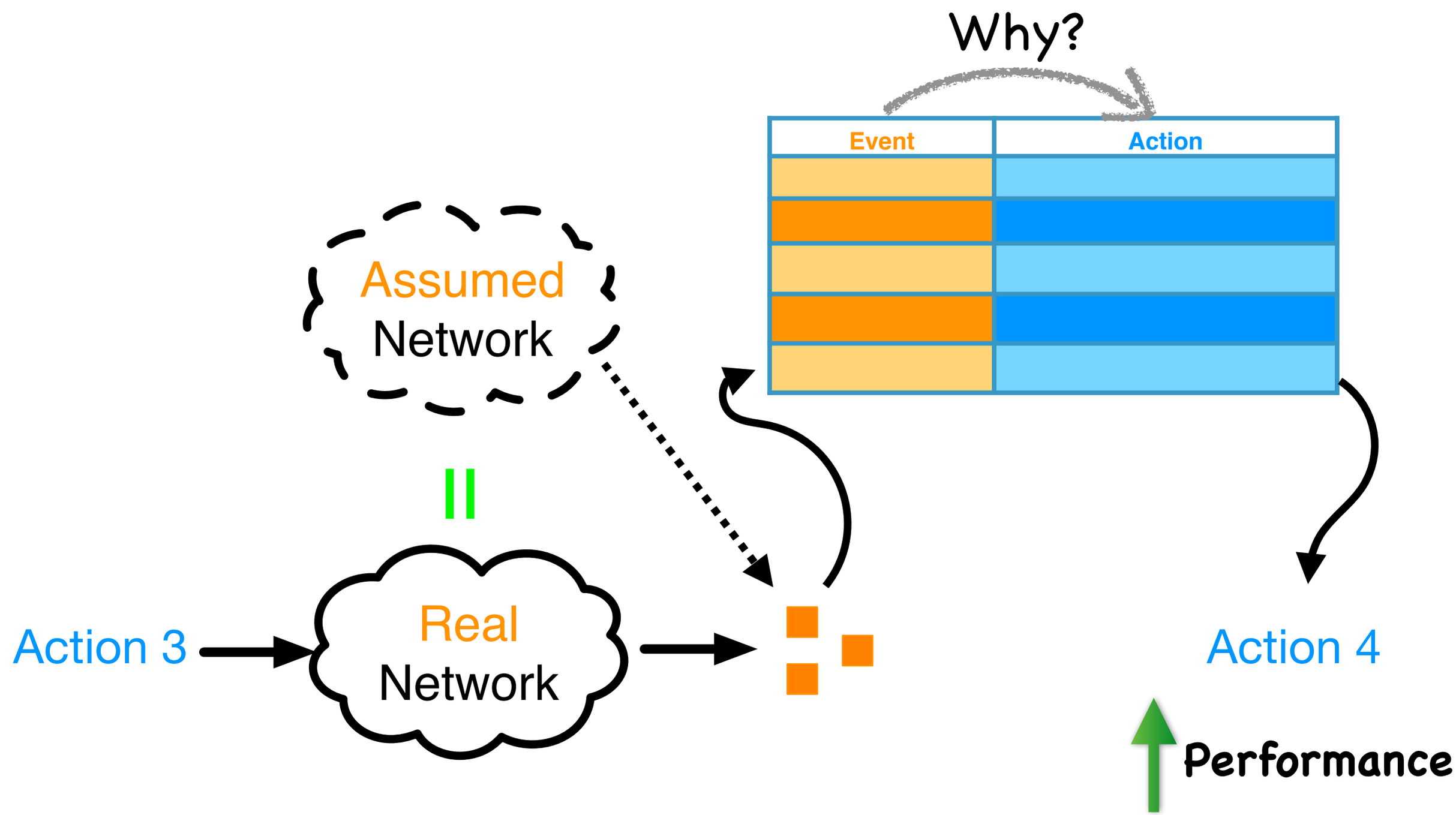


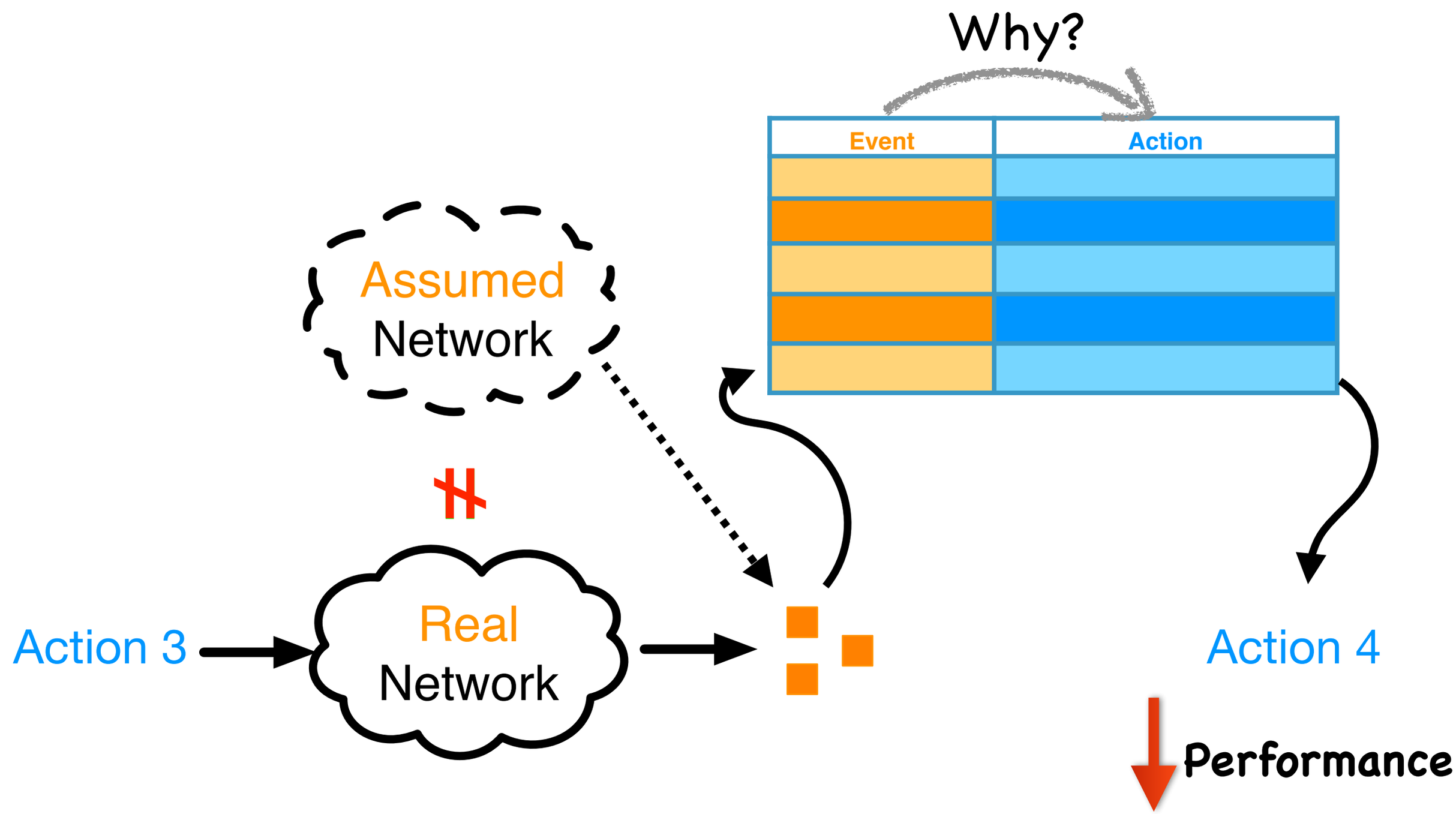

Event
Action
Action 3
Network
Action 4


Event
Action
Action 3
Network
Action 4


Why?
Event
Action
Action 3
Network
Action 4


Why?
Event
Action
Assumed
Network
Action 3
Network
Action 4


Why?
Event
Action
Assumed
Network
Real
Network
Action 3
Action 4
Performance


Why?
Event
Action
Assumed
Network
≠
Action 3
Real
Network
Action 4
Performance

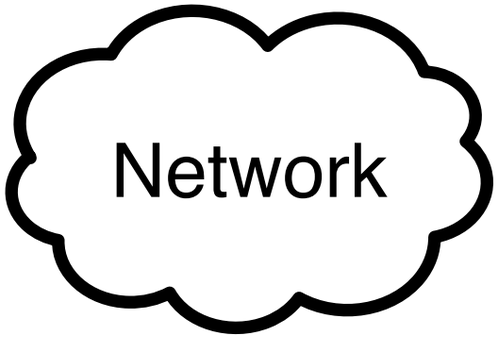
Flow f sends at R

Event	Action

Flow f sends at R

Event	Action
Packet Loss	

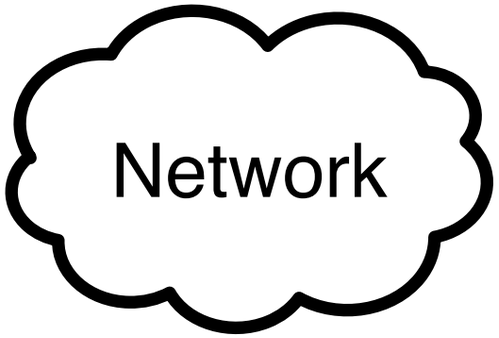
Flow f sends at R



Event	Action
Packet Loss	

Flow f sends at R

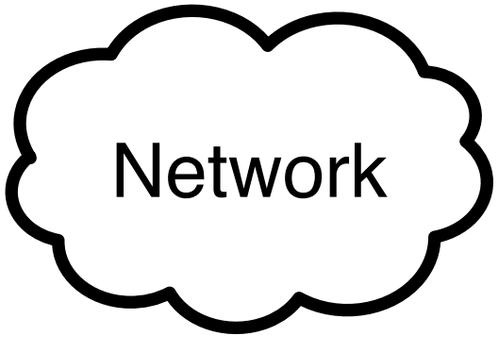
Event	Action
Packet Loss	



f causes
most congestion

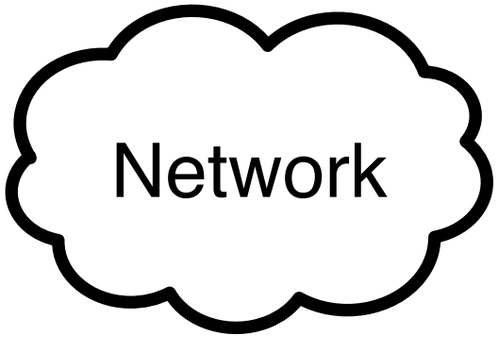
Flow f sends at R

Event	Action
Packet Loss	Dec R a lot



f causes
most congestion

Flow f sends at R



Event	Action
Packet Loss	Dec R a lot

shallow buffer overflow

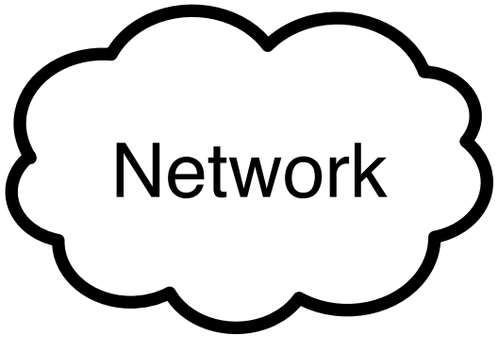
Flow f sends at R

Event	Action
Packet Loss	~~Dec R a lot~~

Network

shallow buffer overflow

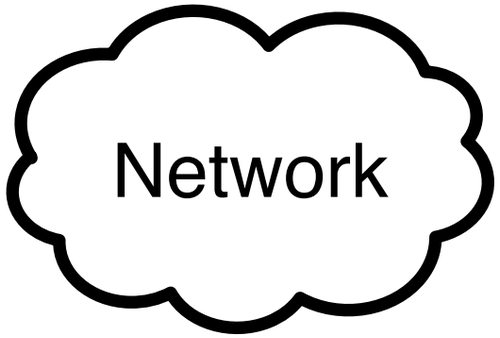
Flow f sends at R



Event	Action
Packet Loss	~~Dec R a lot~~ Dec R a little

shallow buffer overflow

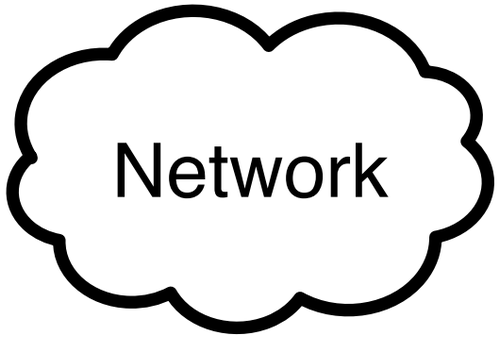
Flow f sends at R



Event	Action
Packet Loss	~~Dec R a lot~~ Dec R a little

other high rate flow
causing congestion

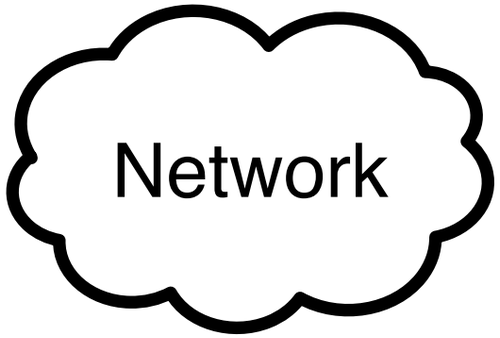
Flow f sends at R



Event	Action
Packet Loss	~~Dec R a lot~~ ~~Dec R a little~~

other high rate flow causing congestion

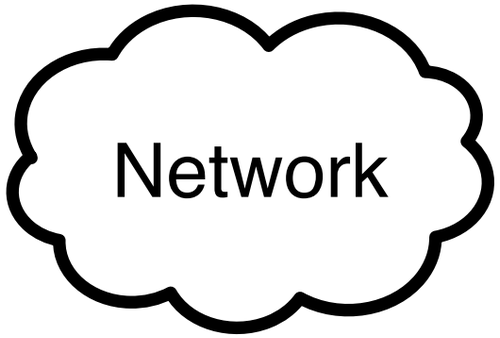
Flow f sends at R



Event	Action
Packet Loss	~~Dec R a lot~~ ~~Dec R a little~~ Maintain R

other high rate flow causing congestion

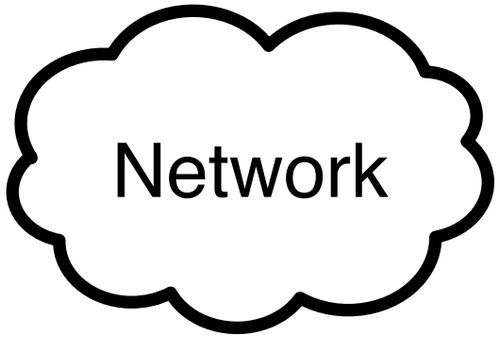
Flow f sends at R



Event	Action
Packet Loss	~~Dec R a lot~~ ~~Dec R a little~~ Maintain R

loss is random

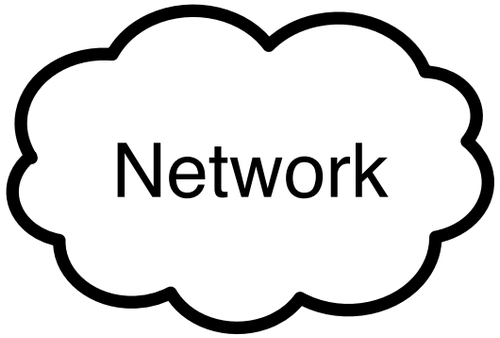
Flow f sends at R



Event	Action
Packet Loss	~~Dec R a lot~~ ~~Dec R a little~~ ~~Maintain R~~

loss is random

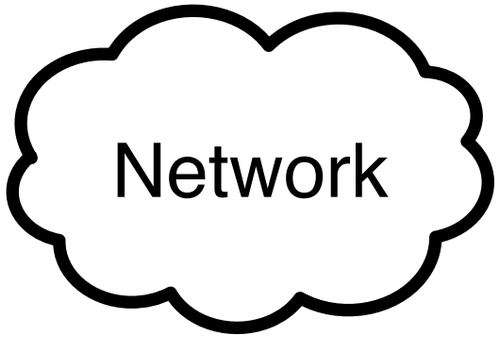
Flow f sends at R



Event	Action
Packet Loss	~~Dec R a lot~~ ~~Dec R a little~~ ~~Maintain R~~ Increase R

loss is random

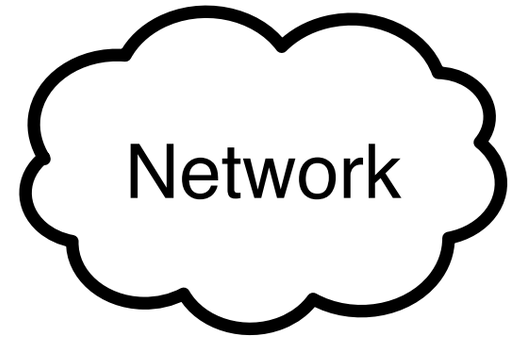
Flow f sends at R



Event	Action
Packet Loss	~~Dec R a lot~~ ~~Dec R a little~~ ~~Maintain R~~ ~~Increase R~~

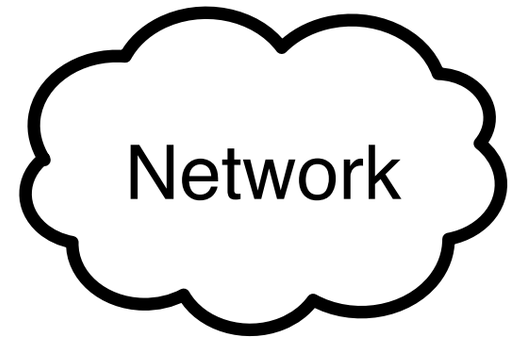
loss is random

Flow f sends at R



Event	Action
Packet Loss	~~Dec R a lot~~ ~~Dec R a little~~ ~~Maintain R~~ ~~Increase R~~

Flow f sends at R

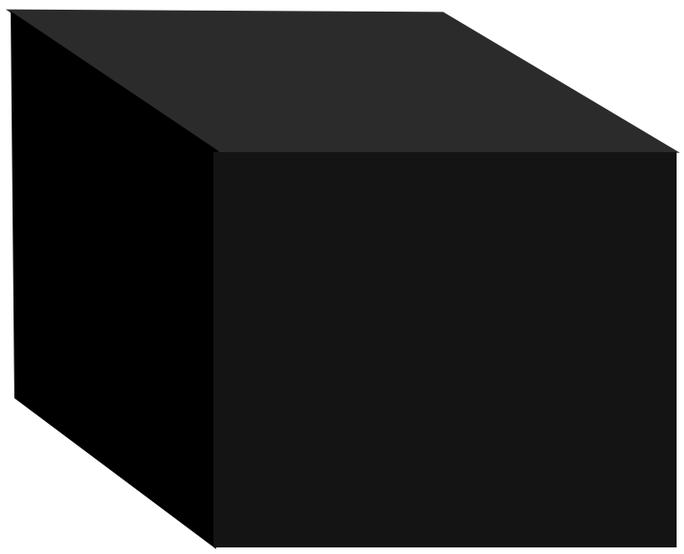


Event	Action
Pac	~~Dec R a lot~~
	~~Maintain R~~
	~~Increase R~~

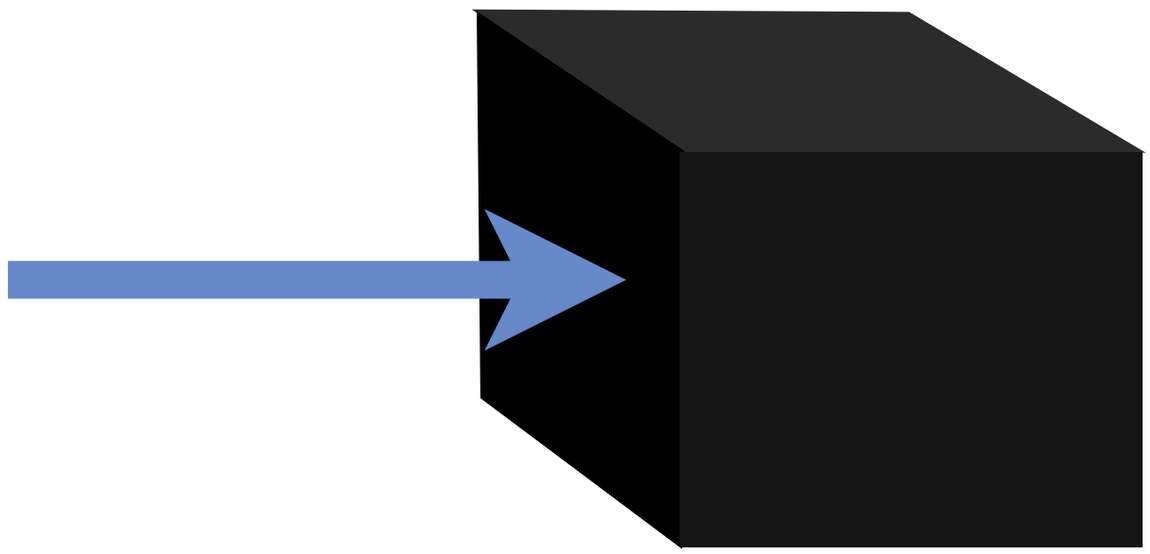
No event-control mapping optimal for all network scenarios

What is the right rate to send?

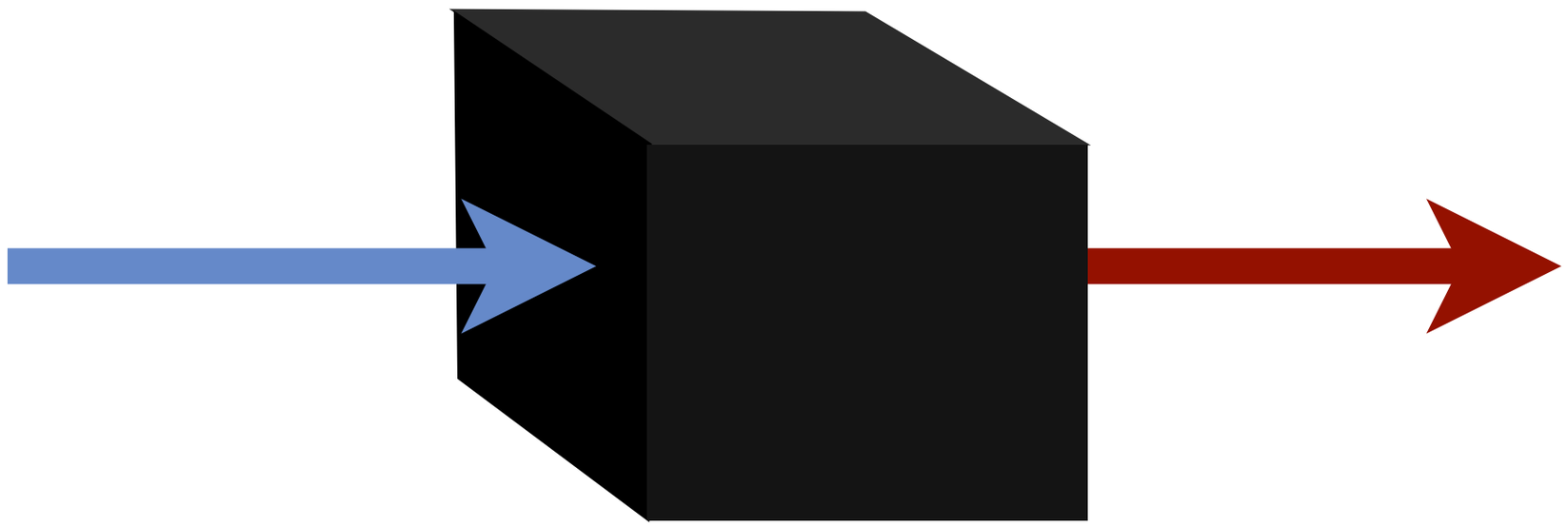
What is the right rate to send?



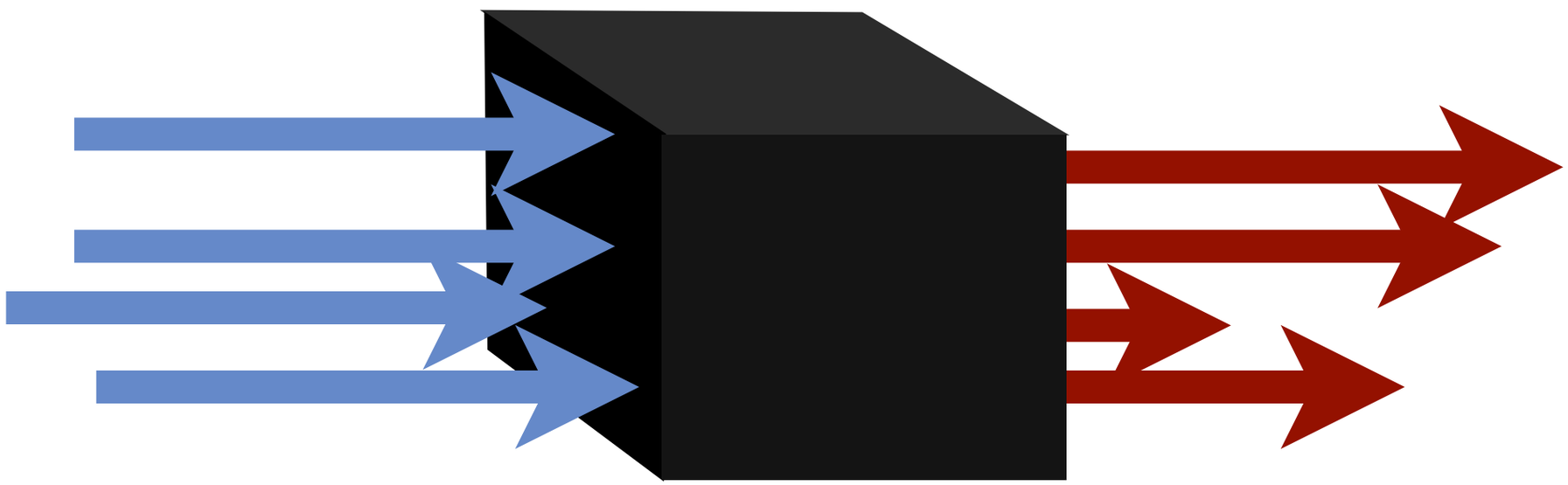
What is the right rate to send?



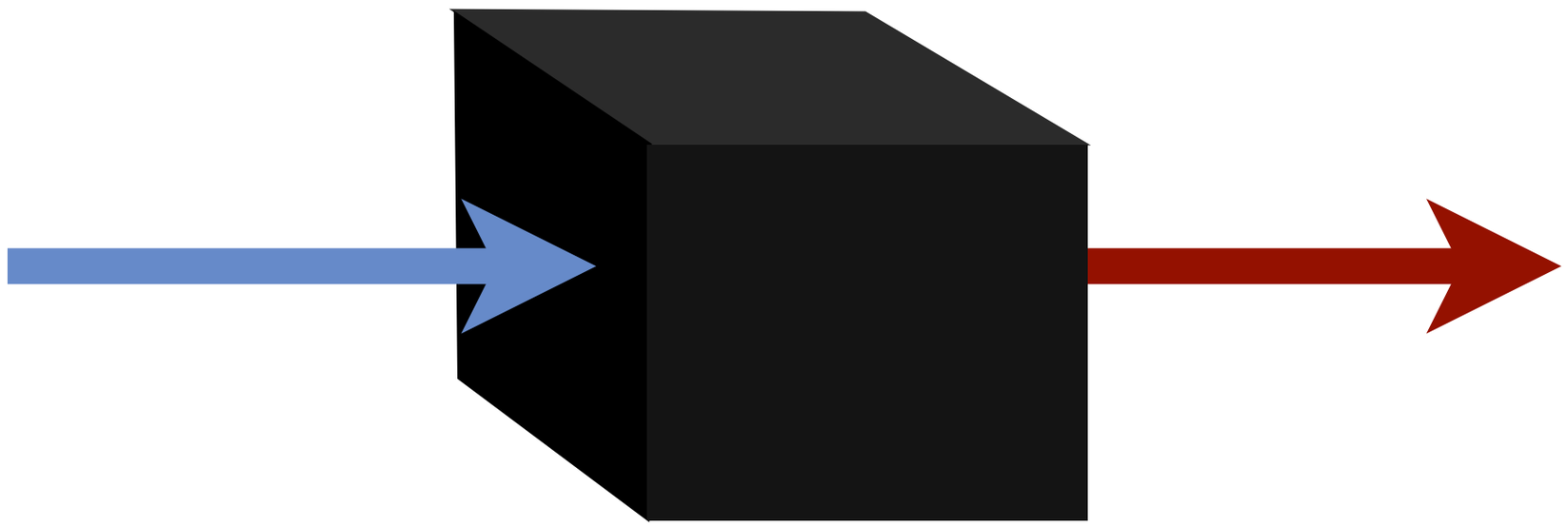
What is the right rate to send?



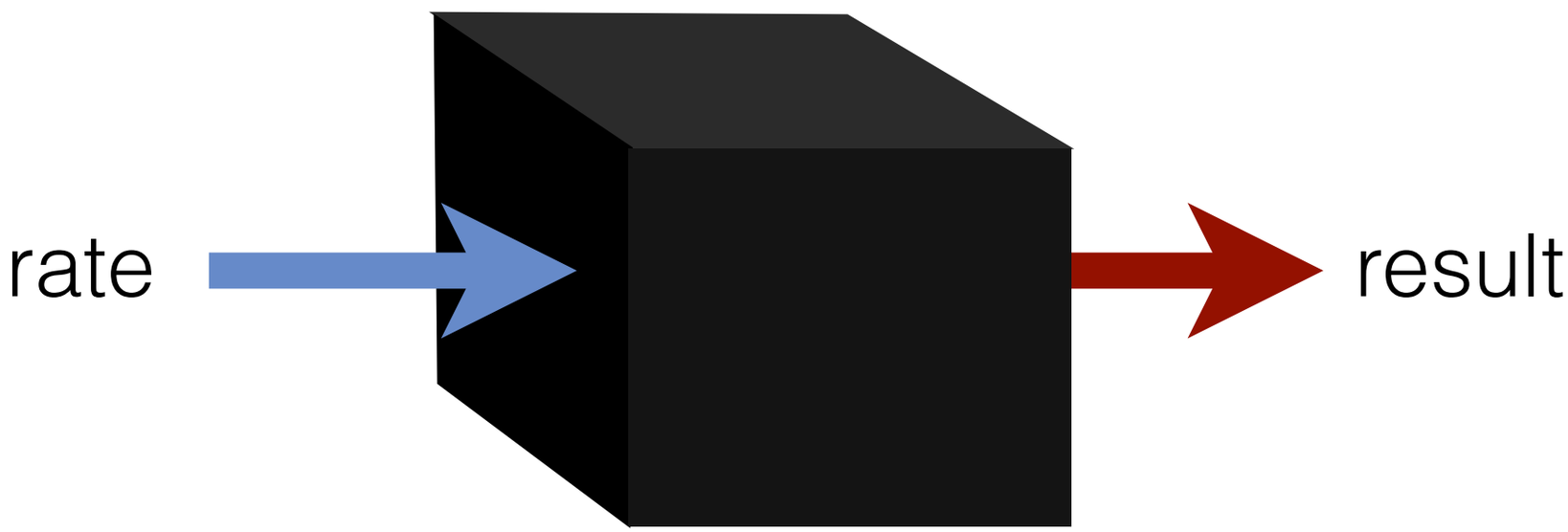
What is the right rate to send?



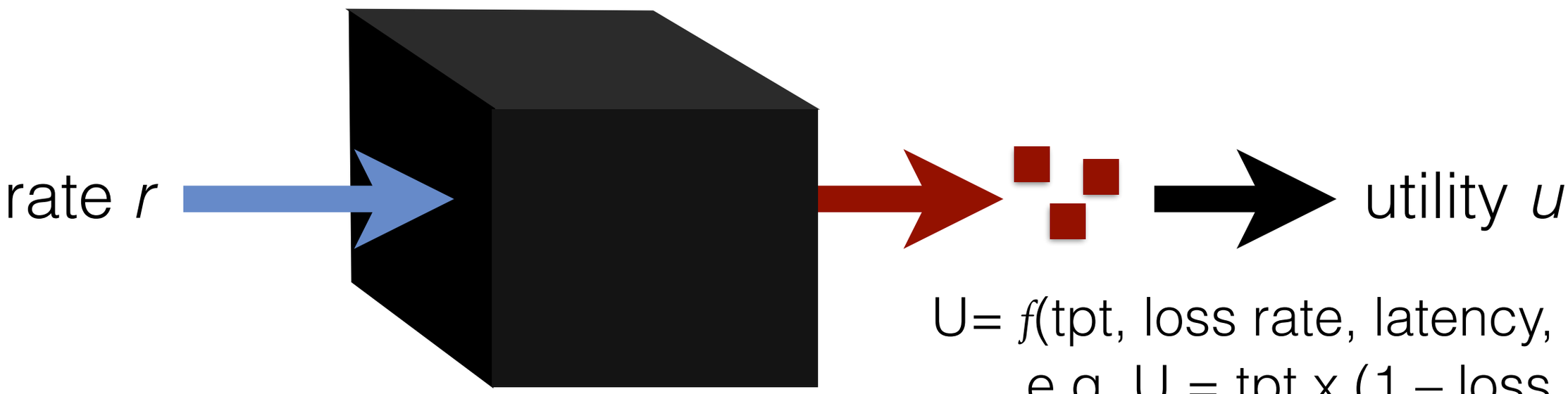
What is the right rate to send?



What is the right rate to send?



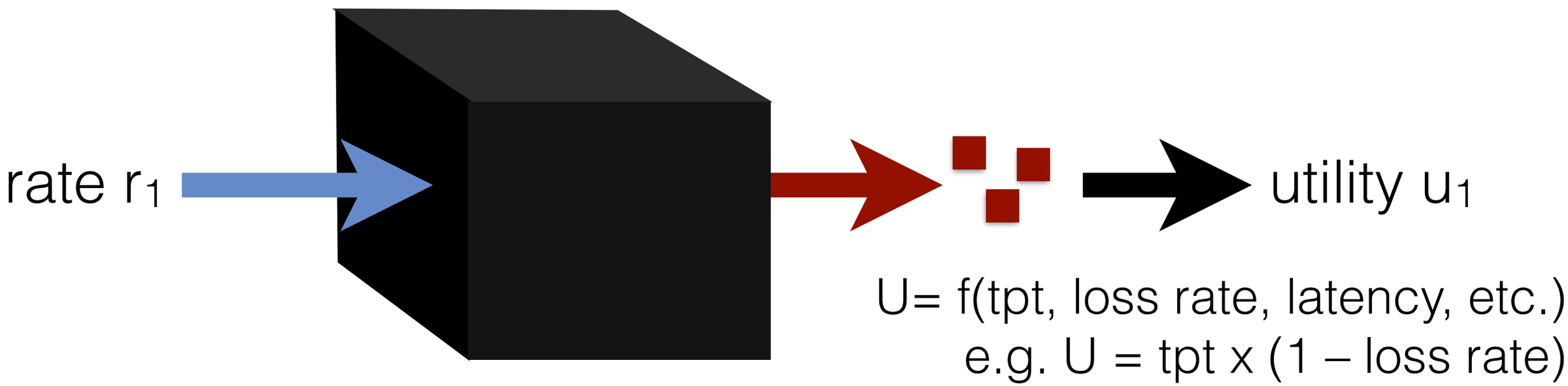
What is the right rate to send?



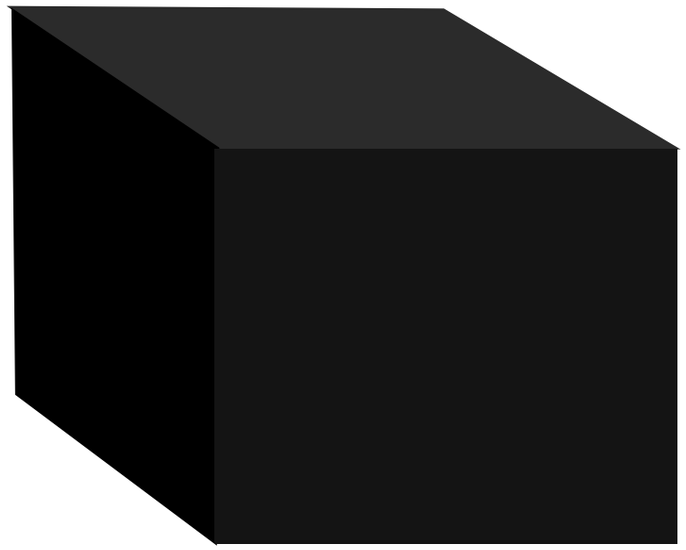
U= *f*(tpt, loss rate, latency, etc.)

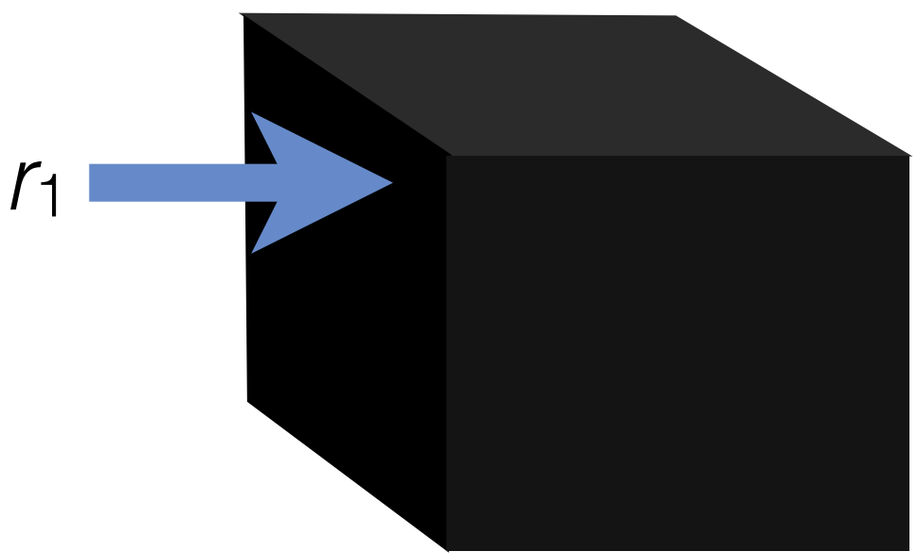
e.g. U = tpt x (1 – loss rate)

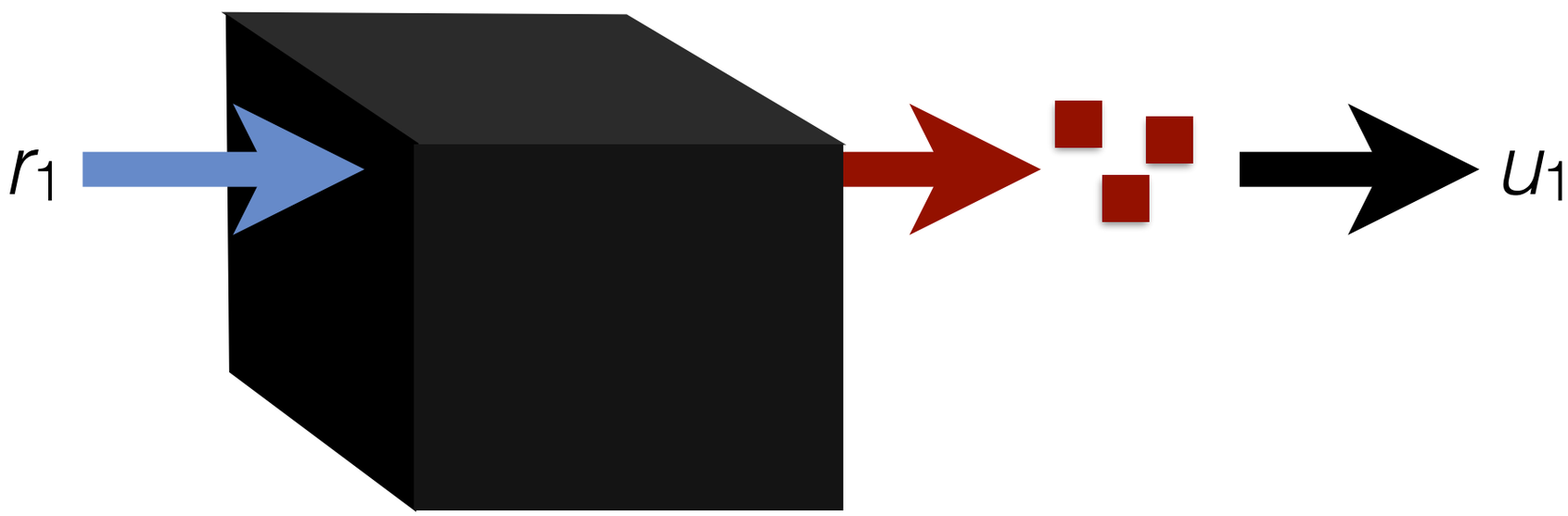
What is the right rate to send?

