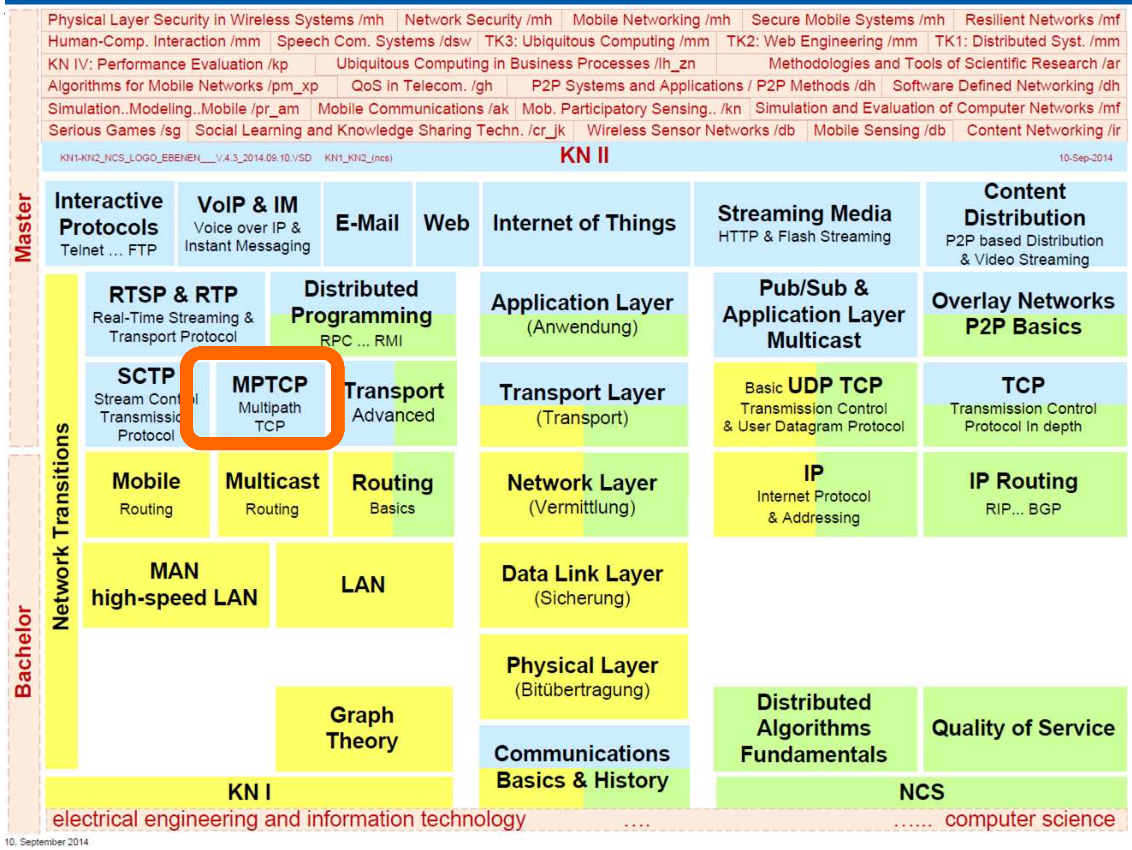
Communication Networks 2

Multipath TCP



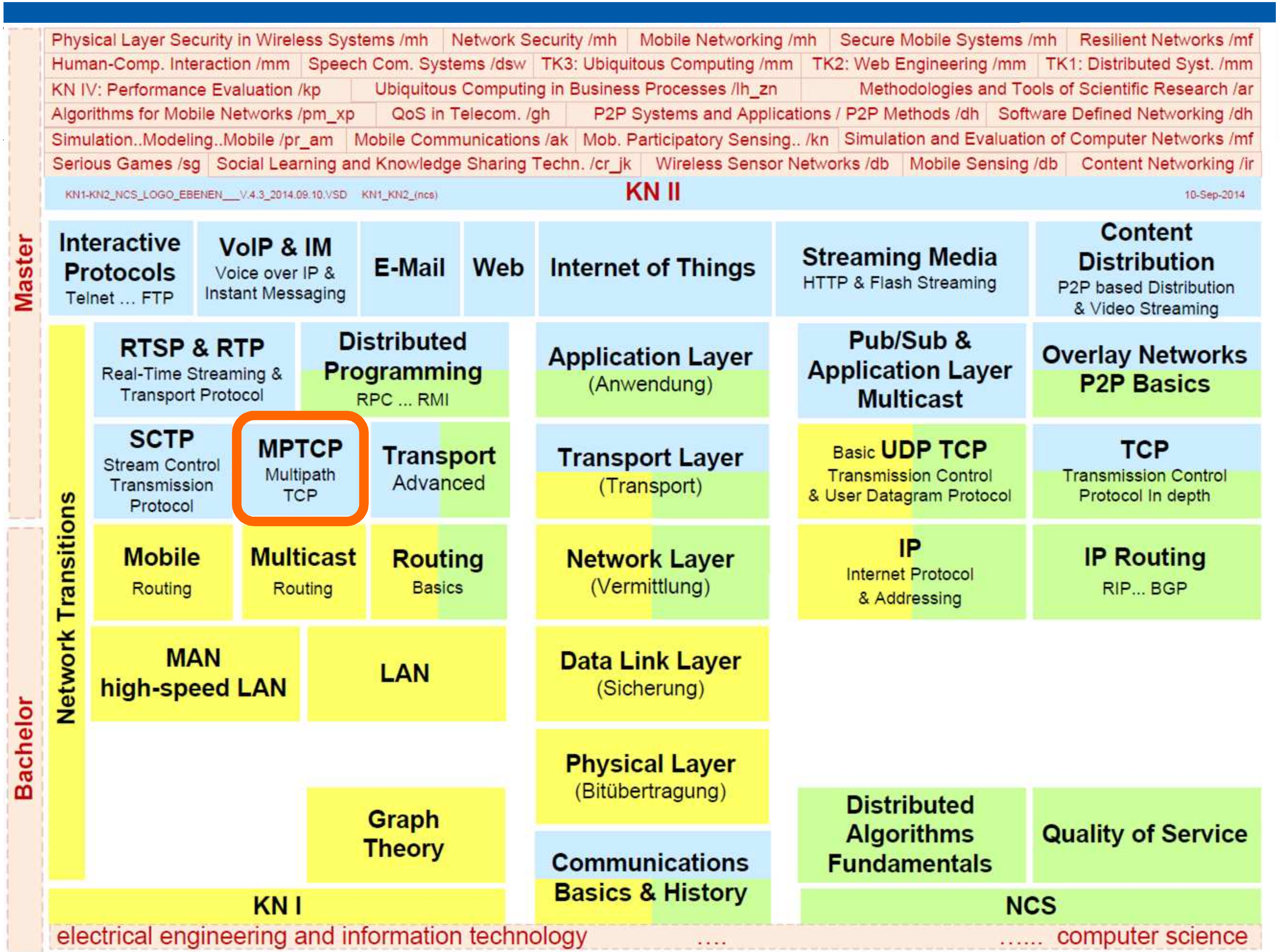
Slides adopted from

Olivier Bonaventure
«Decoupling TCP from IP with Multipath TCP»
http://inl.info.ucl.ac.be

Prof. Dr.-Ing. **Ralf Steinmetz**
KOM - Multimedia Communications Lab

Overview

TECHNISCHE UNIVERSITÄT DARMSTADT



10. September 2014

The motivations for Multipath TCP

The changing Internet

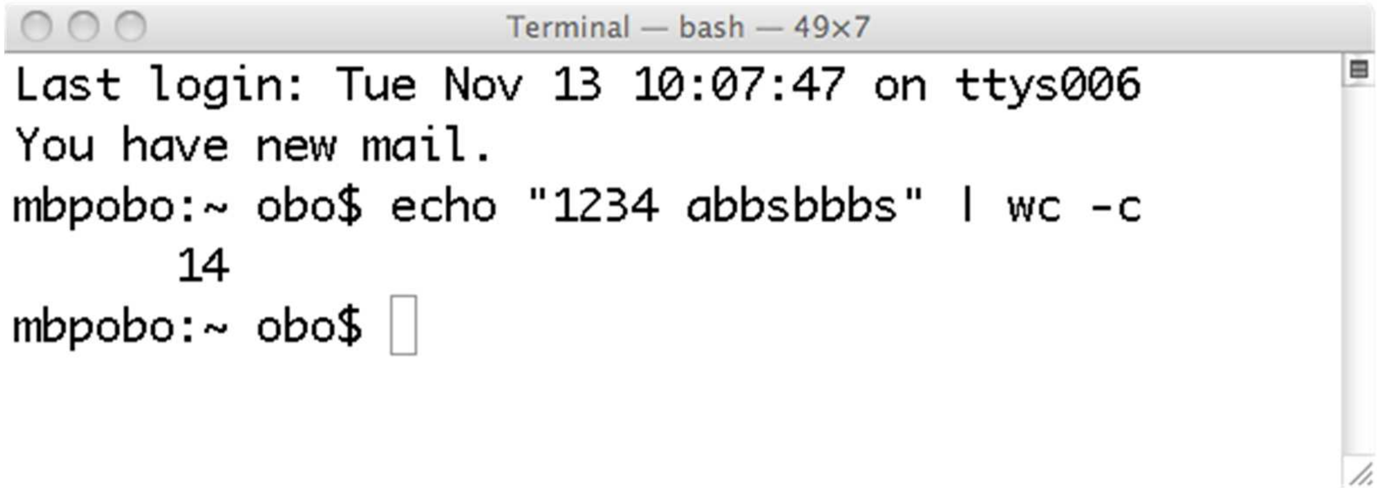
The Multipath TCP Protocol

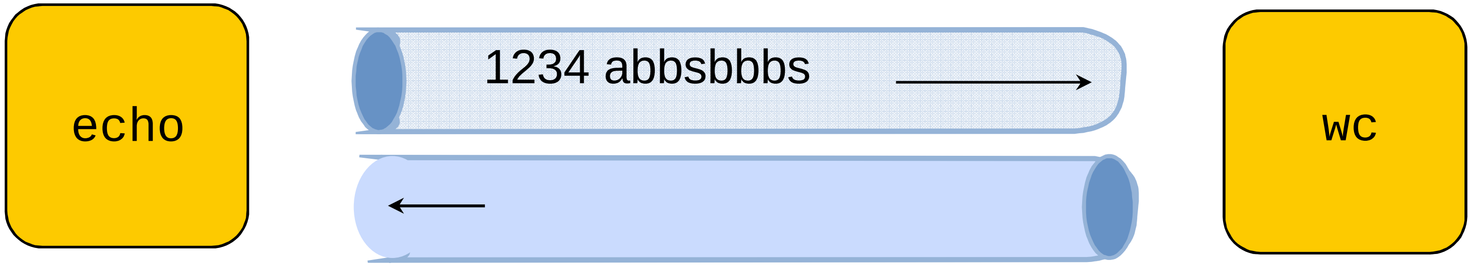
Multipath TCP use cases

1.1 The Origins of TCP

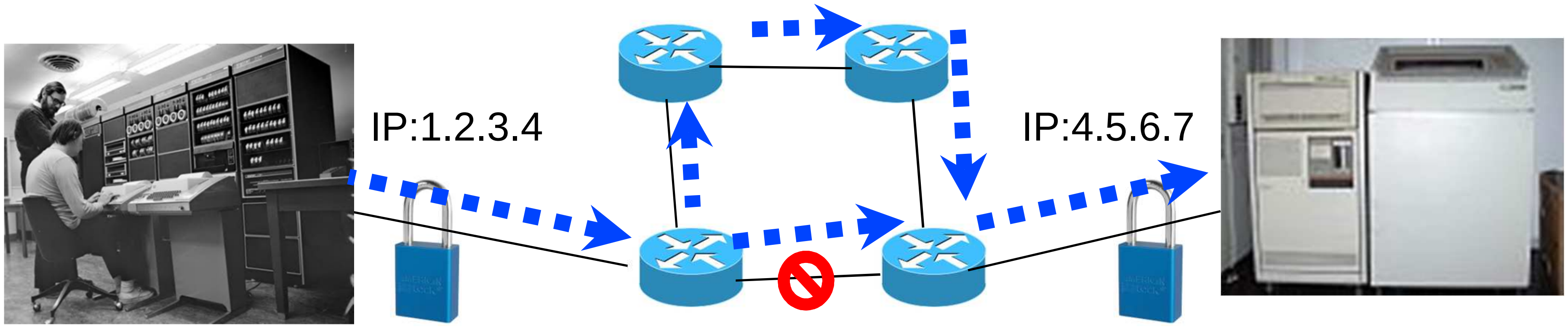
Source : http://spectrum.ieee.org/computing/software/the-strange-birth-and-long-life-of-unix