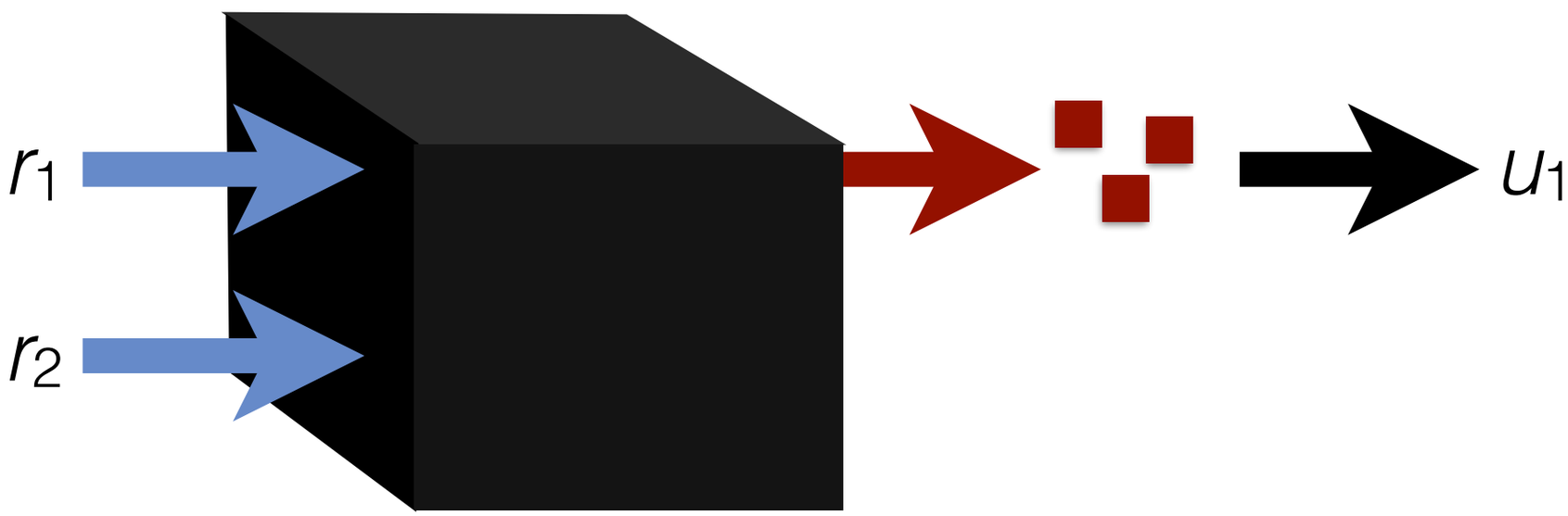


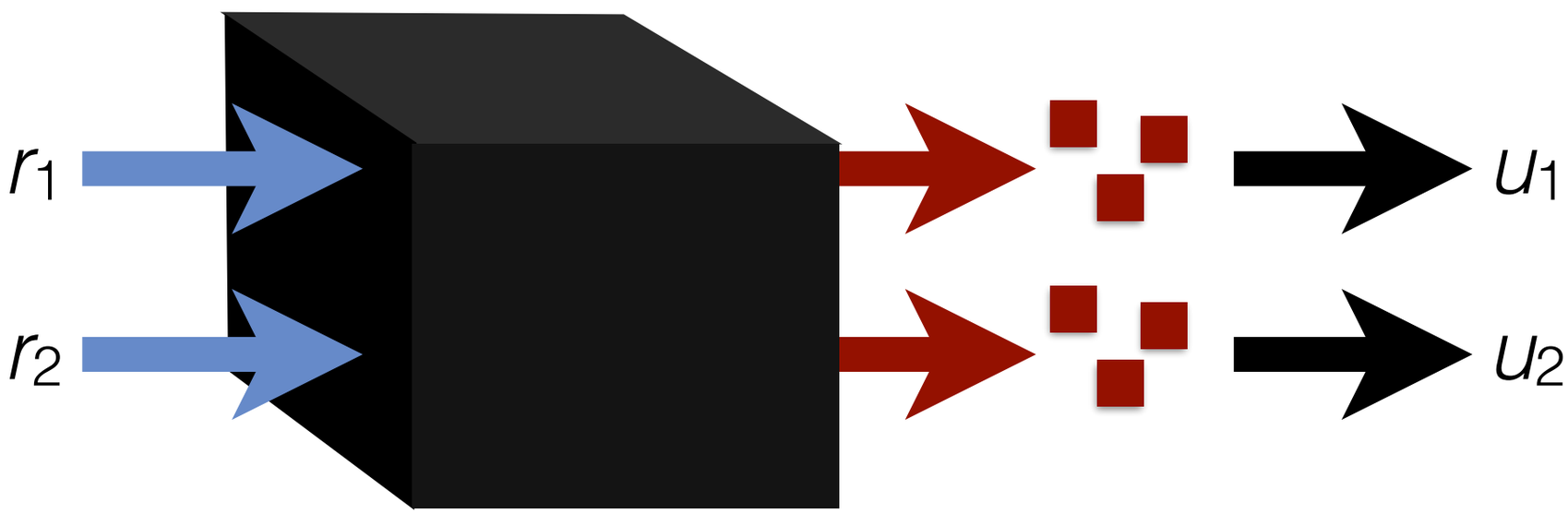
No matter how complex the network,
rate r —> utility u

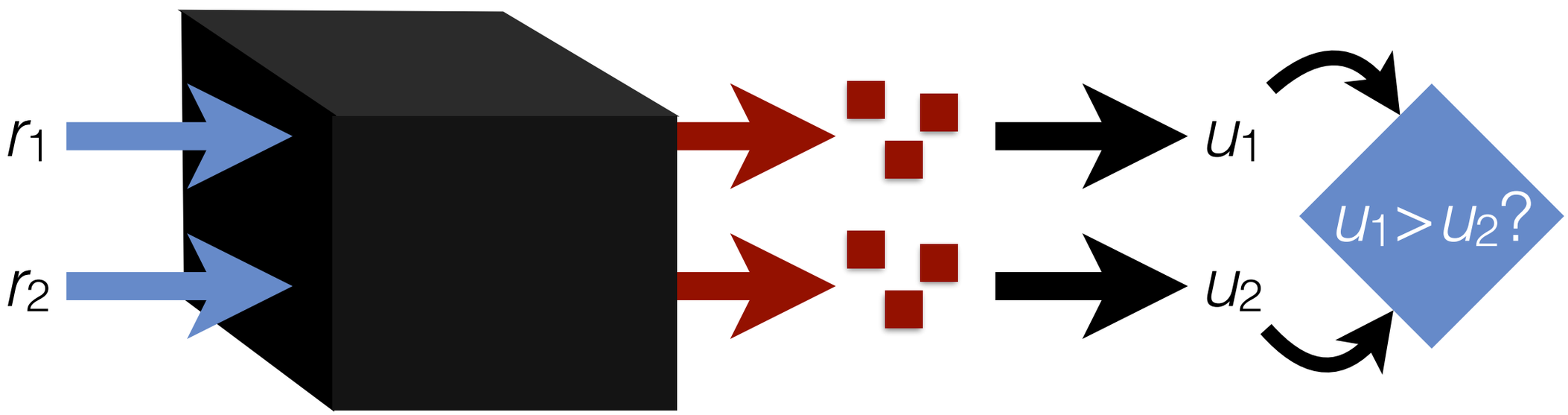


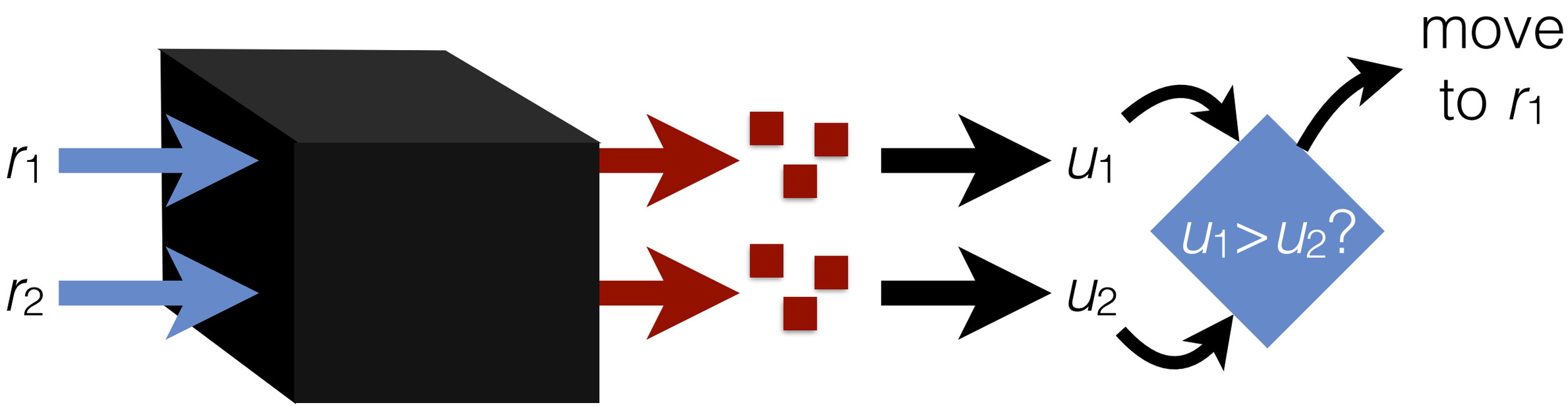

r_1

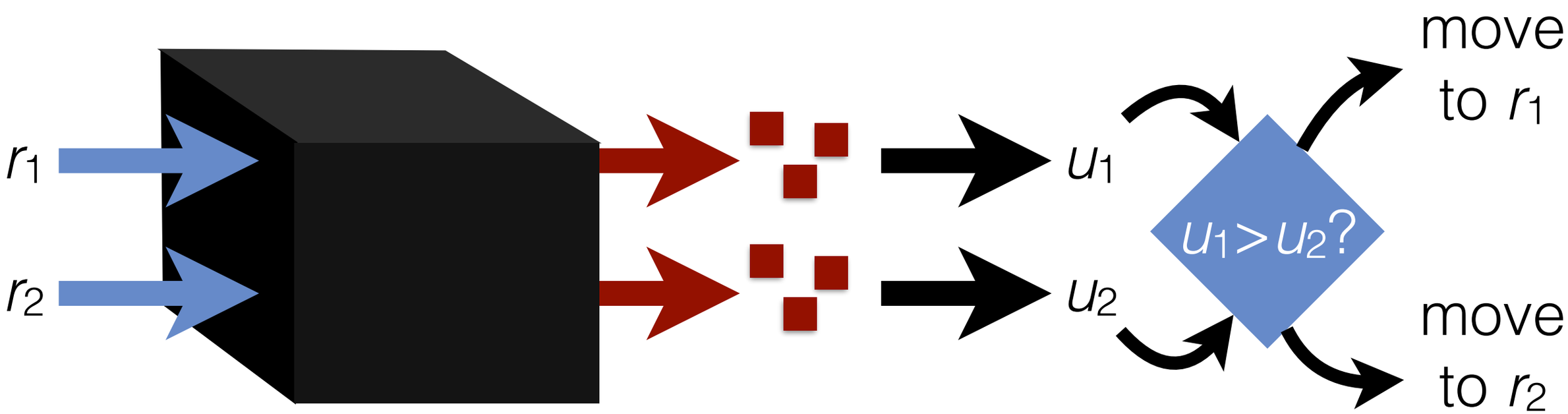

r_1
u_1


r_1
r_2
u_1

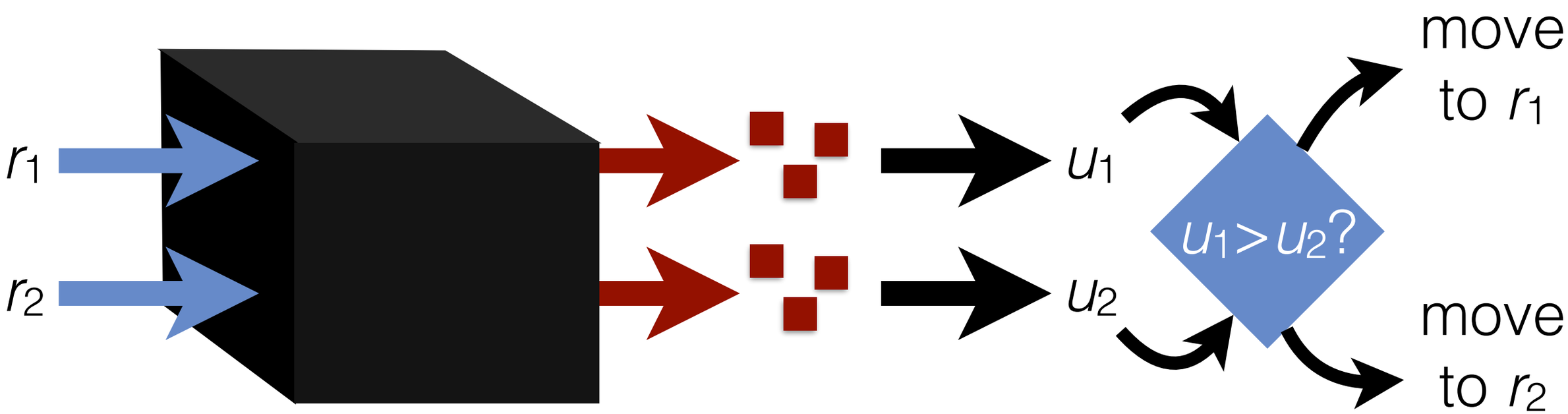

r_1
r_2
u_1
u_2


r_1
r_2
u_1
u_2
$u_1 > u_2$?

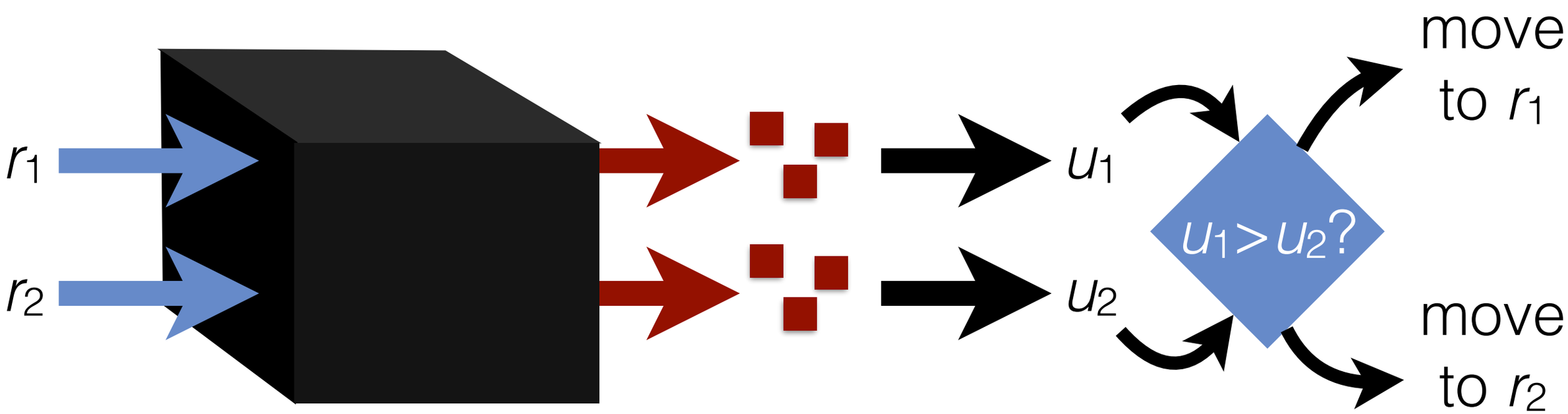

r_1
r_2
u_1
u_2
$u_1 > u_2$?
move to r_1


r_1
r_2
u_1
u_2
$u_1 > u_2$?
move to r_1
move to r_2

Performance-oriented Congestion Control

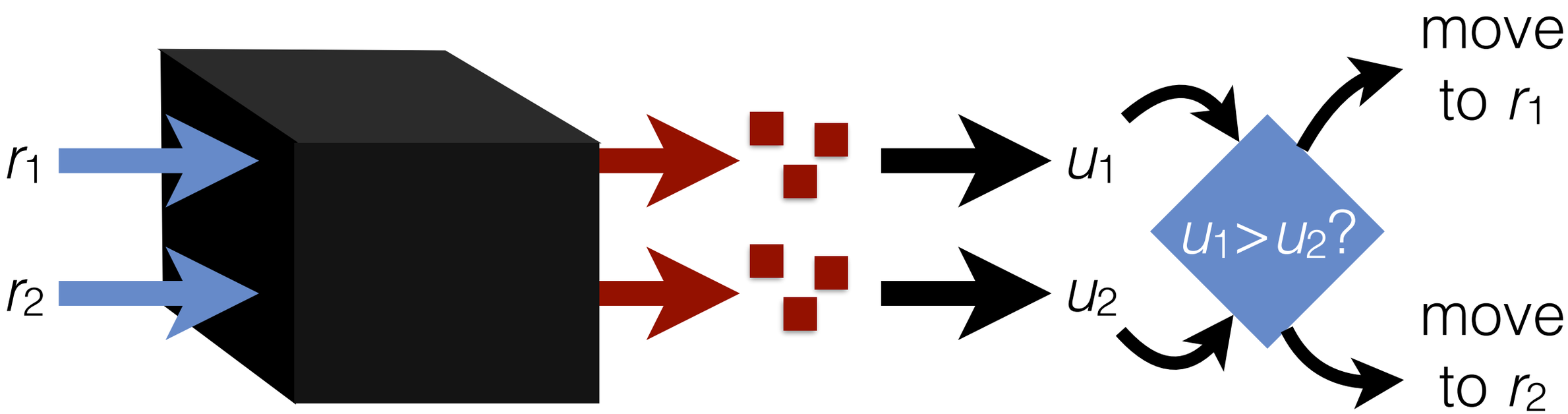


Performance-oriented Congestion Control



Observe real performance

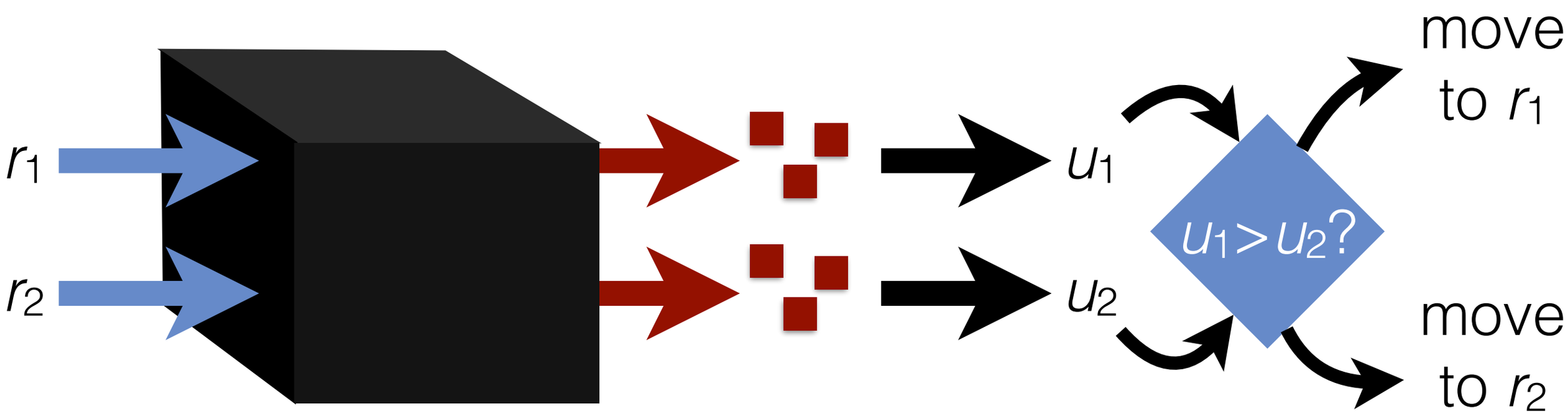
Performance-oriented Congestion Control



Observe real performance

Control based on empirical evidence

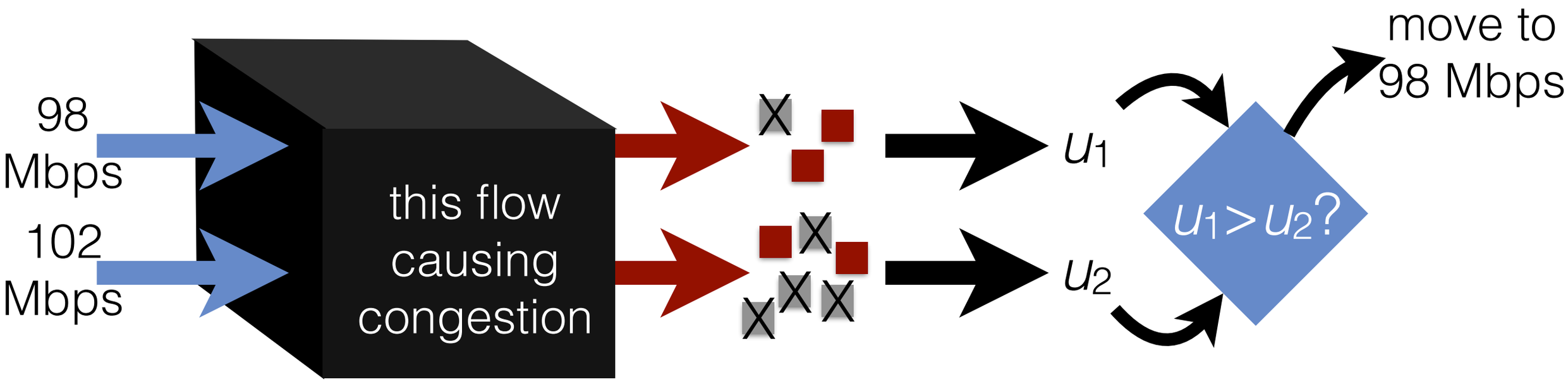
Performance-oriented Congestion Control

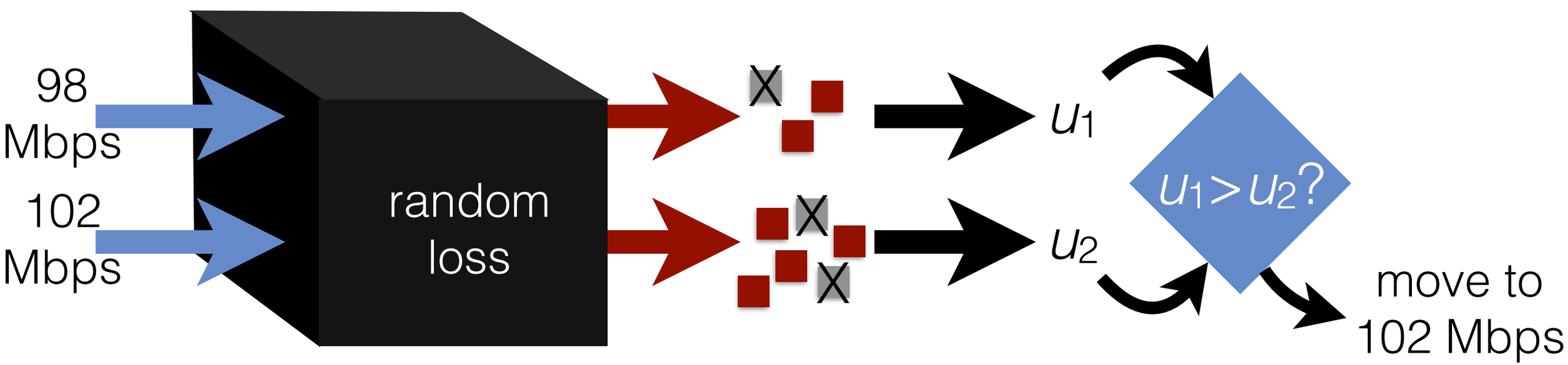


Observe real performance

Control based on empirical evidence

yields **Consistent high performance**


98
Mbps
102
Mbps
this flow
causing
congestion
u_1
u_2
$u_1 > u_2$?
move to
98 Mbps

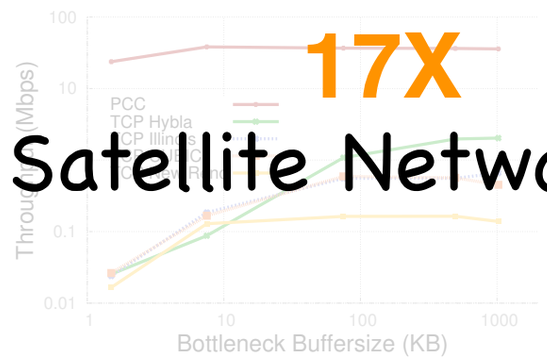

98
Mbps
102
Mbps
random
loss
u_1
u_2
$u_1 > u_2$?
move to
102 Mbps

Consistent High Performance

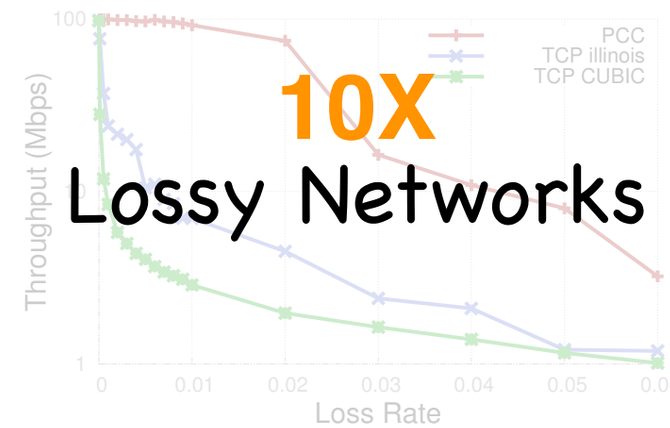
Table 1: PCC significantly outperforms TCP in inter-data center environments. RTT [illegible]ec; throughput in Mbps.

Transmission Pair	RTT	PCC	SABUL	CUBIC	Illinois
GPO → NYSERNet	[illegible]	[illegible]	[illegible]	129	326
GPO → Missouri	[illegible]	[illegible]	[illegible]	80.7	90.1
GPO → Illinois	35.4	766	664	84.5	102
NYSERNet → Missouri	47.4	816	662	108	109
Wisconsin → Illinois	9.01	801	700	547	562
GPO → Wisc.	38.0	783	487	79.3	120
NYSERNet → Wisc.	38.3	791	673	134	134
Missouri → Wisc.	20.9	807	698	259	262
NYSERNet → Illinois	36.1	808	674	141	141

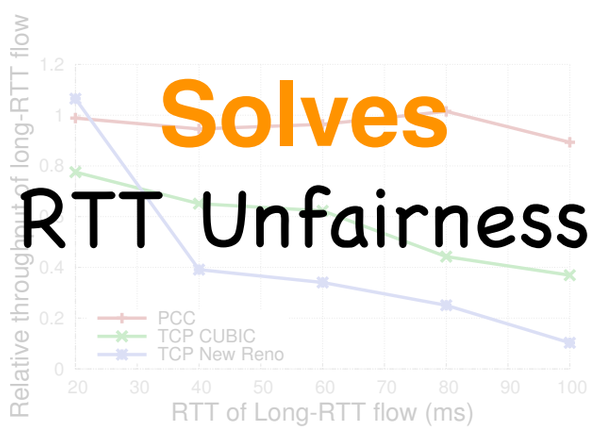
4X
InterDC



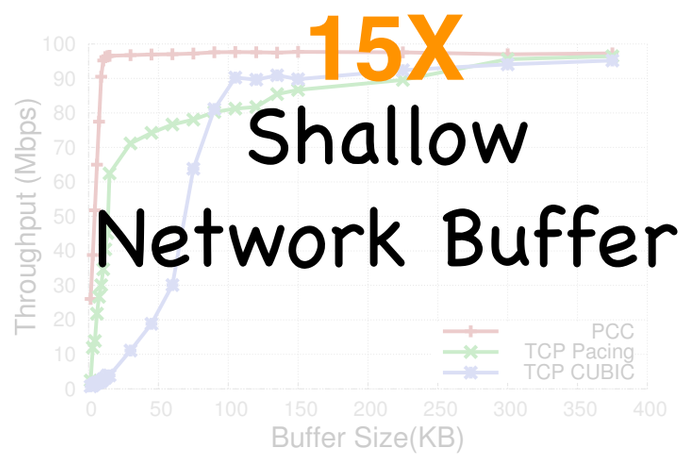
17X
Satellite Networks



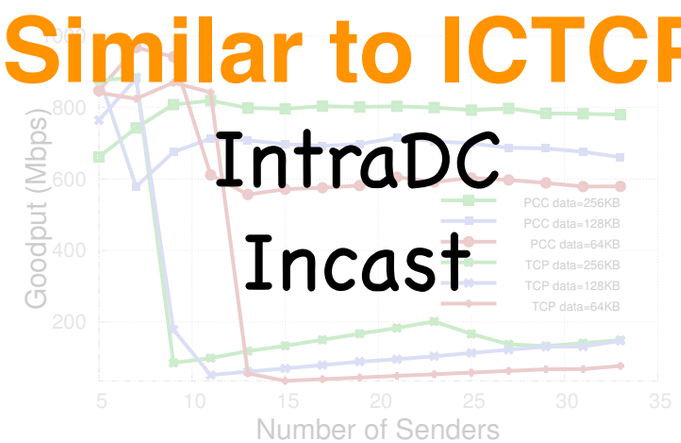
10X
Lossy Networks



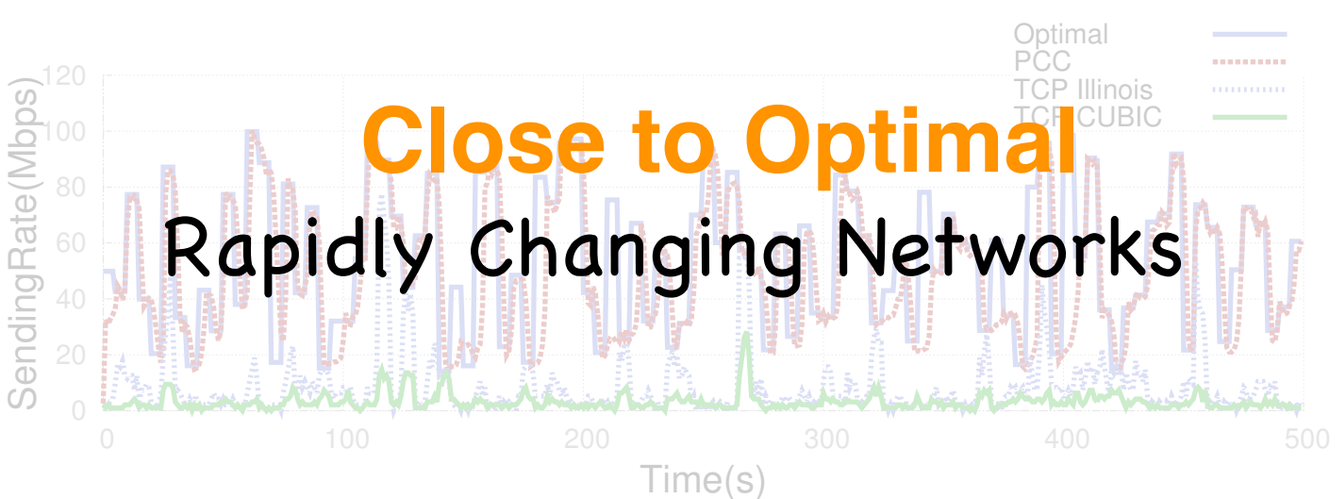
Solves
RTT Unfairness



15X
Shallow
Network Buffer



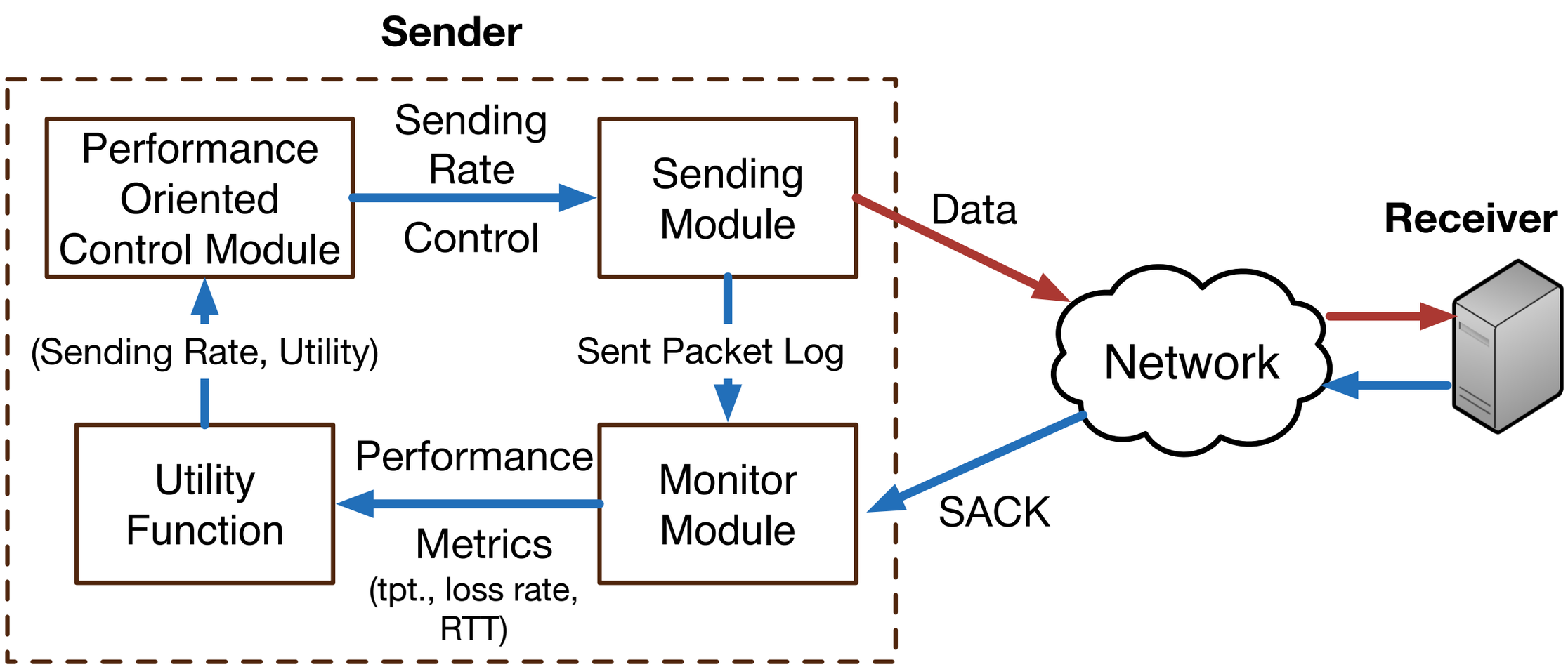
Similar to ICTCP
IntraDC
Incast



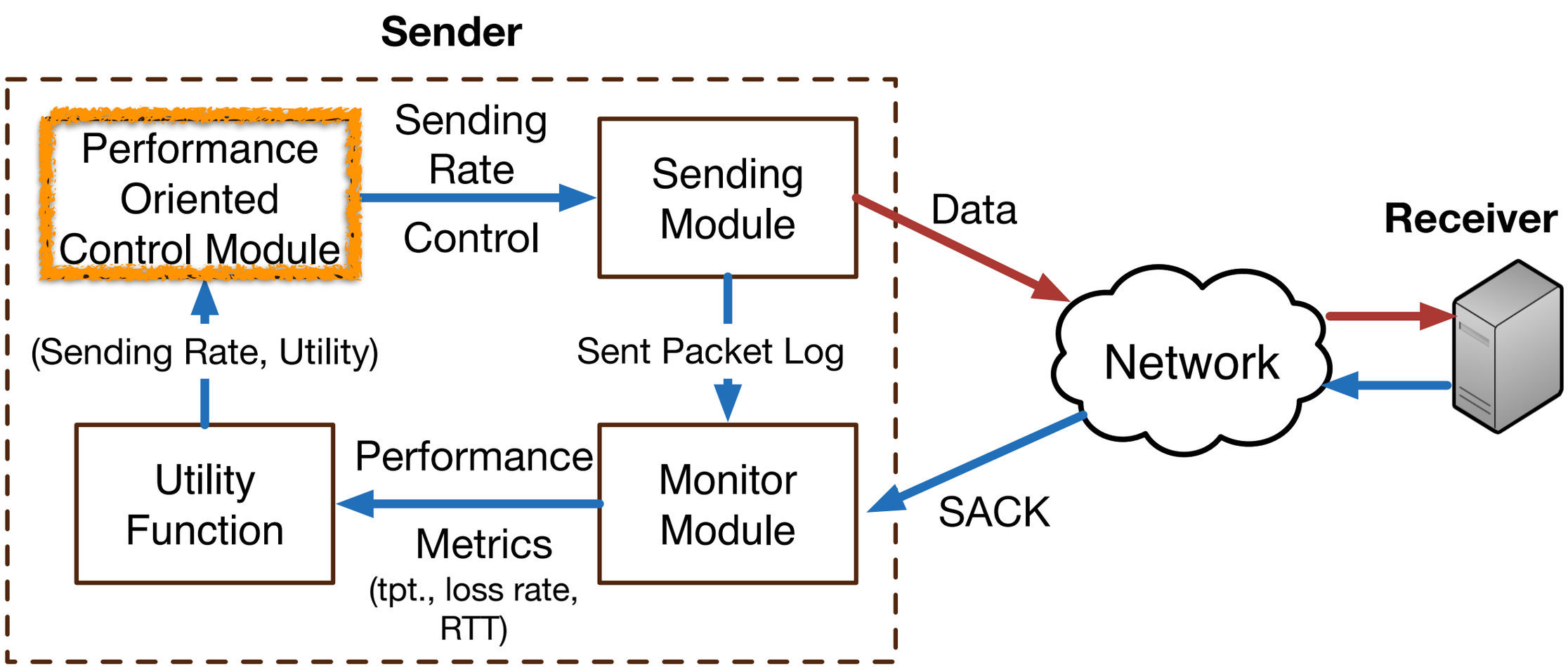
Close to Optimal
Rapidly Changing Networks