The Unix pipe model





The TCP bytestream model



Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Endhosts have evolved

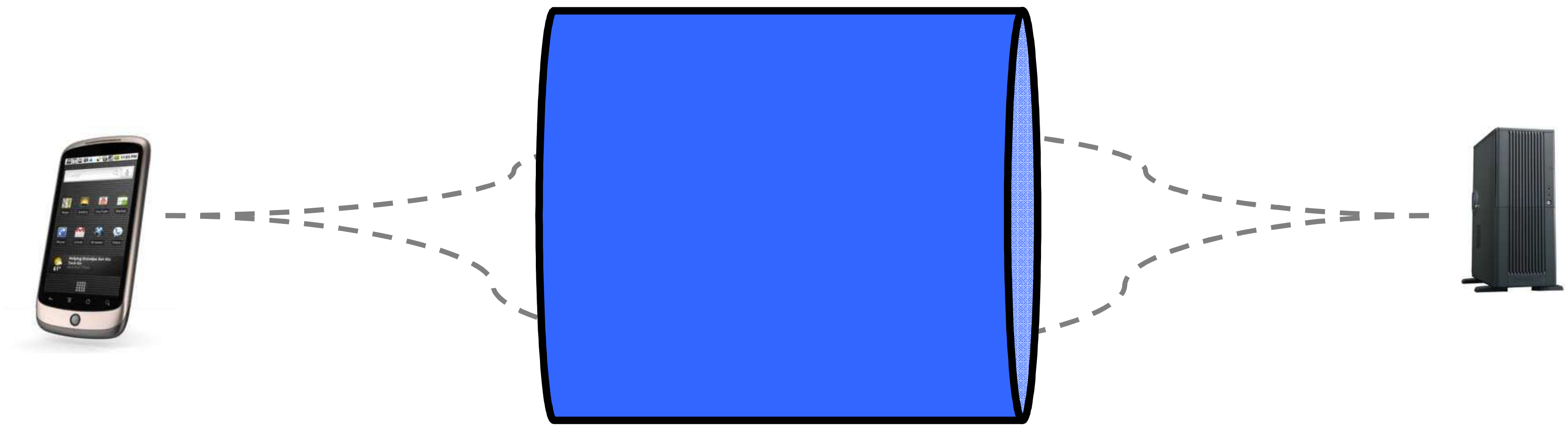




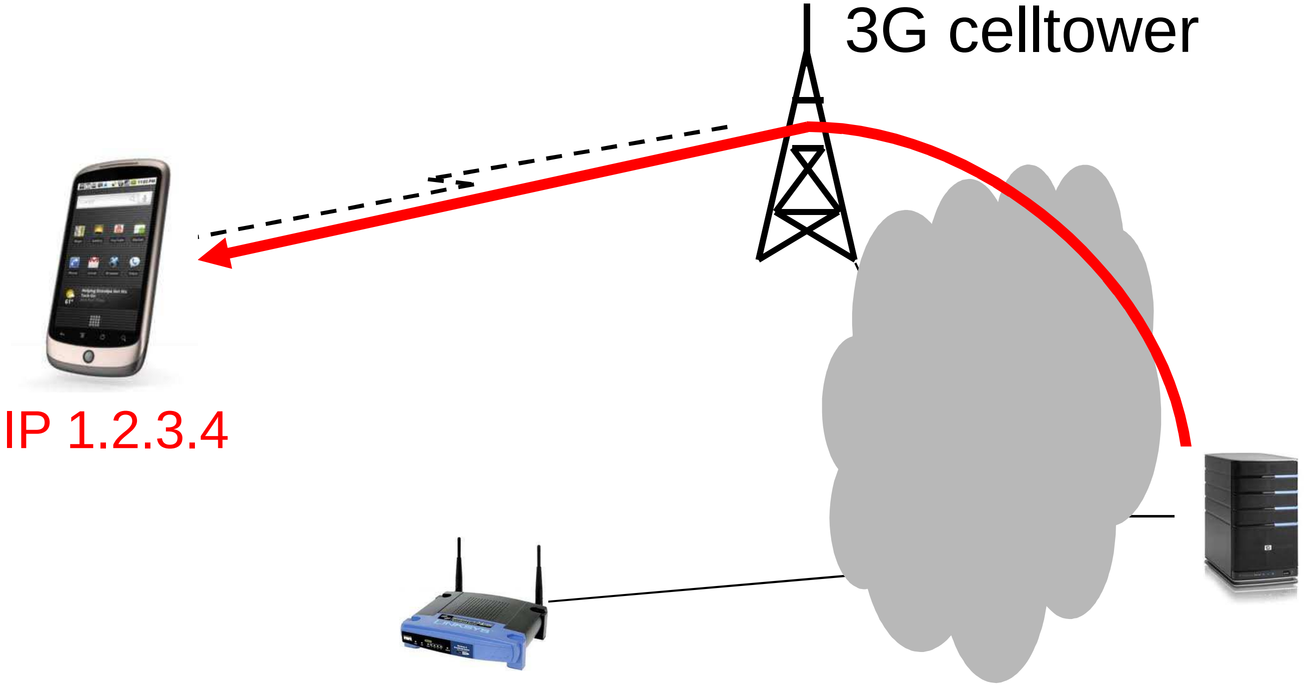



Mobile devices have multiple wireless interfaces

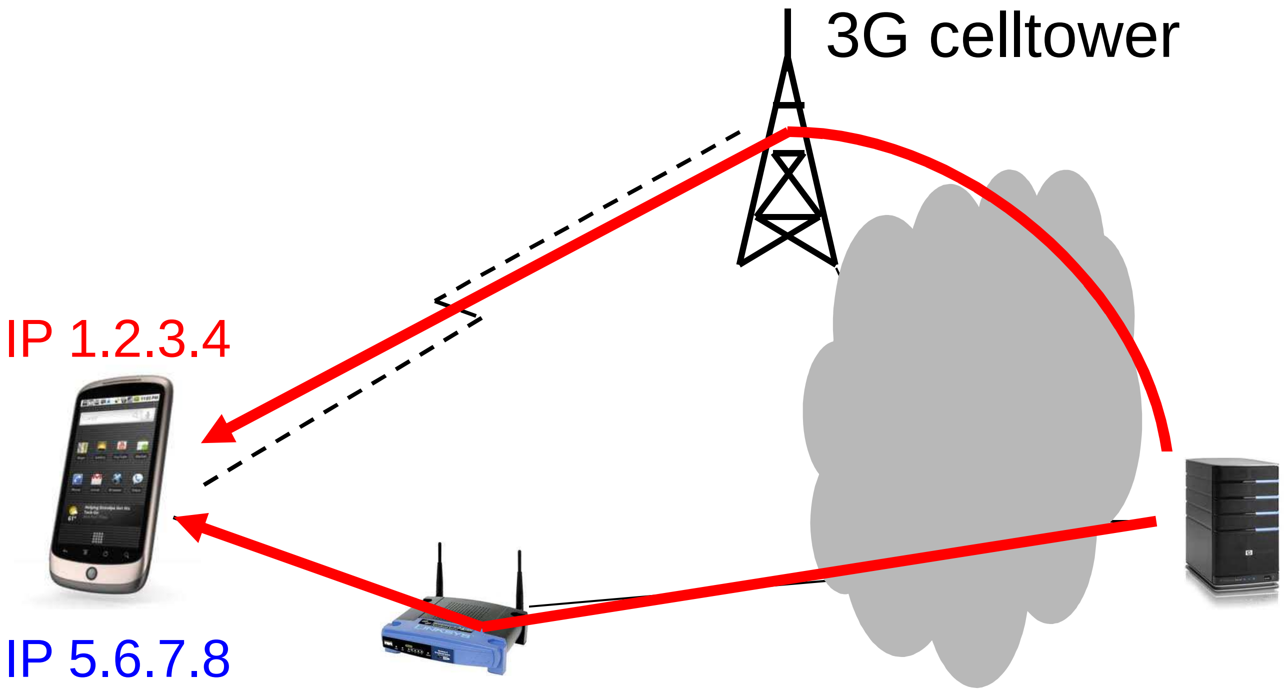
User Expectations



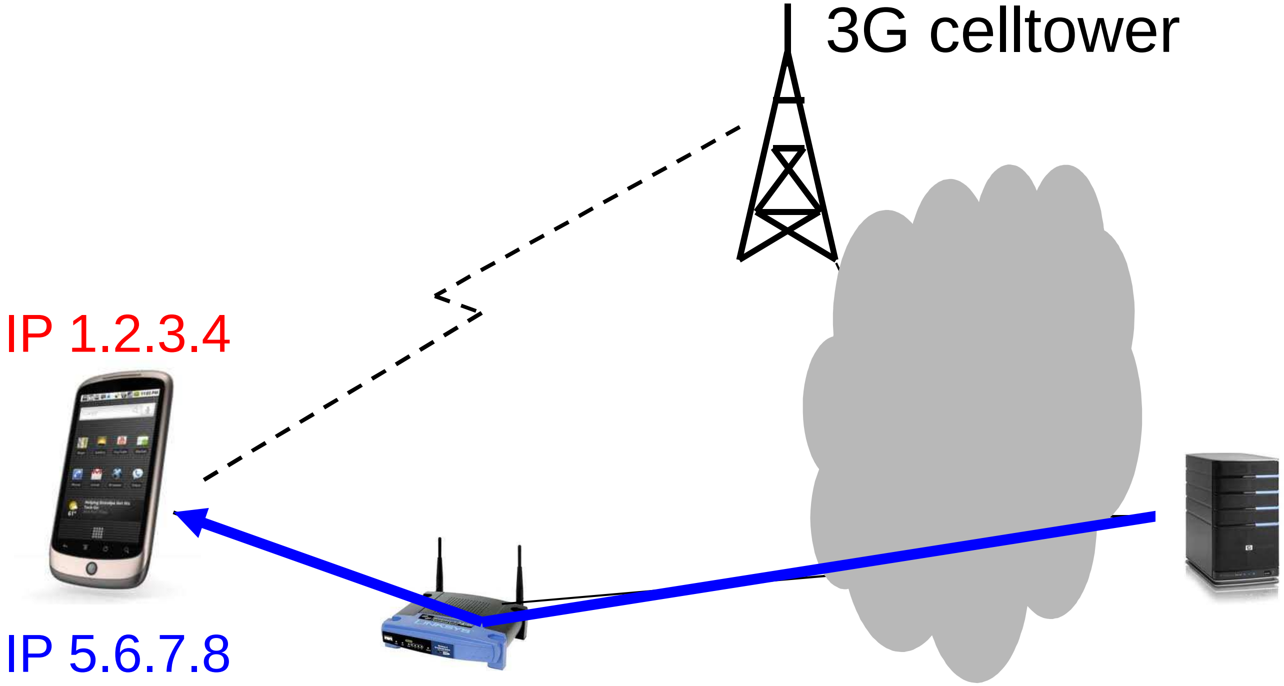
What Technology provides



What Technology provides



What Technology provides



When IP addresses change TCP connections have to be re-established !

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Datacenters

The motivations for Multipath TCP

The changing Internet

The Multipath TCP Protocol

Multipath TCP use cases

The Internet architecture that we explain to our students

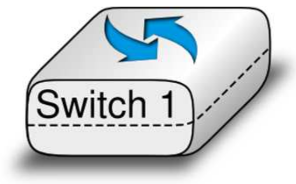
Application

Transport

Network

Datalink

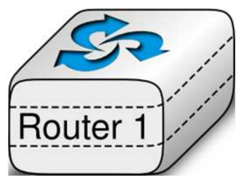
Physical



Datalink

Physical

Physical



Network

Datalink

Physical

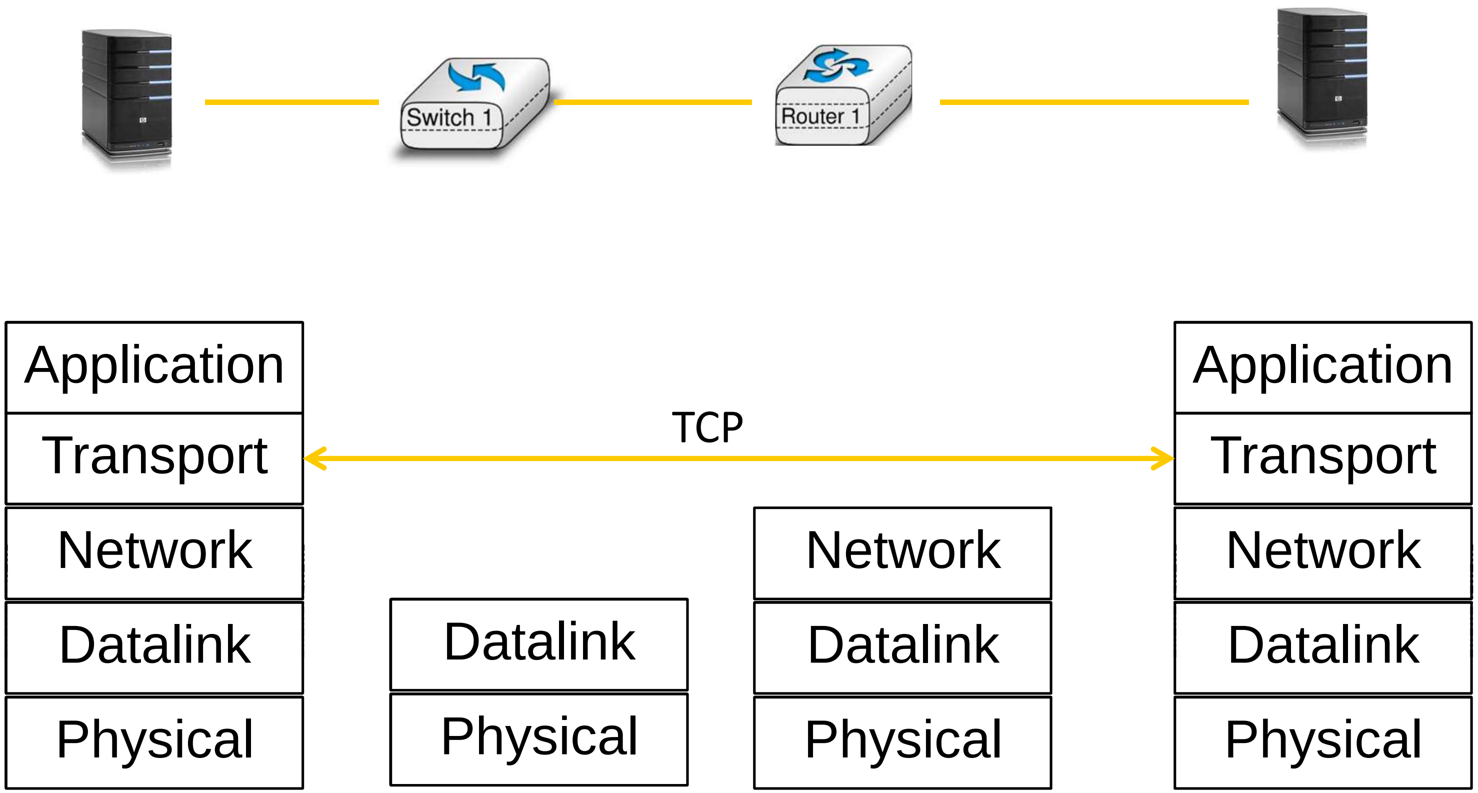
O. Bonaventure, Computer networking : Principles, Protocols and Practice, open ebook, http://inl.info.ucl.ac.be/cnp3

A typical "academic" network

Application			Application
Transport			Transport
Network		Network	Network
Datalink	Datalink	Datalink	Datalink
Physical	Physical	Physical	Physical

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

The end-to-end principle



Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

In reality

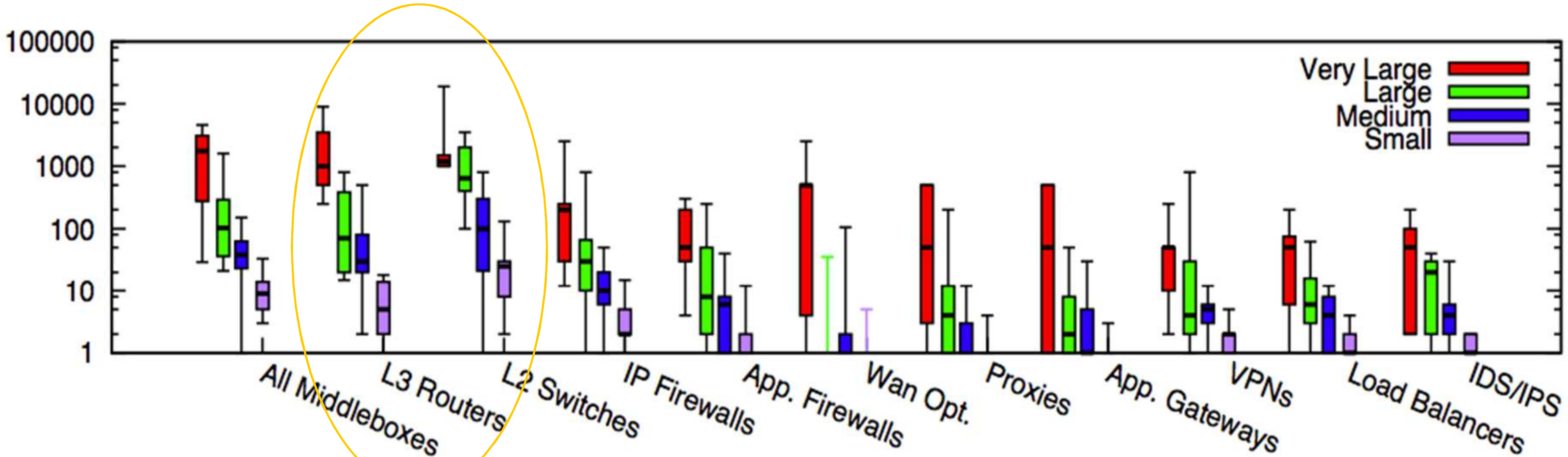
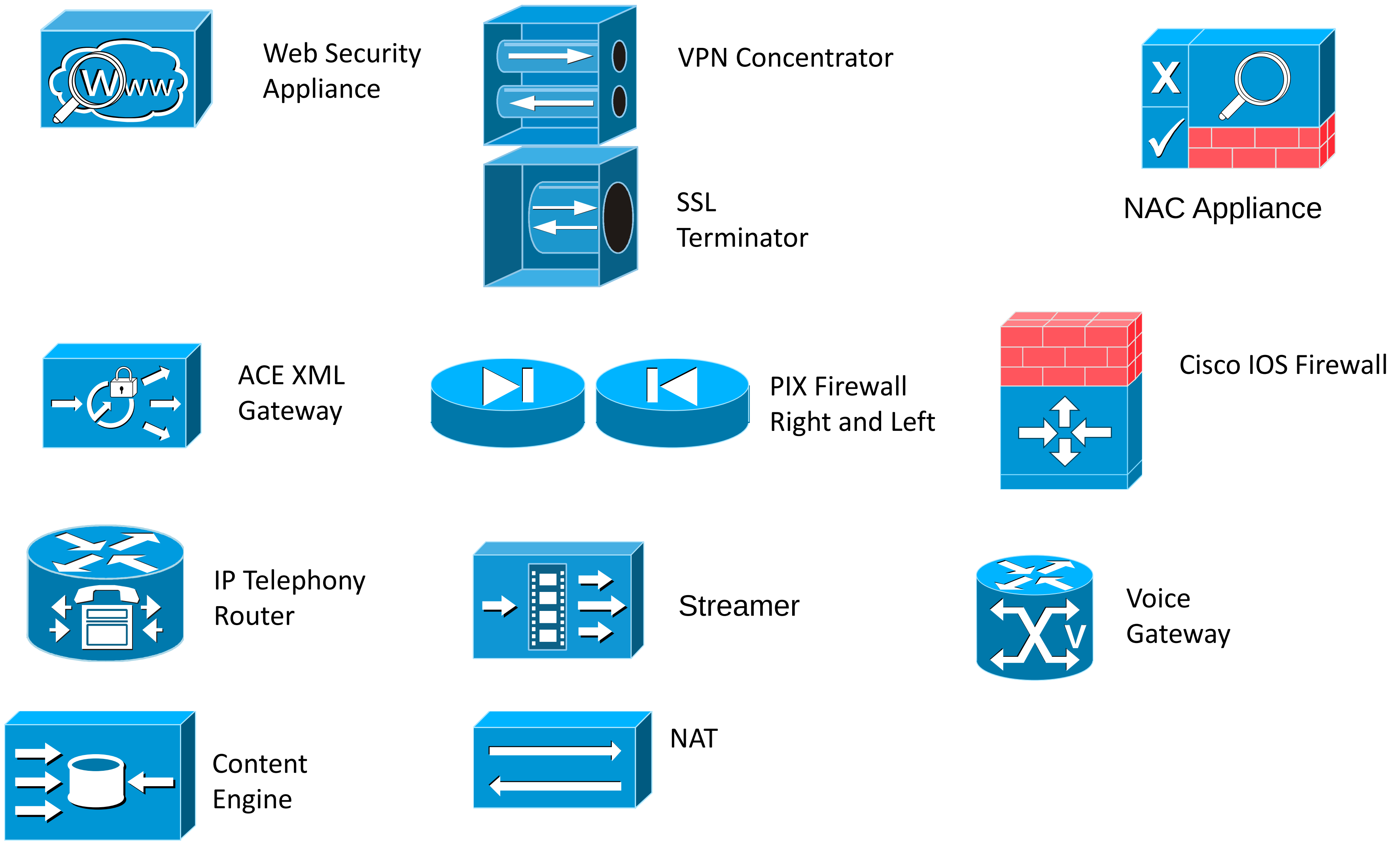


Figure 1: Box plot of middlebox deployments for small (fewer than 1k hosts), medium (1k-10k hosts), large (10k-100k hosts), and very large (more than 100k hosts) enterprise networks. Y-axis is in log scale.

- almost as many middleboxes as routers
- various types of middleboxes are deployed

Sherry, Justine, et al. "*Making middleboxes someone else's problem: Network processing as a cloud service*." Proceedings of the ACM SIGCOMM 2012 conference. ACM, 2012.

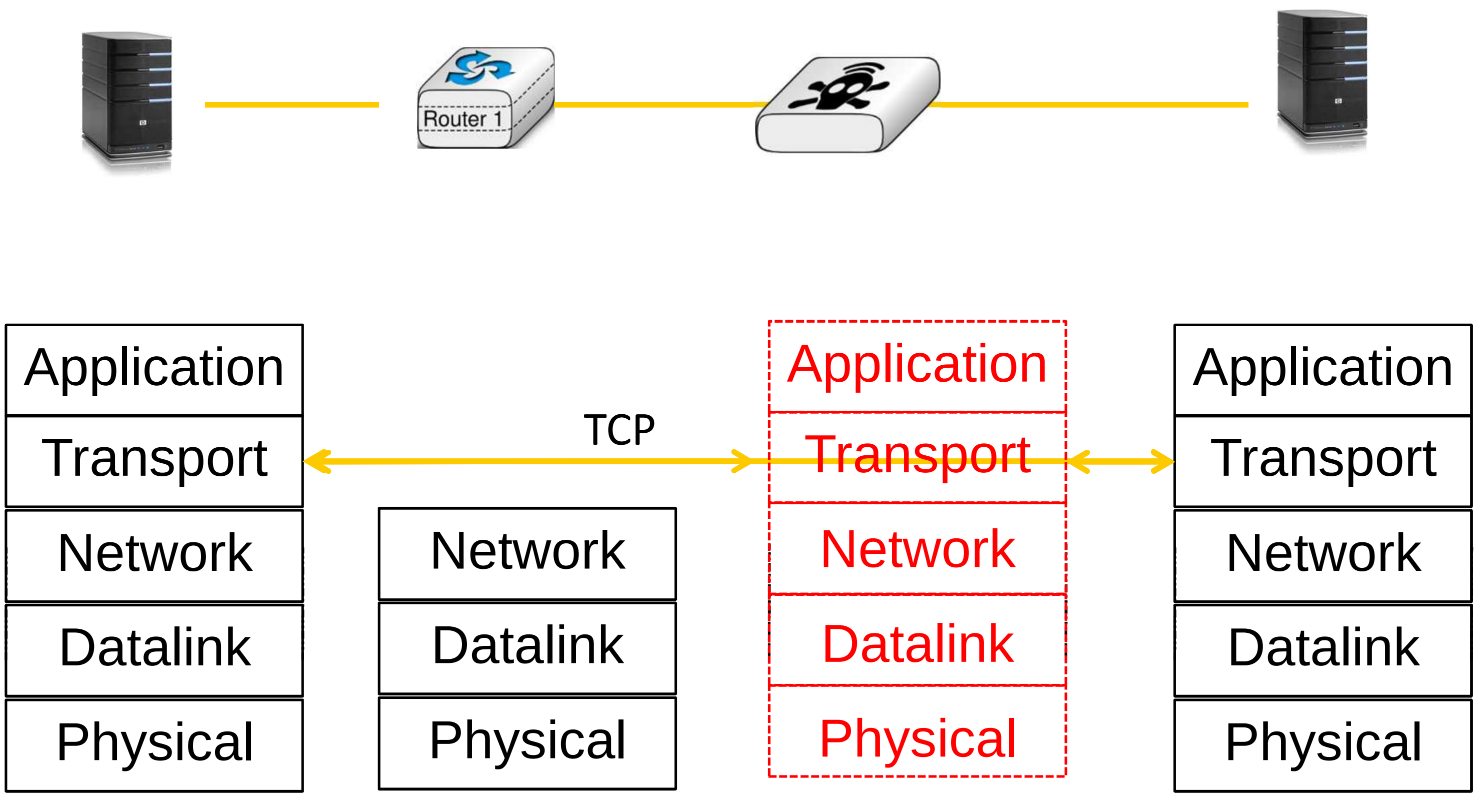
A Middlebox Zoo



http://www.cisco.com/web/about/ac50/ac47/2.html

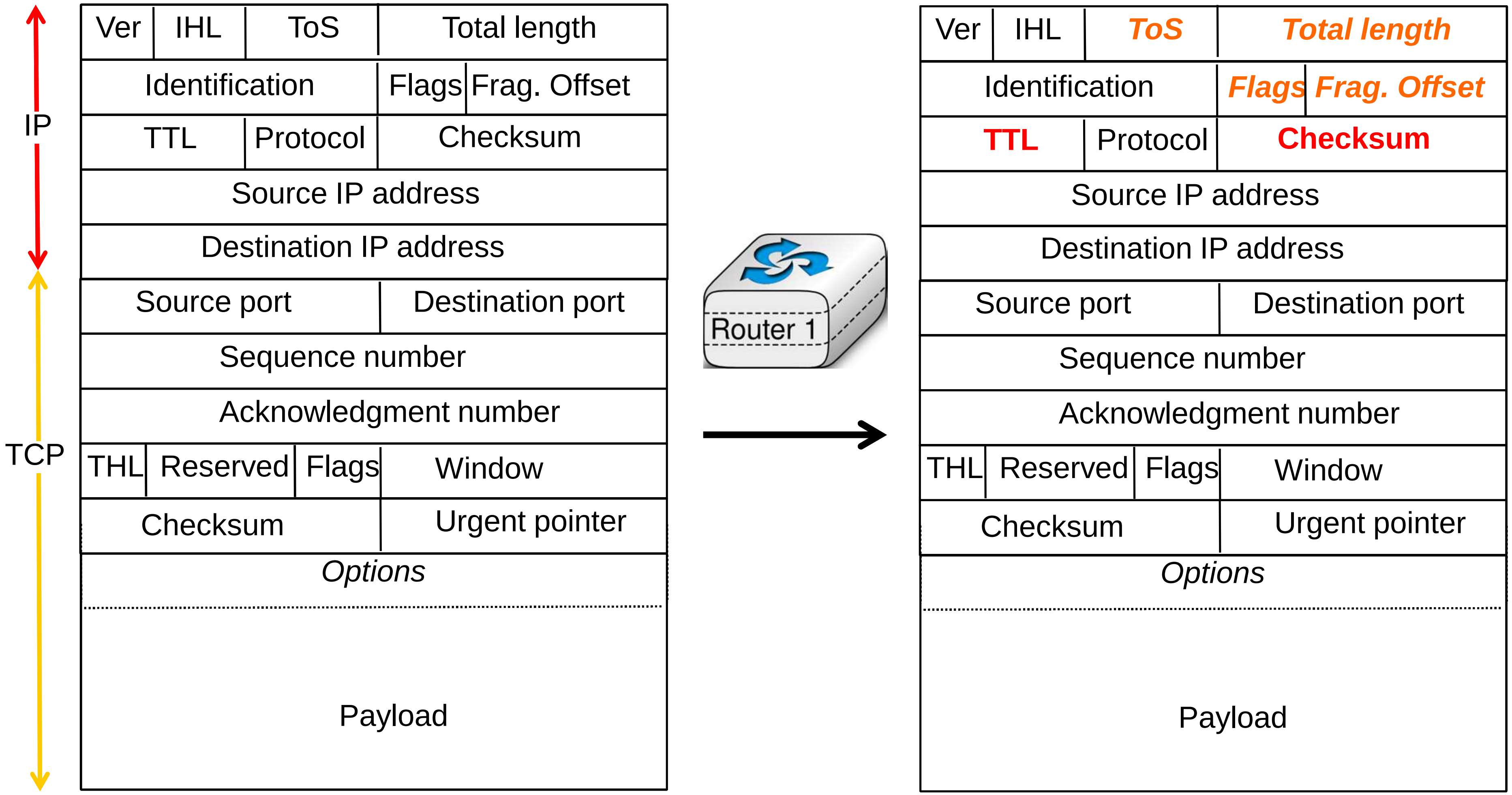
How to model those Middleboxes ?