5X in median
Global Commercial
Internet

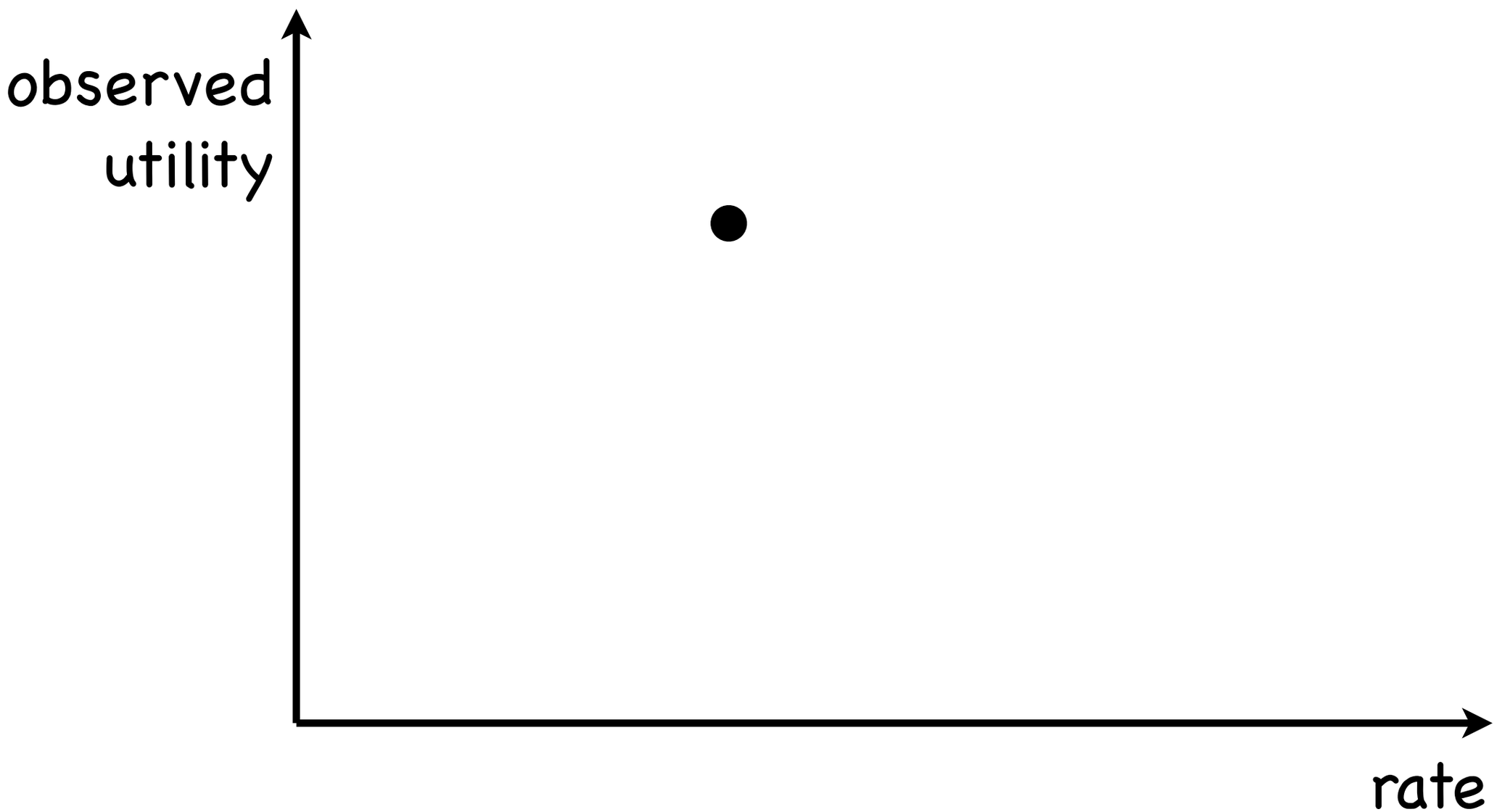
Software Components



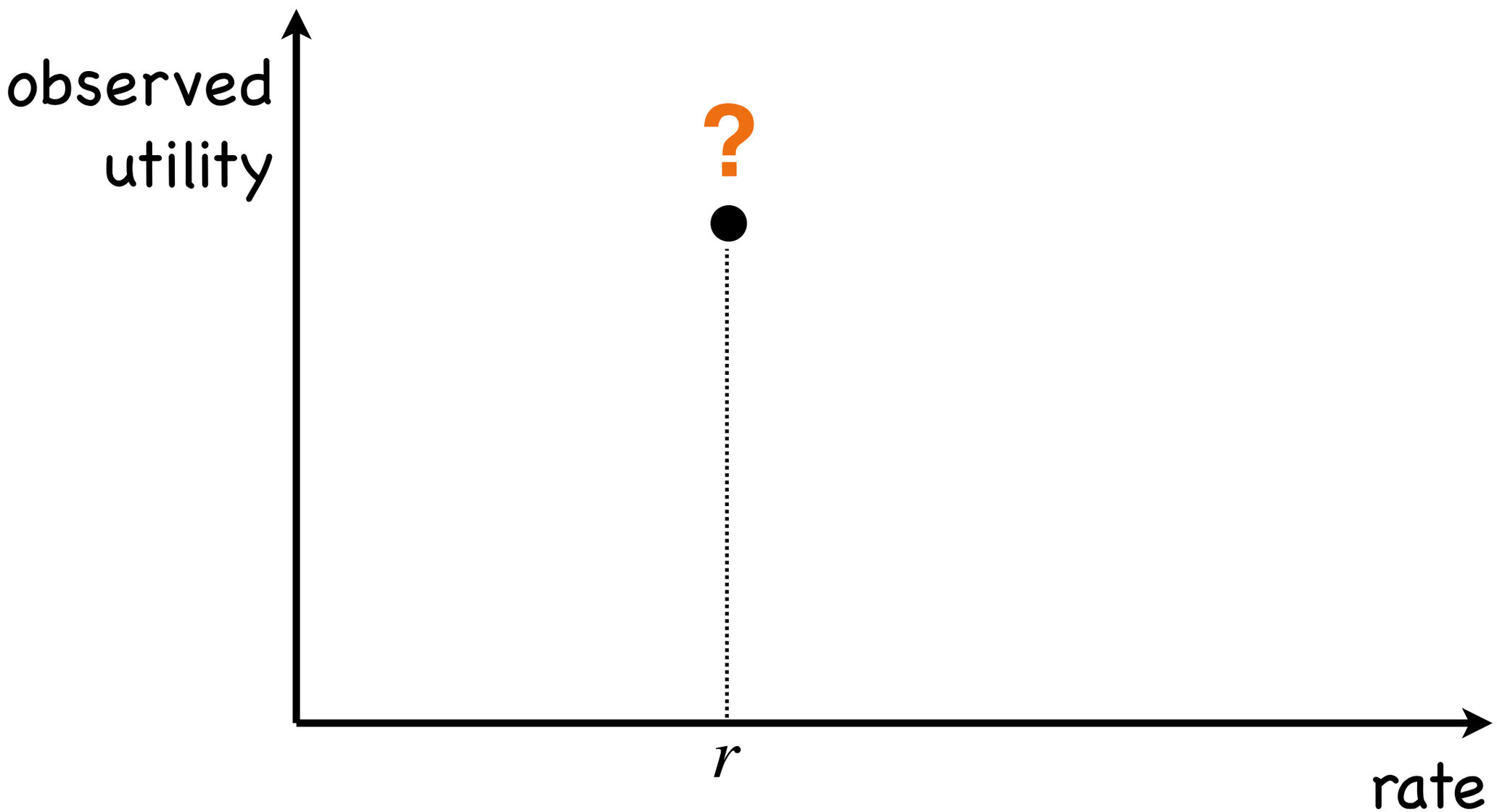
Software Components



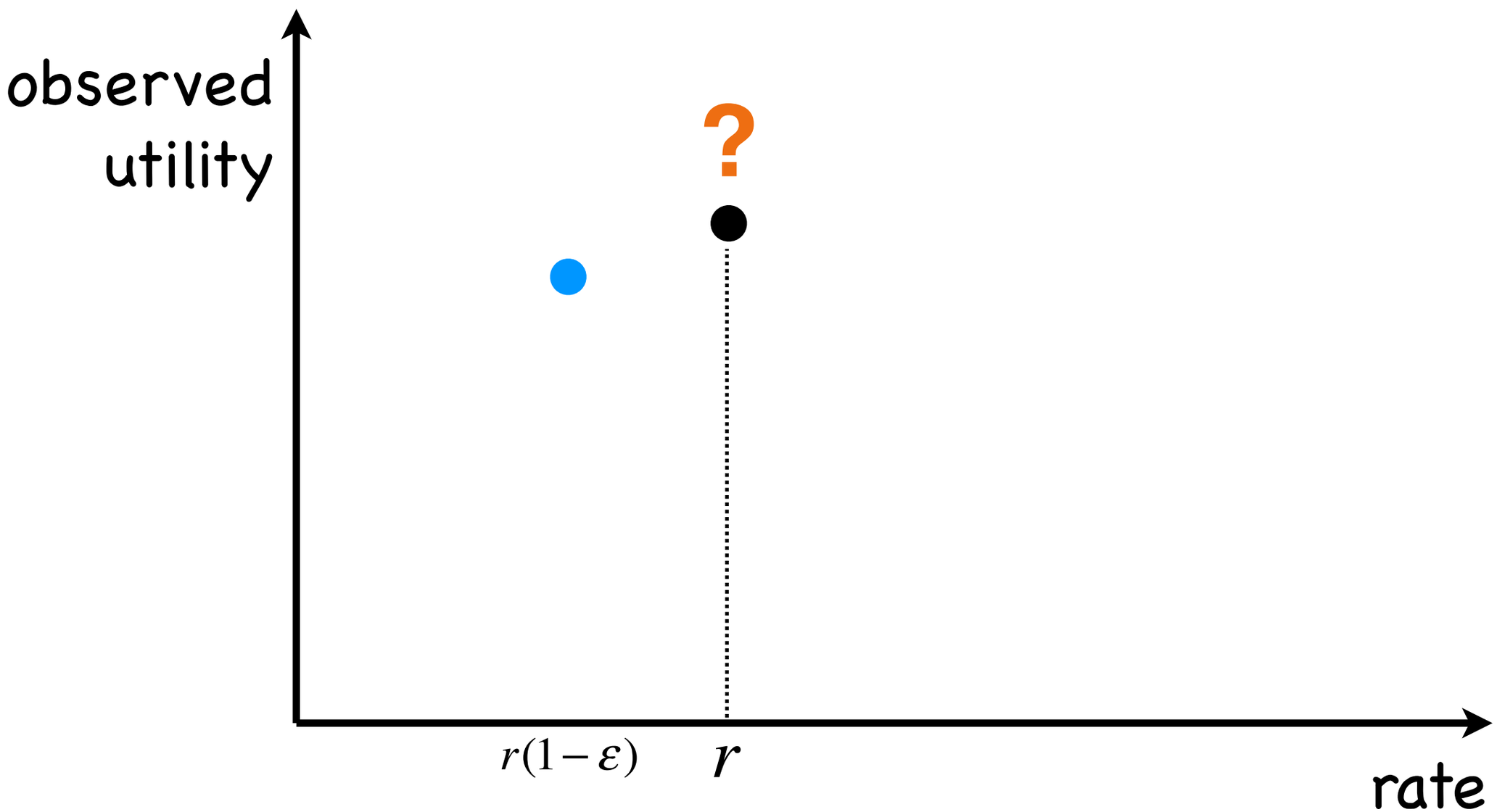
Performance Oriented Control



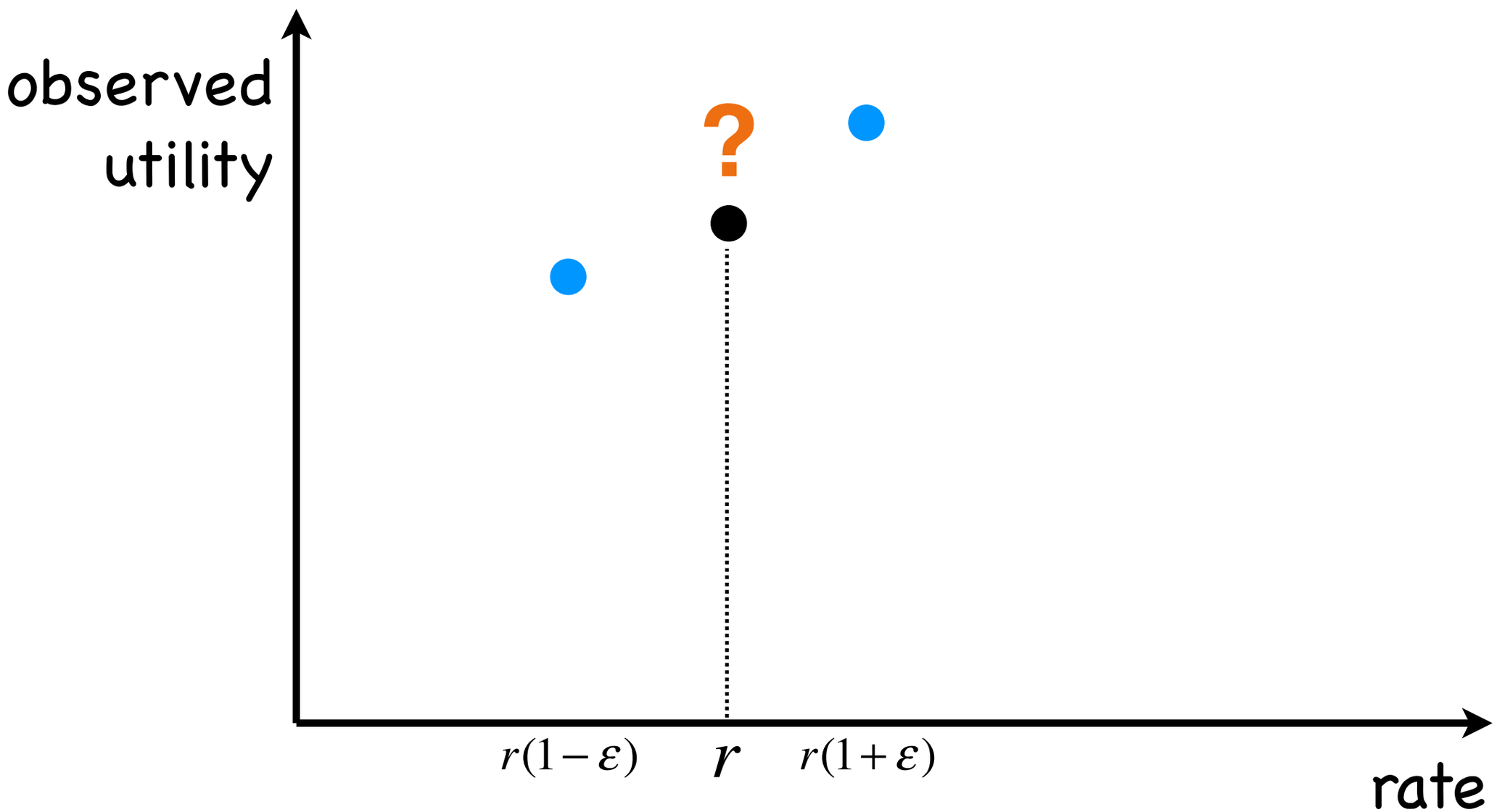
Performance Oriented Control



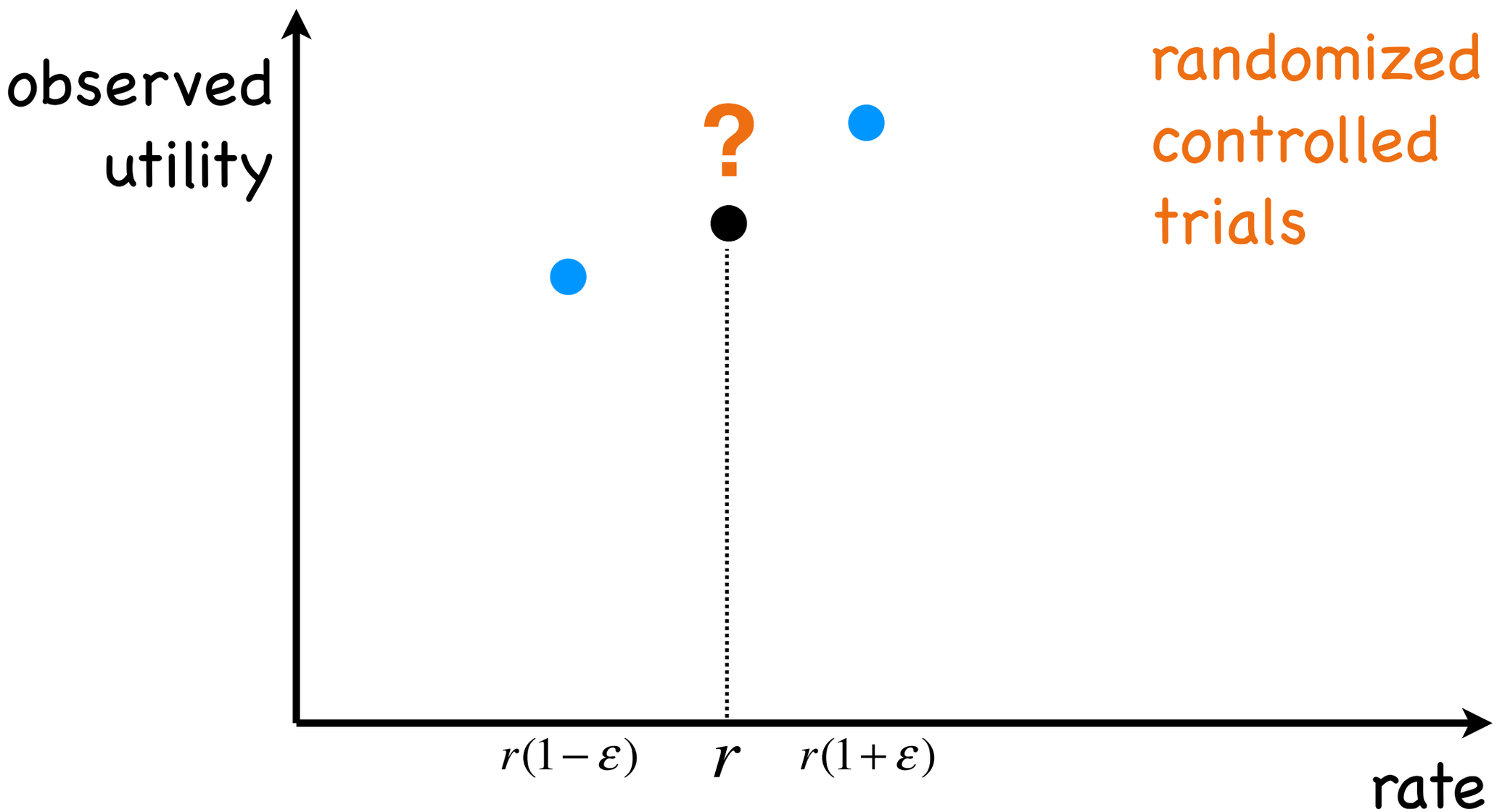
Performance Oriented Control



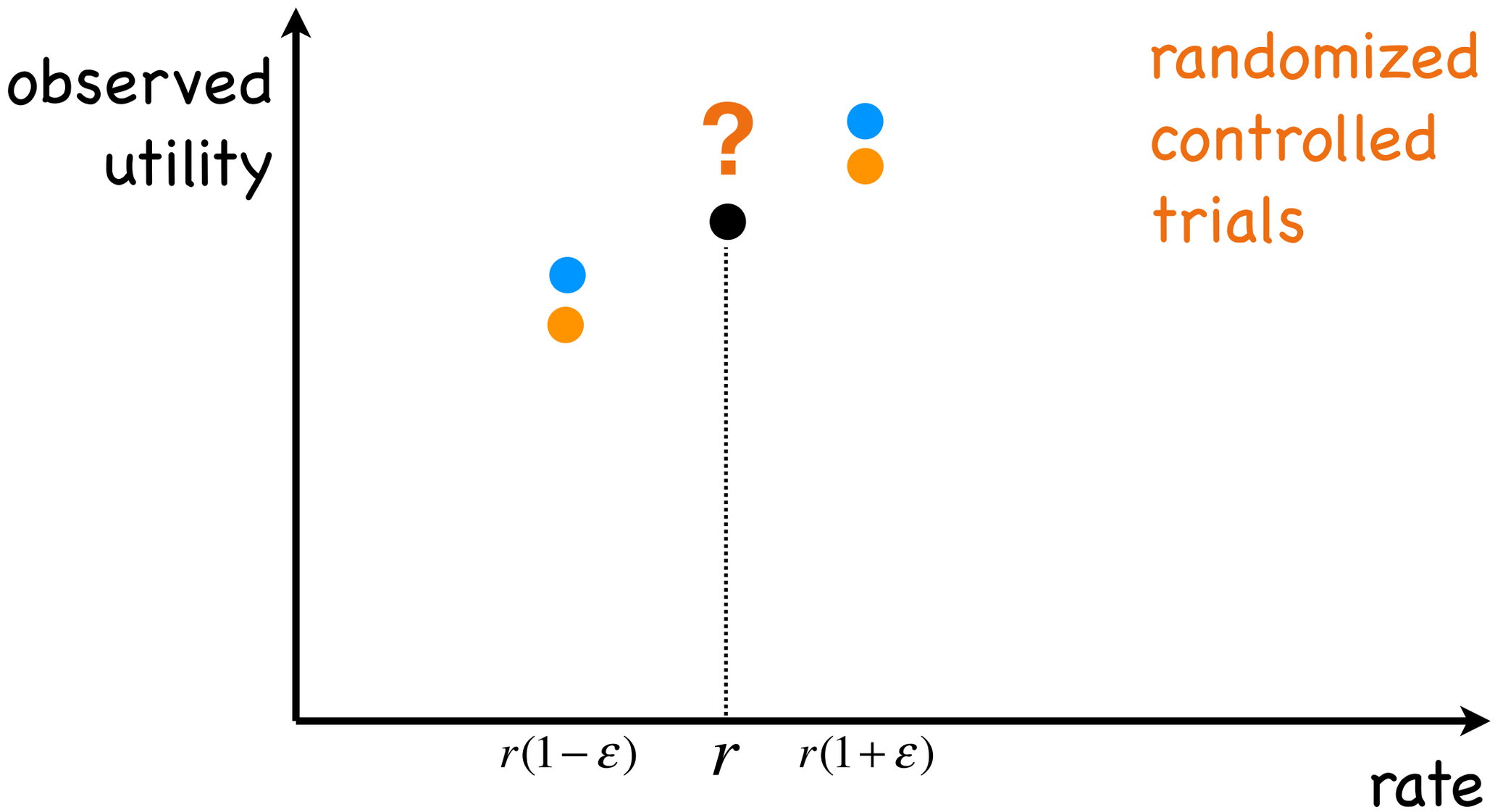
Performance Oriented Control



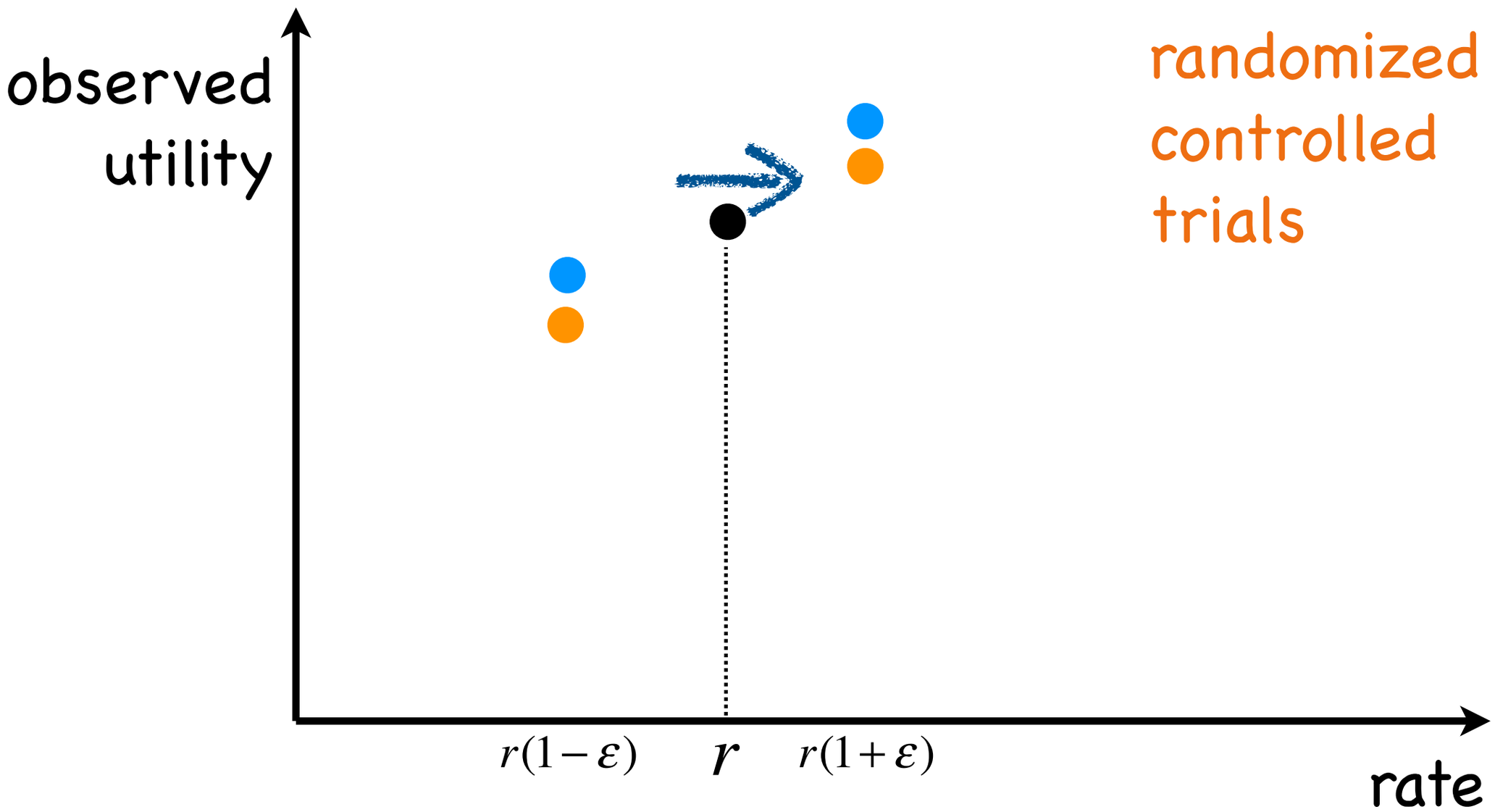
Performance Oriented Control



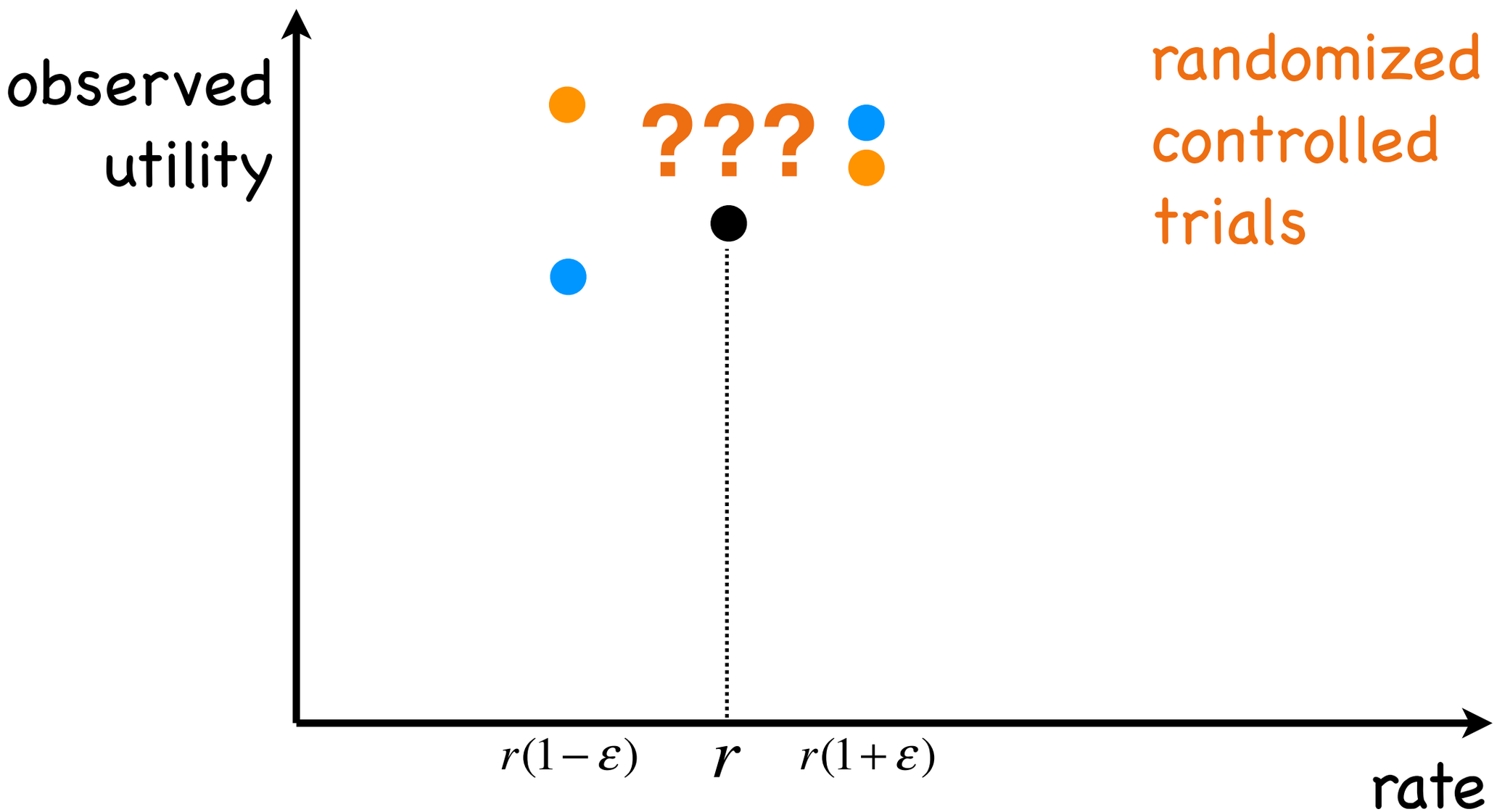
Performance Oriented Control



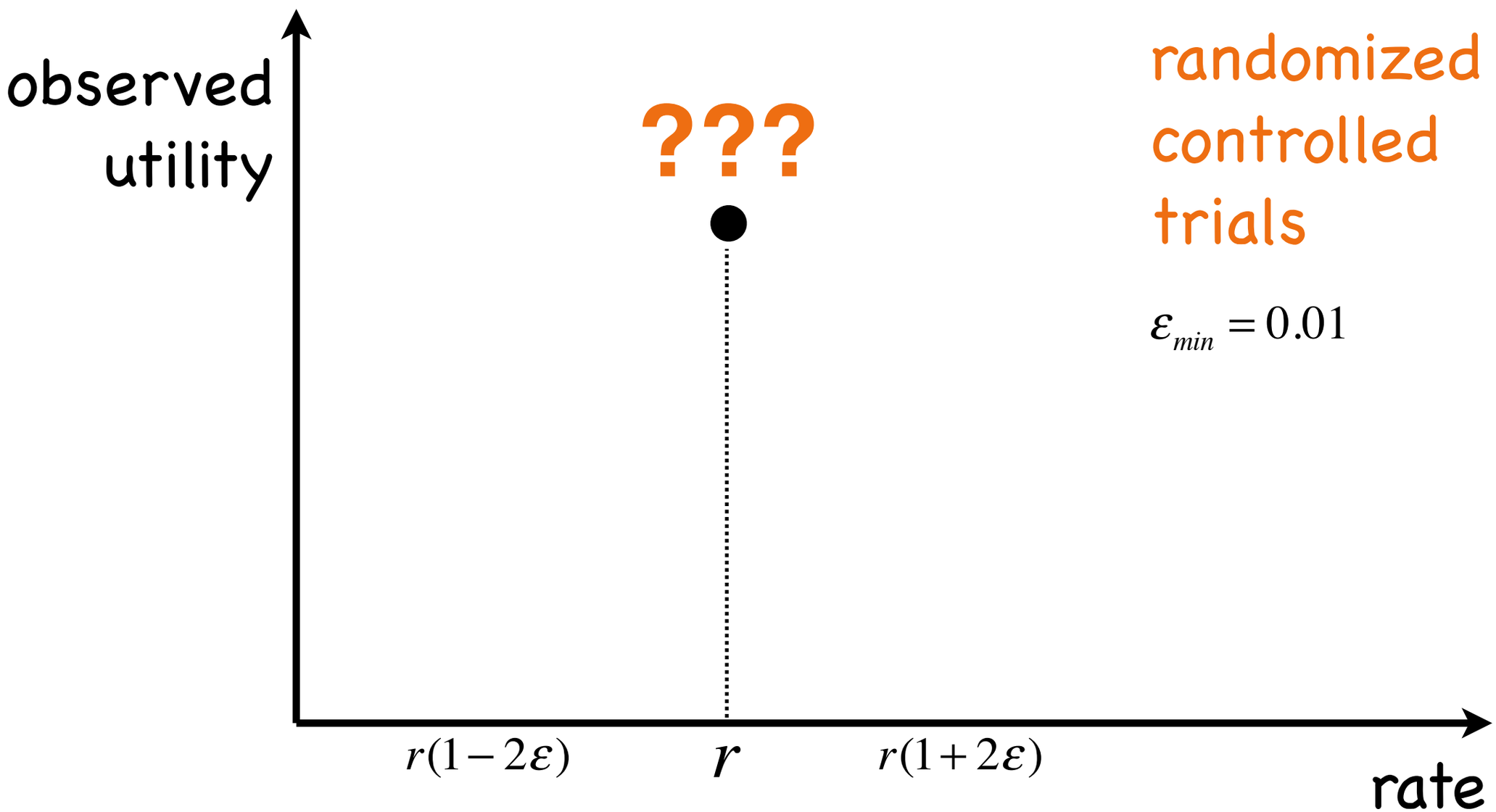
Performance Oriented Control



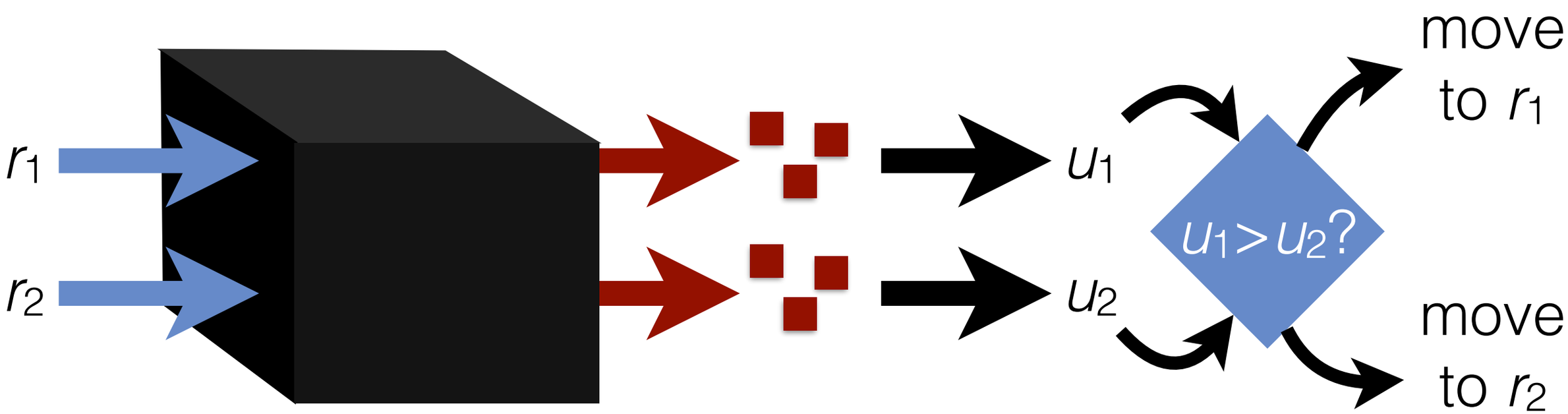
Performance Oriented Control



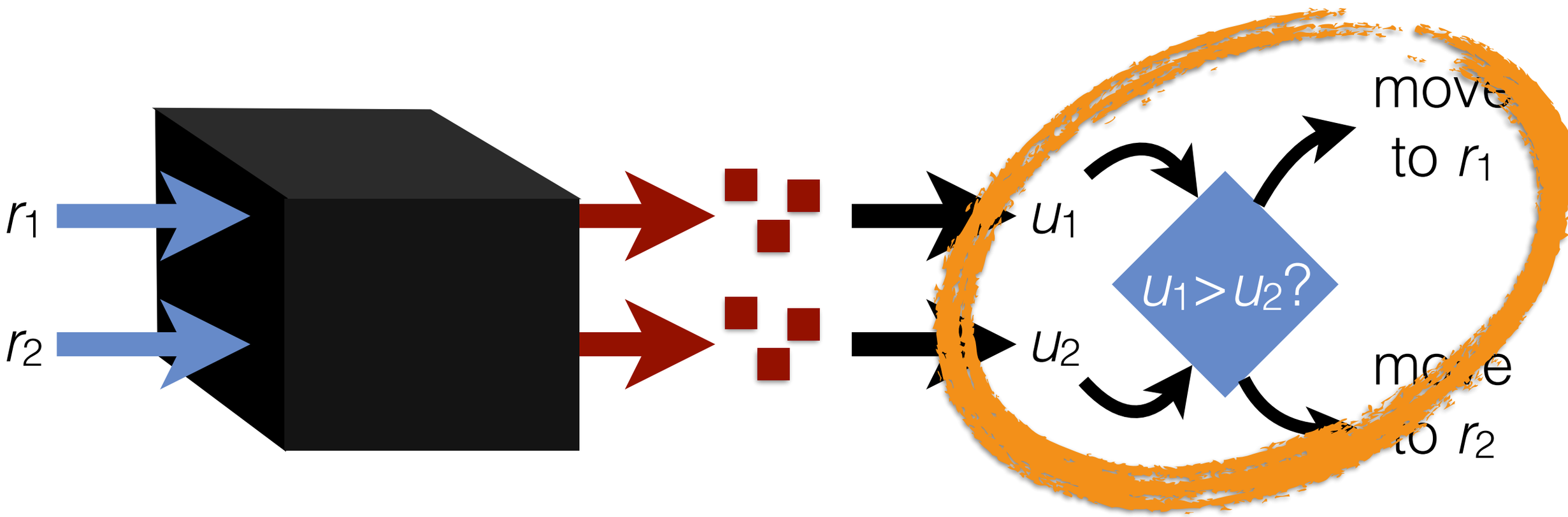
Performance Oriented Control



Where is Congestion Control?

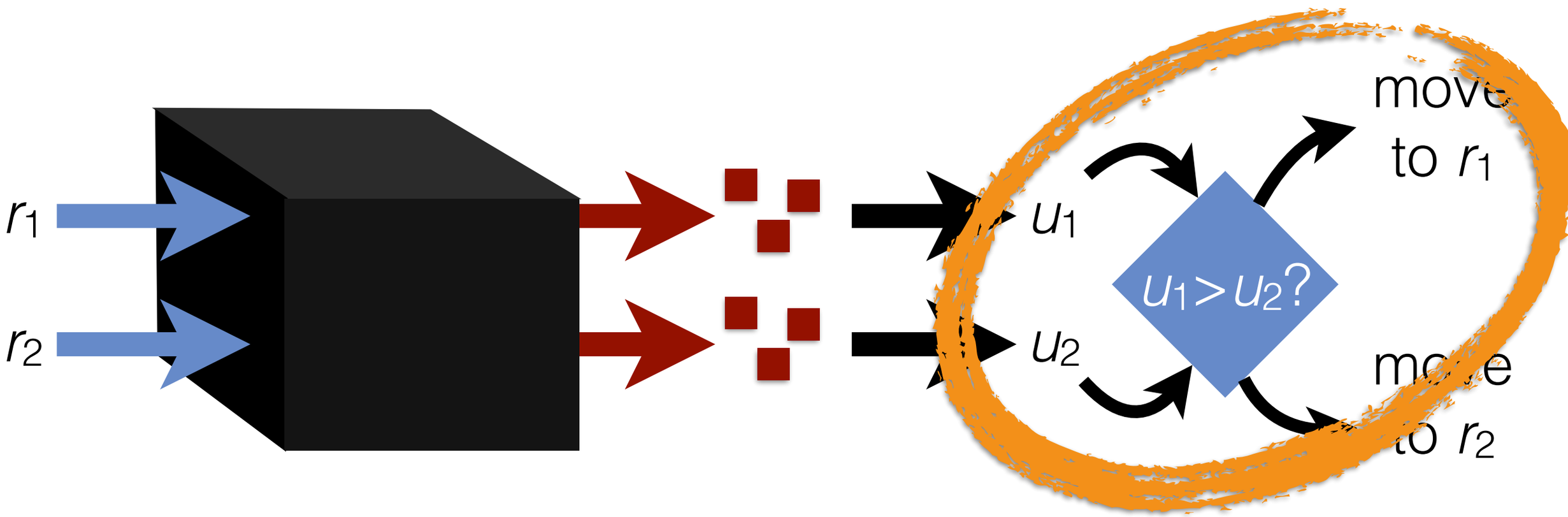


Where is Congestion Control?



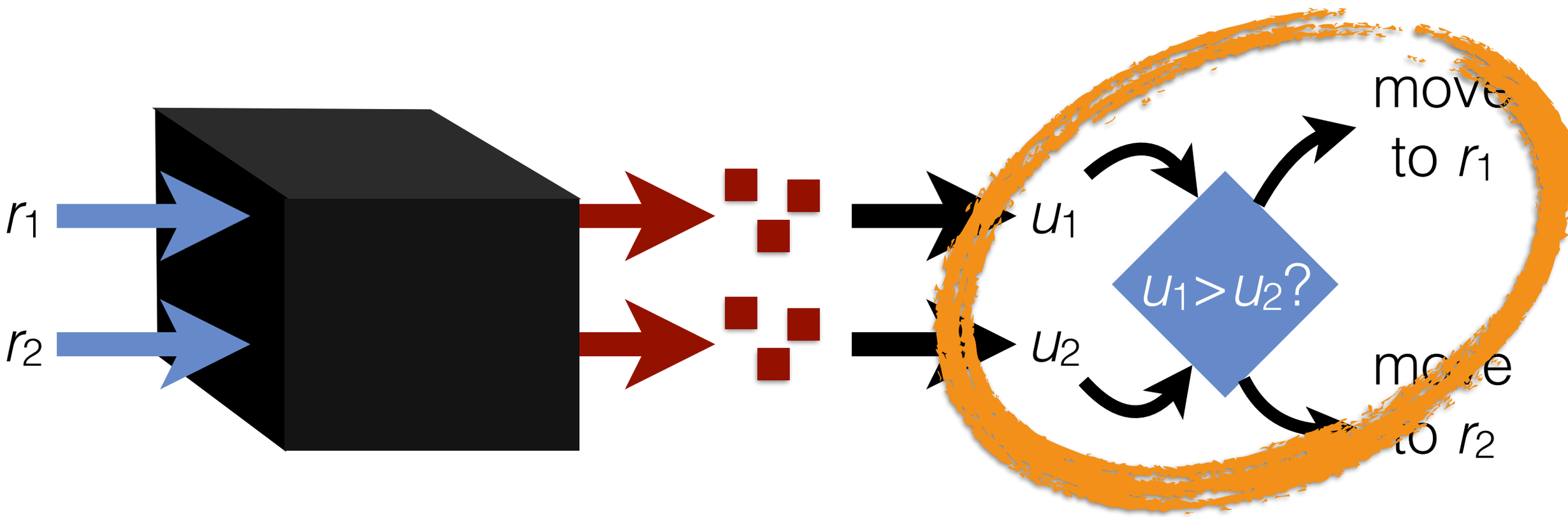
Selfishly maximizing utility

Where is Congestion Control?



Selfishly maximizing utility
=> non-cooperative game

Where is Congestion Control?



Selfishly maximizing utility
=> non-cooperative game

Do we converge to a fair Nash equilibrium?