- **In the official architecture, they do not exist**
- **In reality...**



Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

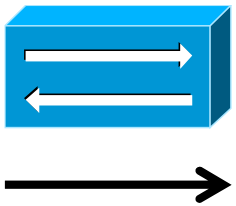
TCP segments processed by a router



Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

TCP segments processed by a NAT

Ver	IHL	ToS	Total length
Identification		Flags	Frag. Offset
TTL	Protocol	Checksum	
Source IP address			
Destination IP address			
Source port		Destination port	
Sequence number			
Acknowledgment number			
THL	Reserved	Flags	Window
Checksum		Urgent pointer	
Options			
Payload			



Ver	IHL	***ToS***	***Total length***
Identification		***Flags***	***Frag. Offset***
TTL	Protocol	**Checksum**	
Source IP address			
Destination IP address			
Source port		**Destination port**	
Sequence number			
Acknowledgment number			
THL	Reserved	Flags	Window
Checksum		Urgent pointer	
Options			
Payload			

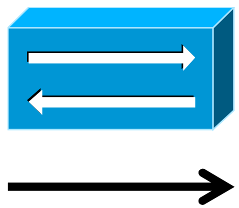
TCP segments processed by a NAT (2)

active mode ftp behind a NAT

```
220 ProFTPD 1.3.3d Server (BELNET FTPD Server) [193.190.67.15]
ftp_login: user `<null>' pass `<null>' host `ftp.belnet.be'
Name (ftp.belnet.be:obo): anonymous
---> USER anonymous
331 Anonymous login ok, send your complete email address as your password
Password:
---> PASS XXXX
---> PORT 192,168,0,7,195,120
200 PORT command successful
---> LIST
150 Opening ASCII mode data connection for file list
lrw-r--r--   1 ftp      ftp            6 Jun  1  2011 pub -> mirror
226 Transfer complete
```

TCP segments processed by an ALG running on a NAT

Ver	IHL	ToS	Total length
Identification		Flags	Frag. Offset
TTL	Protocol	Checksum	
Source IP address			
Destination IP address			
Source port		Destination port	
Sequence number			
Acknowledgment number			
THL	Reserved	Flags	Window
Checksum		Urgent pointer	
Options			
Payload			



Ver	IHL	***ToS***	***Total length***
Identification		***Flags***	***Frag. Offset***
TTL	Protocol	**Checksum**	
Source IP address			
Destination IP address			
Source port		**Destination port**	
Sequence number			
Acknowledgment number			
THL	Reserved	Flags	Window
Checksum		Urgent pointer	
Options			
Payload			

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

The motivations for Multipath TCP

The changing Internet

➡ **The Multipath TCP Protocol**

Multipath TCP use cases

Multipath TCP is an *evolution* of TCP

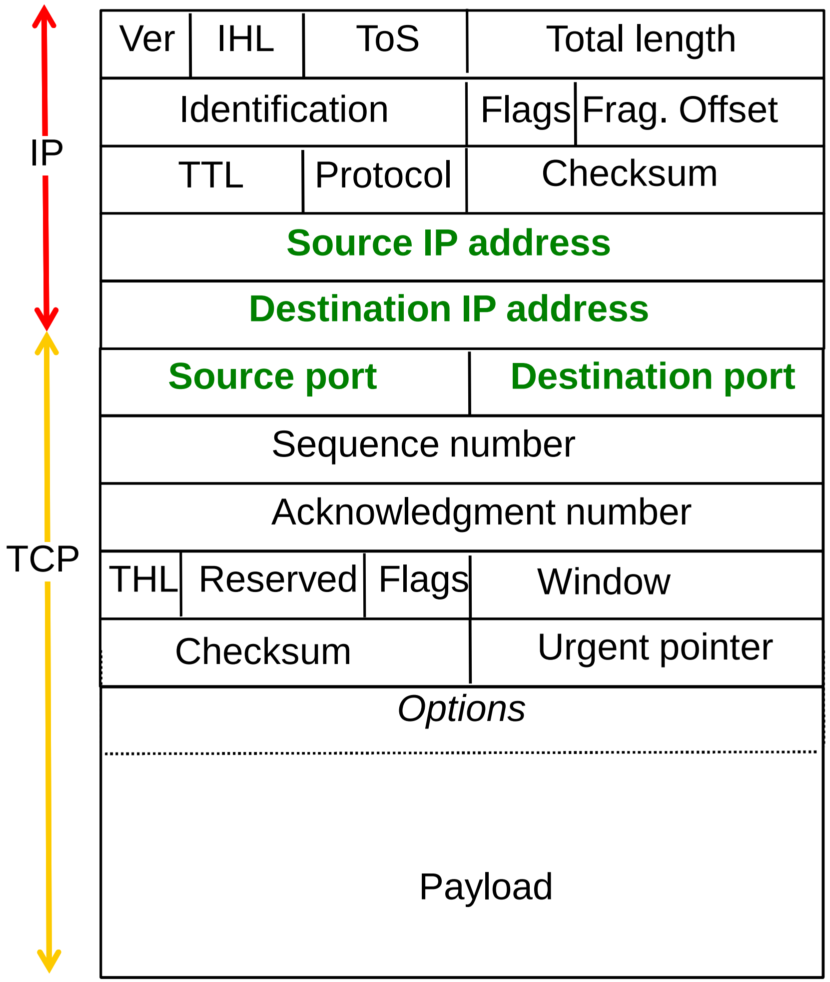
Design objectives

- Support unmodified applications
- Work over today's networks
- Works in all networks where regular TCP works

RFC 6824: TCP Extensions for Multipath Operation with Multiple Addresses
RFC 6182: Architectural Guidelines for Multipath TCP Development
RFC 6356: Coupled Congestion Control for Multipath Transport Protocols

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Identification of a TCP connection

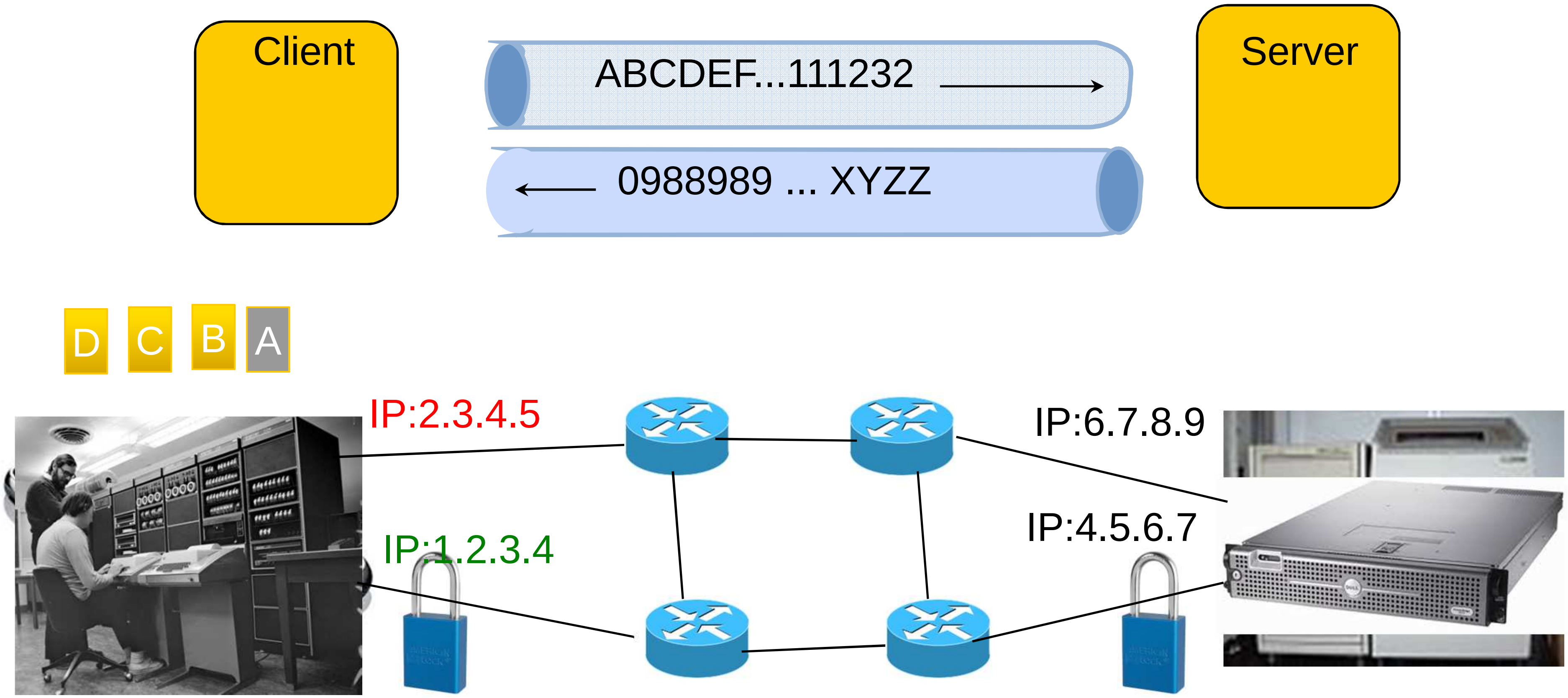


Four tuple

- IP_{source}
- IP_{dest}
- $Port_{source}$
- $Port_{dest}$

All TCP segments contain the four tuple

The *new* bytestream model



Control plane

- How to manage a Multipath TCP connection that uses several paths ?

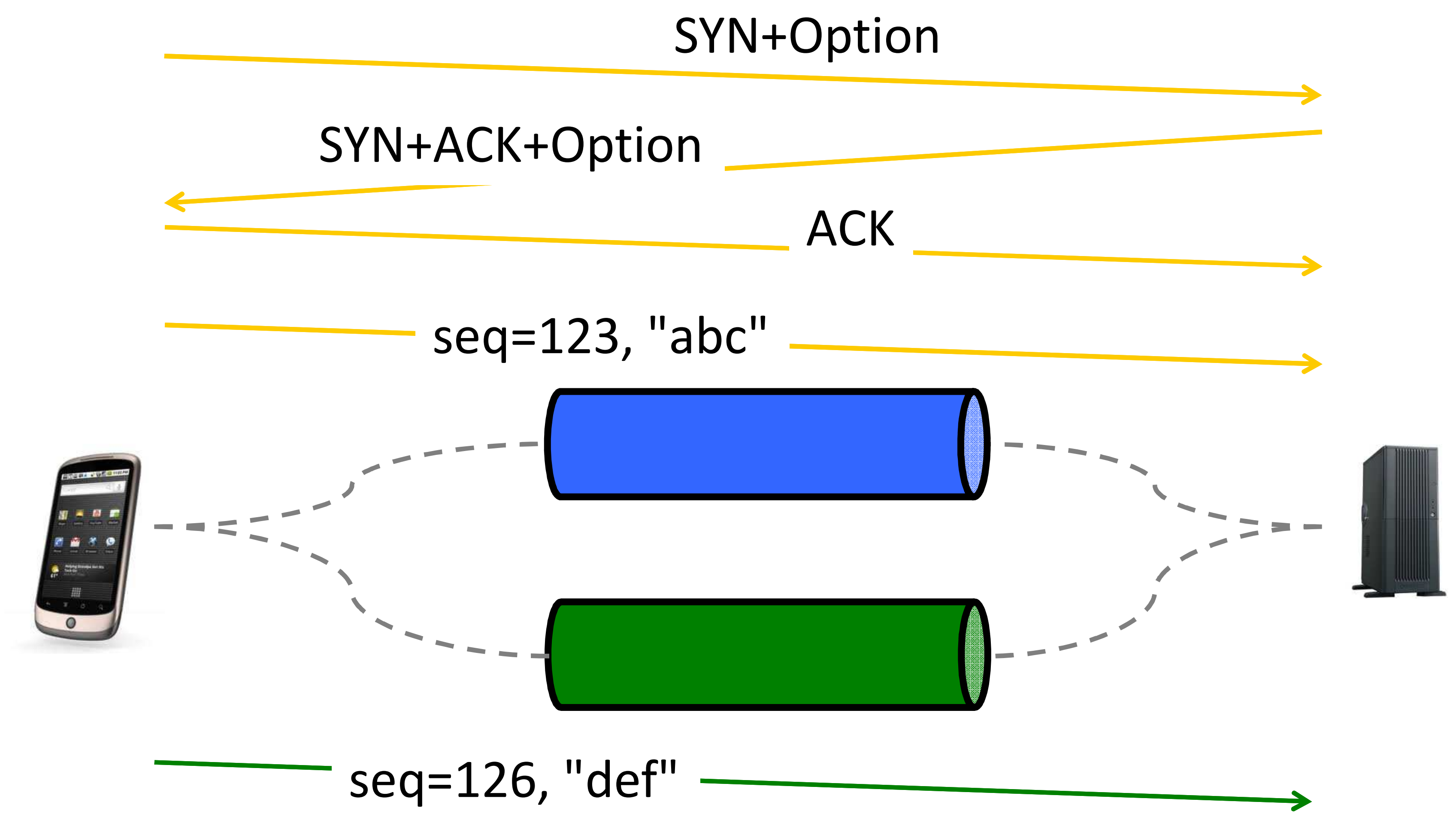
Data plane

- How to transport data ?

Congestion control

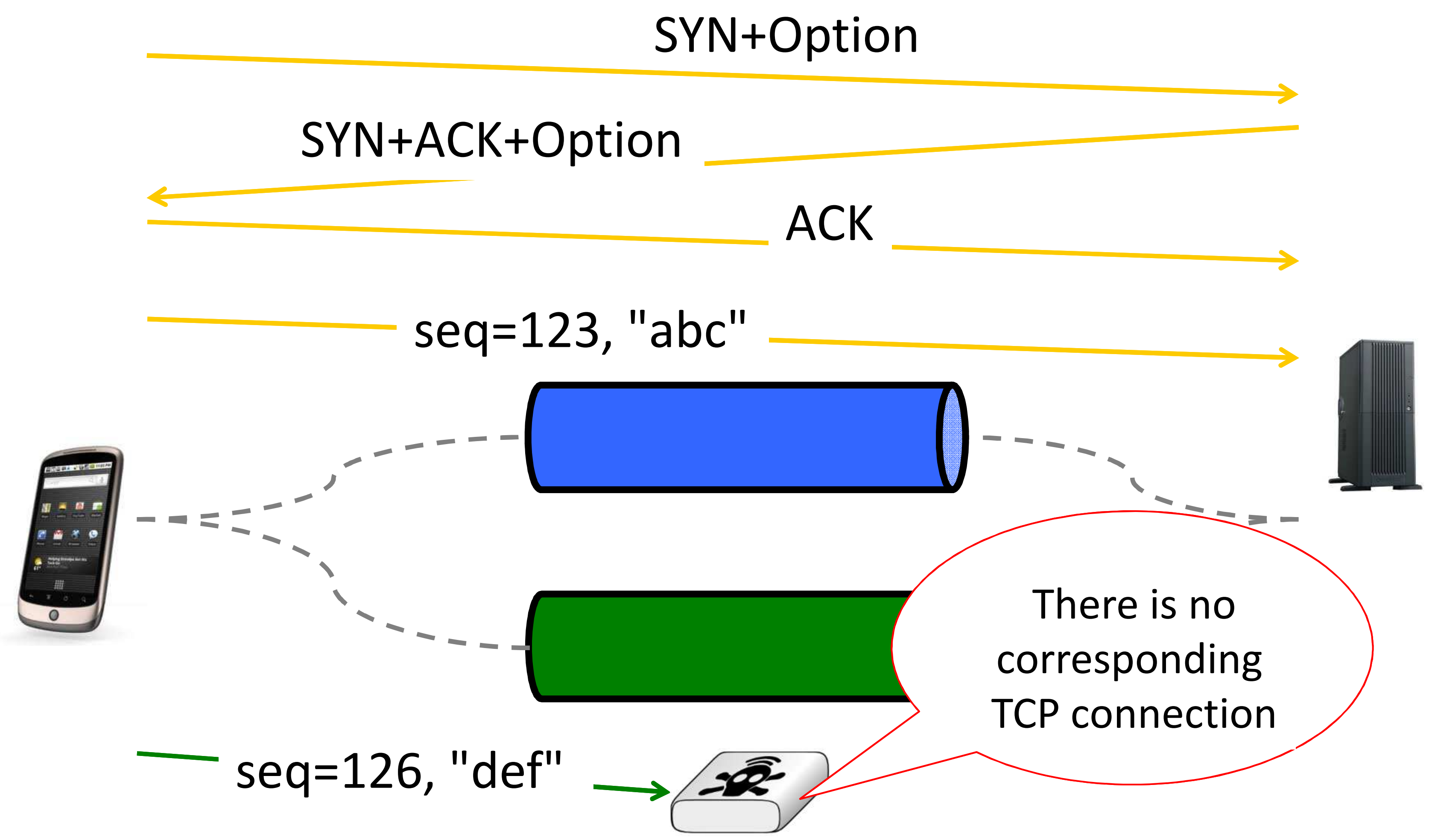
- How to control congestion over multiple paths ?

A naïve Multipath TCP



Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

A naïve Multipath TCP
In today's Internet ?

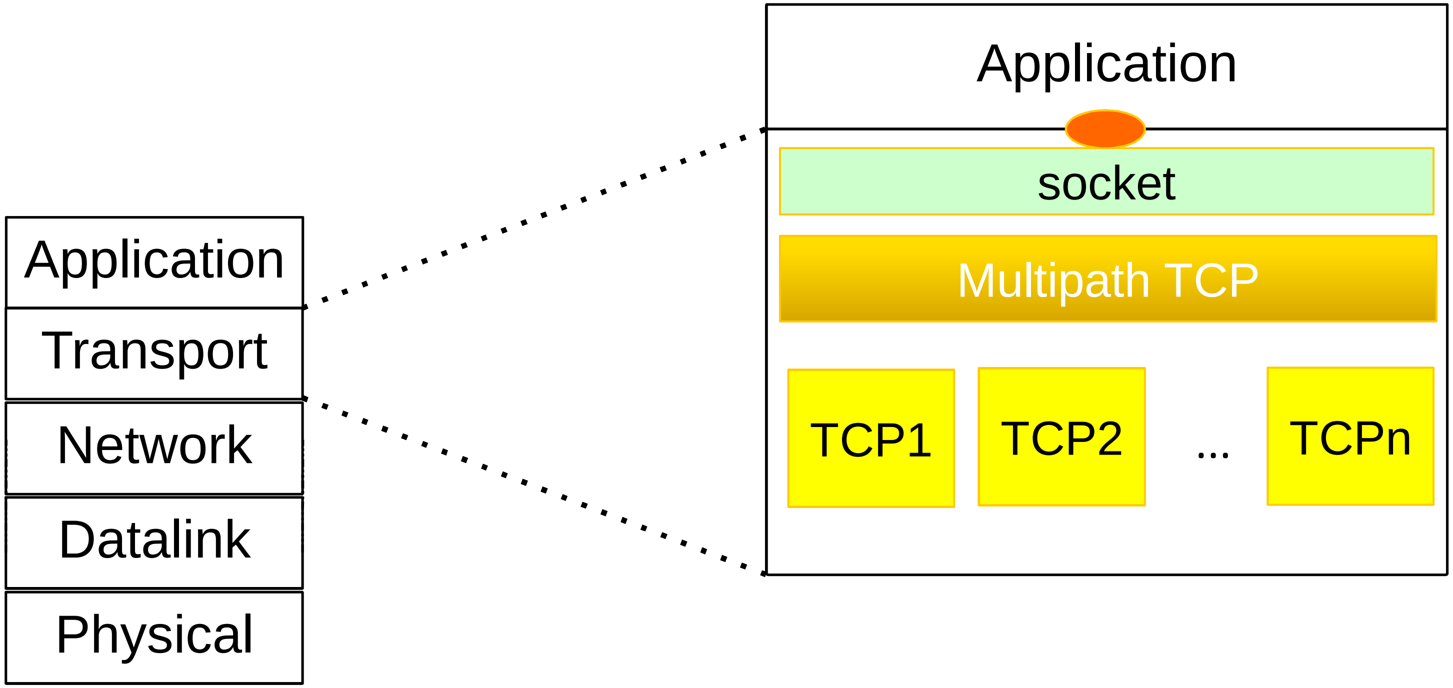


Design decision

- *A Multipath TCP connection is composed of one of more regular TCP subflows that are combined*
- Each host maintains state that glues the TCP subflows that compose a Multipath TCP connection together
- Each TCP subflow is sent over a single path and appears like a **regular TCP** connection along this path

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Multipath TCP and the architecture



A. Ford, C. Raiciu, M. Handley, S. Barre, and J. Iyengar, "Architectural guidelines for multipath TCP development", RFC6182 2011.

What is a *regular* TCP connection ?

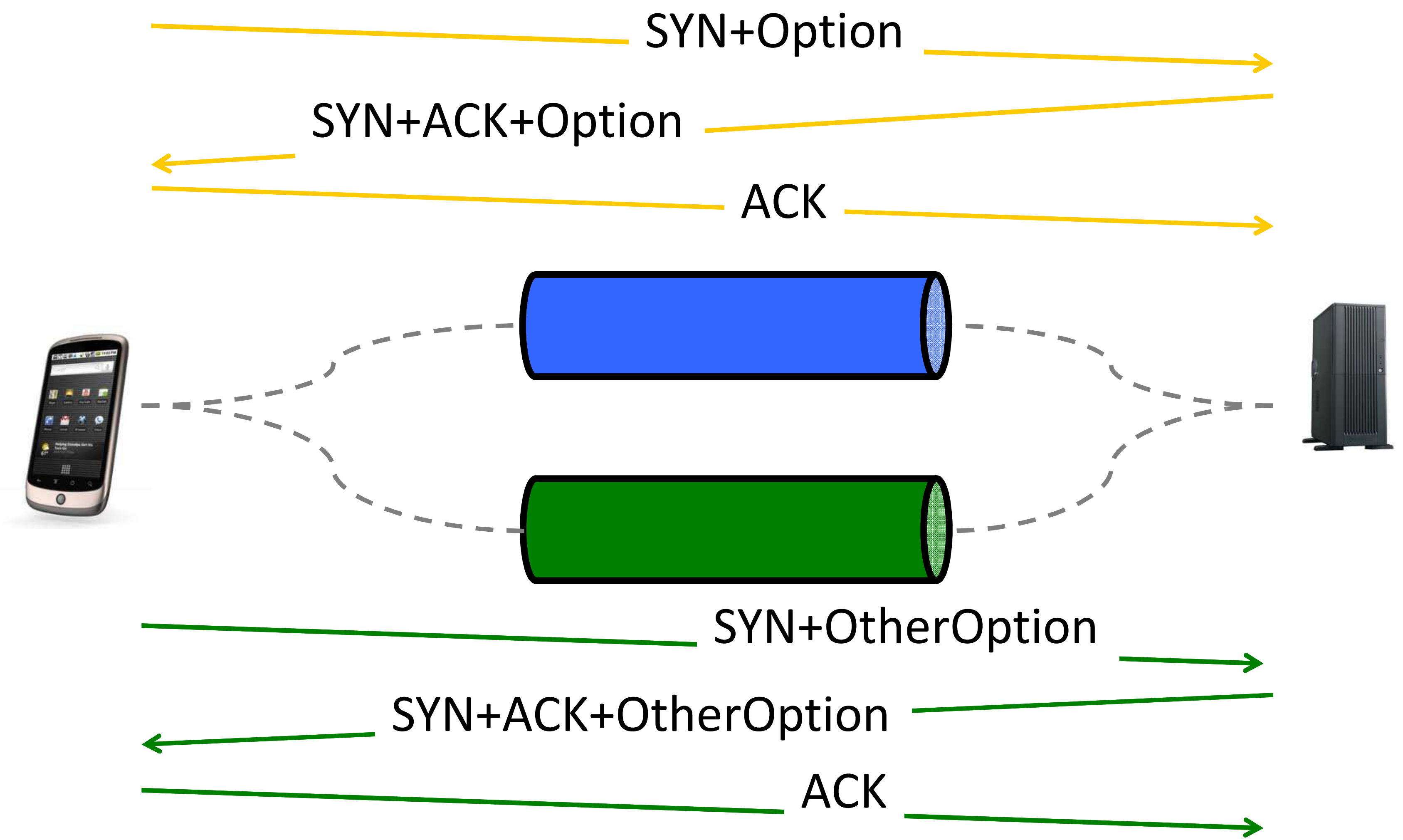
- It starts with a three-way handshake
  - SYN segments may contain special options

- All data segments are sent in sequence
  - There is no gap in the sequence numbers