A class of utility functions converge to a fair and efficient Nash Equilibrium

A class of utility functions converge to a fair and efficient Nash Equilibrium

$$u_i(x) = T_i \qquad - x_i * L_i$$

A class of utility functions converge to a fair and efficient Nash Equilibrium

$$u_i(x) = T_i - x_i * L_i$$

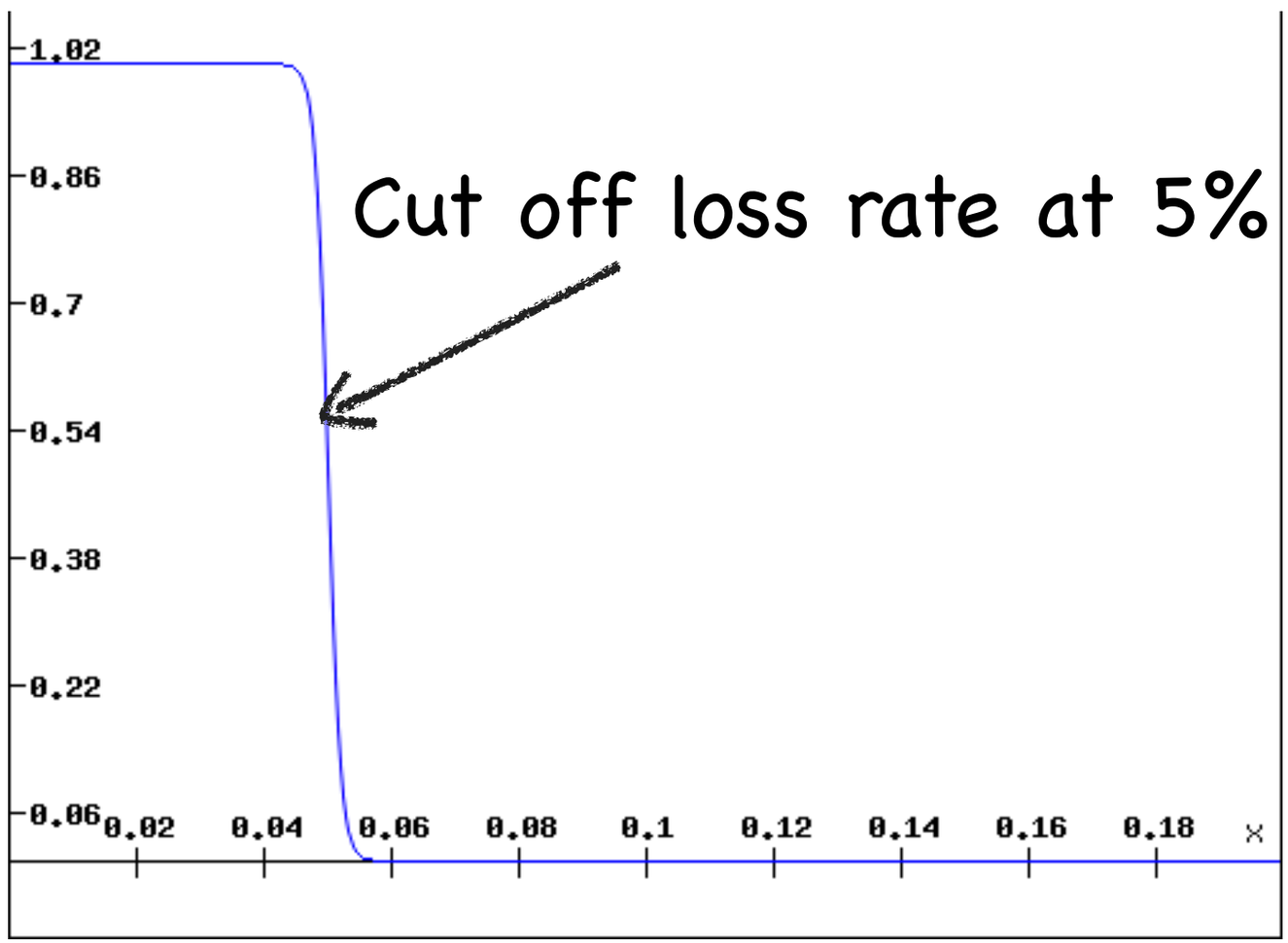
throughput

sending rate

observed loss rate

A class of utility functions converge to a fair and efficient Nash Equilibrium

$$u_i(x) = T_i * sigmoid(L_i - 0.05) - x_i * L_i$$



Loss Rate

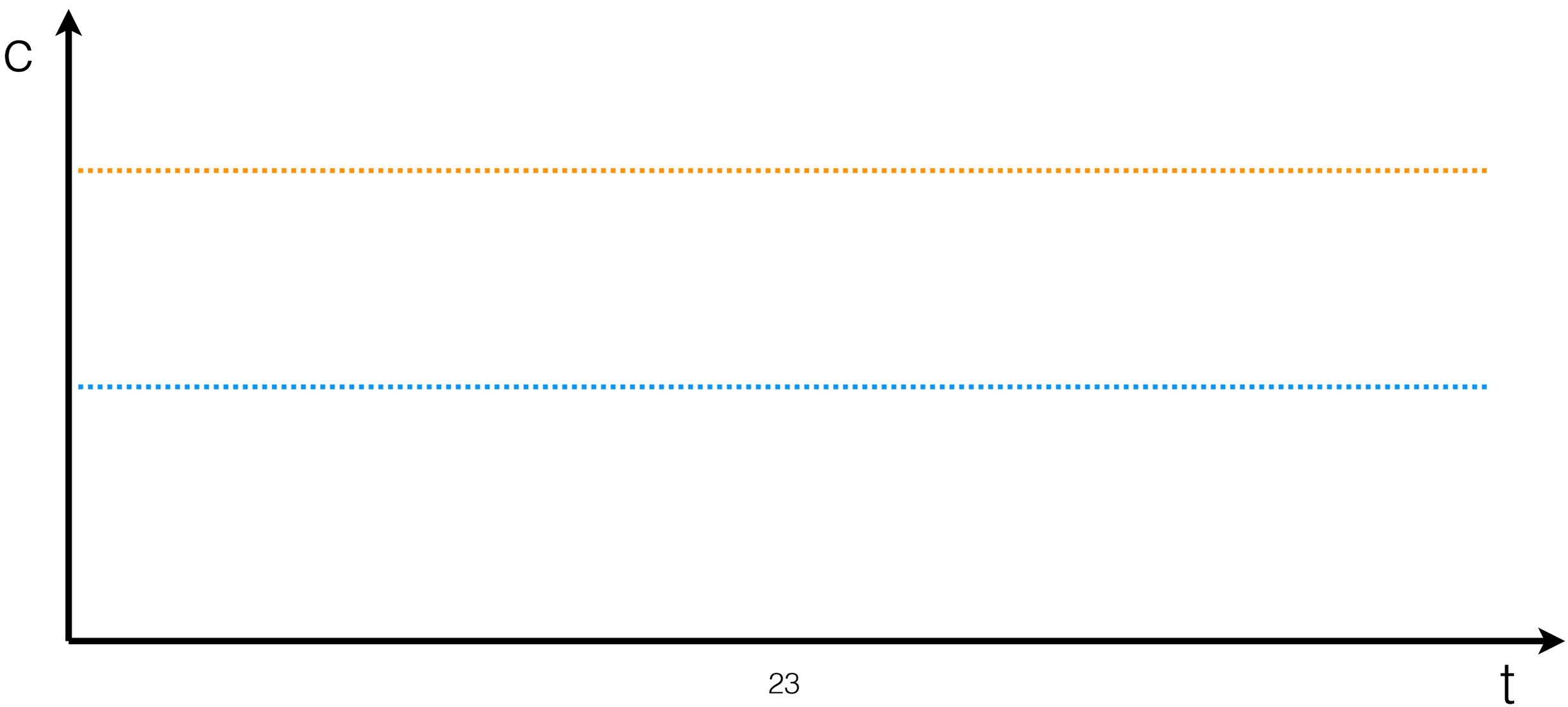
TCP Dynamics

TCP Dynamics

AIMD as a "hack" to asymptotic fairness

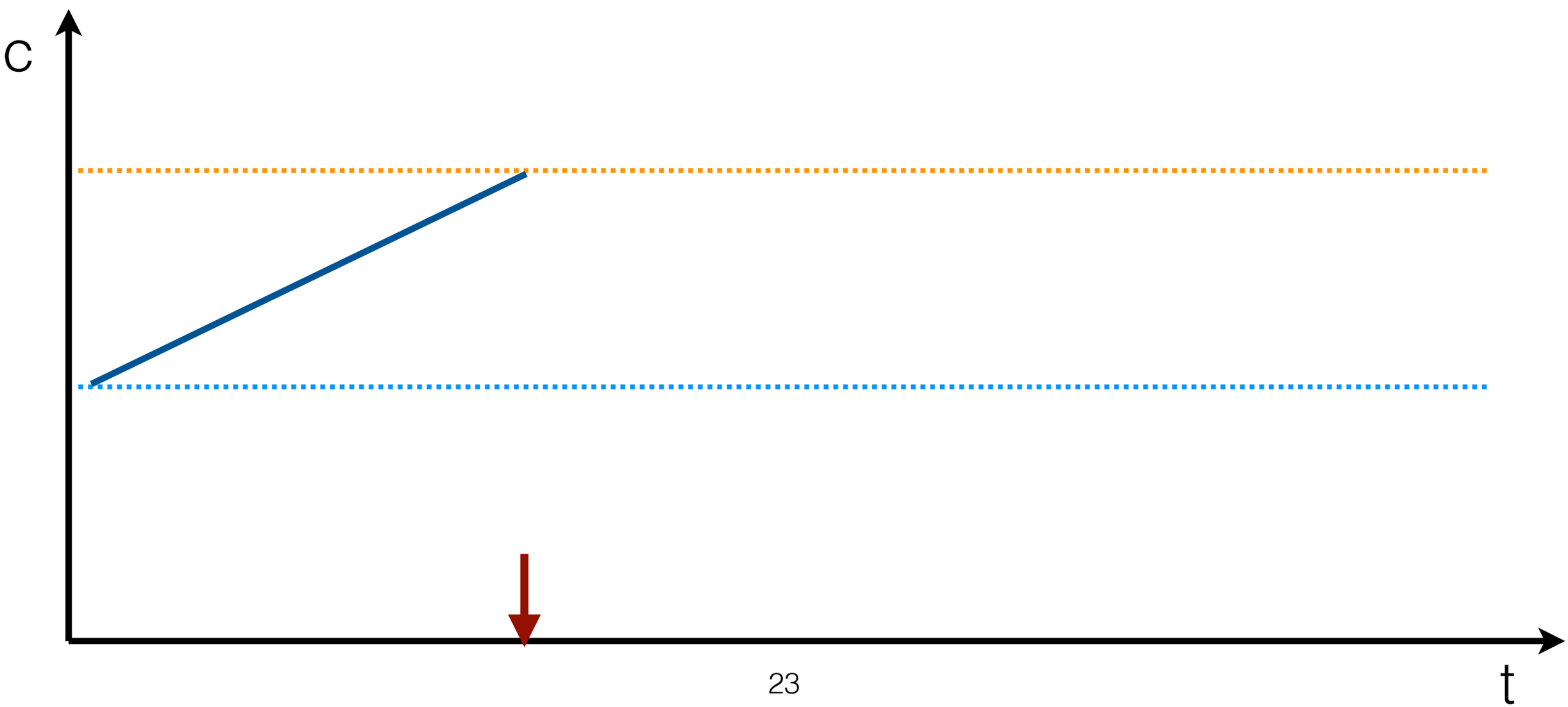
TCP Dynamics

AIMD as a "hack" to asymptotic fairness



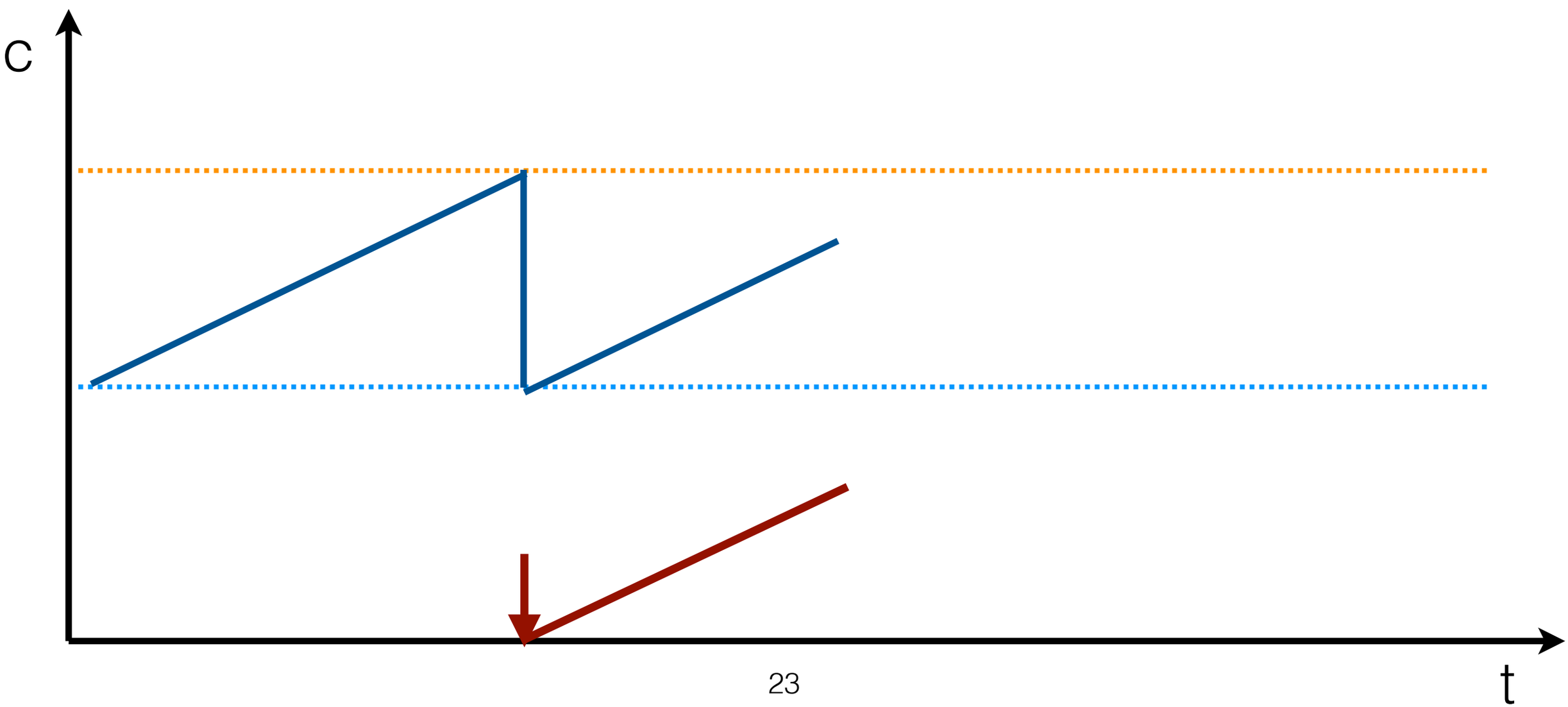
TCP Dynamics

AIMD as a "hack" to asymptotic fairness



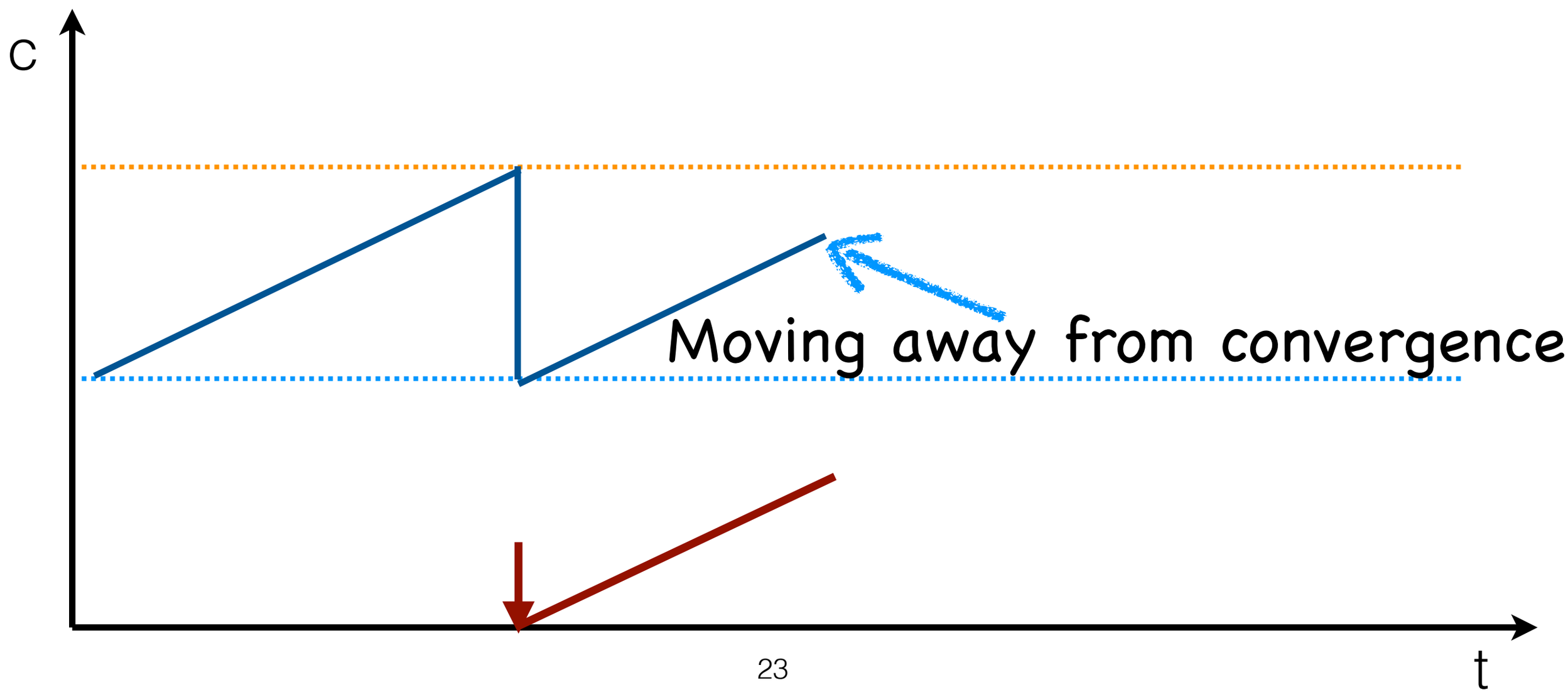
TCP Dynamics

AIMD as a "hack" to asymptotic fairness

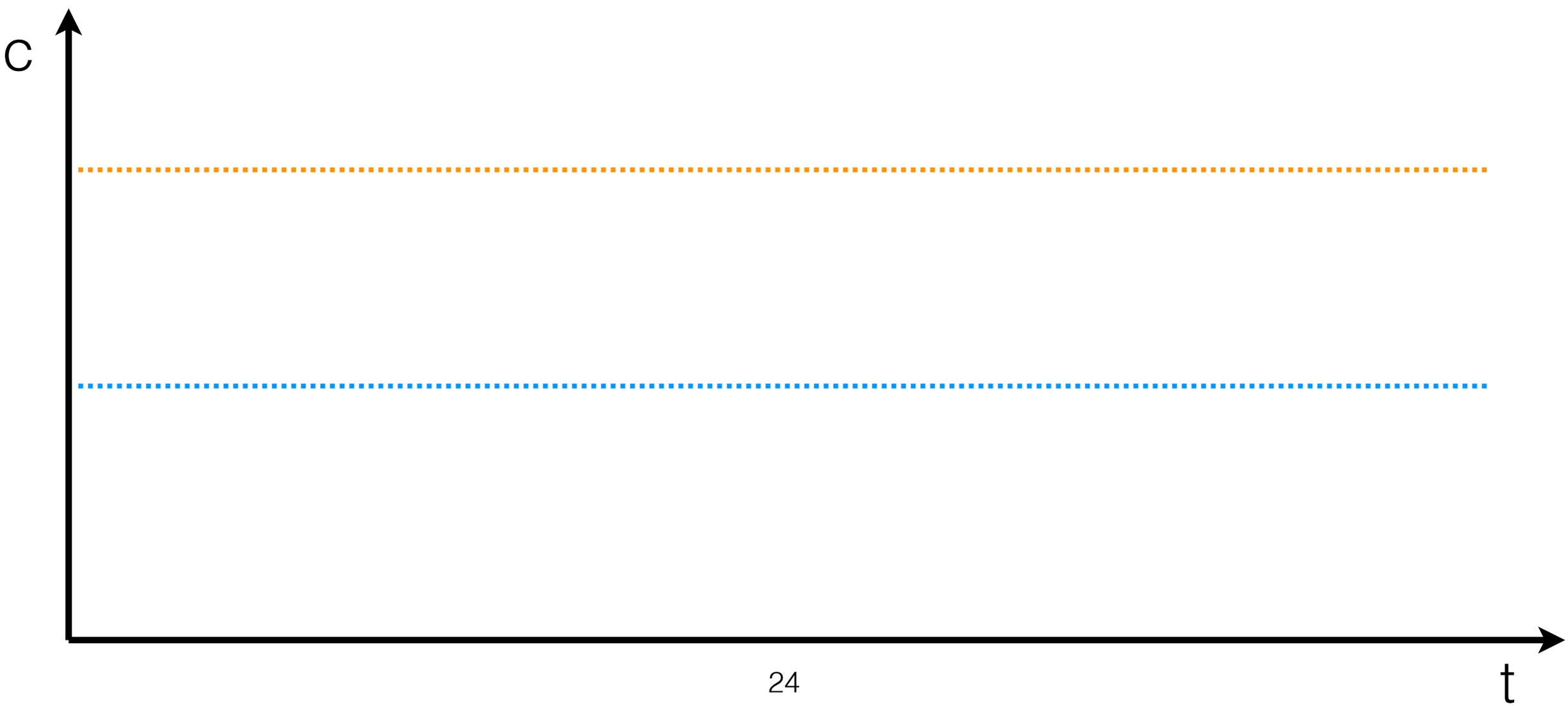


TCP Dynamics

AIMD as a "hack" to asymptotic fairness

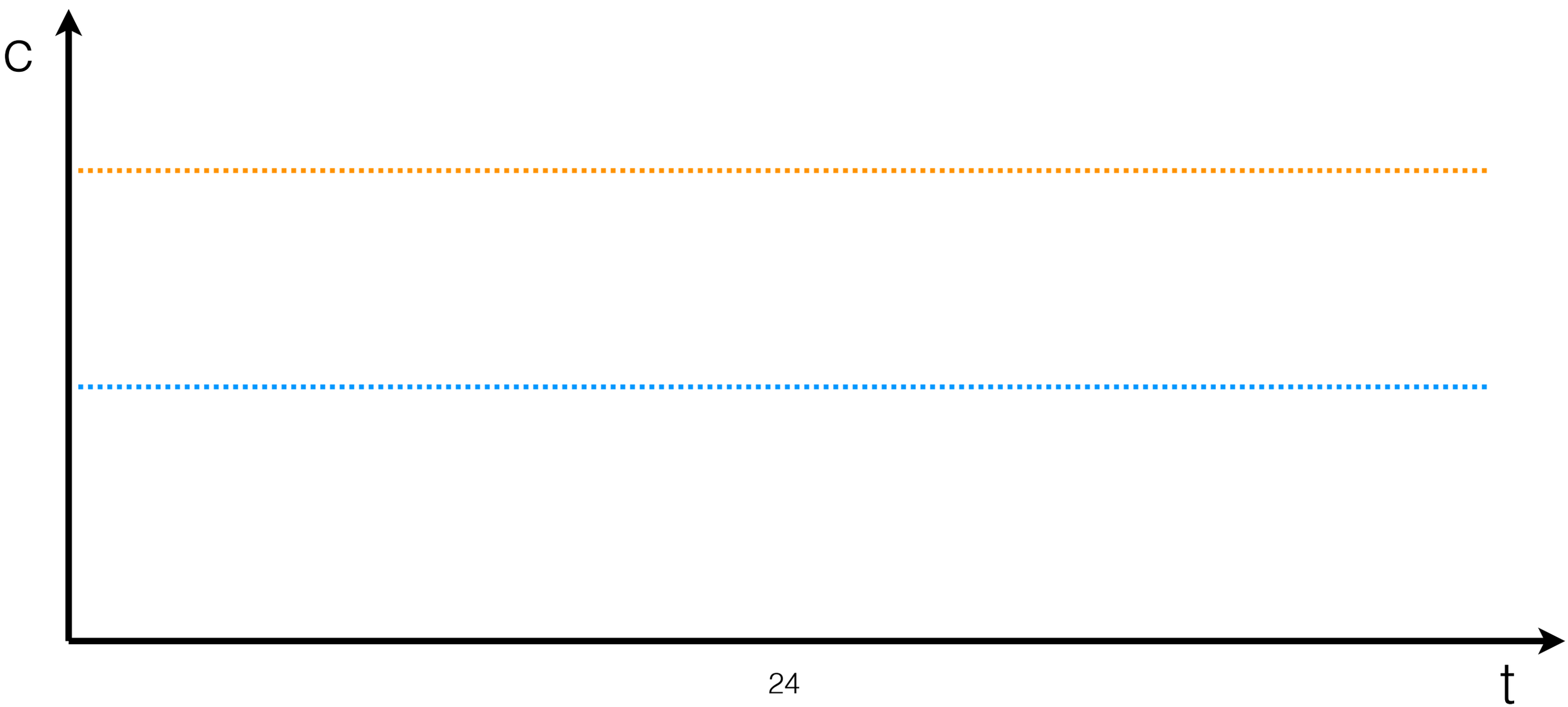


PCC Dynamics



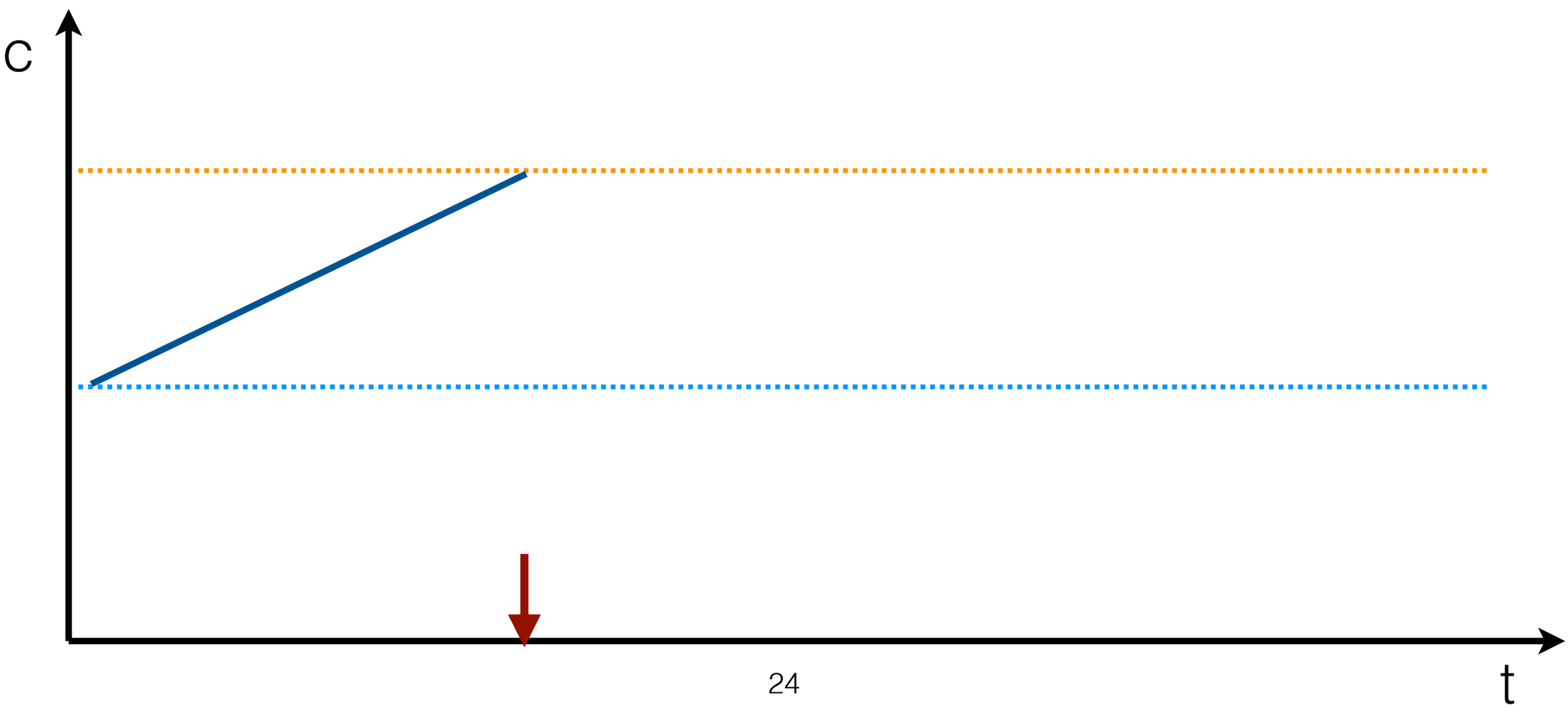
PCC Dynamics

PCC does not need AIMD because it looks at real performance



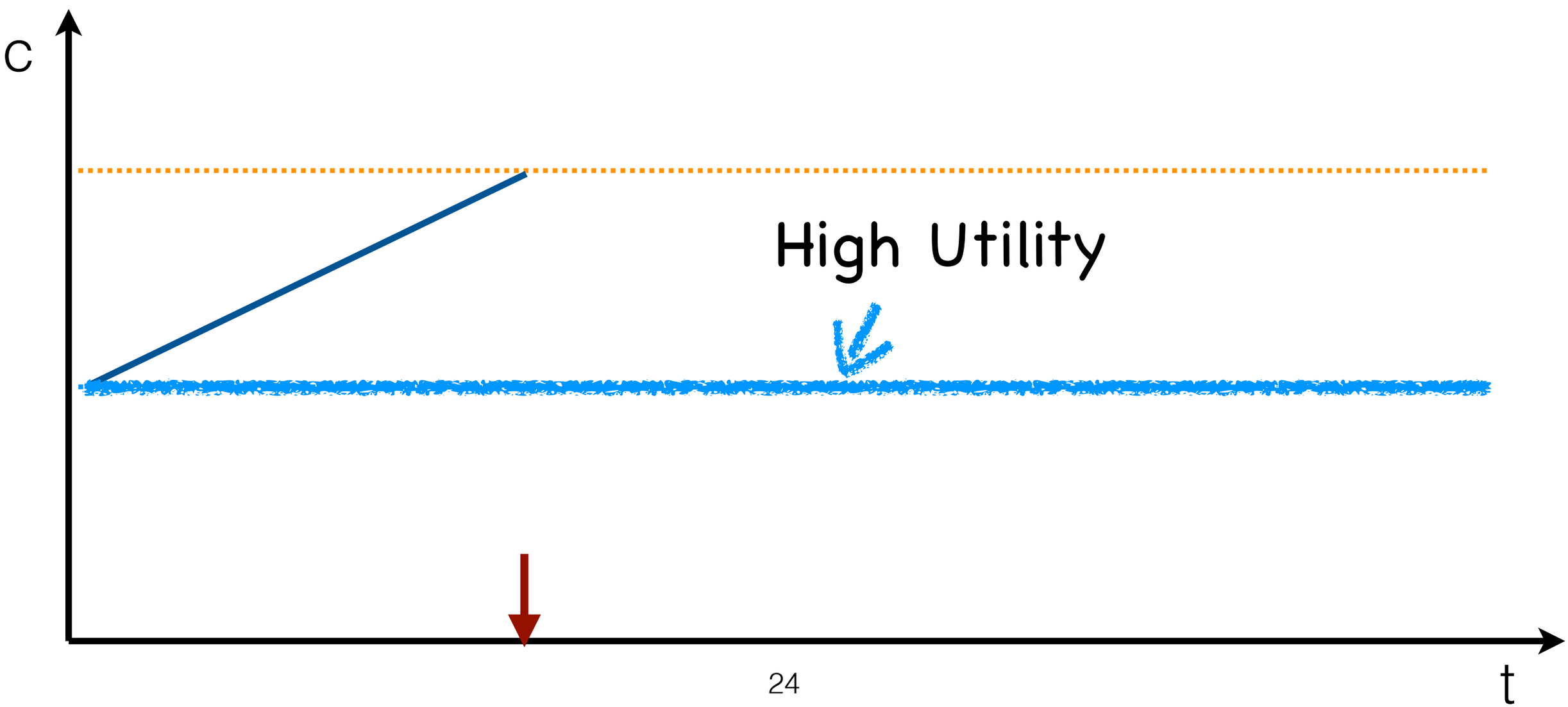
PCC Dynamics

PCC does not need AIMD because it looks at real performance



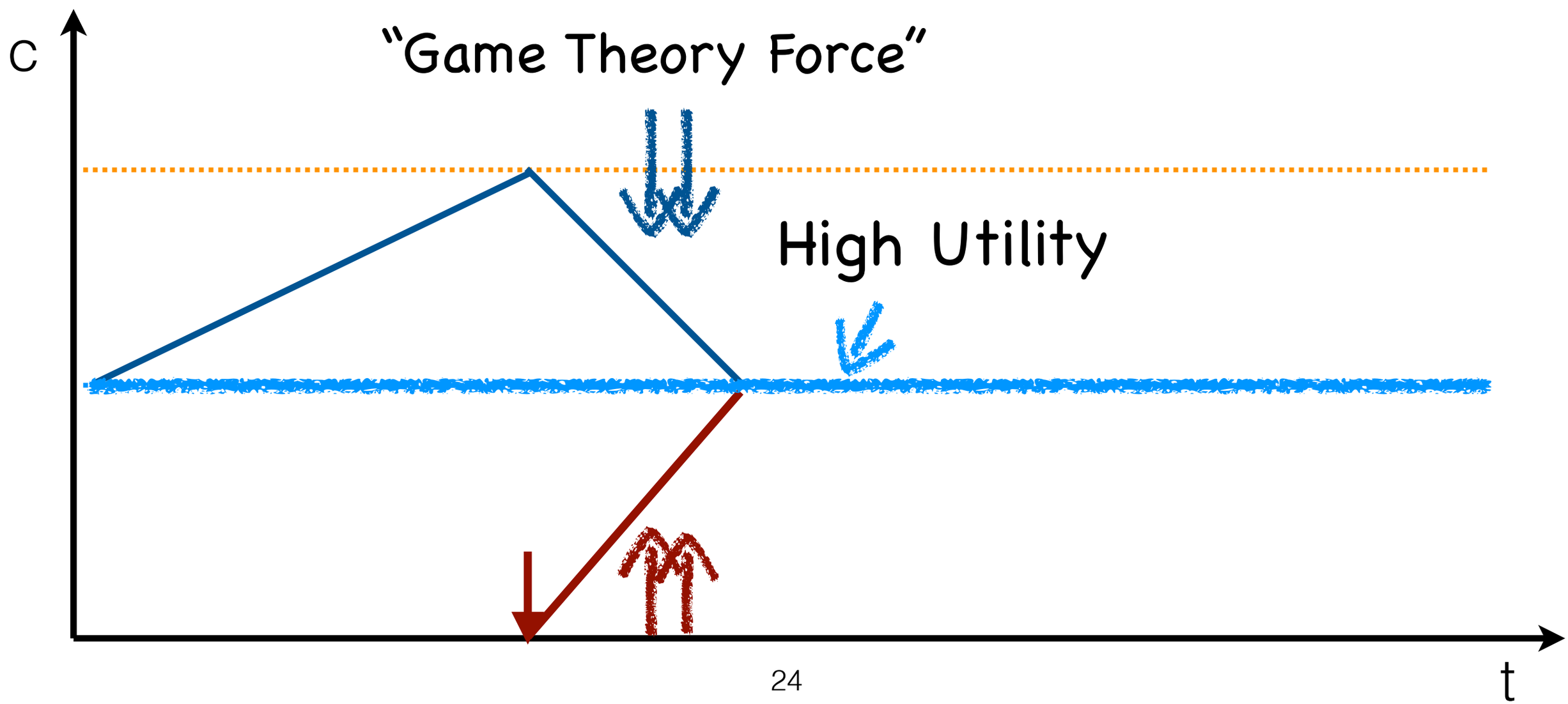
PCC Dynamics

PCC does not need AIMD because it looks at real performance



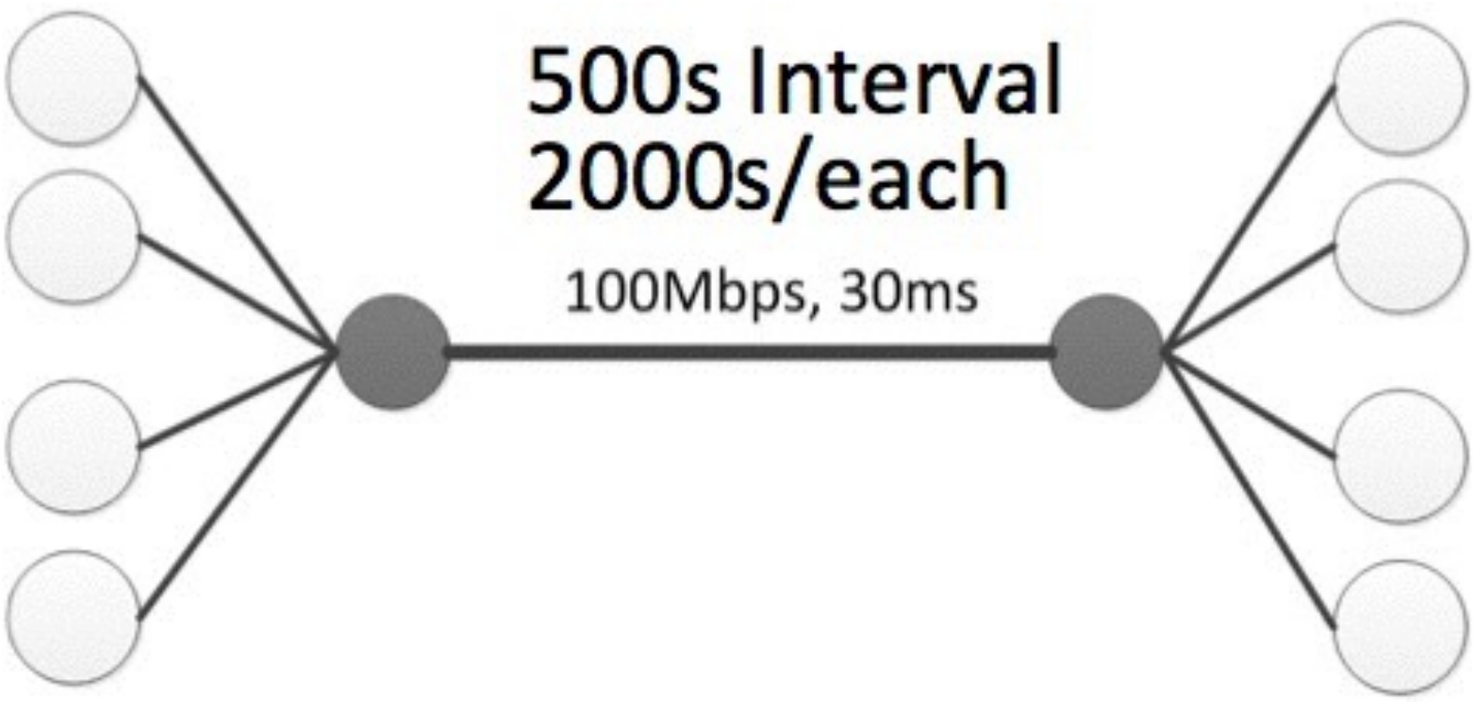
PCC Dynamics

PCC does not need AIMD because it looks at real performance

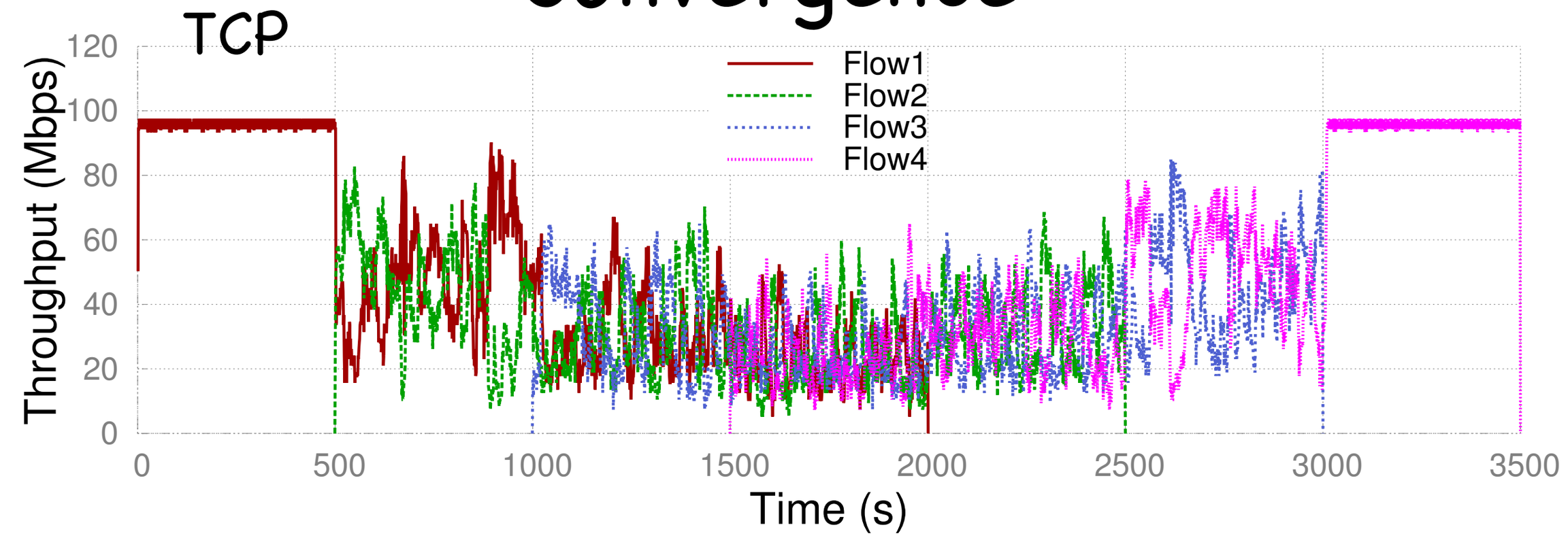


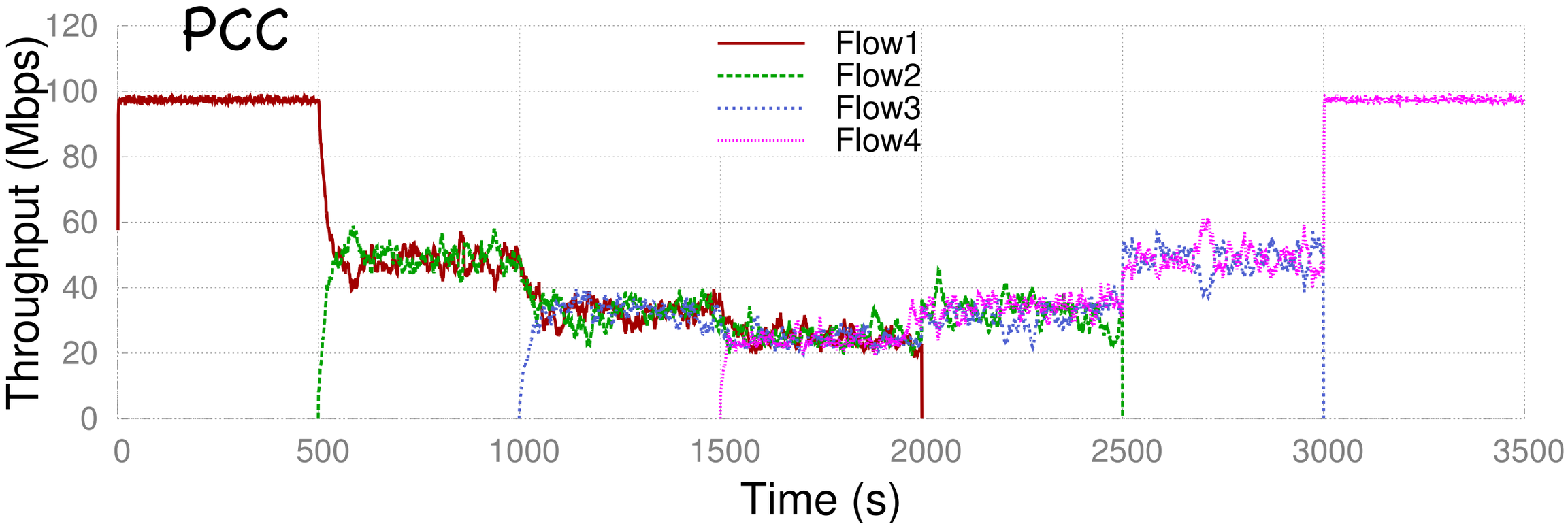
Convergence

Convergence



Convergence





Deployment

- No hardwired support, packet header, protocol change needed
- Where to deploy
  - CDN backbone, Inter-data center, dedicated scientific nw
  - "In the Wild"?

TCP Friendliness

TCP Friendliness

- Is PCC TCP-friendly?

TCP Friendliness

- Is PCC TCP-friendly?

TCP Friendliness

TCP Friendliness

- PCC's default utility function is not TCP Friendly

TCP Friendliness

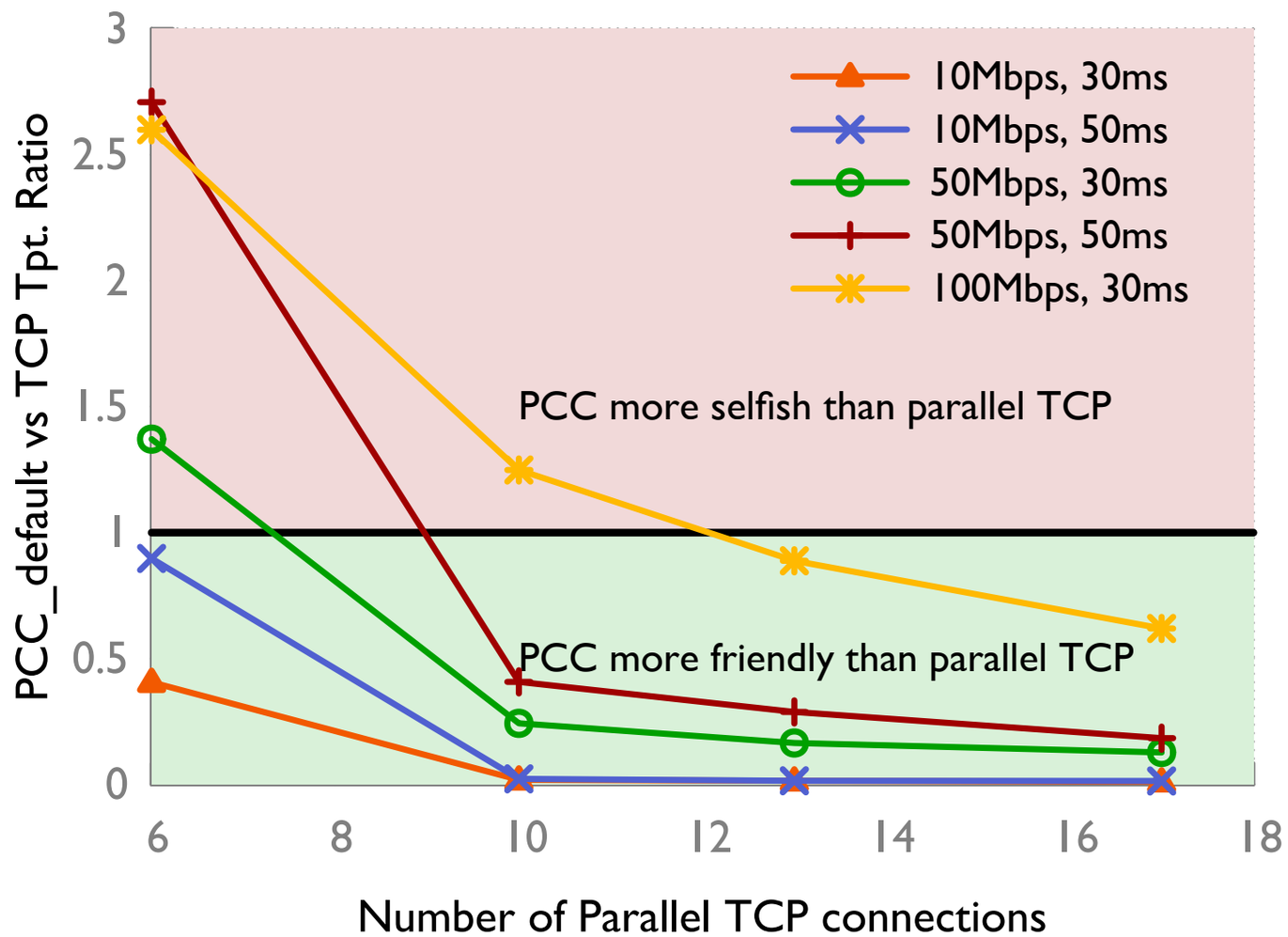
- PCC's default utility function is not TCP Friendly

But not **that** bad

TCP Friendliness

- PCC's default utility function is not TCP Friendly

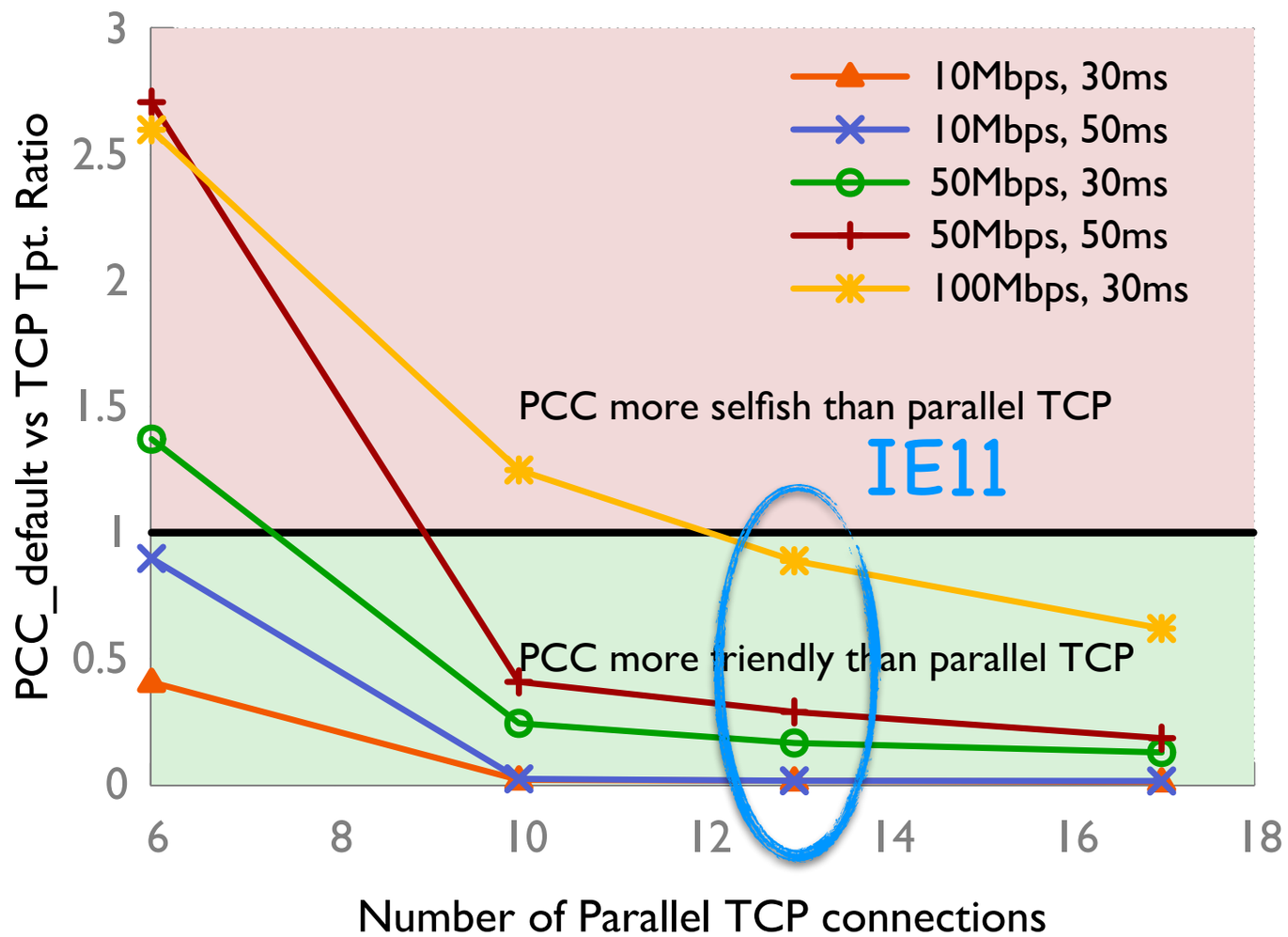
But not **that** bad



TCP Friendliness

- PCC's default utility function is not TCP Friendly

But not **that** bad



TCP Friendliness

- PCC's default utility function is not TCP Friendly
- Different utility functions can be a solution

TCP Friendliness

- PCC's default utility function is not TCP Friendly
- Different utility functions can be a solution

TCP vs TCP TCP vs PCC

TCP Friendliness

- PCC's default utility function is not TCP Friendly
- Different utility functions can be a solution

TCP vs TCP → TCP vs PCC

TCP Friendliness

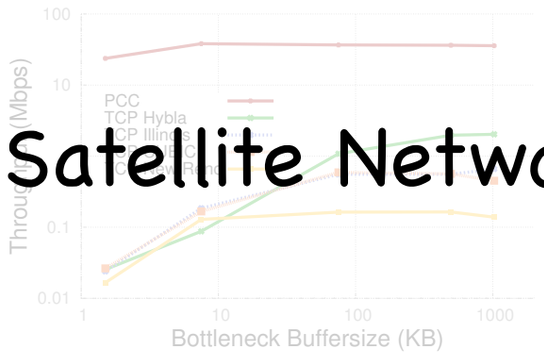
- PCC's default utility function is not TCP Friendly
- Different utility functions can be a solution

TCP vs TCP → TCP vs PCC

		30ms	60ms	90ms
$\beta = 10$	10Mbit/s	0.94	0.75	0.67
	50Mbit/s	0.74	0.73	0.81
	90Mbit/s	0.89	0.91	1.01
$\beta = 100$	10Mbit/s	0.71	0.58	0.63
	50Mbit/s	0.56	0.58	0.54
	90Mbit/s	0.63	0.62	0.88

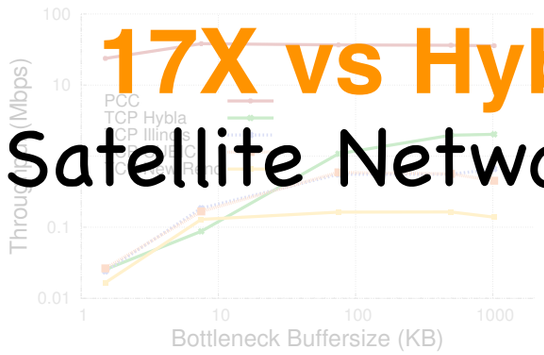
Consistent High Performance

Consistent High Performance



Satellite Networks

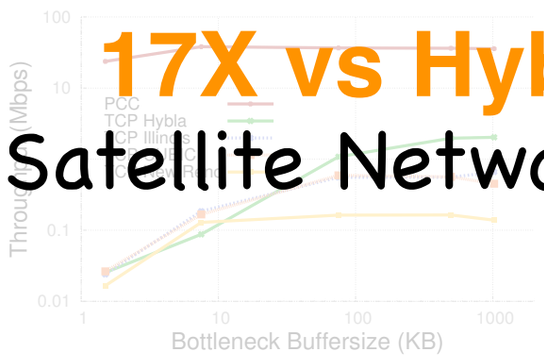
Consistent High Performance



17X vs Hybla

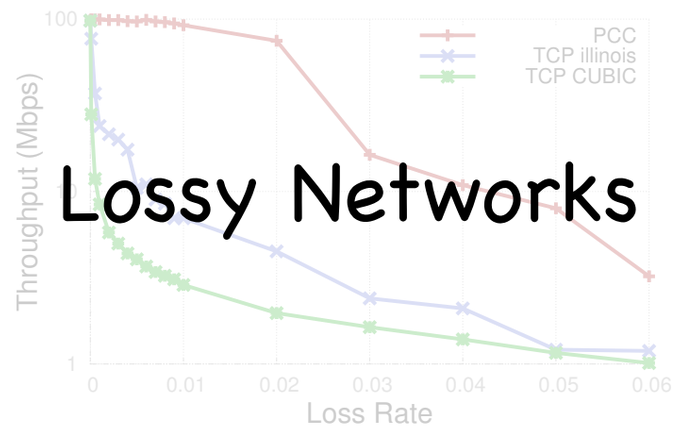
Satellite Networks

Consistent High Performance



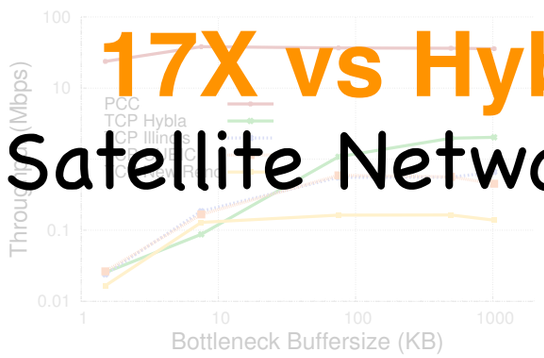
17X vs Hybla

Satellite Networks



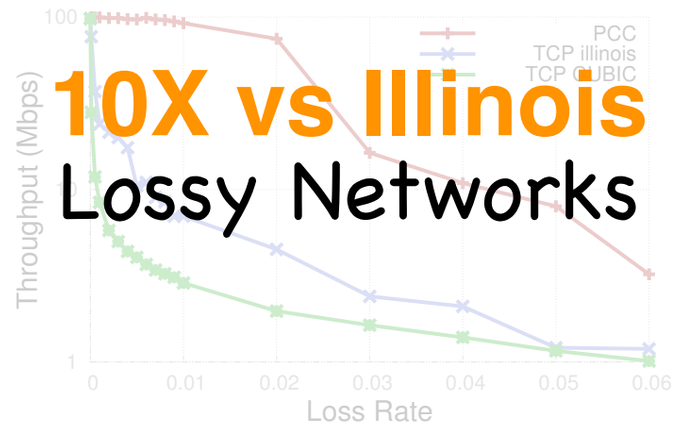
Lossy Networks

Consistent High Performance



17X vs Hybla

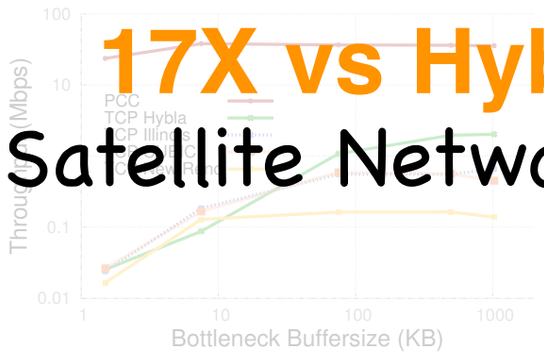
Satellite Networks



10X vs Illinois

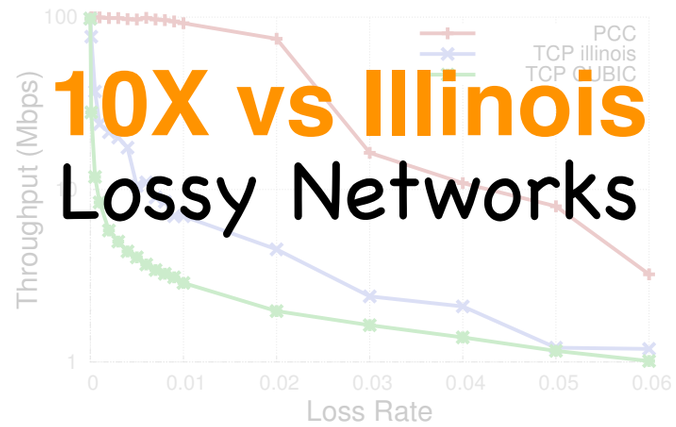
Lossy Networks

Consistent High Performance



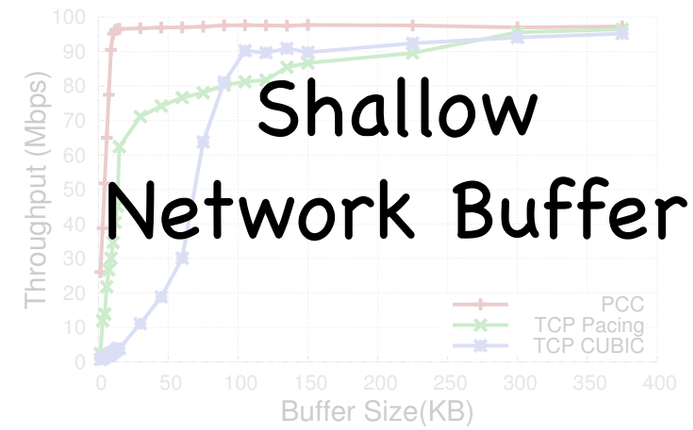
17X vs Hybla

Satellite Networks



10X vs Illinois

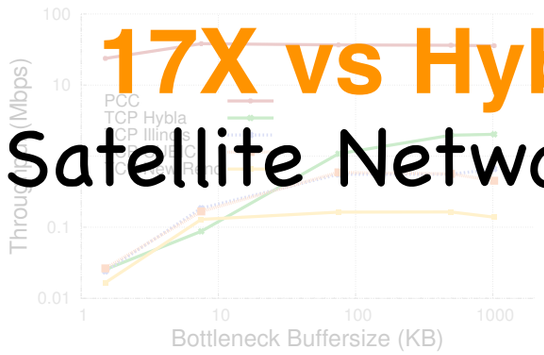
Lossy Networks



Shallow

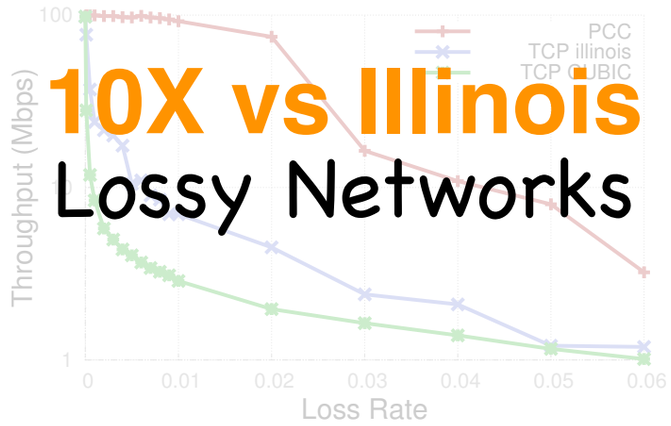
Network Buffer

Consistent High Performance



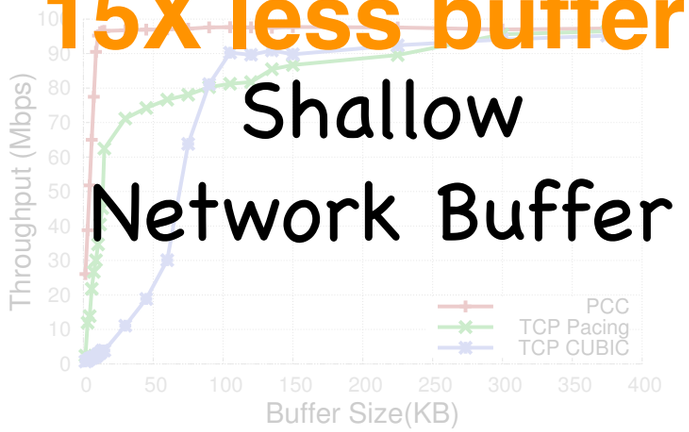
17X vs Hybla

Satellite Networks



10X vs Illinois

Lossy Networks

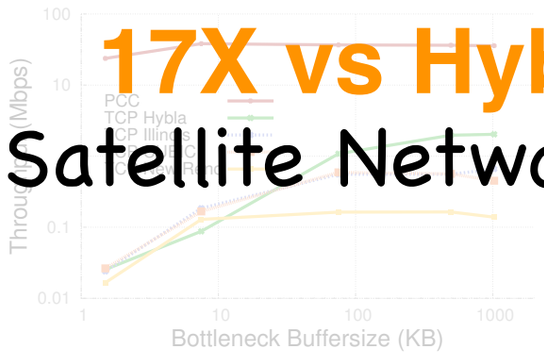


15X less buffer

Shallow

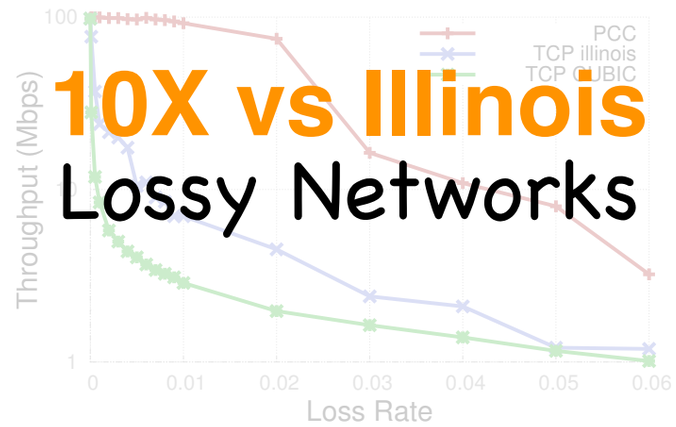
Network Buffer

Consistent High Performance



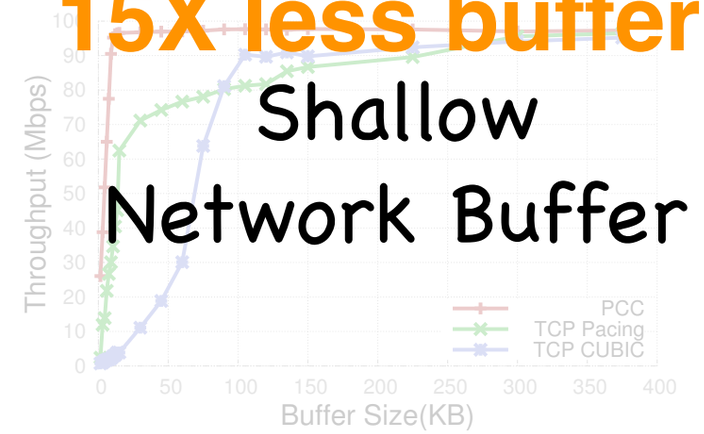
17X vs Hybla

Satellite Networks



10X vs Illinois

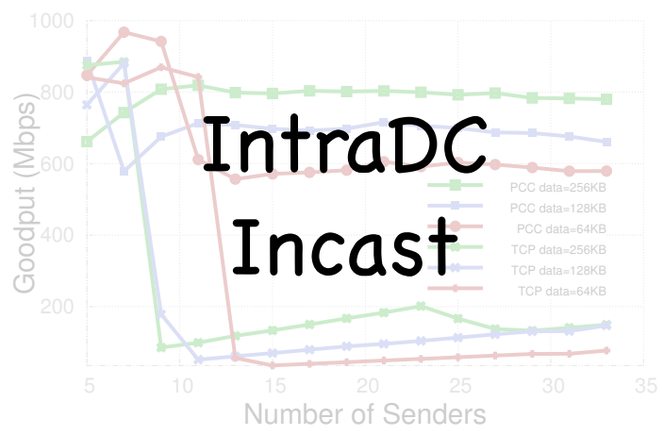
Lossy Networks



15X less buffer

Shallow

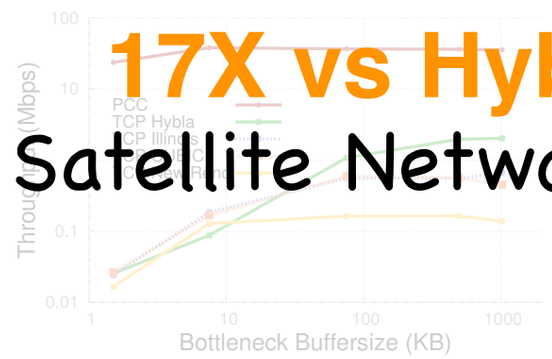
Network Buffer



IntraDC

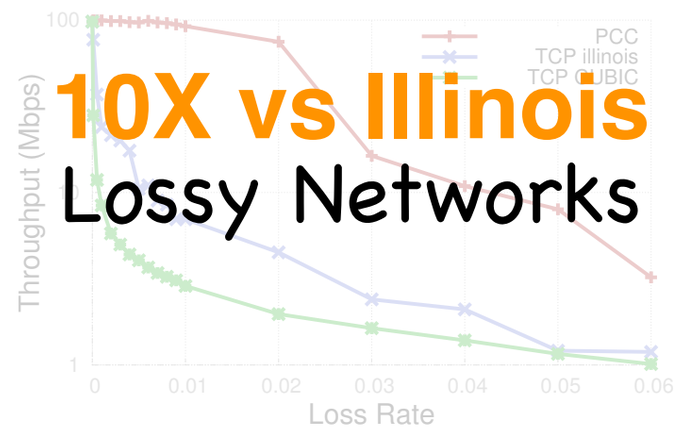
Incast

Consistent High Performance



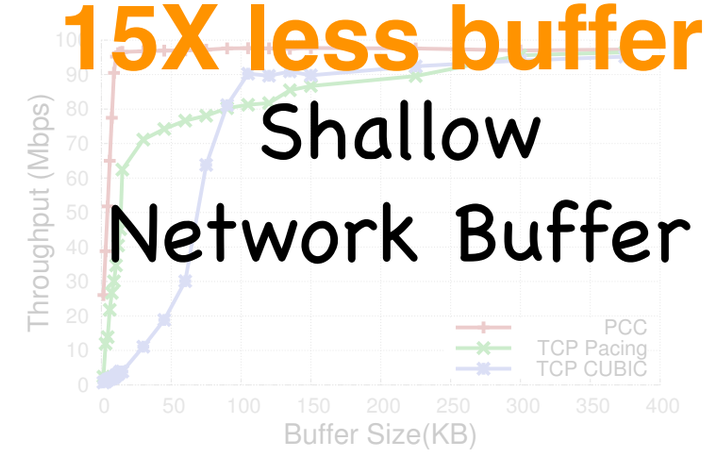
17X vs Hybla

Satellite Networks



10X vs Illinois

Lossy Networks

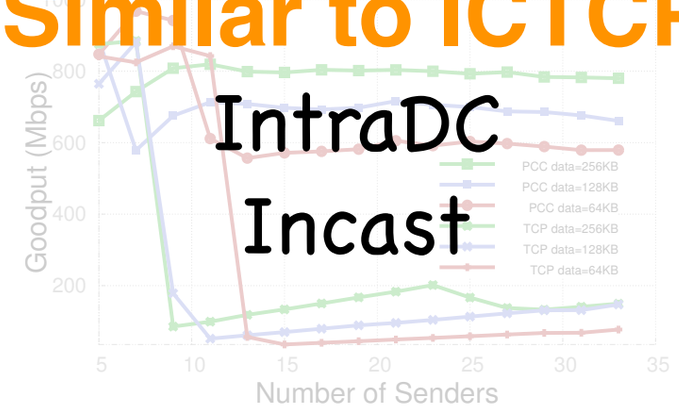


15X less buffer

Shallow

Network Buffer

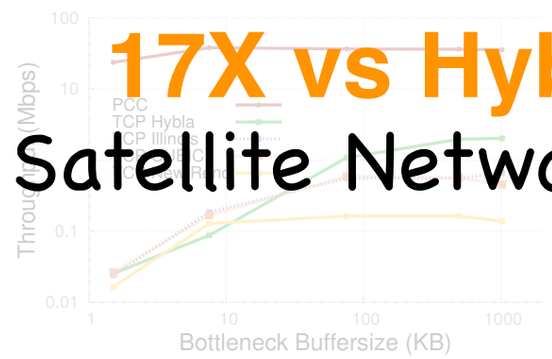
Similar to ICTCP



IntraDC

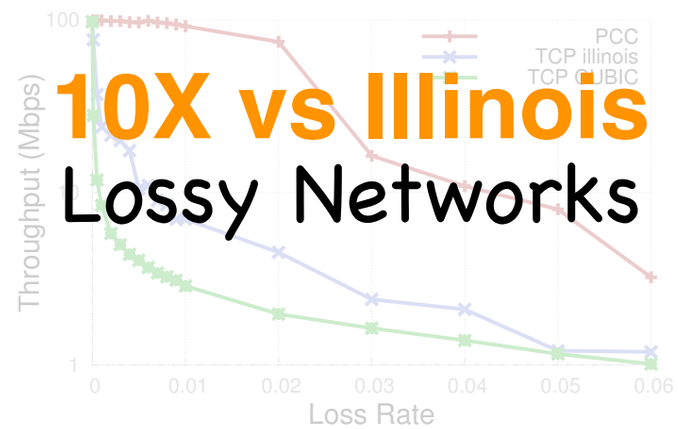
Incast

Consistent High Performance



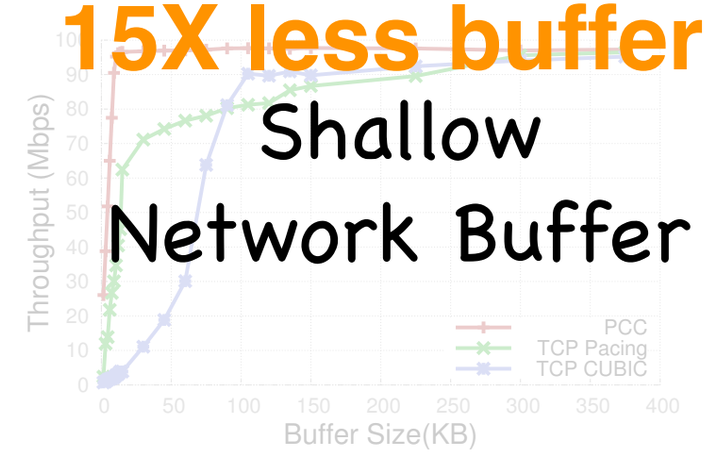
17X vs Hybla

Satellite Networks



10X vs Illinois

Lossy Networks

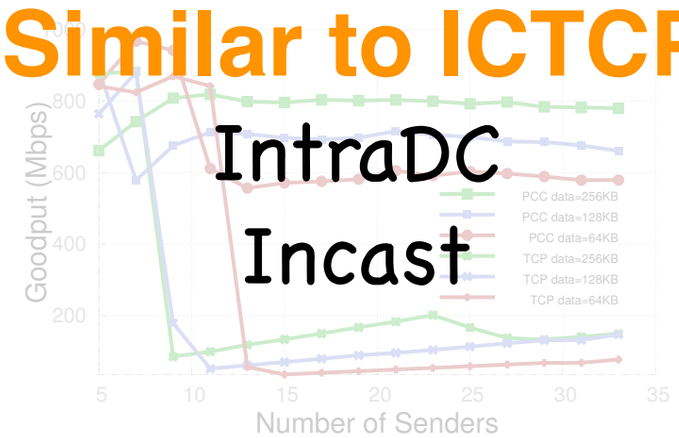


15X less buffer

Shallow

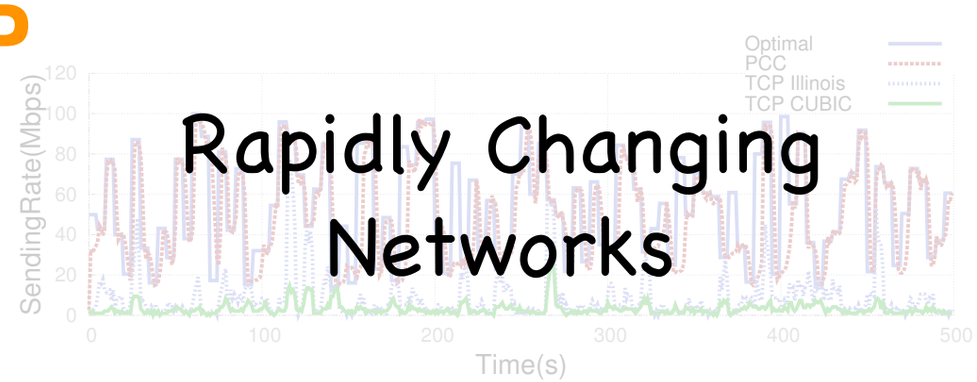
Network Buffer

Similar to ICTCP



IntraDC

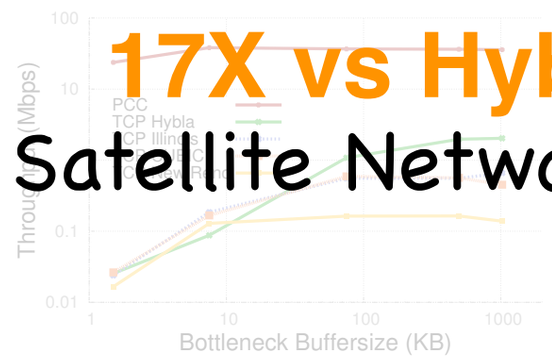
Incast



Rapidly Changing

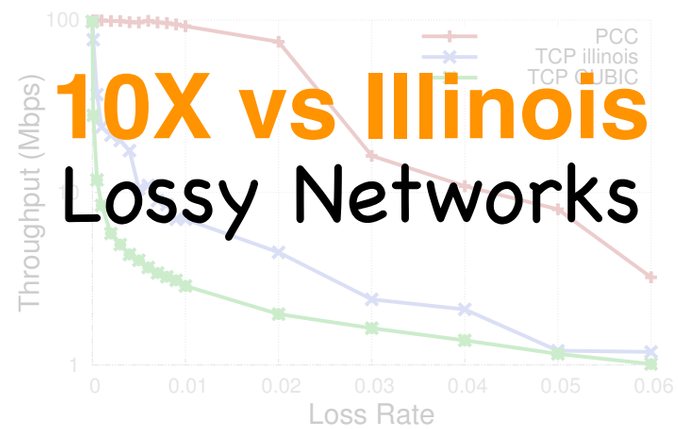
Networks

Consistent High Performance



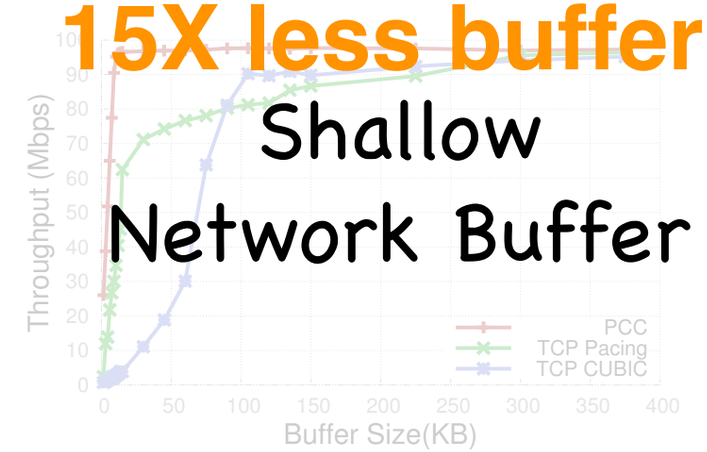
17X vs Hybla

Satellite Networks



10X vs Illinois

Lossy Networks

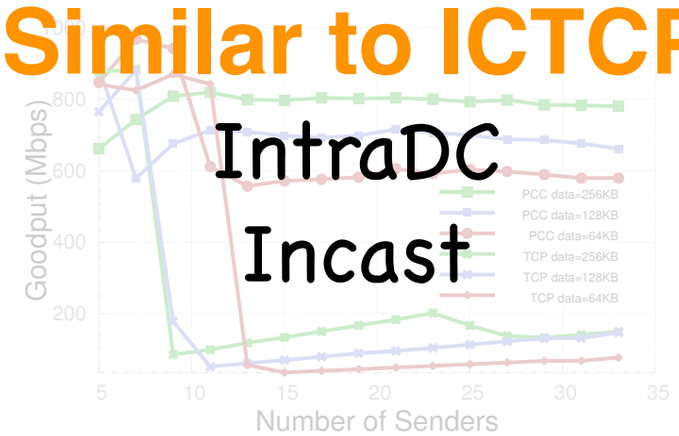


15X less buffer

Shallow

Network Buffer

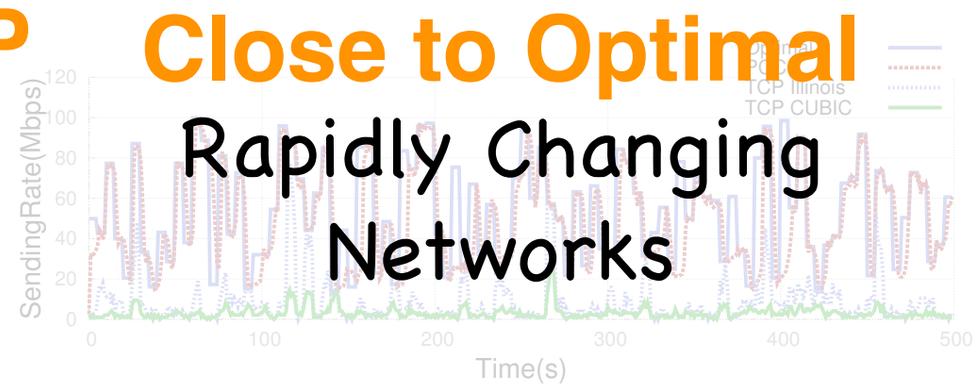
Similar to ICTCP



IntraDC

Incast

Close to Optimal



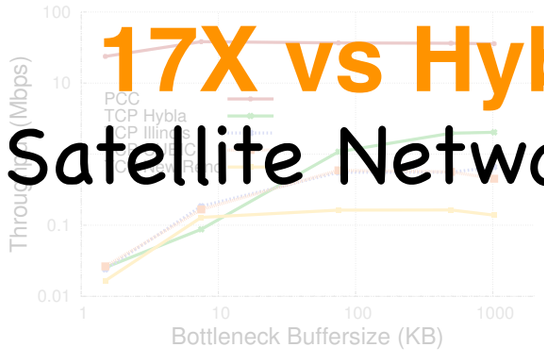
Rapidly Changing

Networks

Consistent High Performance

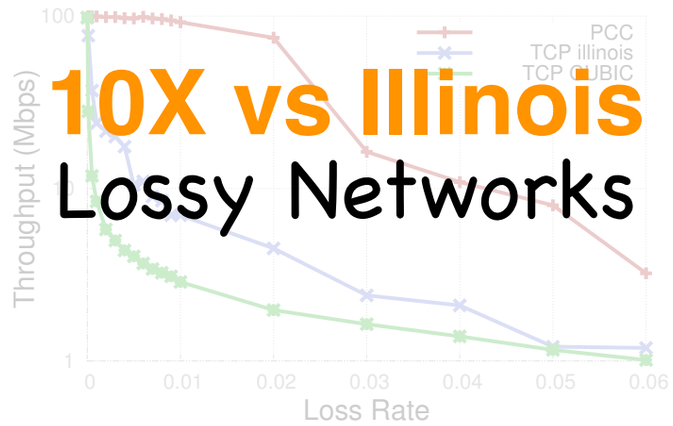
17X vs Hybla

Satellite Networks



10X vs Illinois

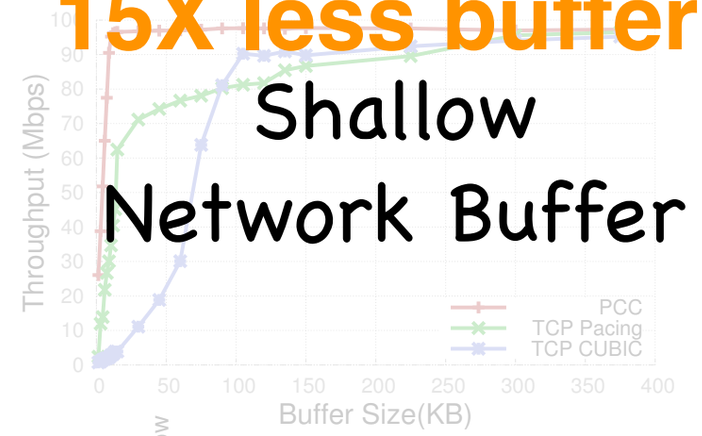
Lossy Networks



15X less buffer

Shallow

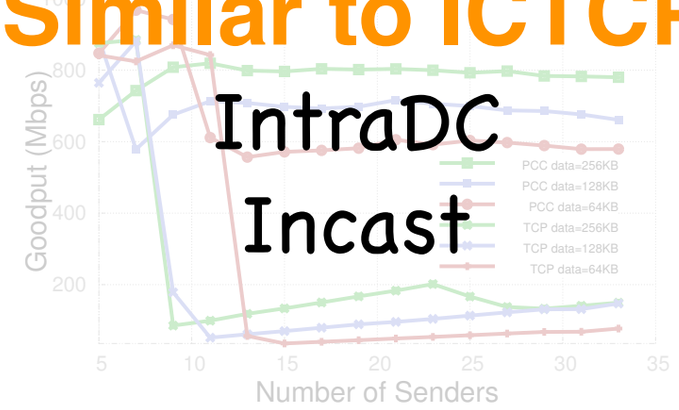
Network Buffer



Similar to ICTCP

IntraDC

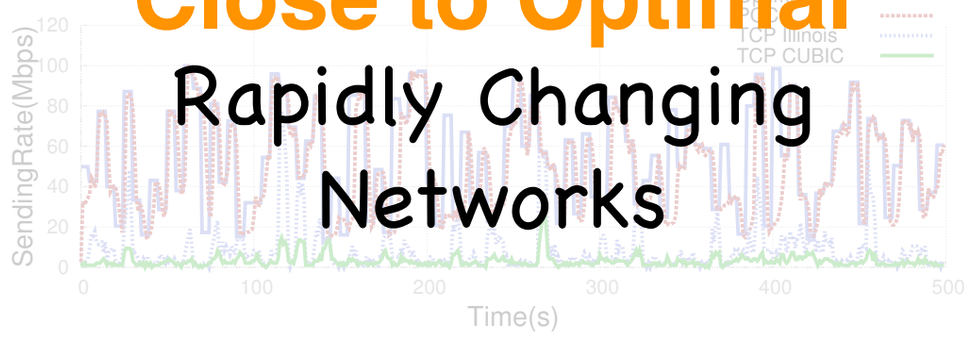
Incast



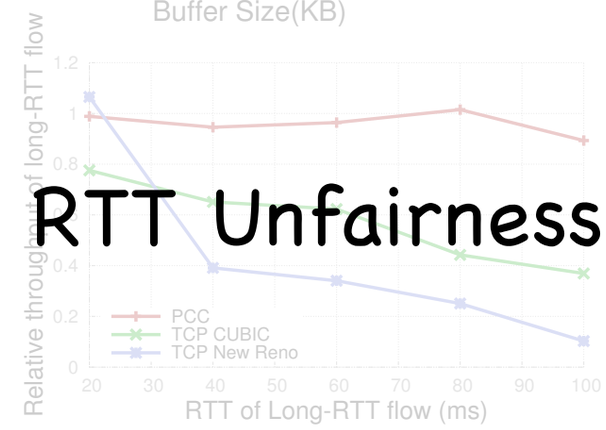
Close to Optimal

Rapidly Changing

Networks



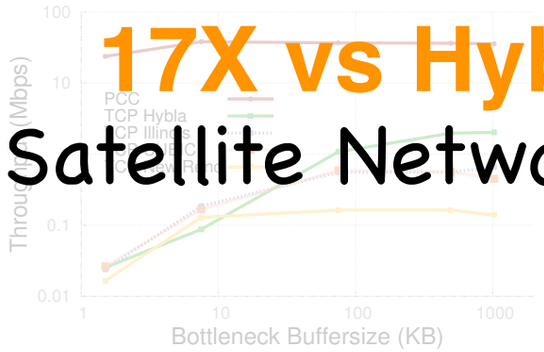
RTT Unfairness



Consistent High Performance

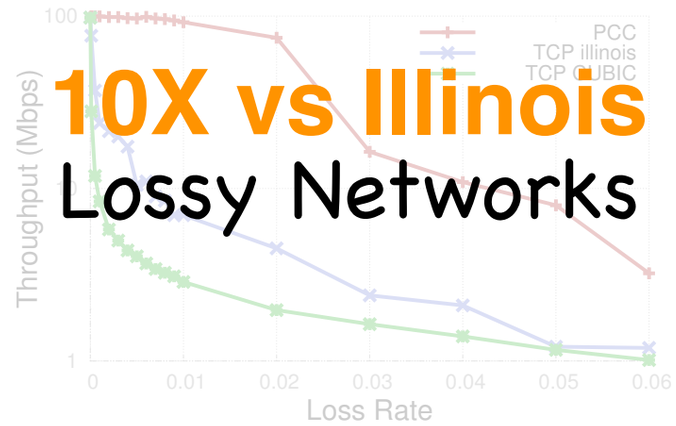
17X vs Hybla

Satellite Networks



10X vs Illinois

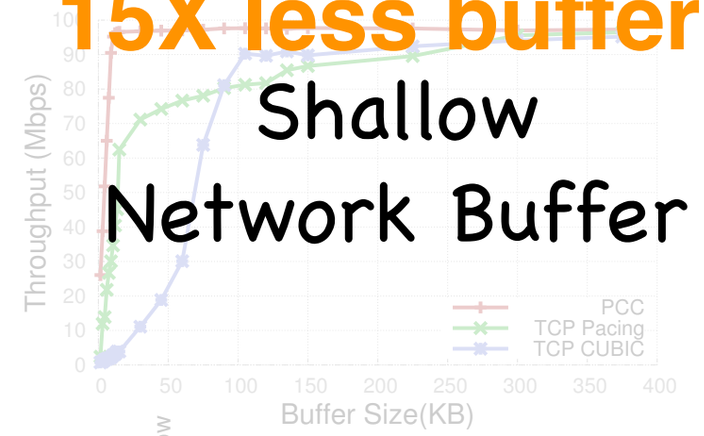
Lossy Networks



15X less buffer

Shallow

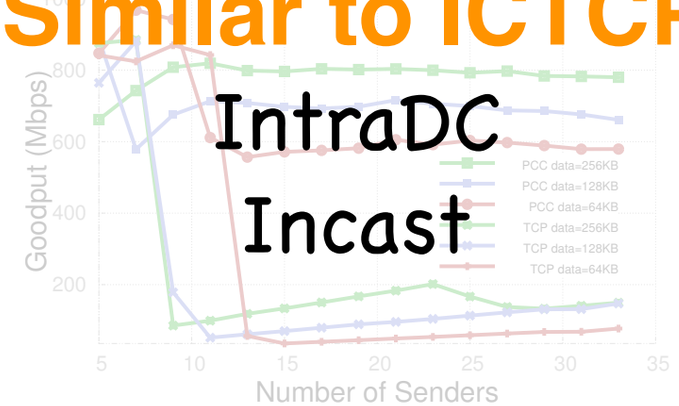
Network Buffer



Similar to ICTCP

IntraDC

Incast



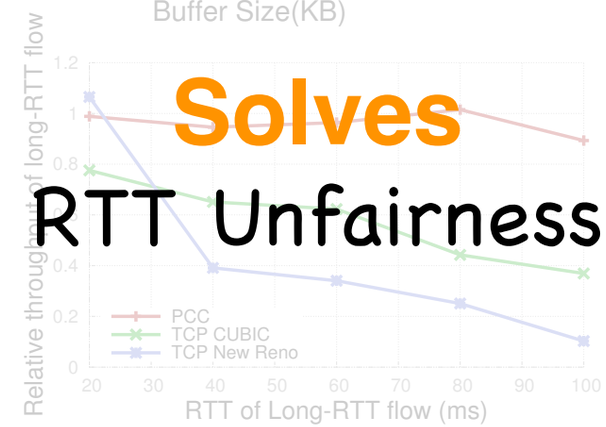
Close to Optimal

Rapidly Changing

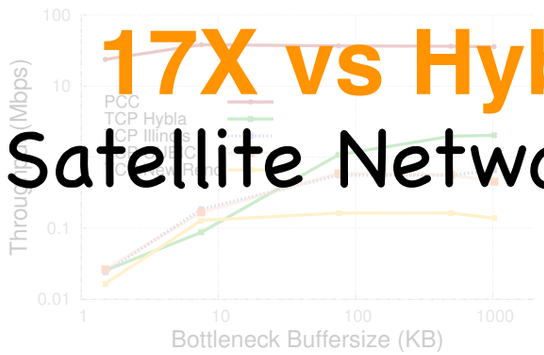
Networks

Solves

RTT Unfairness

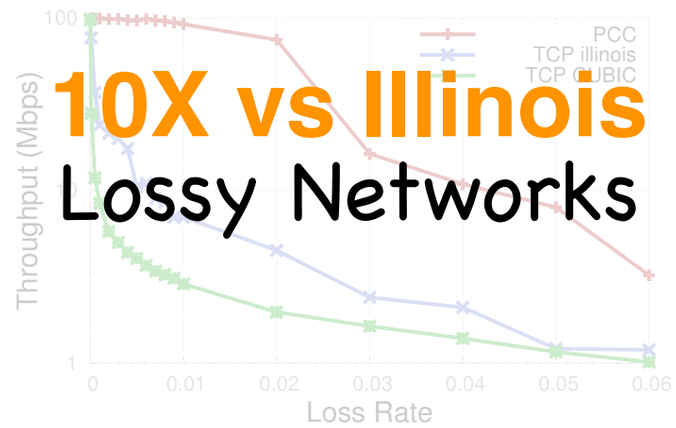


Consistent High Performance



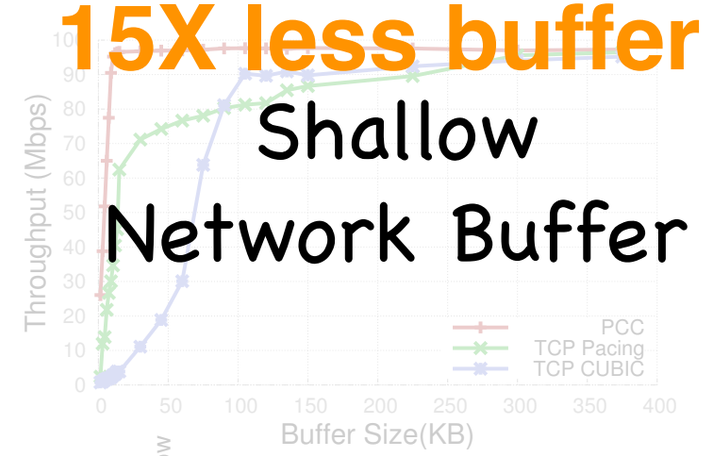
17X vs Hybla

Satellite Networks



10X vs Illinois

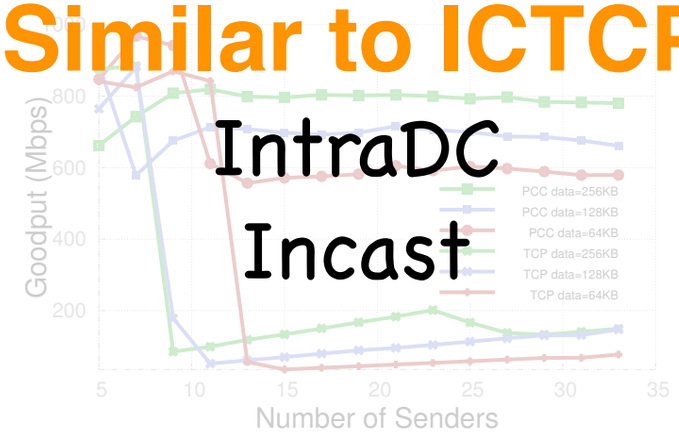
Lossy Networks



15X less buffer

Shallow

Network Buffer



Similar to ICTCP

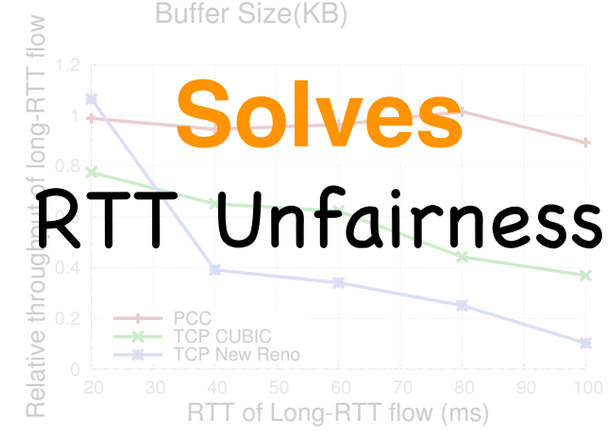
IntraDC

Incast

Close to Optimal

Rapidly Changing

Networks



Solves

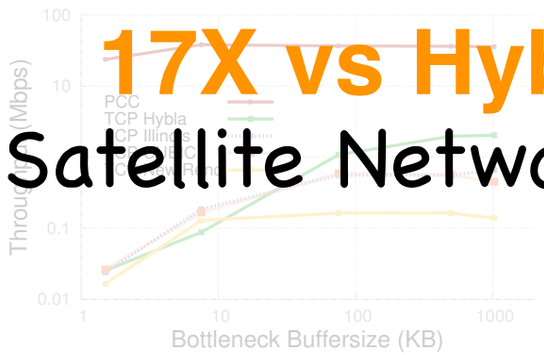
RTT Unfairness

Table 1: PCC significantly outperforms TCP in inter-data center environments. RTT is in msec; throughput in Mbps.

Transmission Pair	RTT	PCC	SABUL	CUBIC	Illinois
GPO → NYSERNet	[illegible]	[illegible]	[illegible]	129	326
GPO → Missouri	[illegible]	[illegible]	[illegible]	80.7	90.1
GPO → Illinois	35.4	766	664	84.5	102
NYSERNet → Missouri	47.4	816	662	108	109
Wisconsin → Illinois	9.01	801	700	547	562
GPO → Wisc.	38.0	783	487	79.3	120
NYSERNet → Wisc.	38.3	791	673	134	134
Missouri → Wisc.	20.9	807	698	259	262
NYSERNet → Illinois	36.1	808	674	141	141

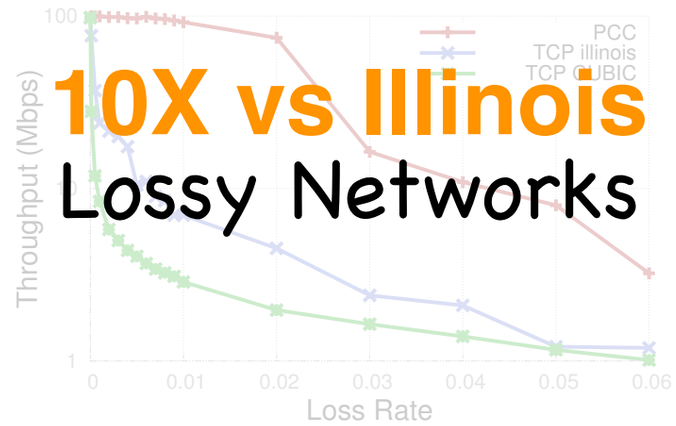
InterDC

Consistent High Performance



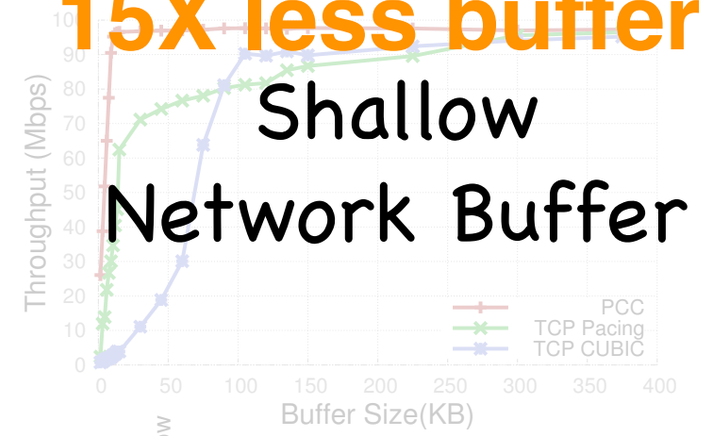
17X vs Hybla

Satellite Networks



10X vs Illinois

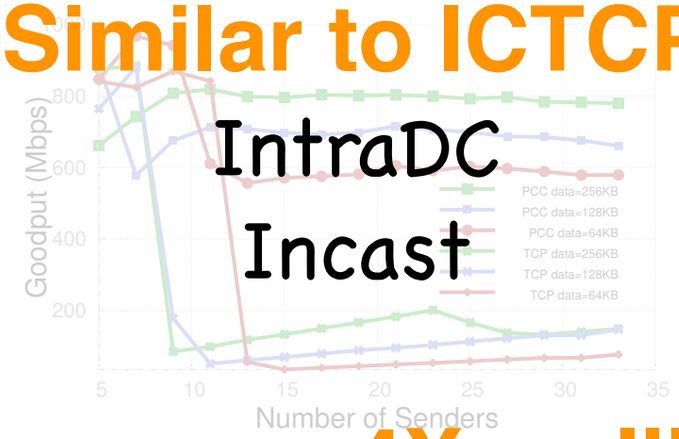
Lossy Networks



15X less buffer

Shallow

Network Buffer



Similar to ICTCP

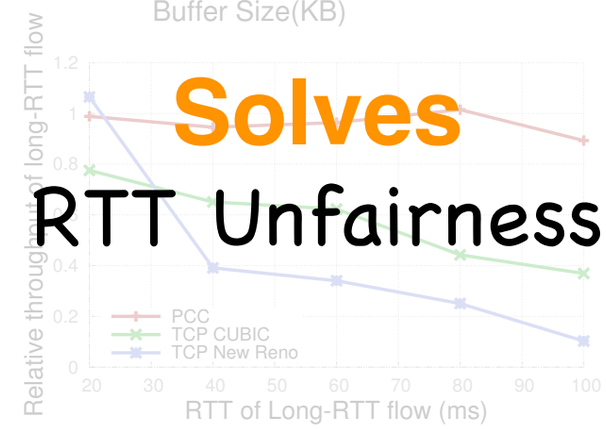
IntraDC

Incast

Close to Optimal

Rapidly Changing

Networks



Solves

RTT Unfairness

4X vs Illinois

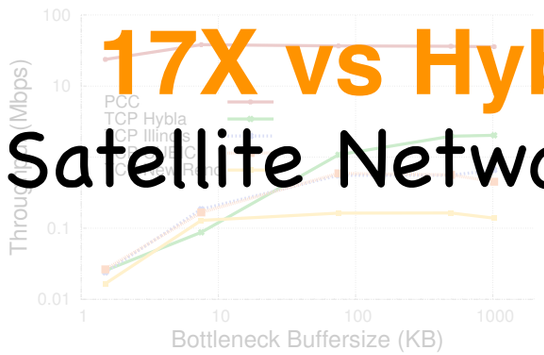
1.23X vs UDT

InterDC

Table 1: PCC significantly outperforms TCP in inter-data center environments. RTT is in msec; throughput in Mbps.

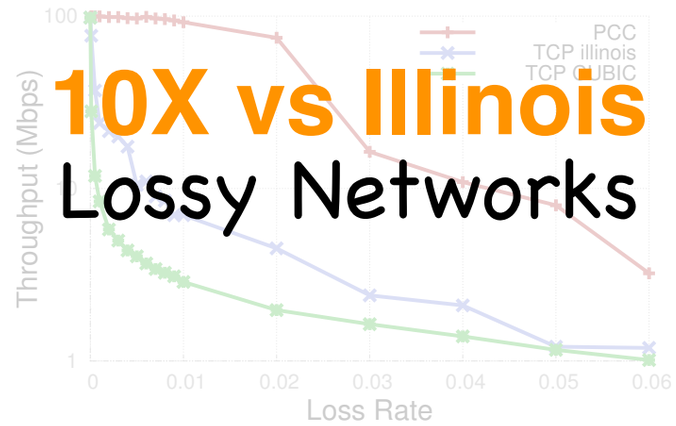
Transmission	RTT	PCC	PABUL	CUBIC	Illinois
GPO → NYSERNet	12.1	396	818	129	326
GPO → Missouri	47.3	801	752	80.7	90.1
GPO → Illinois	35.4	766	664	84.5	102
NYSERNet → Missouri	47.4	816	662	108	109
Wisconsin → Illinois	9.01	801	700	547	562
GPO → Wisc.	38.0	783	487	79.3	120
NYSERNet → Wisc.	38.3	791	673	134	134
Missouri → Wisc.	20.9	807	698	259	262
NYSERNet → Illinois	36.1	808	674	141	141

Consistent High Performance



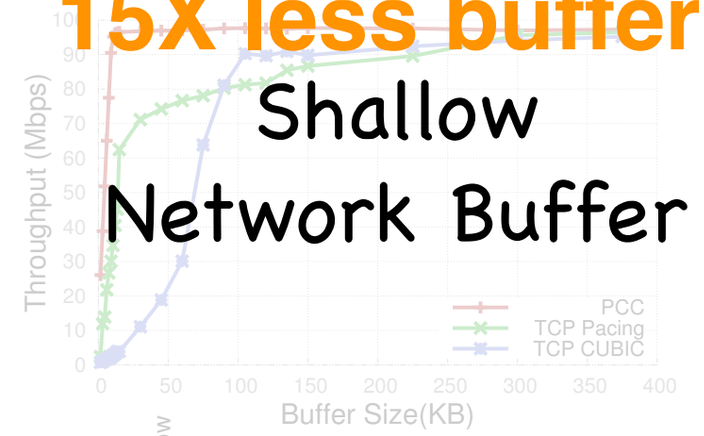
17X vs Hybla

Satellite Networks



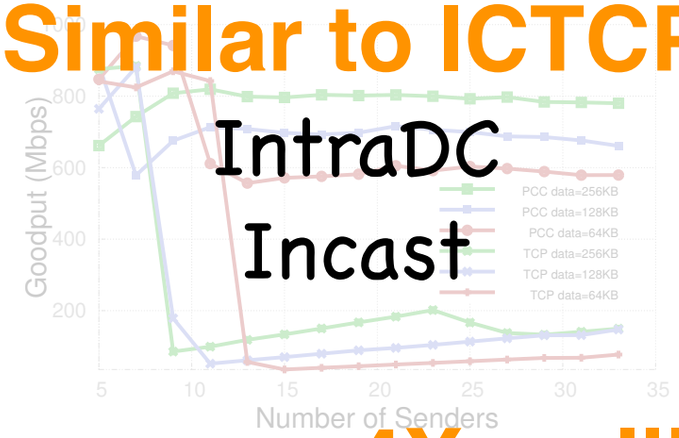
10X vs Illinois

Lossy Networks



15X less buffer

Shallow Network Buffer

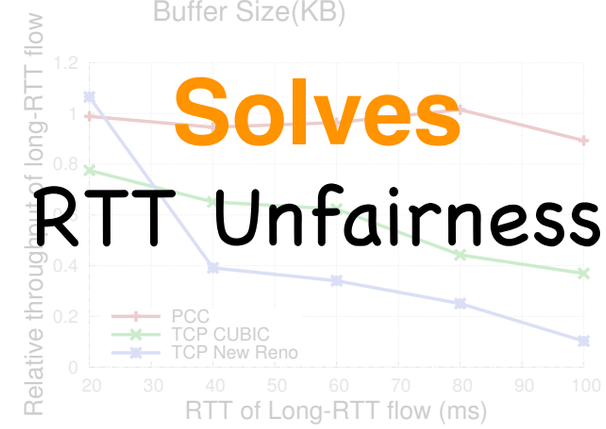


Similar to ICTCP

IntraDC Incast

Close to Optimal

Rapidly Changing Networks



Solves

RTT Unfairness

4X vs Illinois

1.23X vs UDT

InterDC

Table 1: PCC significantly outperforms TCP in inter-data center environments. RTT is in msec; throughput in Mbps.

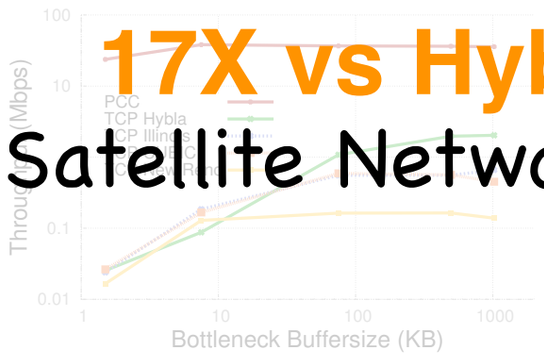
Transmission	RTT	PCC	PABUL	CUBIC	Illinois
GPO → NYSERNet	12.1	818	679	129	326
GPO → Missouri	47.3	624	564	80.7	90.1
GPO → Illinois	35.4	766	664	84.5	102
NYSERNet → Missouri	47.4	816	662	108	109
Wisconsin → Illinois	9.01	801	700	547	562
GPO → Wisc.	38.0	783	487	79.3	120
NYSERNet → Wisc.	38.3	791	673	134	134
Missouri → Wisc.	20.9	807	698	259	262
NYSERNet → Illinois	36.1	808	674	141	141

Global Commercial Internet

Consistent High Performance

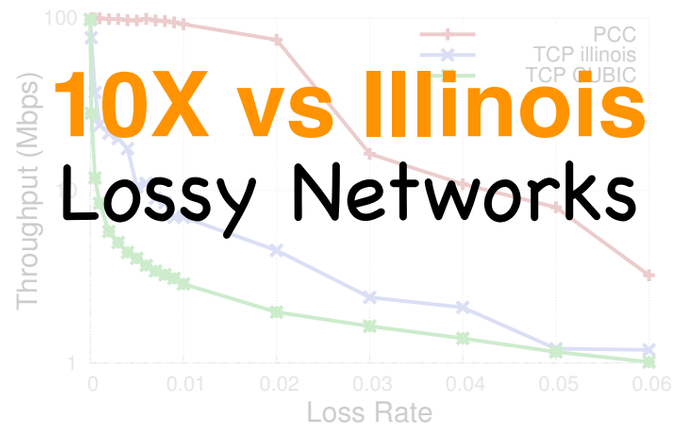
17X vs Hybla

Satellite Networks



10X vs Illinois

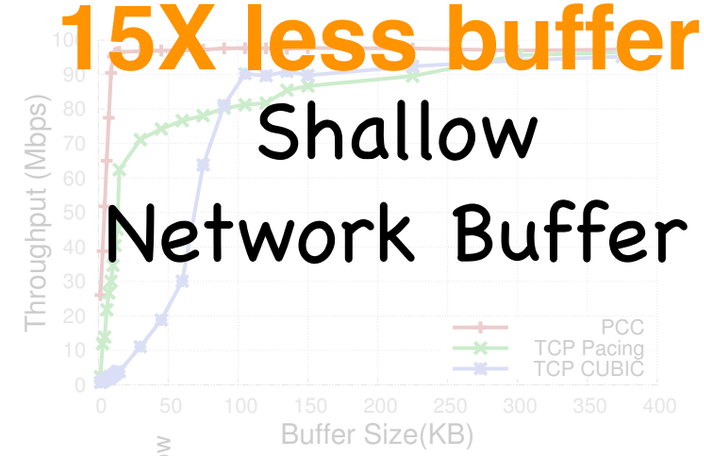
Lossy Networks



15X less buffer

Shallow

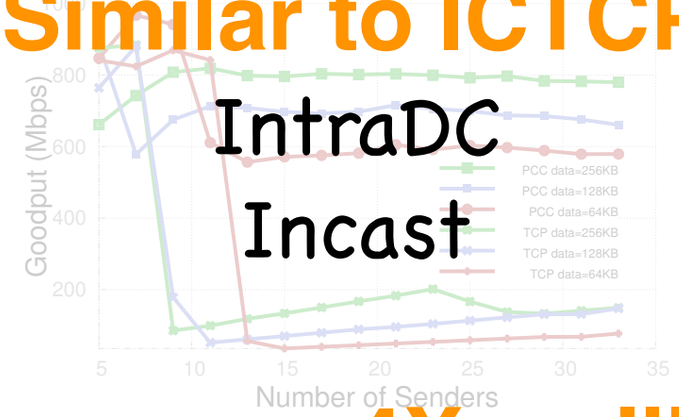
Network Buffer



Similar to ICTCP

IntraDC

Incast



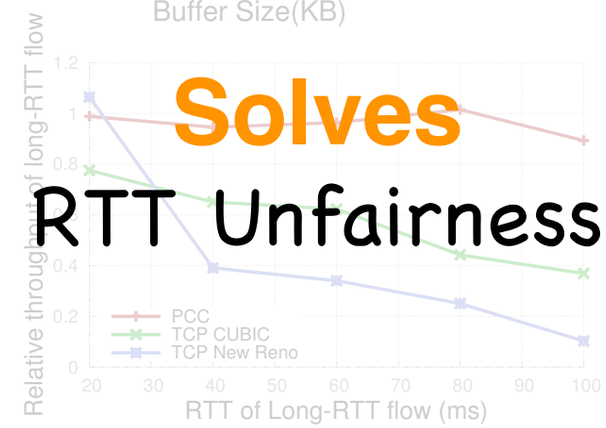
Close to Optimal

Rapidly Changing

Networks

Solves

RTT Unfairness



4X vs Illinois

1.23X vs UDT

InterDC

Table 1: PCC significantly outperforms TCP in inter-data center environments. RTT is in msec; throughput in Mbps.