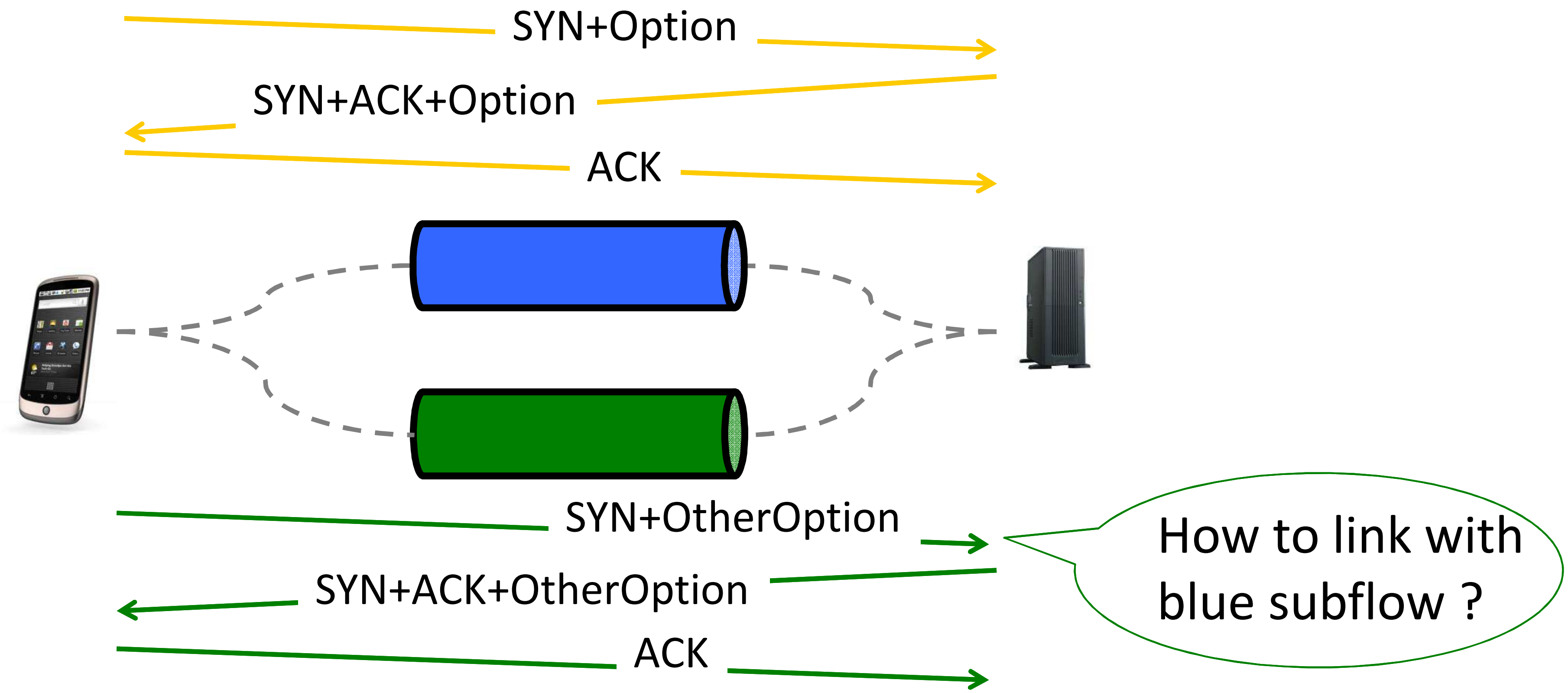
- It is terminated by using FIN or RST

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

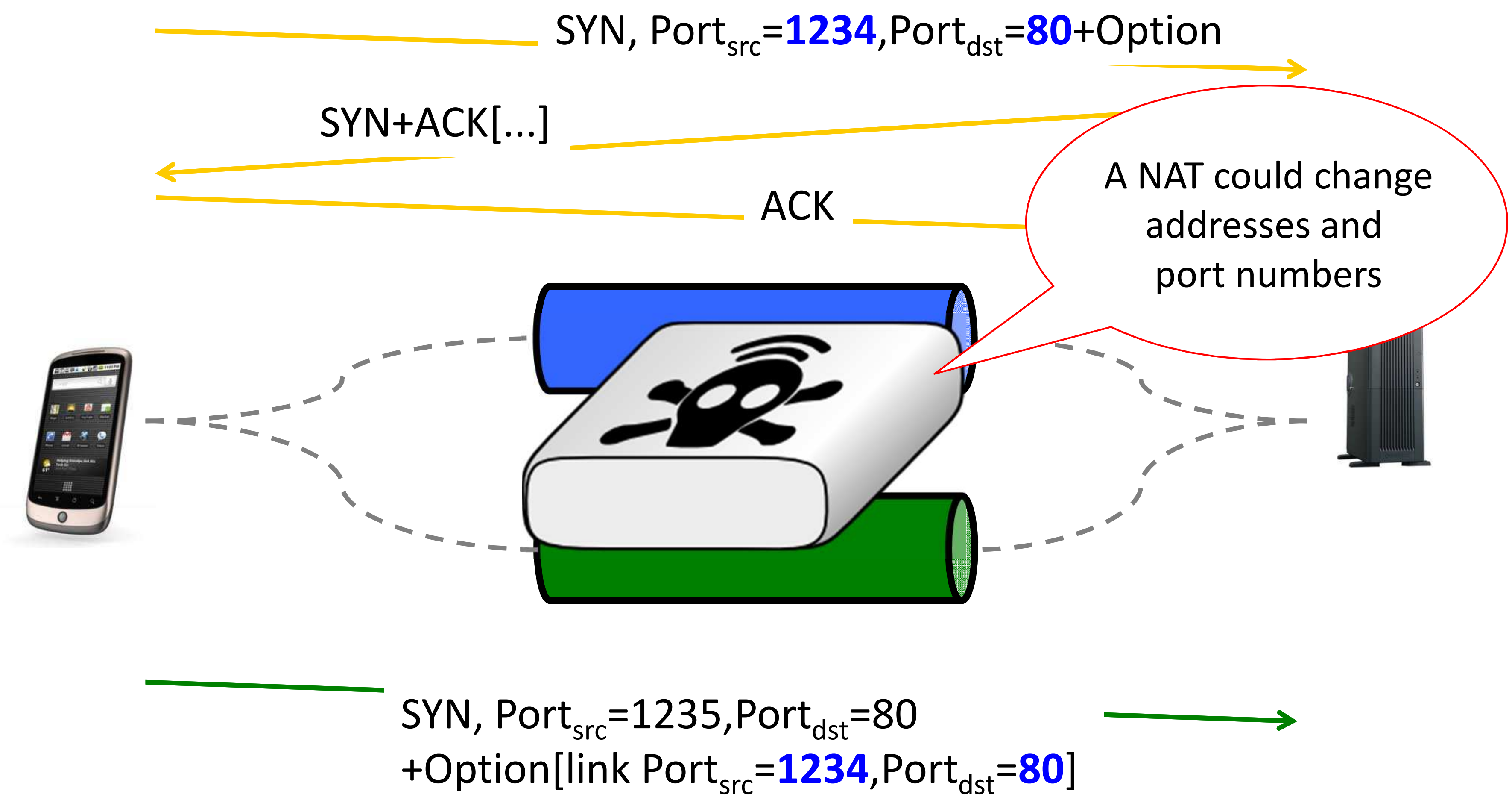
Multipath TCP



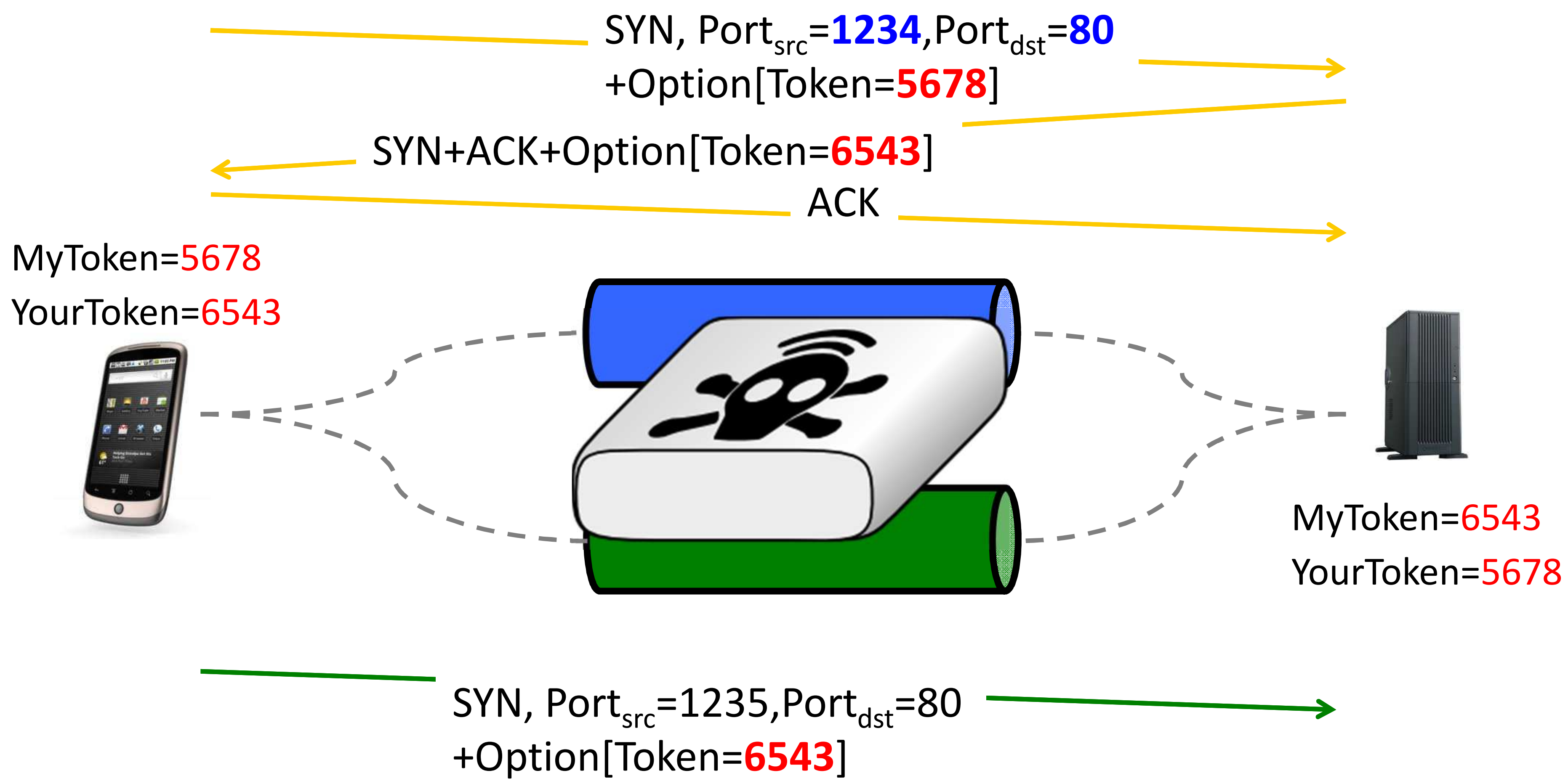
How to combine two TCP subflows ?



How to link TCP subflows ?



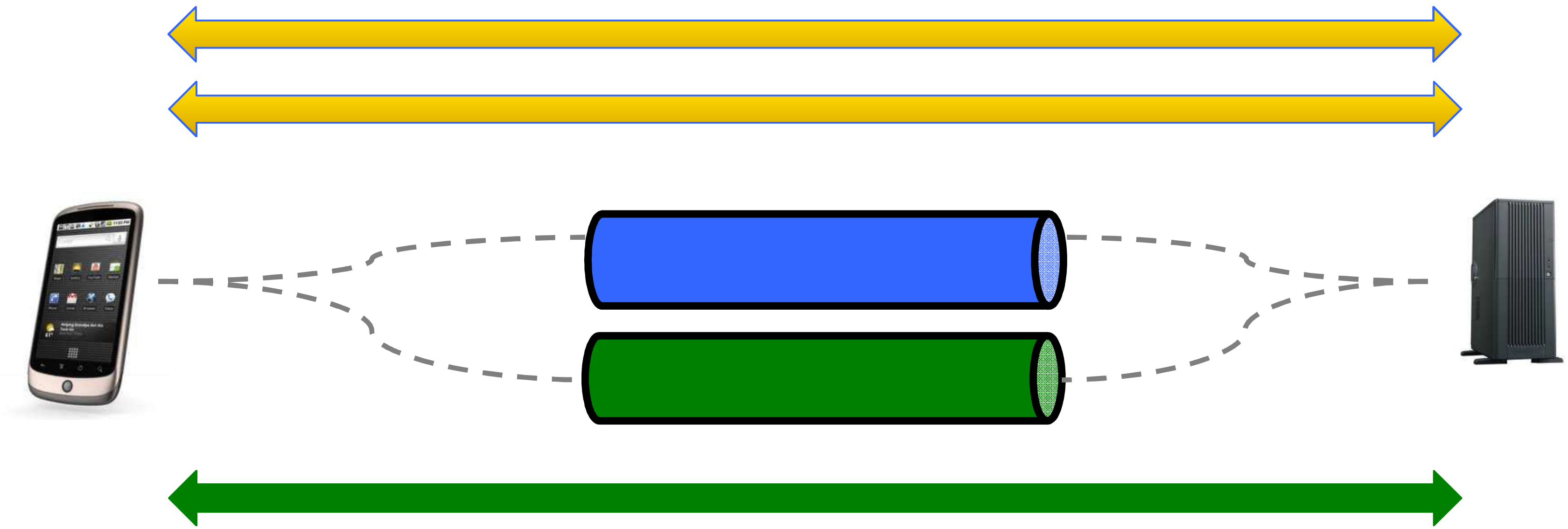
How to link TCP subflows ?



Subflow agility

Multipath TCP supports

- addition of subflows
- removal of subflows



Control plane

- How to manage a Multipath TCP connection that uses several paths ?