Transmission	RTT	PCC	PABUL	CUBIC	Illinois
GPO → NYSERNet	12.1	818	564	129	326
GPO → Missouri	46.5	624	557	80.7	90.1
GPO → Illinois	35.4	766	664	84.5	102
NYSERNet → Missouri	47.4	816	662	108	109
Wisconsin → Illinois	9.01	801	700	547	562
GPO → Wisc.	38.0	783	487	79.3	120
NYSERNet → Wisc.	38.3	791	673	134	134
Missouri → Wisc.	20.9	807	698	259	262
NYSERNet → Illinois	36.1	808	674	141	141

4X median vs CUBIC

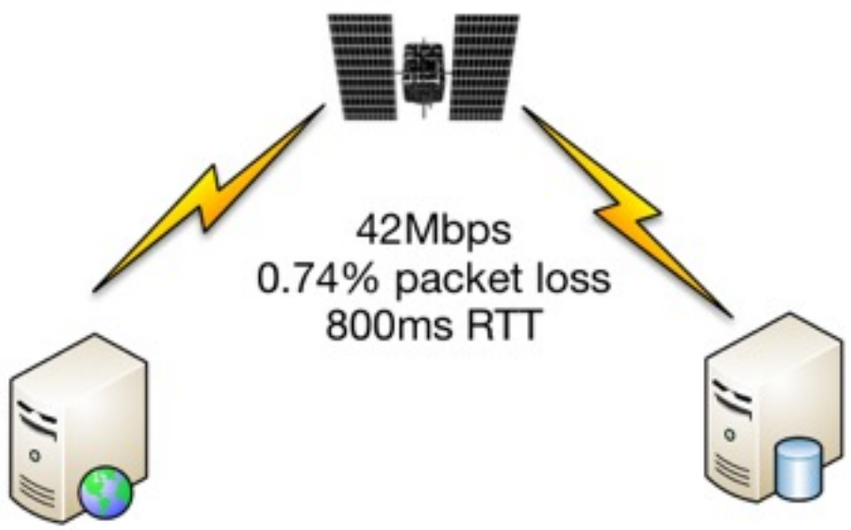
1.48X vs UDT + 4X less loss

Global Commercial

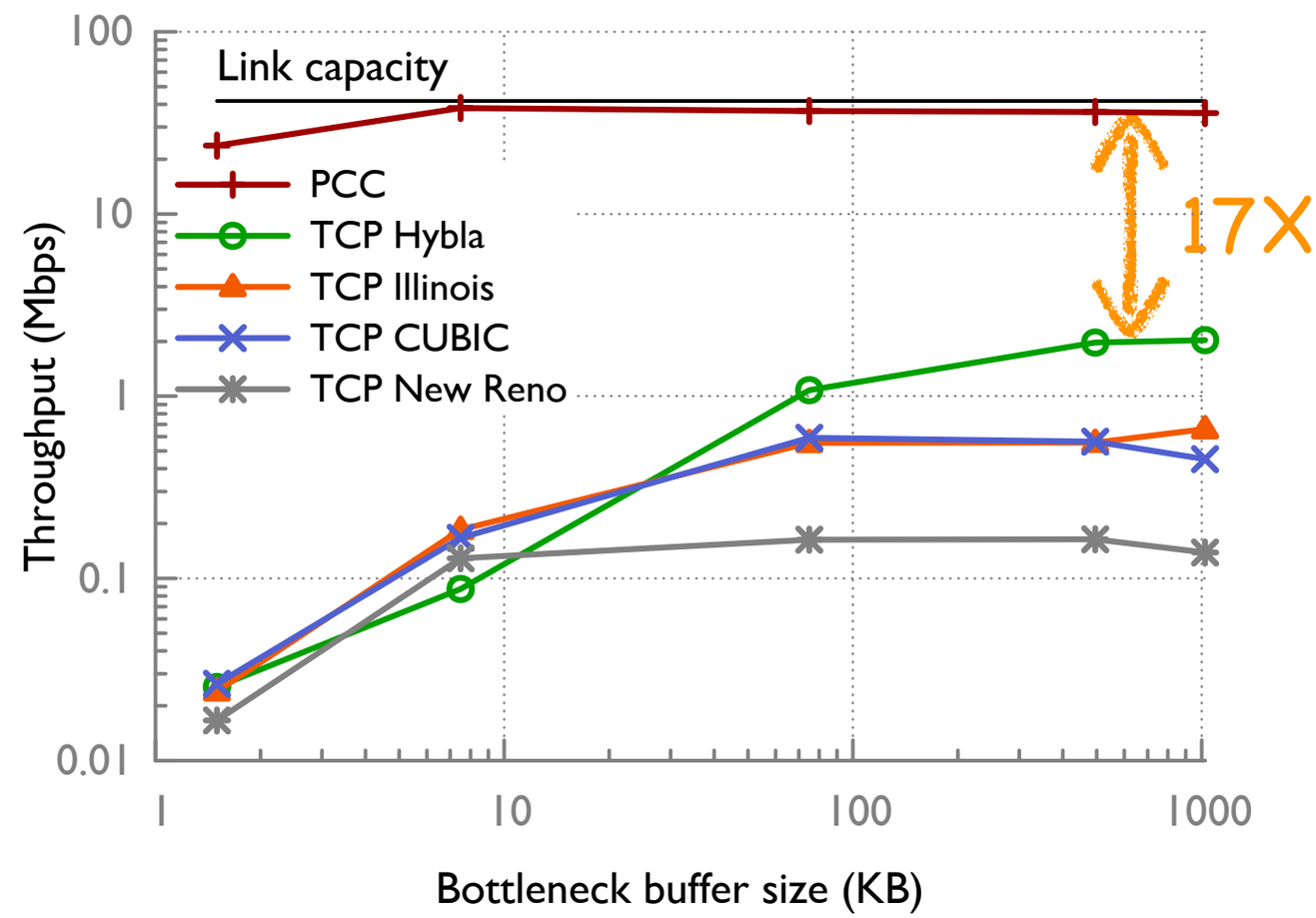
Internet

Consistent High Performance

Satellite Network



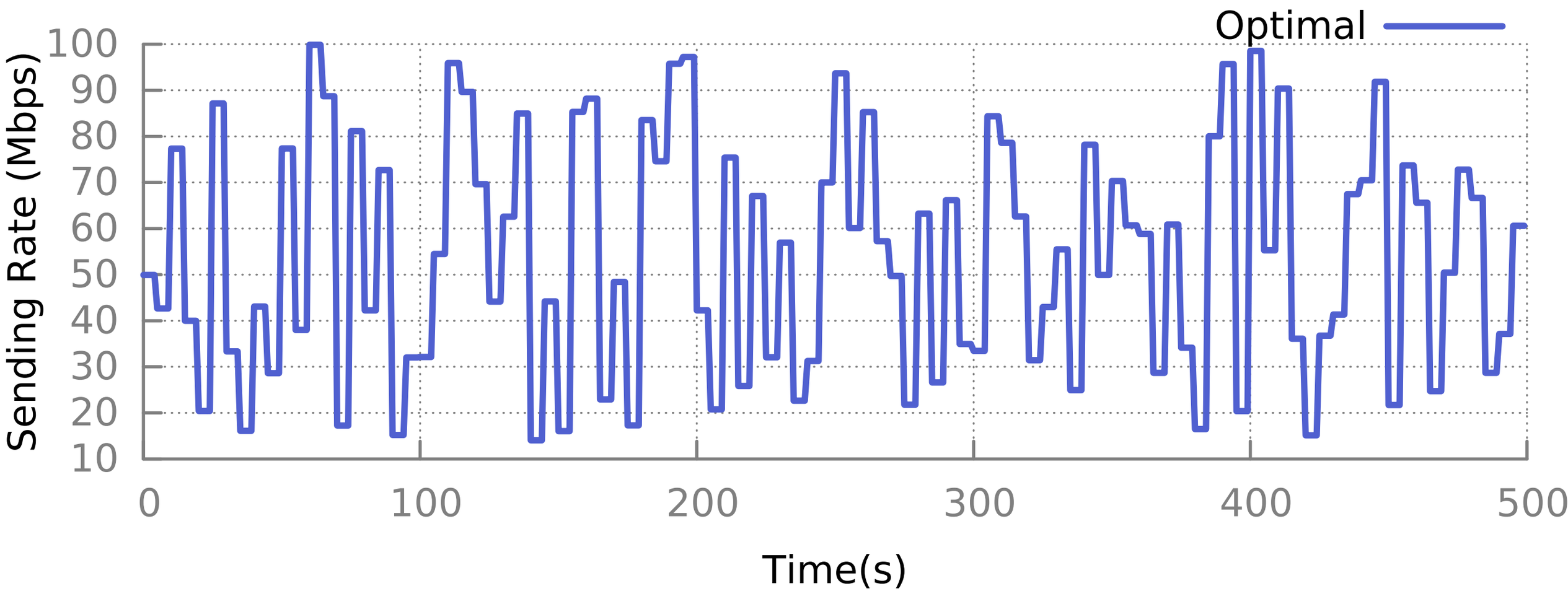
WINDS System



Consistent High Performance

Rapidly Changing Networks

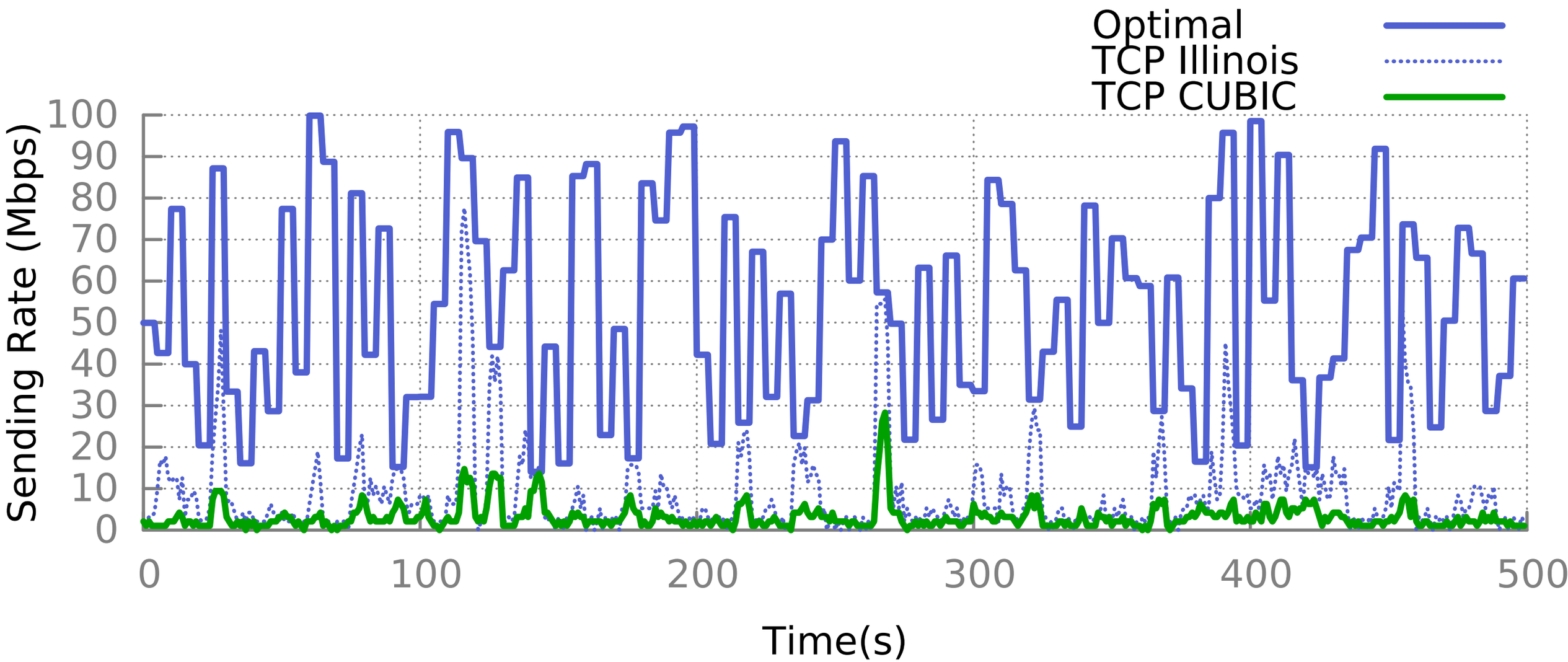
BW: 10-100Mbps; RTT: 10-100ms; Loss Rate: 0-1%
Change every 5 seconds



Consistent High Performance

Rapidly Changing Networks

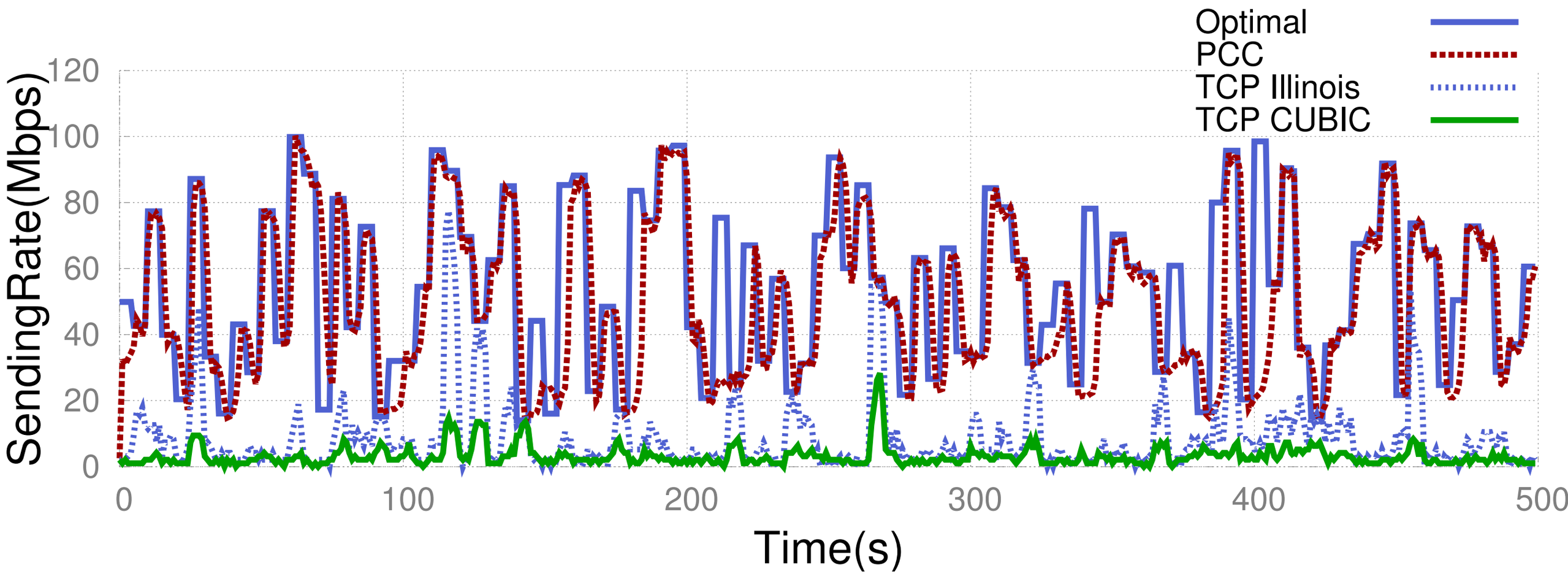
BW: 10-100Mbps; RTT: 10-100ms; Loss Rate: 0-1%
Change every 5 seconds

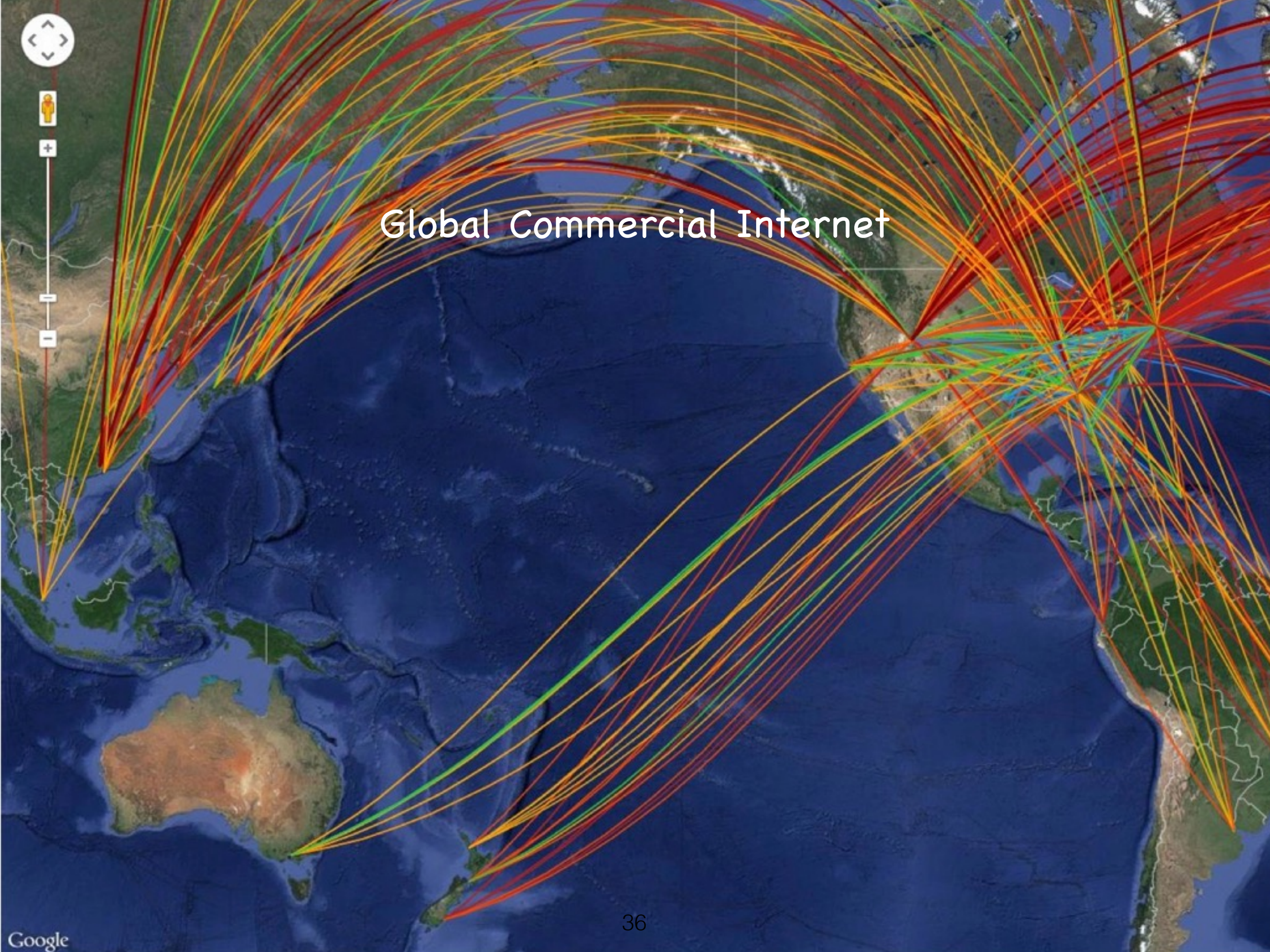


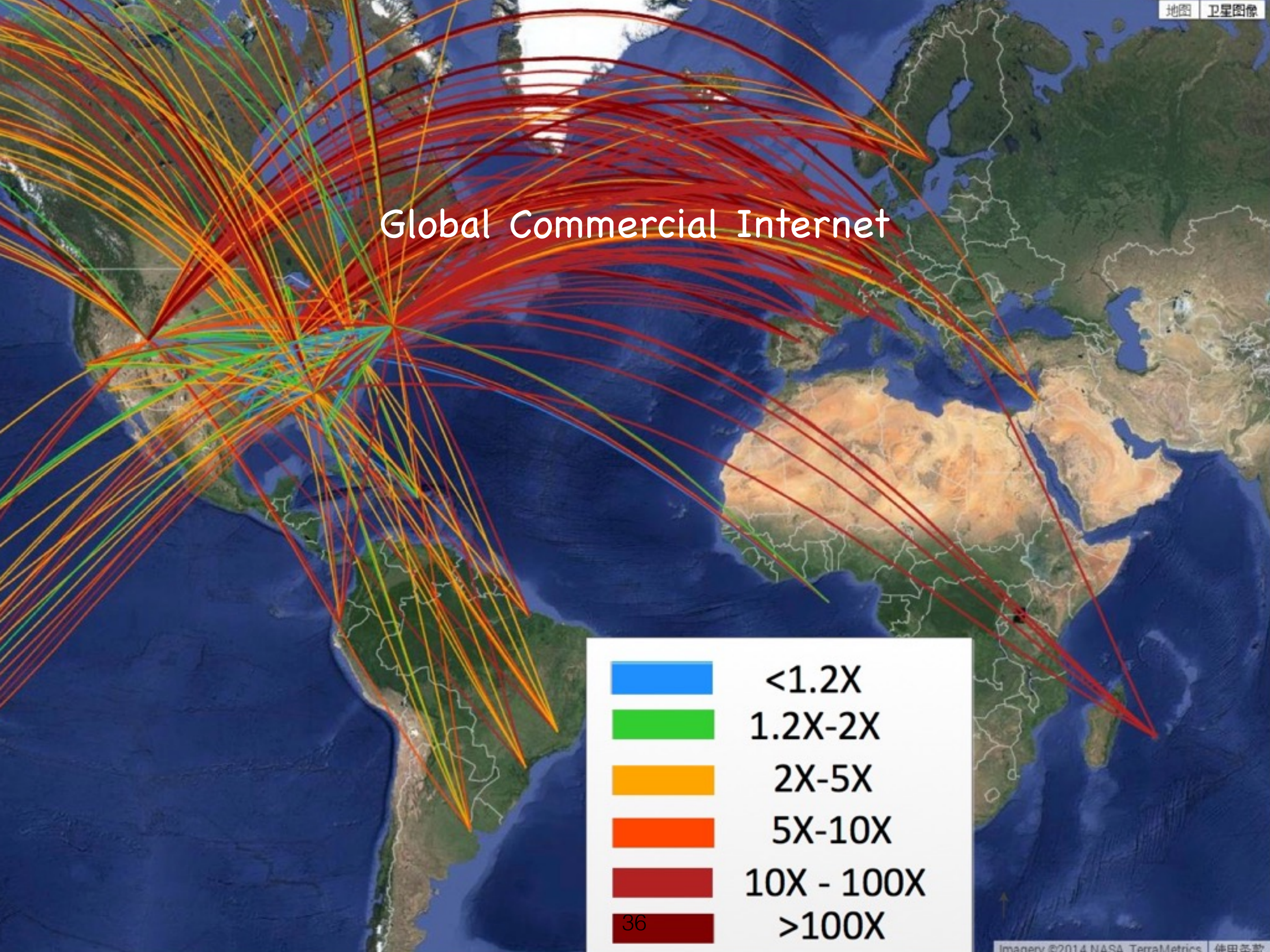
Consistent High Performance

Rapidly Changing Networks

BW: 10-100Mbps; RTT: 10-100ms; Loss Rate: 0-1%
Change every 5 seconds



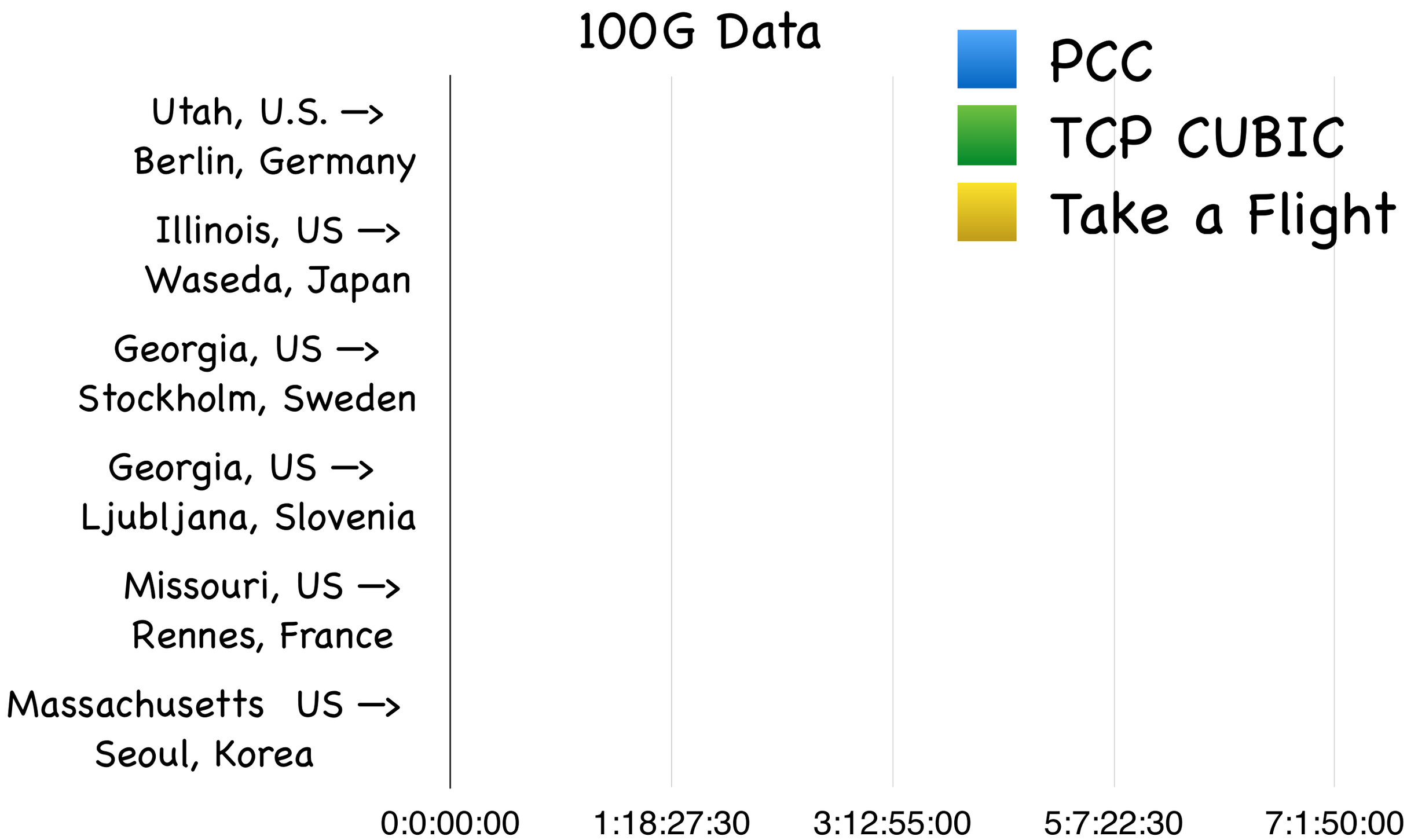

Global Commercial Internet
Google


Global Commercial Internet
<1.2X
1.2X-2X
2X-5X
5X-10X
10X - 100X
>100X
地图
卫星图像
Imagery ©2014 NASA, TerraMetrics

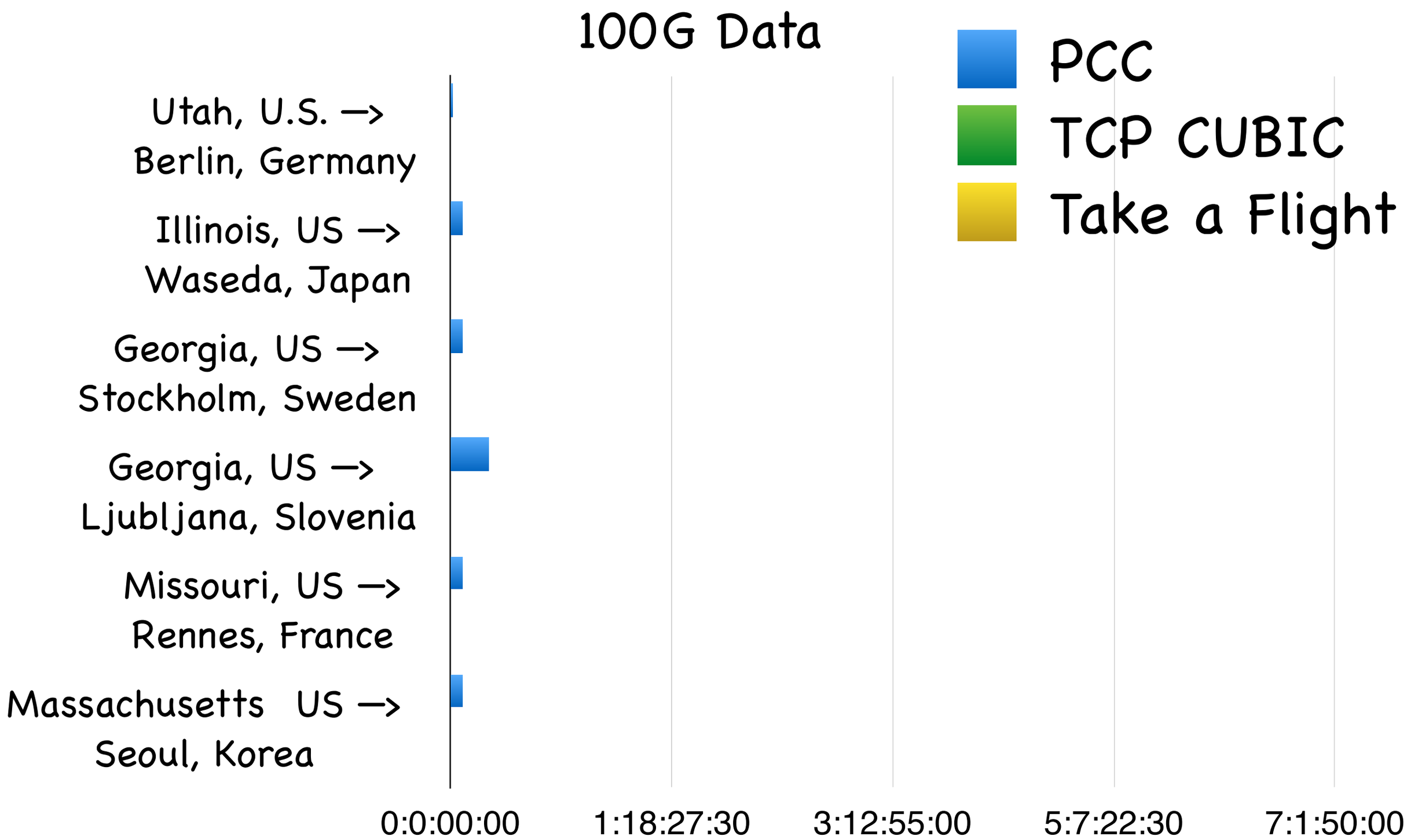
PCC vs TCP vs Take a Flight

100G Data

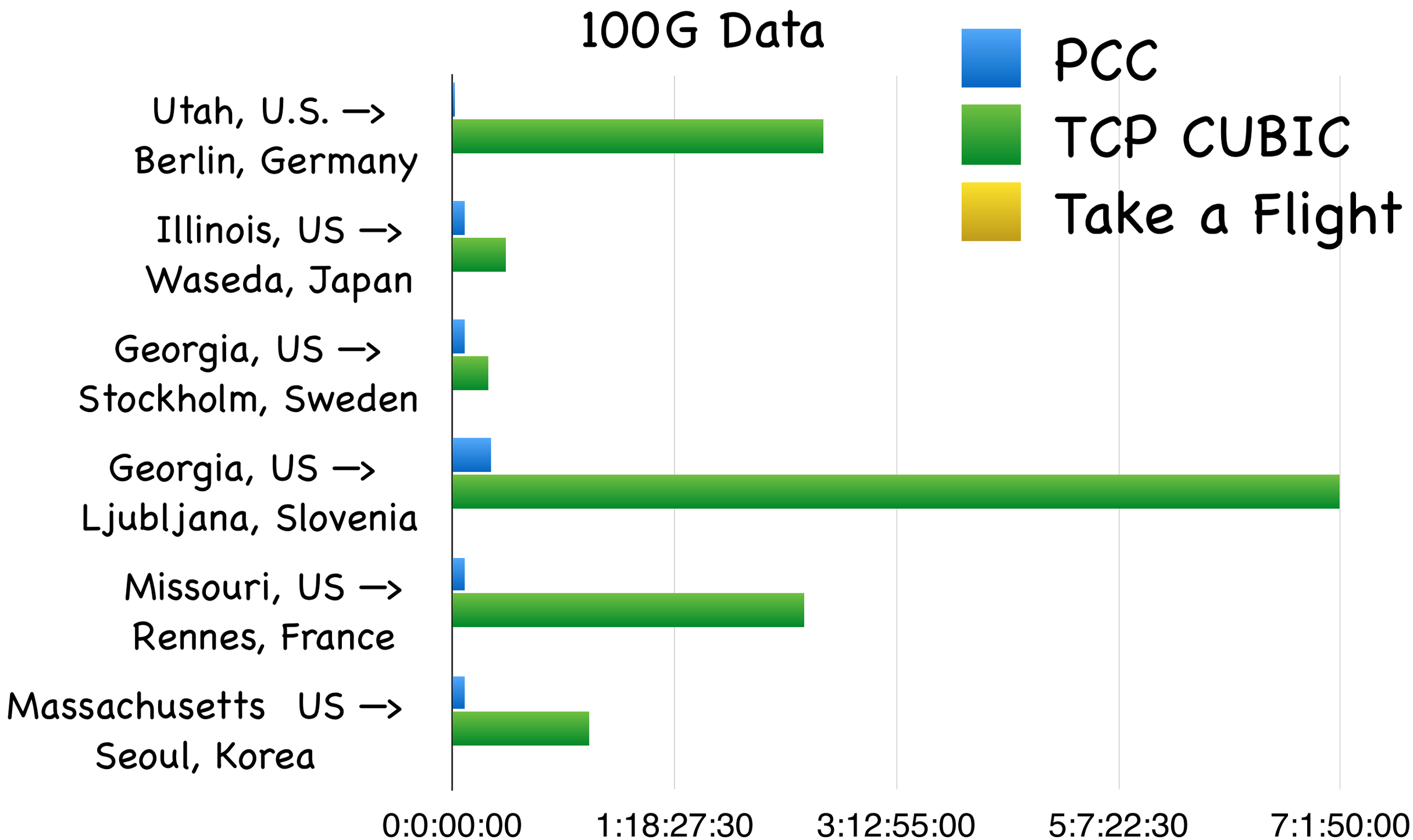
PCC vs TCP vs Take a Flight



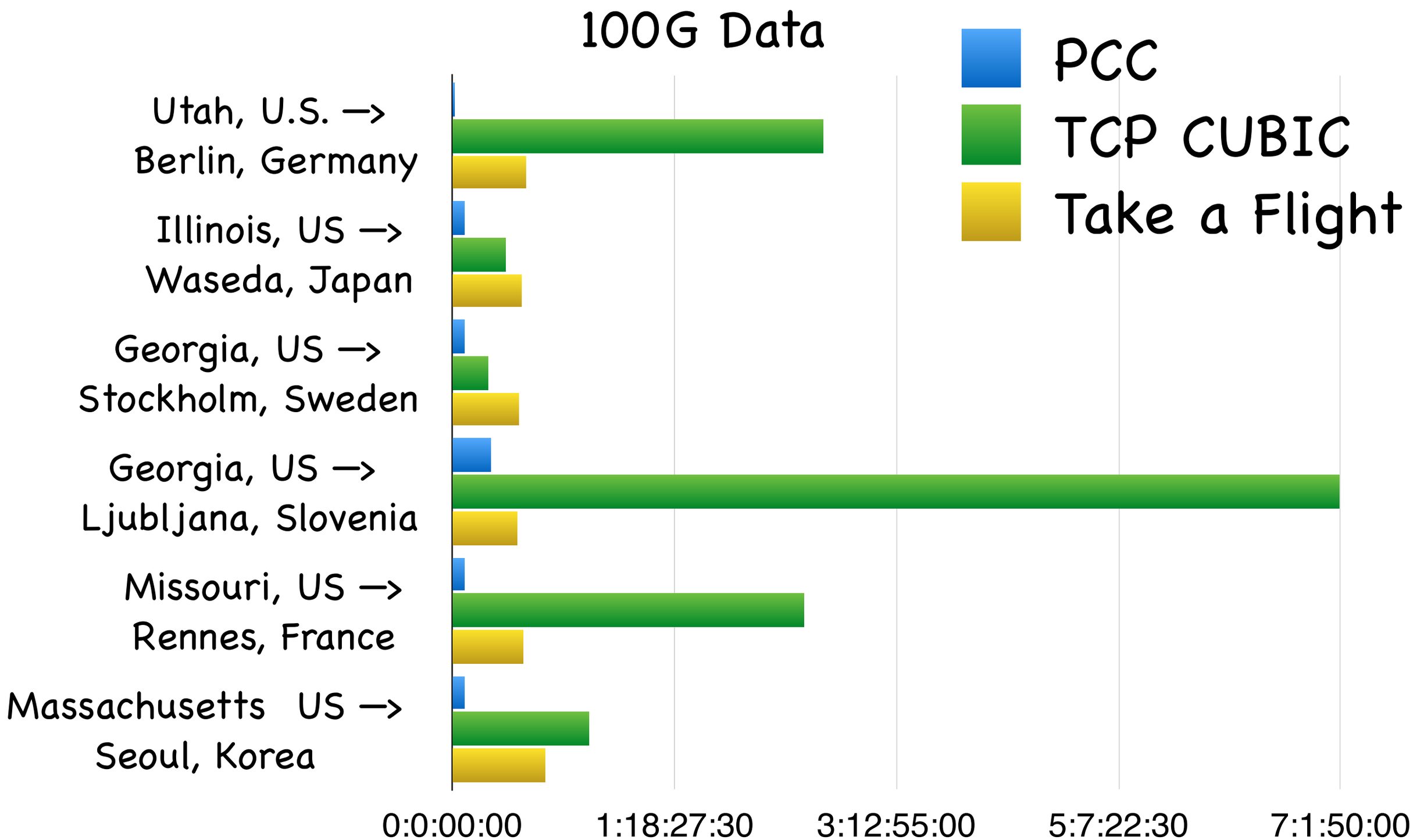
PCC vs TCP vs Take a Flight



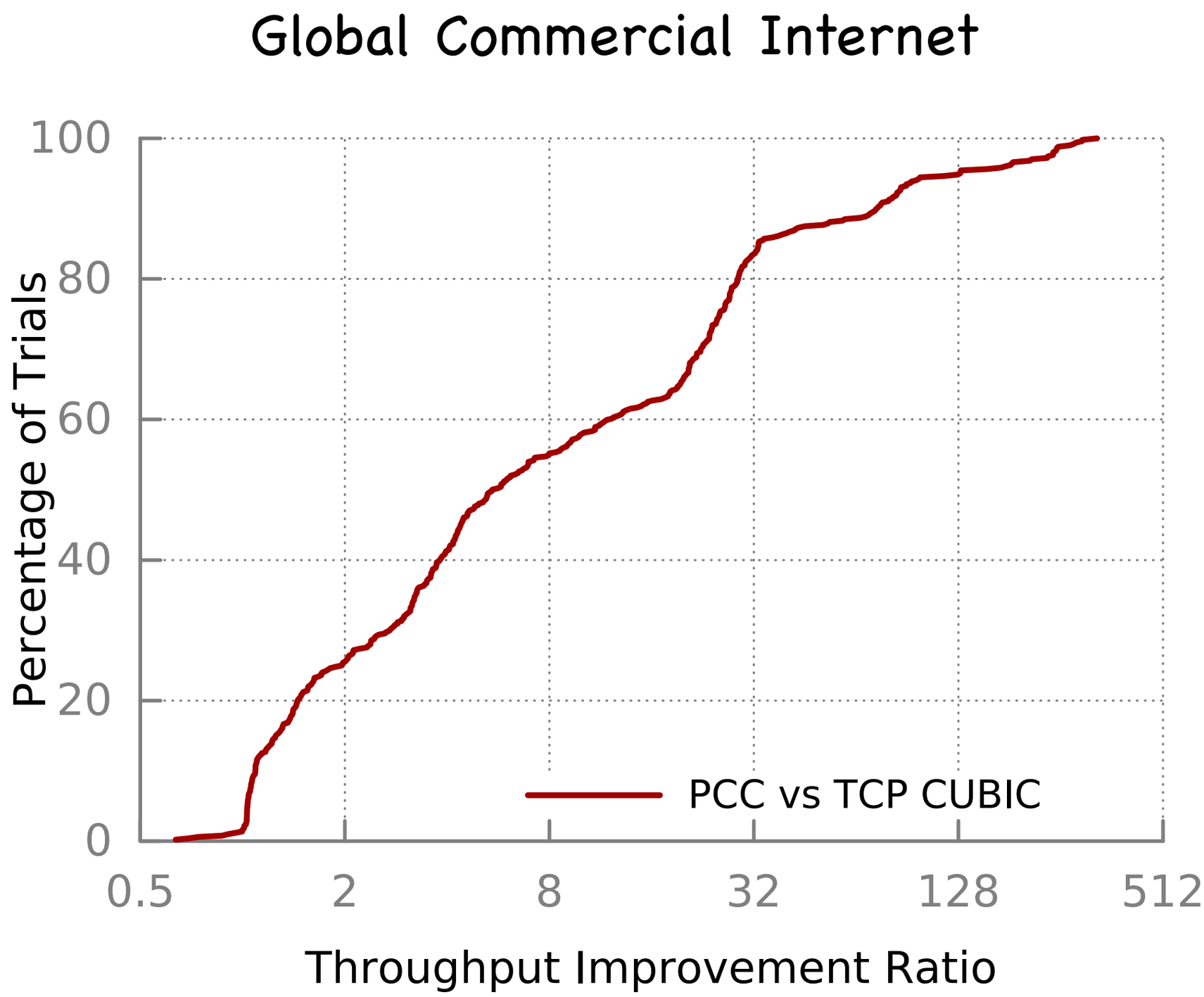
PCC vs TCP vs Take a Flight



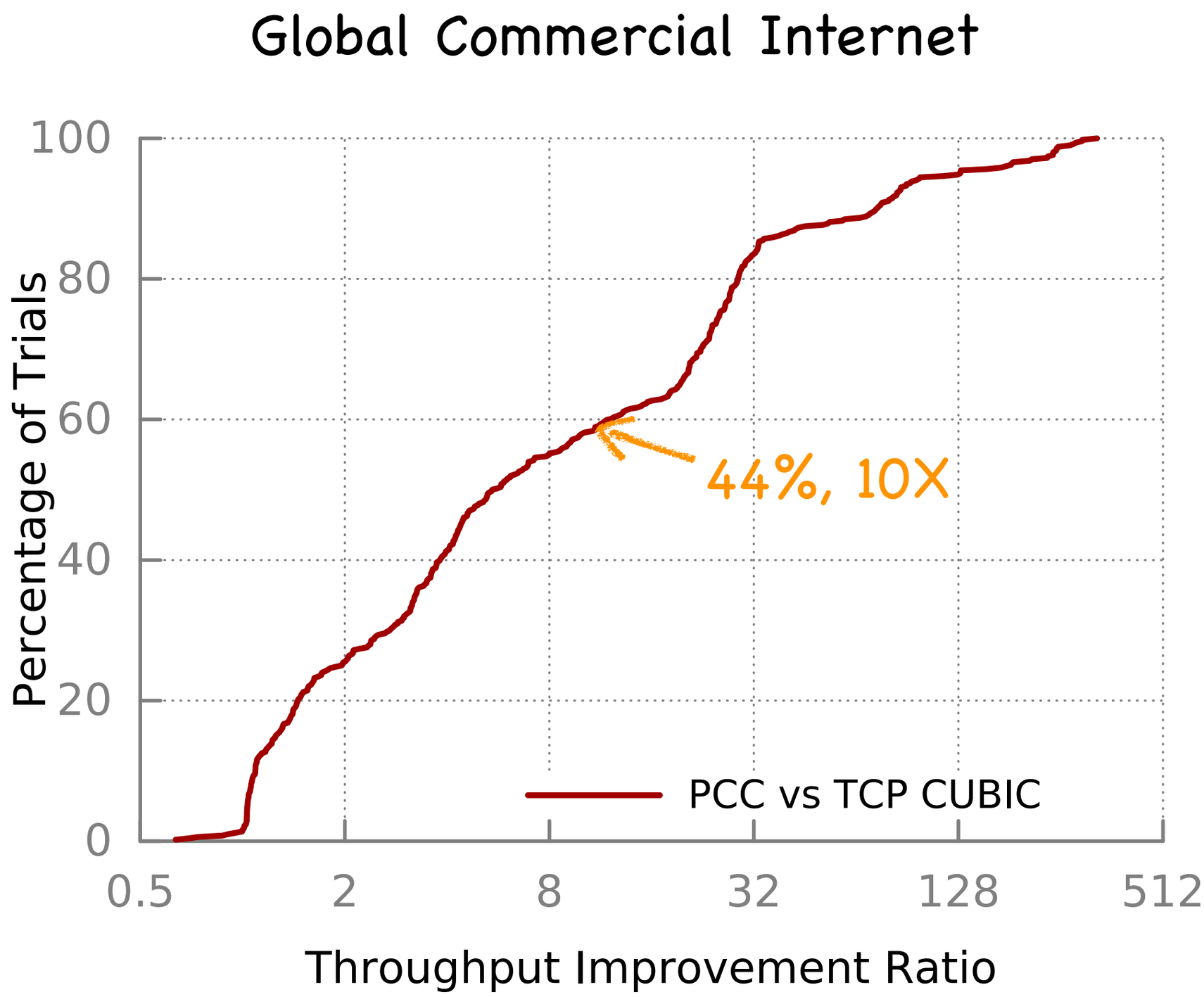
PCC vs TCP vs Take a Flight



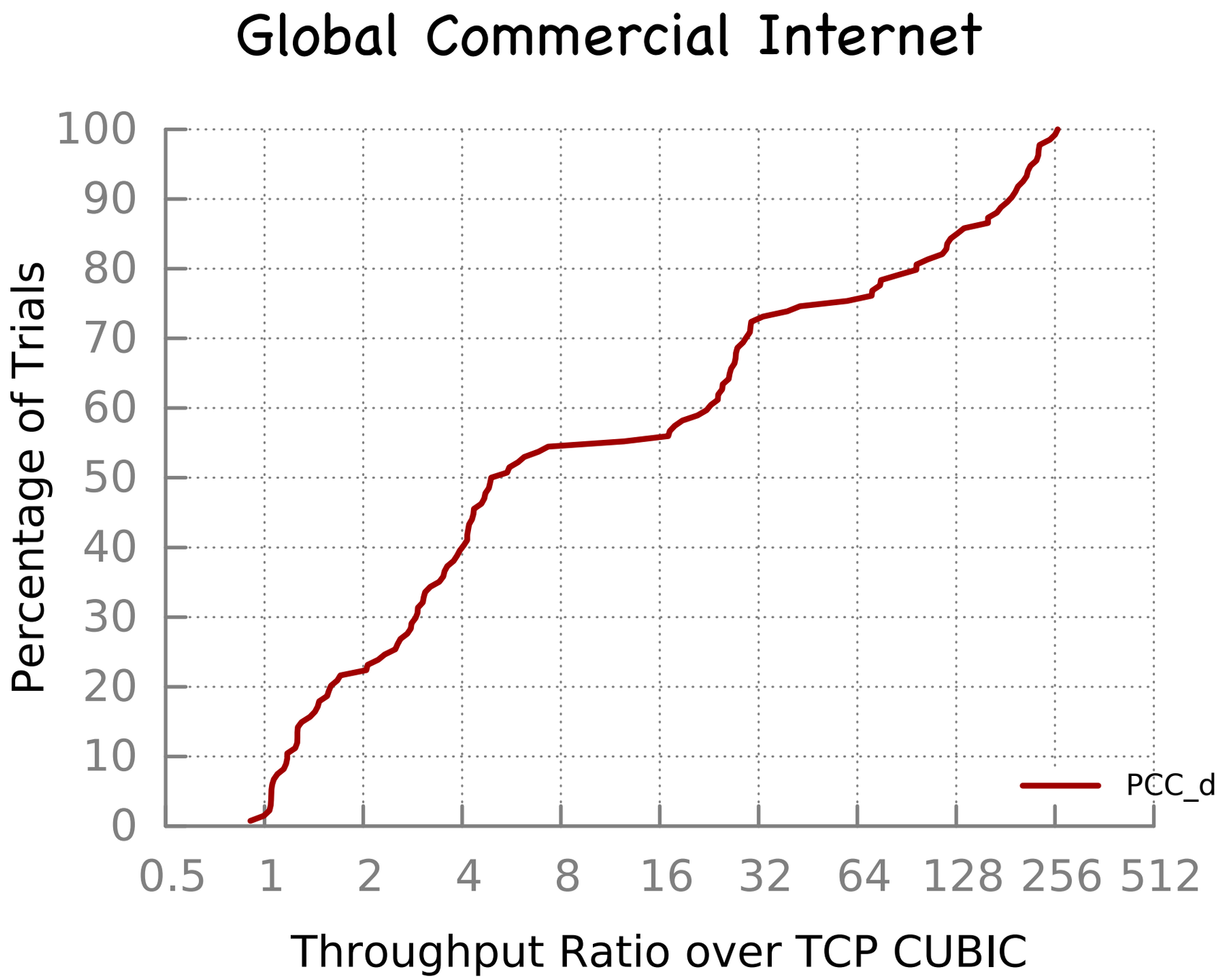
Consistent High Performance



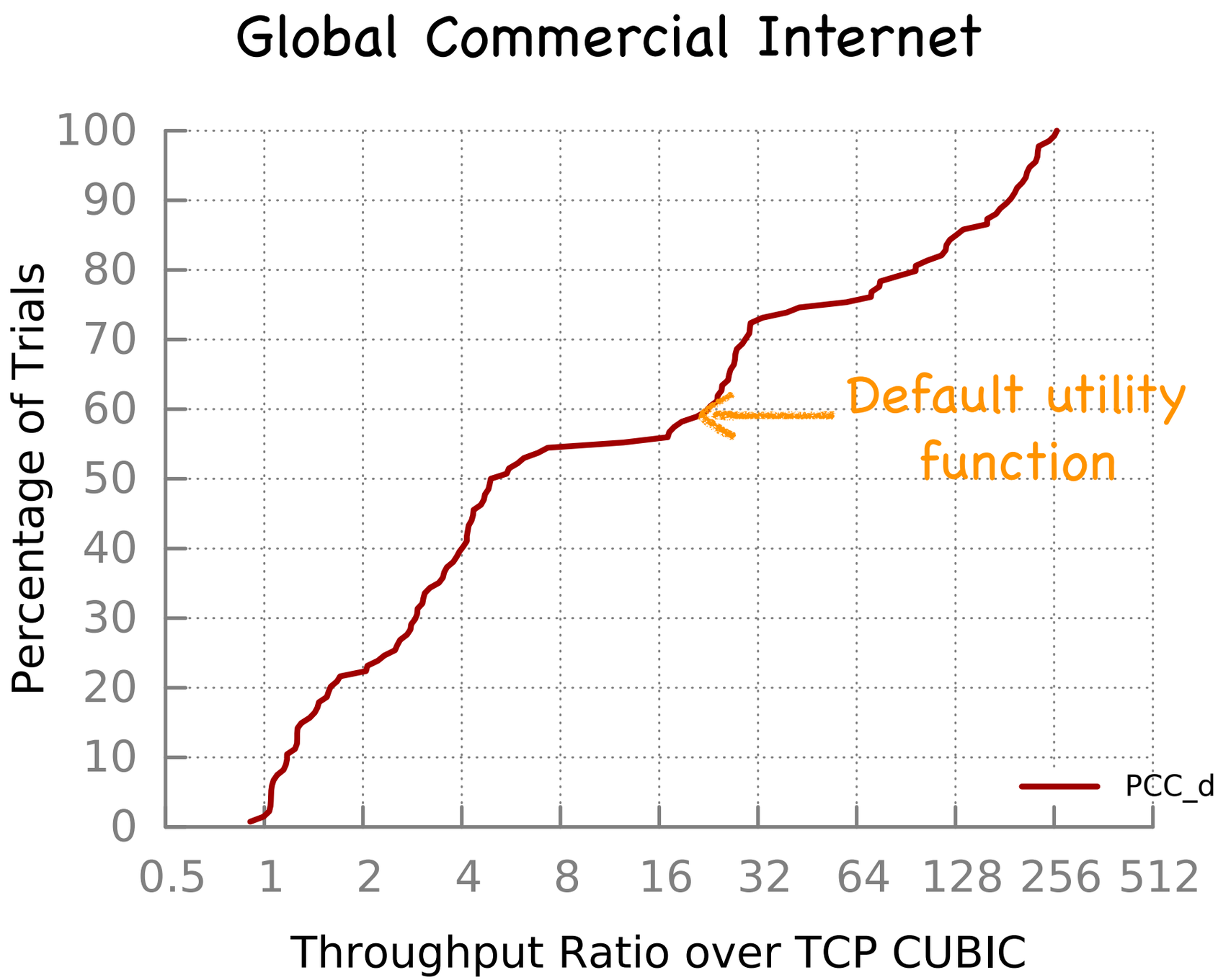
Consistent High Performance



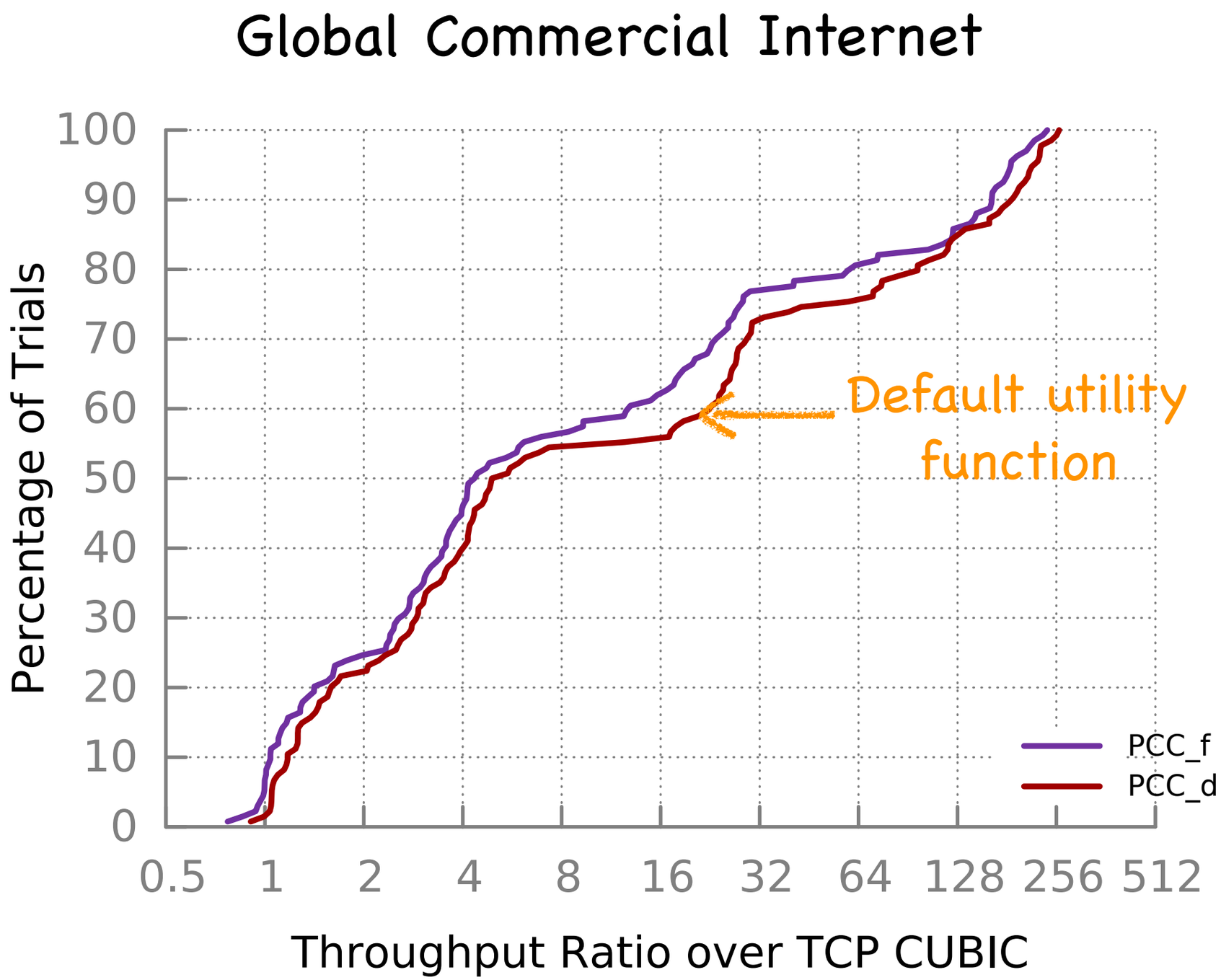
Consistent High Performance



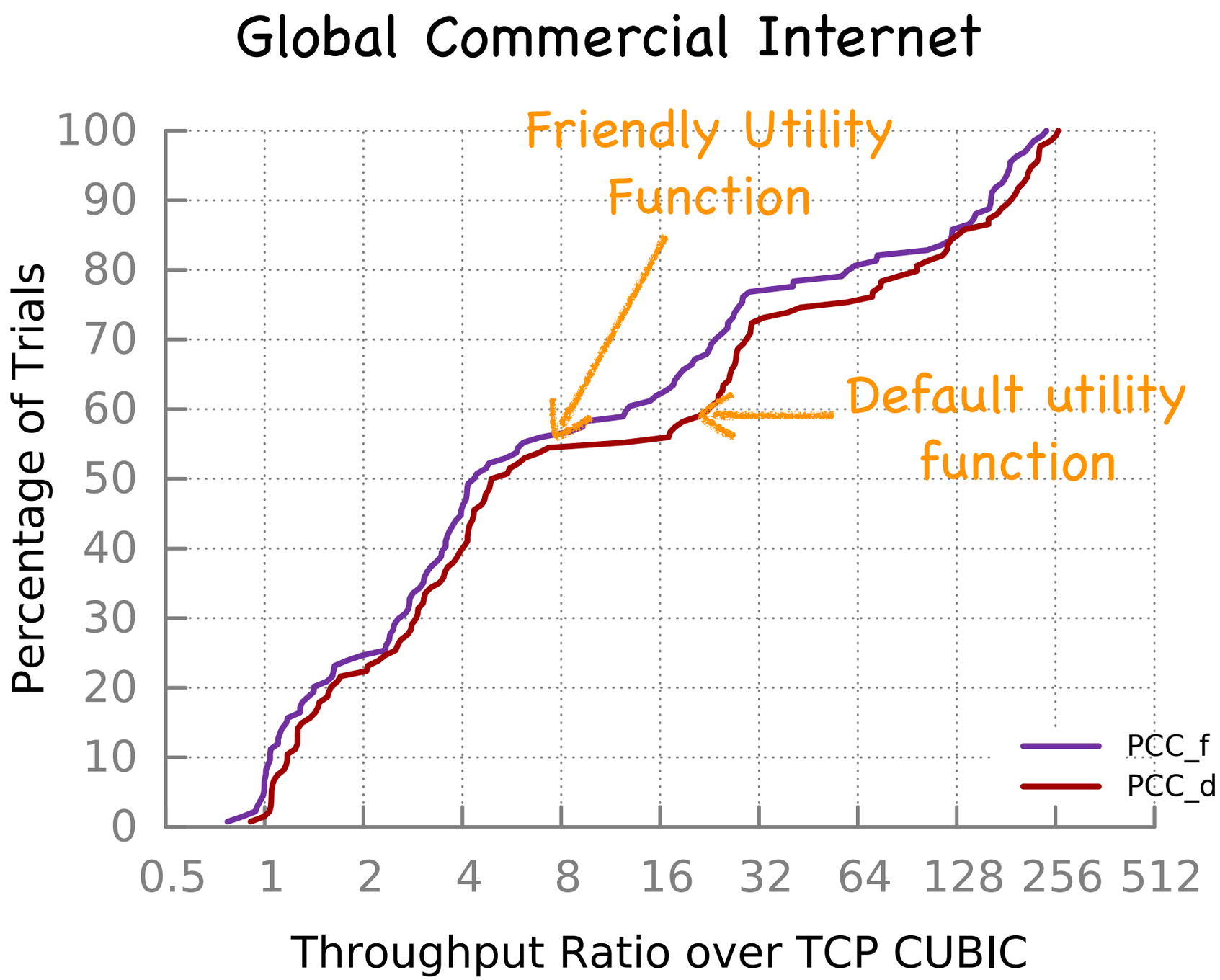
Consistent High Performance



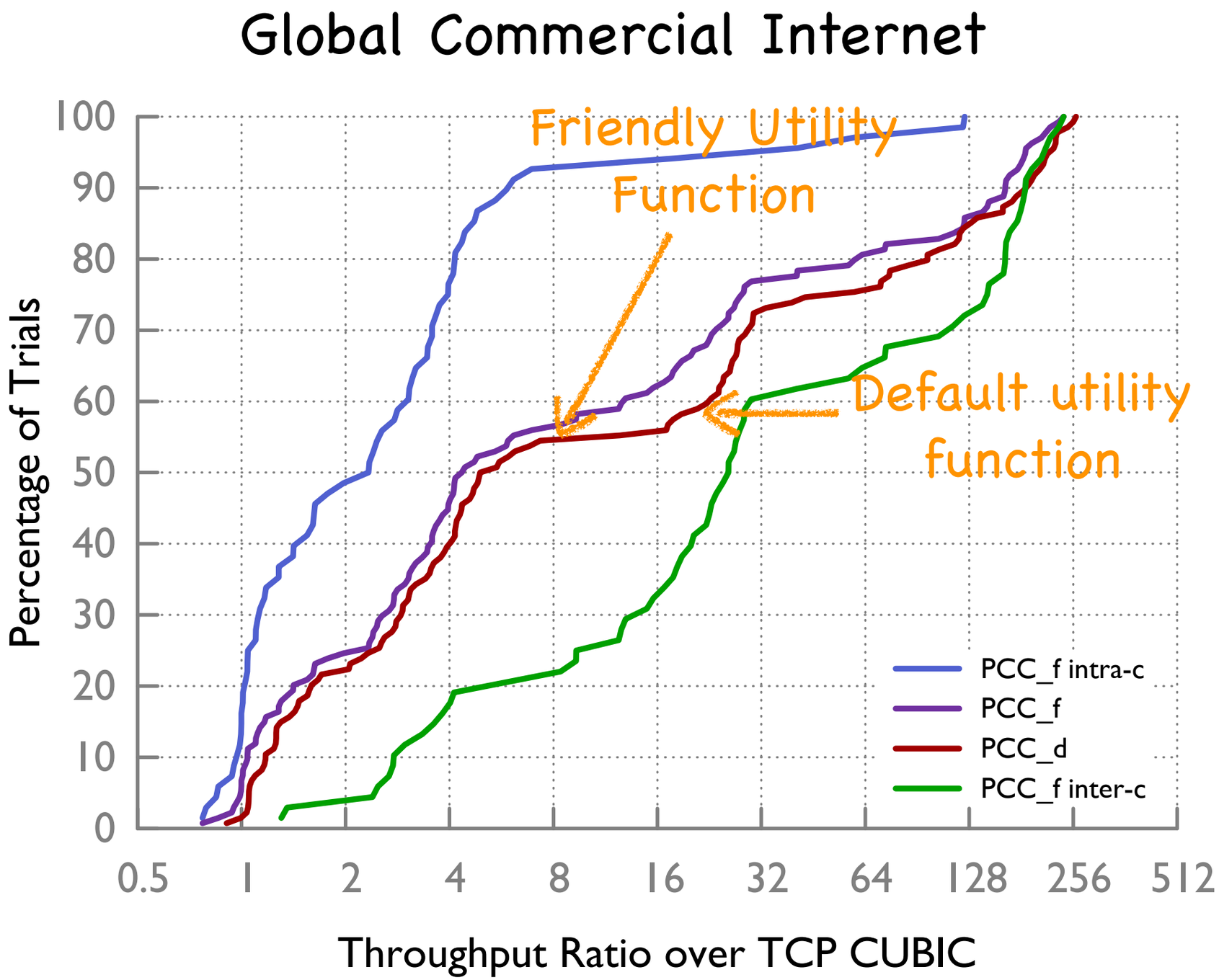
Consistent High Performance



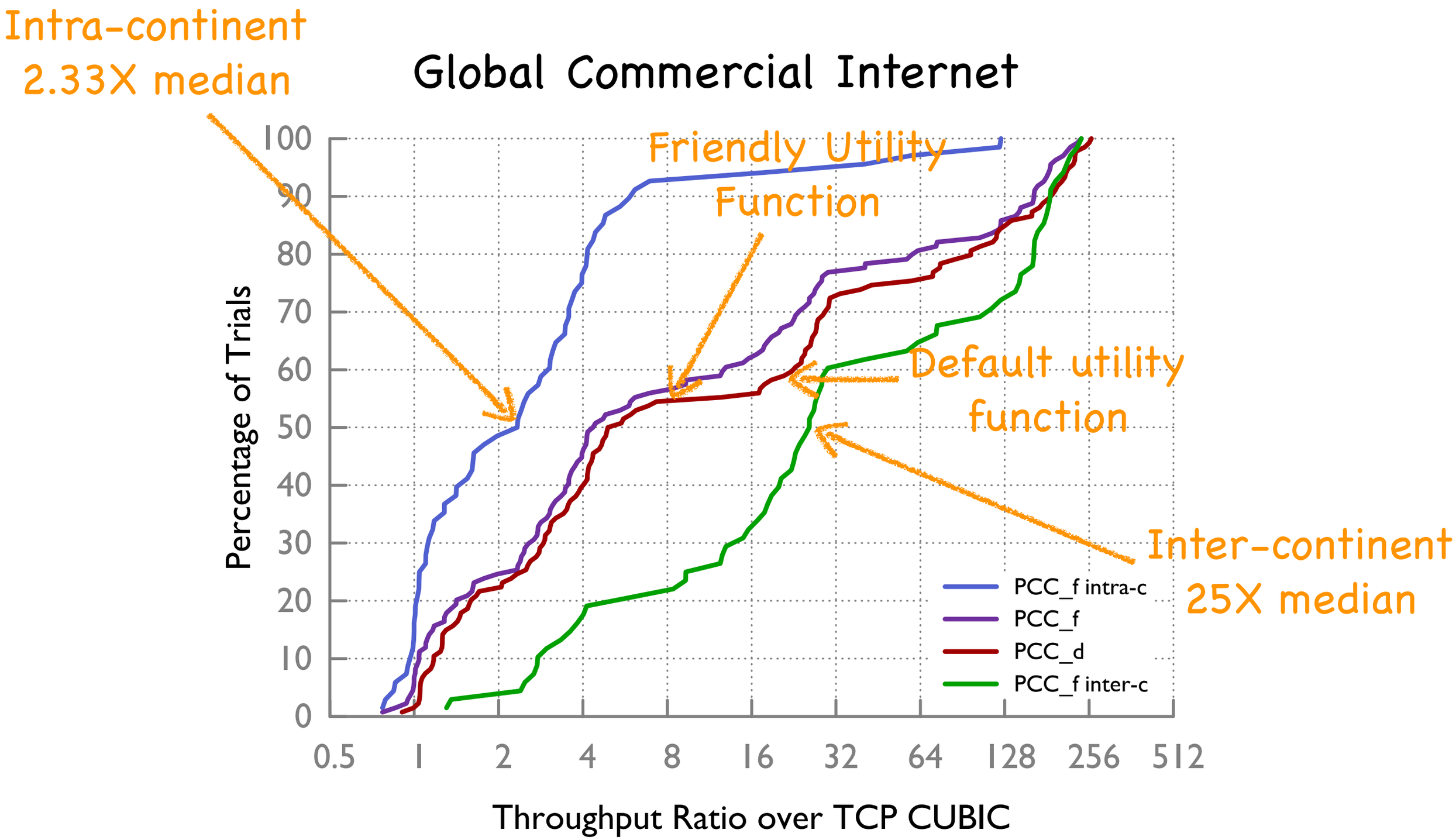
Consistent High Performance



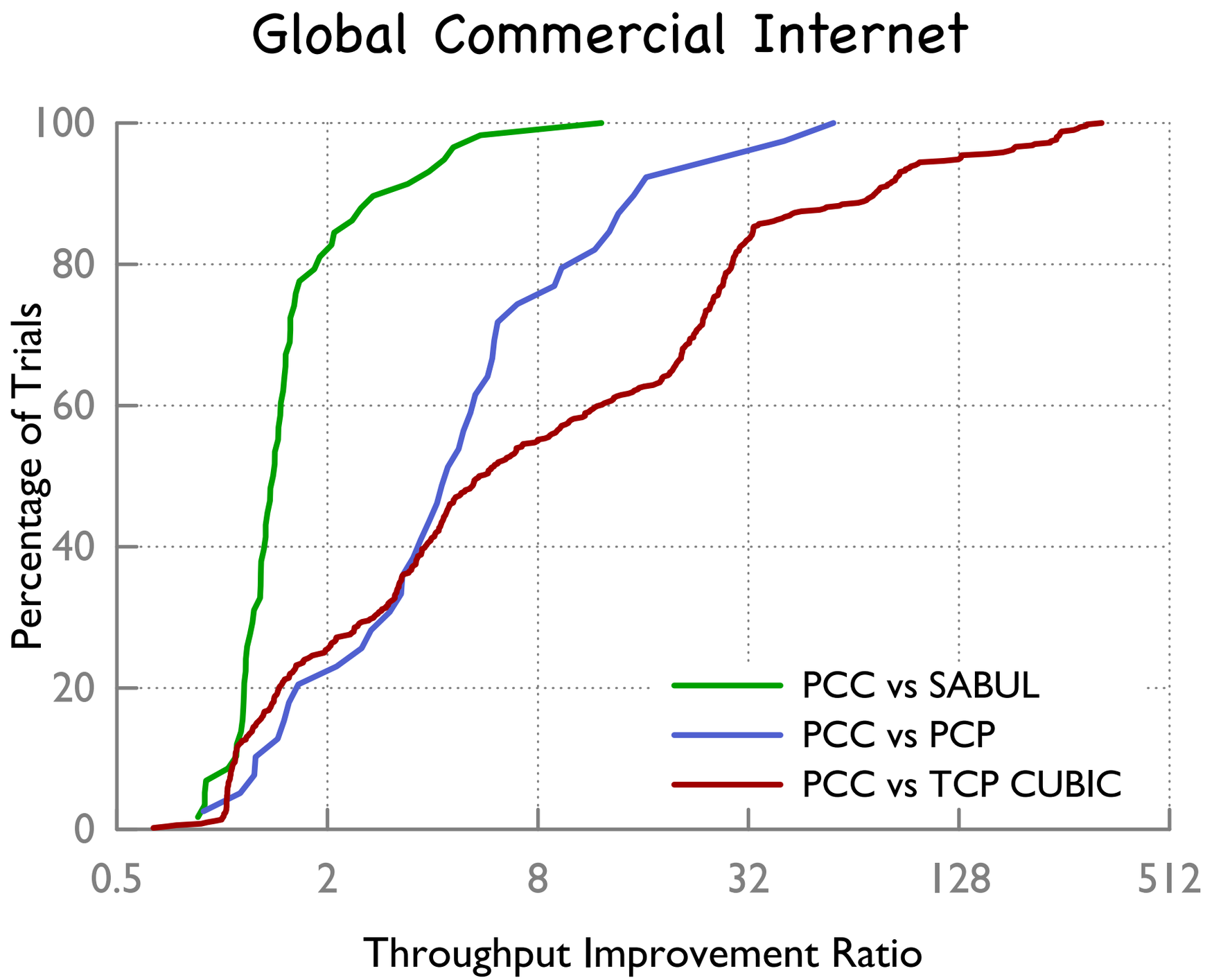
Consistent High Performance



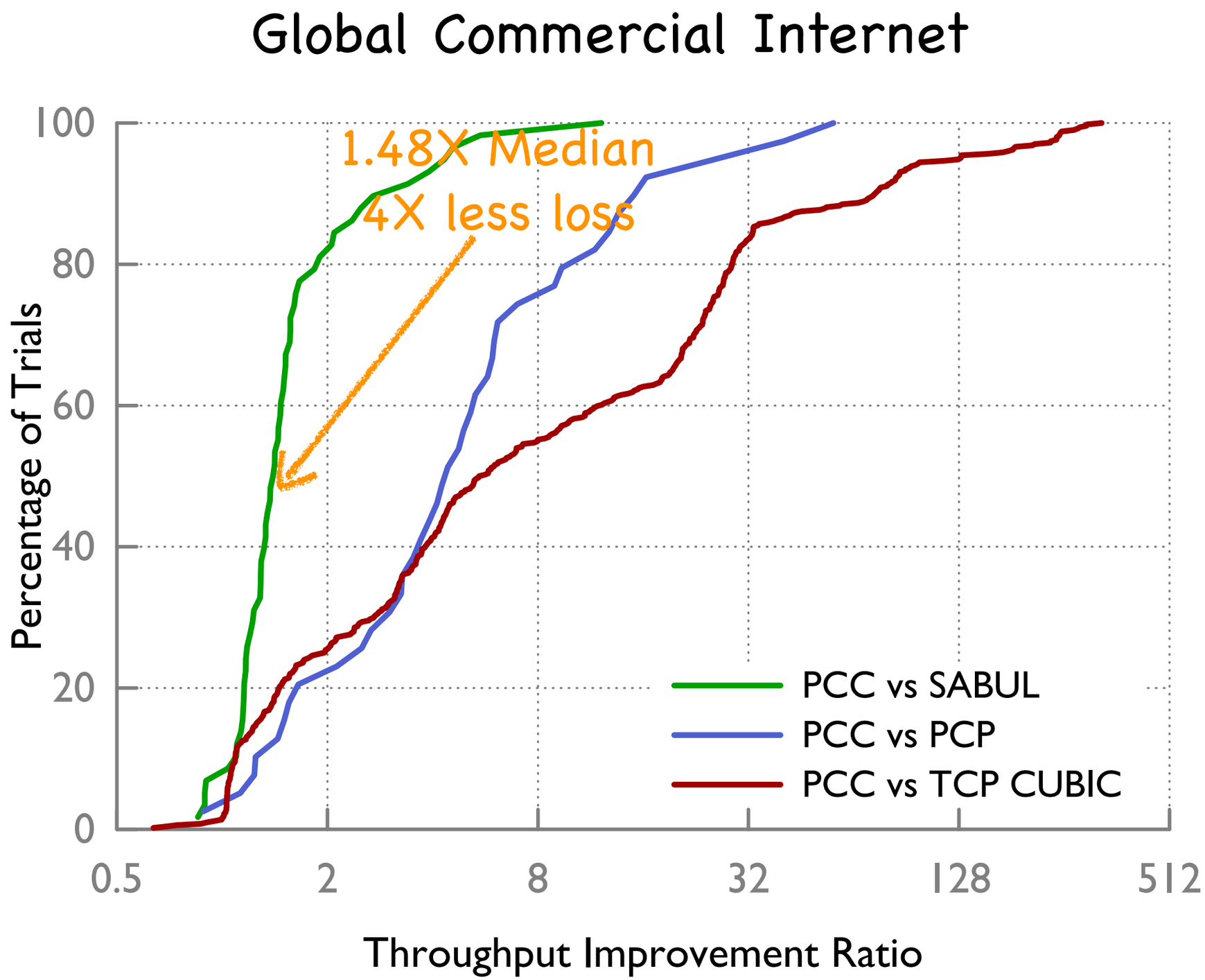
Consistent High Performance



Consistent High Performance



Consistent High Performance



Long list of things we have done but don't have time to talk about

- More stories about the fact that TCP is broken
- Proof of Nash Equilibrium and Convergence
- Concrete Implementation of PCC
  - Performance monitoring
  - Details of learning control algorithm
  - Implementation designs and optimizations
- Performance Evaluation
  - Inter data center networks
  - small buffer networks
  - Reactiveness and stability tradeoff
  - Jain index fairness
  - Benefit of Randomized Control Trials
  - Details of TCP friendliness evaluation
  - Emulated satellite networks
  - Emulated datacenter networks
  - Cure RTT unfairness
  - Does not fundamentally harm short flow FCT
  - Evaluation in the wild vs non-TCP protocols
- Flexibility by pluggable utility function

Long list of things we have done but don't have time to talk about

- More stories about the fact that TCP is broken
- Proof of Nash Equilibrium and Convergence
- Concrete Implementation of PCC
  - Performance monitoring
  - Details of learning control algorithm
  - Implementation designs and optimizations
- Performance Evaluation
  - Inter data center networks
  - small buffer networks
  - Reactiveness and stability tradeoff
  - Jain index fairness
  - Benefit of Randomized Control Trials
  - Details of TCP friendliness evaluation
  - Emulated satellite networks
  - Emulated datacenter networks
  - Cure RTT unfairness
  - Does not fundamentally harm short flow FCT
  - Evaluation in the wild vs non-TCP protocols
- Flexibility by pluggable utility function

11 more!

Same Rate Control Algorithm

Same Rate Control Algorithm

+

Same Rate Control Algorithm

+

Different Utility Function

=

Flexibility

Limitation and Future Work

Limitation and Future Work

- Understanding of Utility Function and Convergence
- Better Control Algorithm
- Better Scalability

Limitation and Future Work

- Understanding of Utility Function and Convergence
- Better Control Algorithm
- Better Scalability

Related Works

Related Works

- TCP Family: TCP exMachina (Remy)

Related Works

- TCP Family: TCP exMachina (Remy)
- Non-TCP protocols: PCP

Related Works

- TCP Family: TCP exMachina (Remy)
- Non-TCP protocols: PCP
- In-network feedback protocols: DCTCP, XCP

- Rate control based on empirically observed performance yields
  - Consistent high performance
  - Better stability than TCP
  - Flexible performance objectives
  - Relatively easy and safe to deploy

speedier.net

speedier.net/pcc

github.com/modong/pcc

Find me to see a demo