Data plane

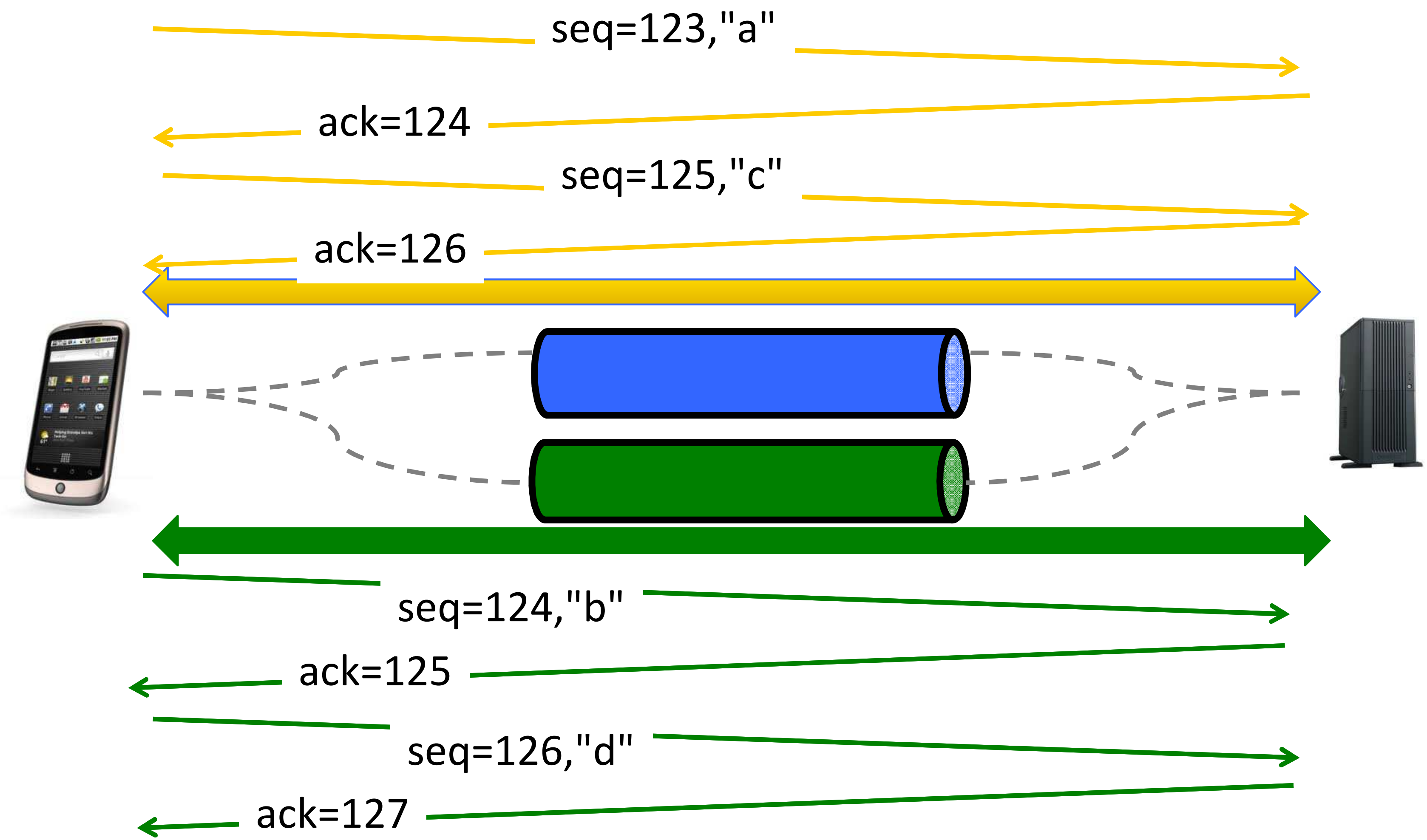
- How to transport data ?

Congestion control

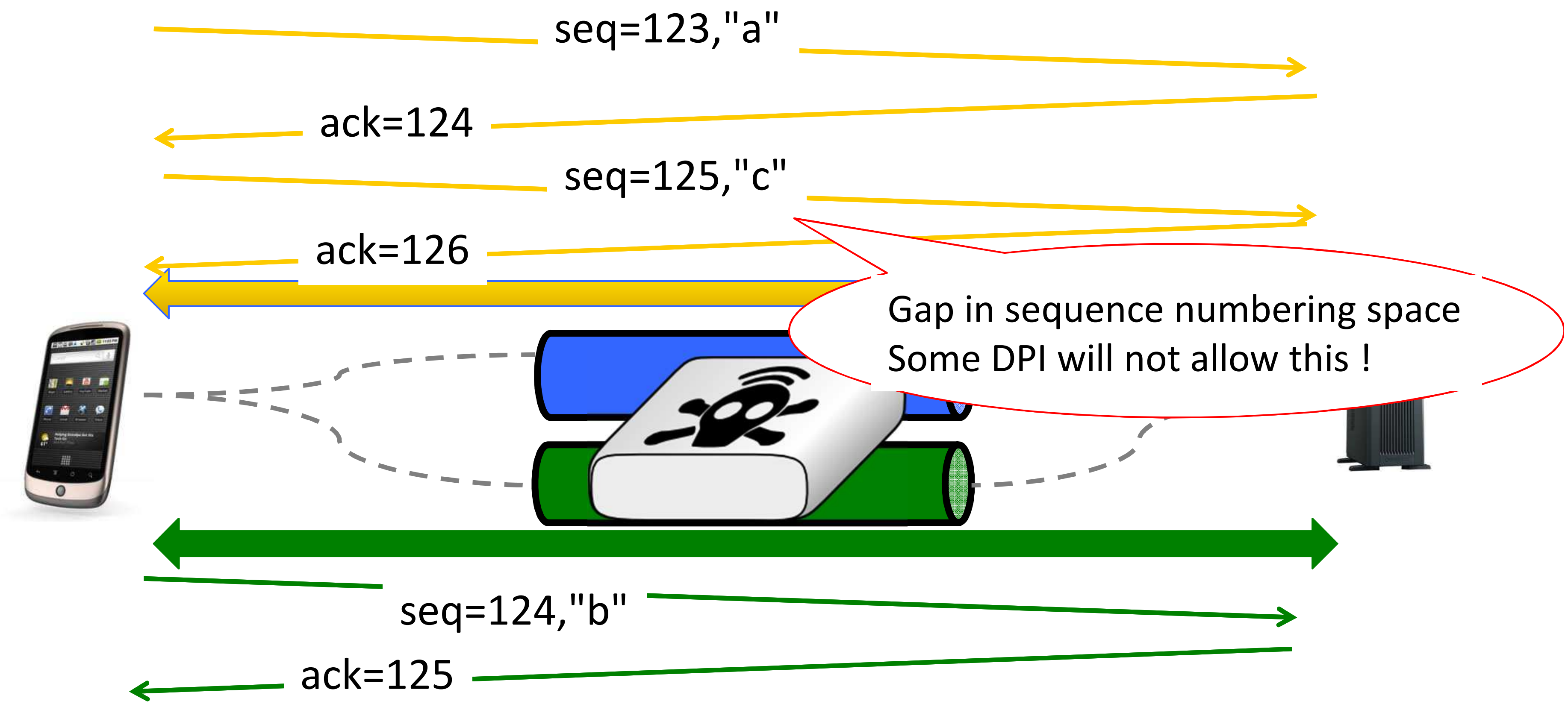
- How to control congestion over multiple paths ?

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

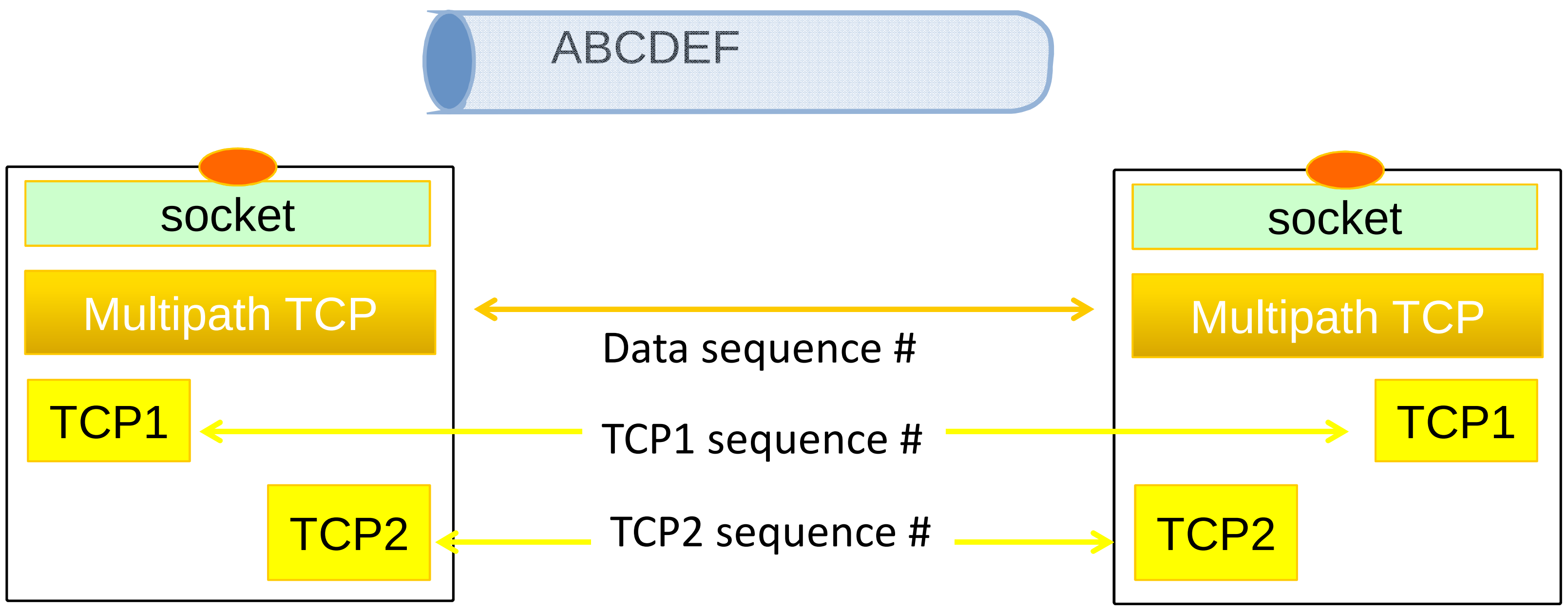
How to transfer data ?



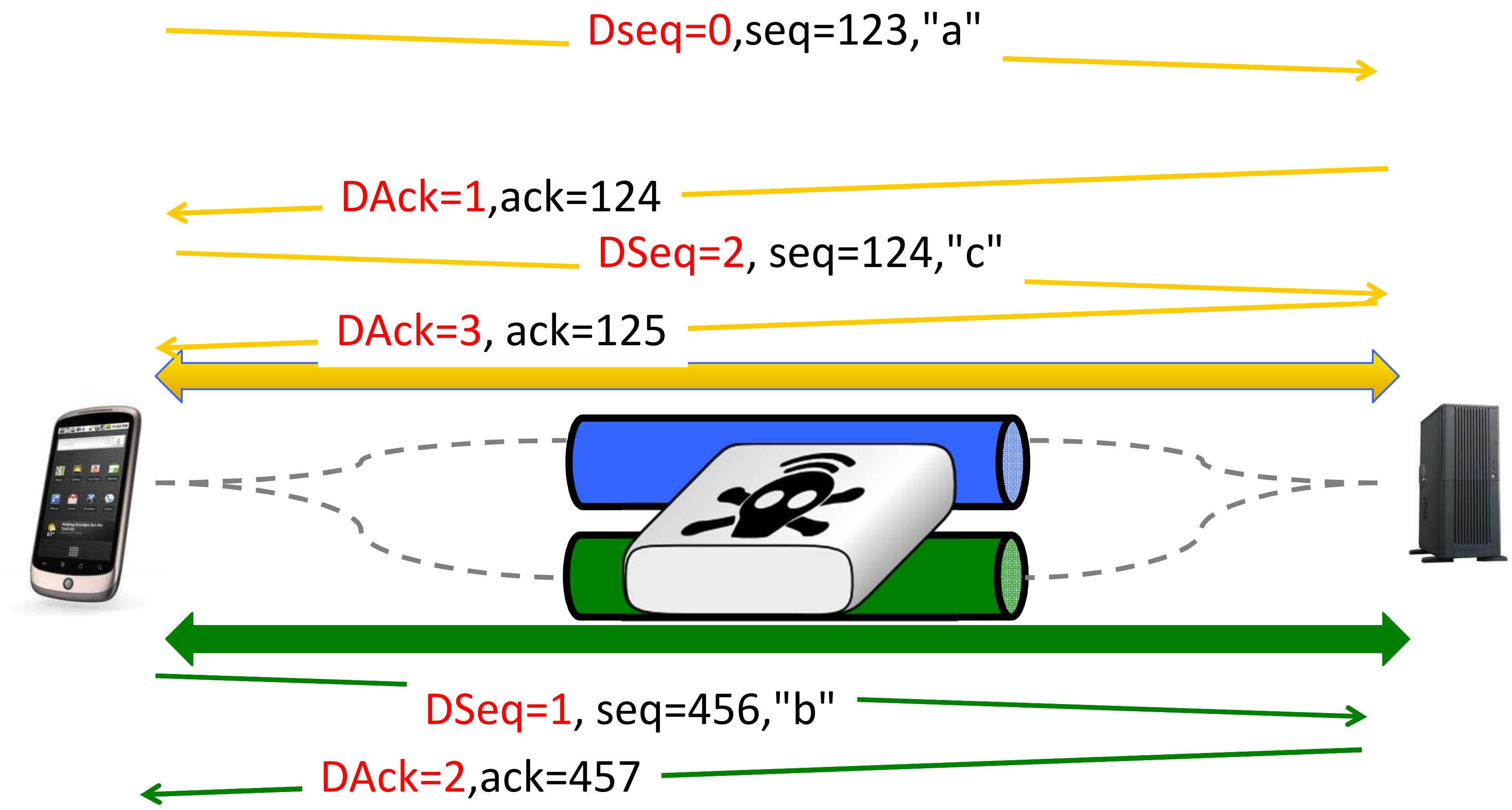
How to transfer data in today's Internet ?



Two levels of sequence numbers

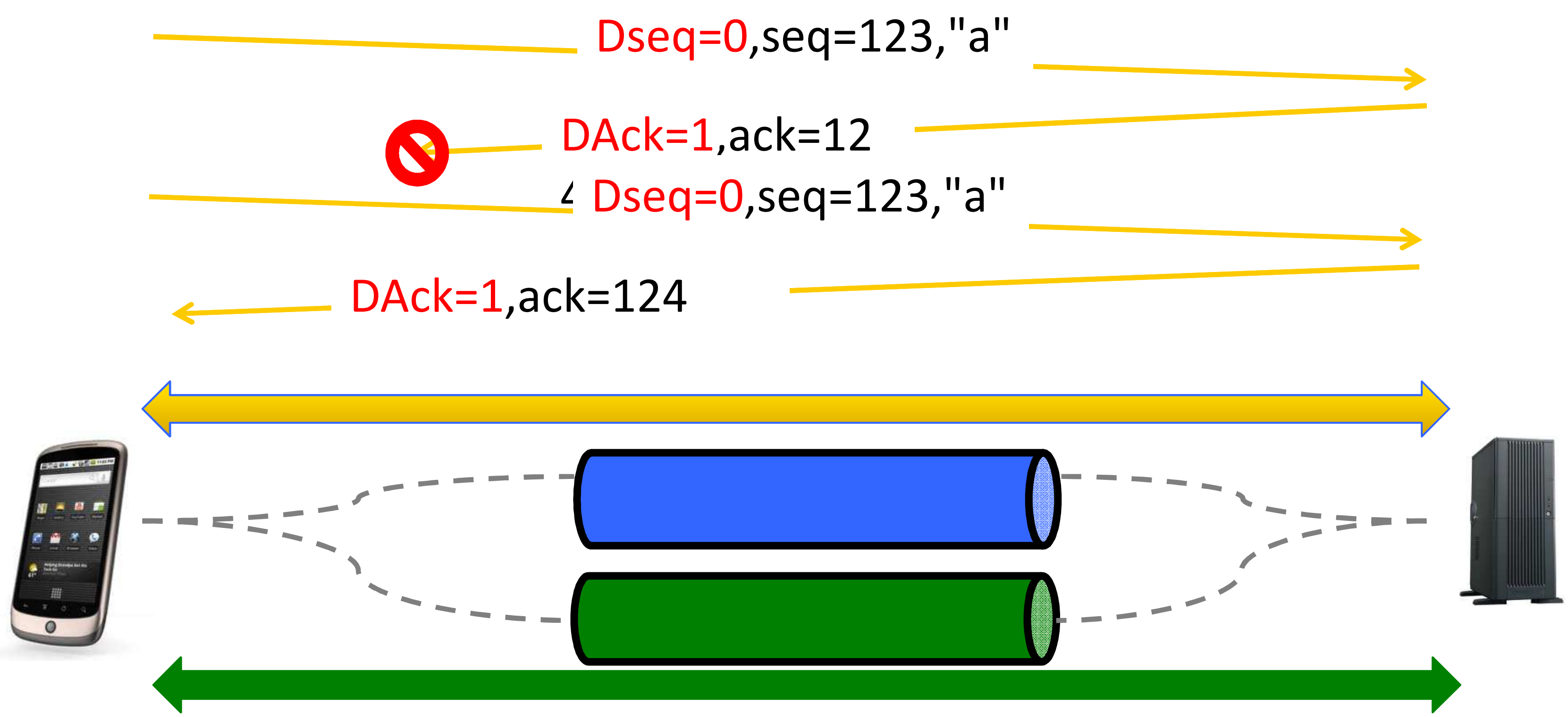


Multipath TCP - Data transfer

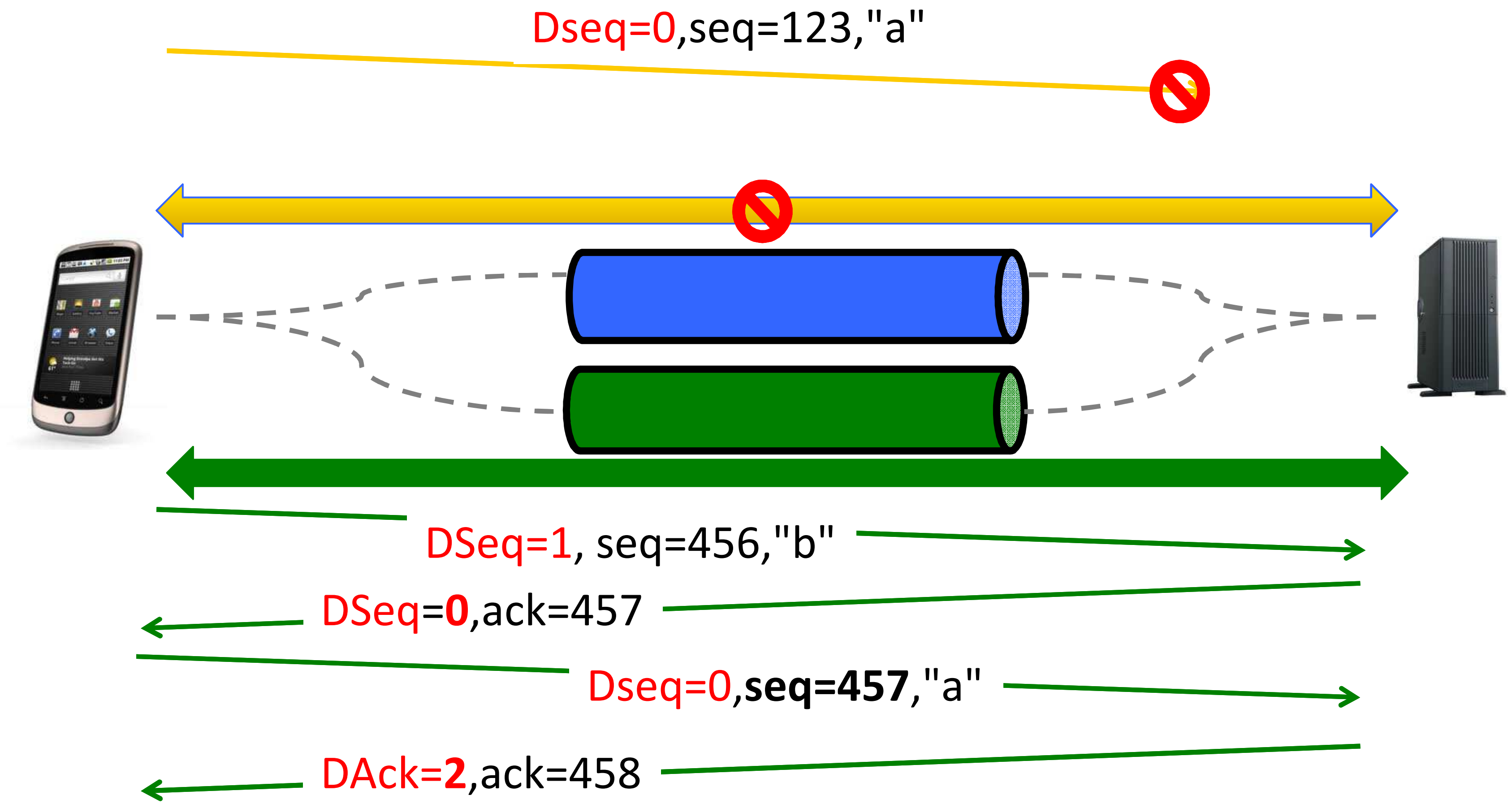


Data losses over one TCP subflow

- Fast retransmit and timeout as in regular TCP



What happens when a TCP subflow fails ?



Retransmission heuristics

Heuristics used by current Linux implementation

- Fast retransmit is performed on the same subflow as the original transmission
- Upon timeout expiration, reevaluate whether the segment could be retransmitted over another subflow

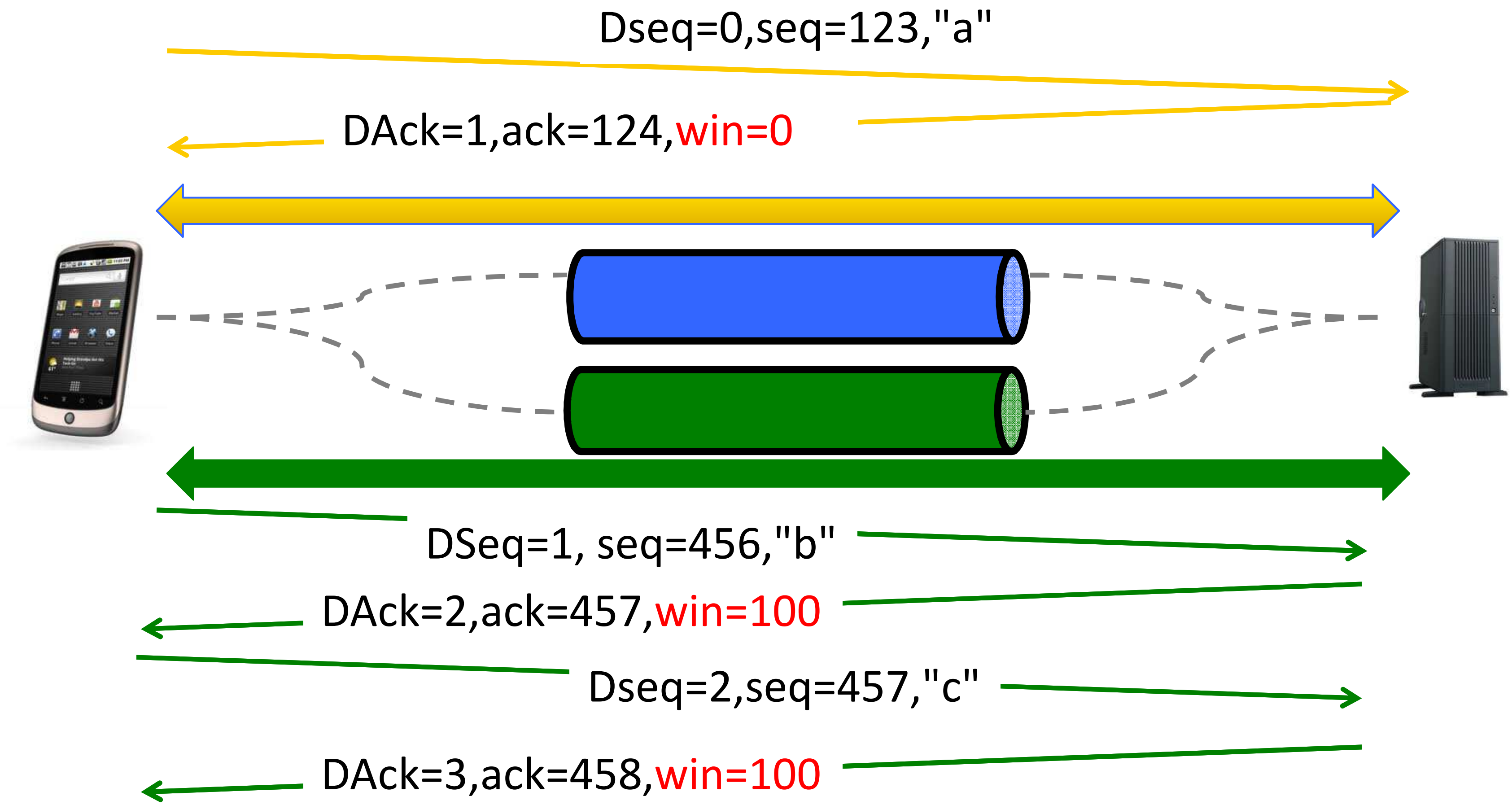
- Upon loss of a subflow, all the unacknowledged data are retransmitted on other subflows

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

How should the window-based flow control be performed ?

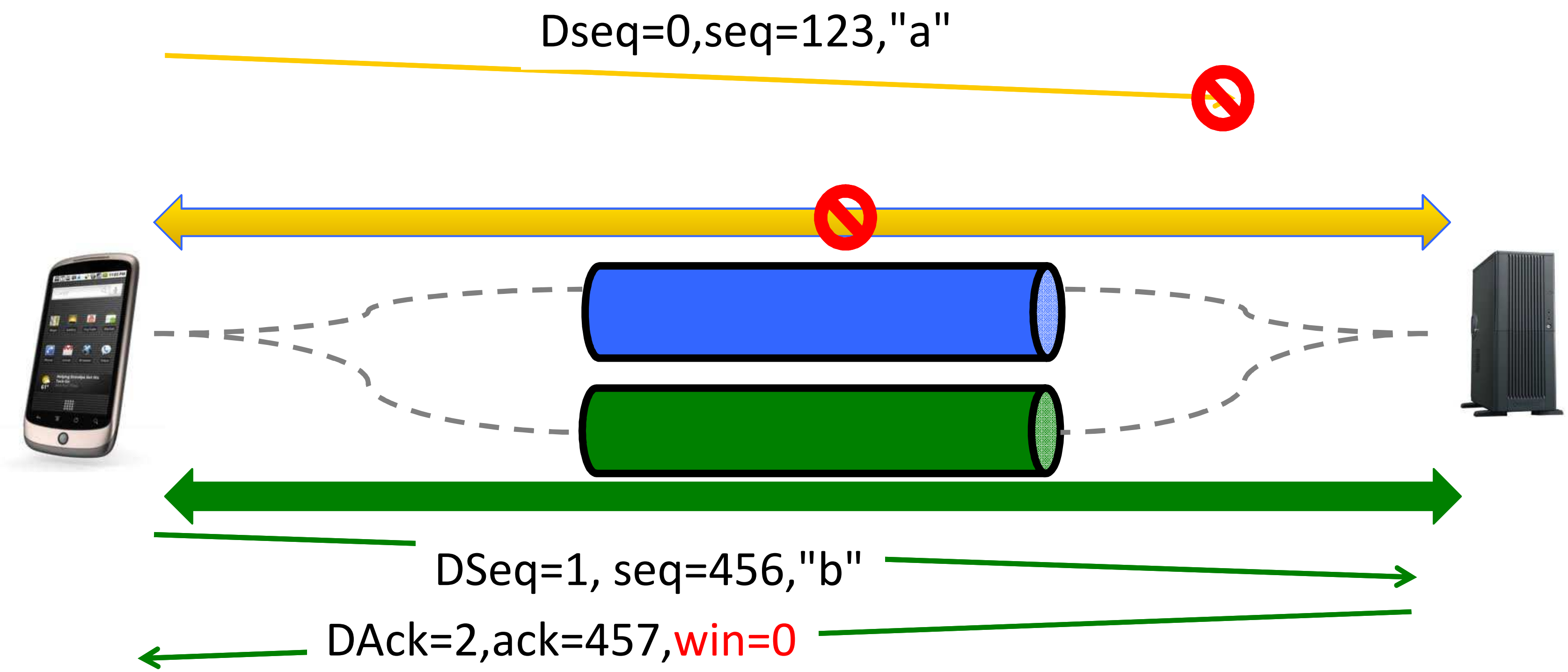
- Independant windows on each TCP subflow
- A single window that is shared among all TCP subflows

Independant windows



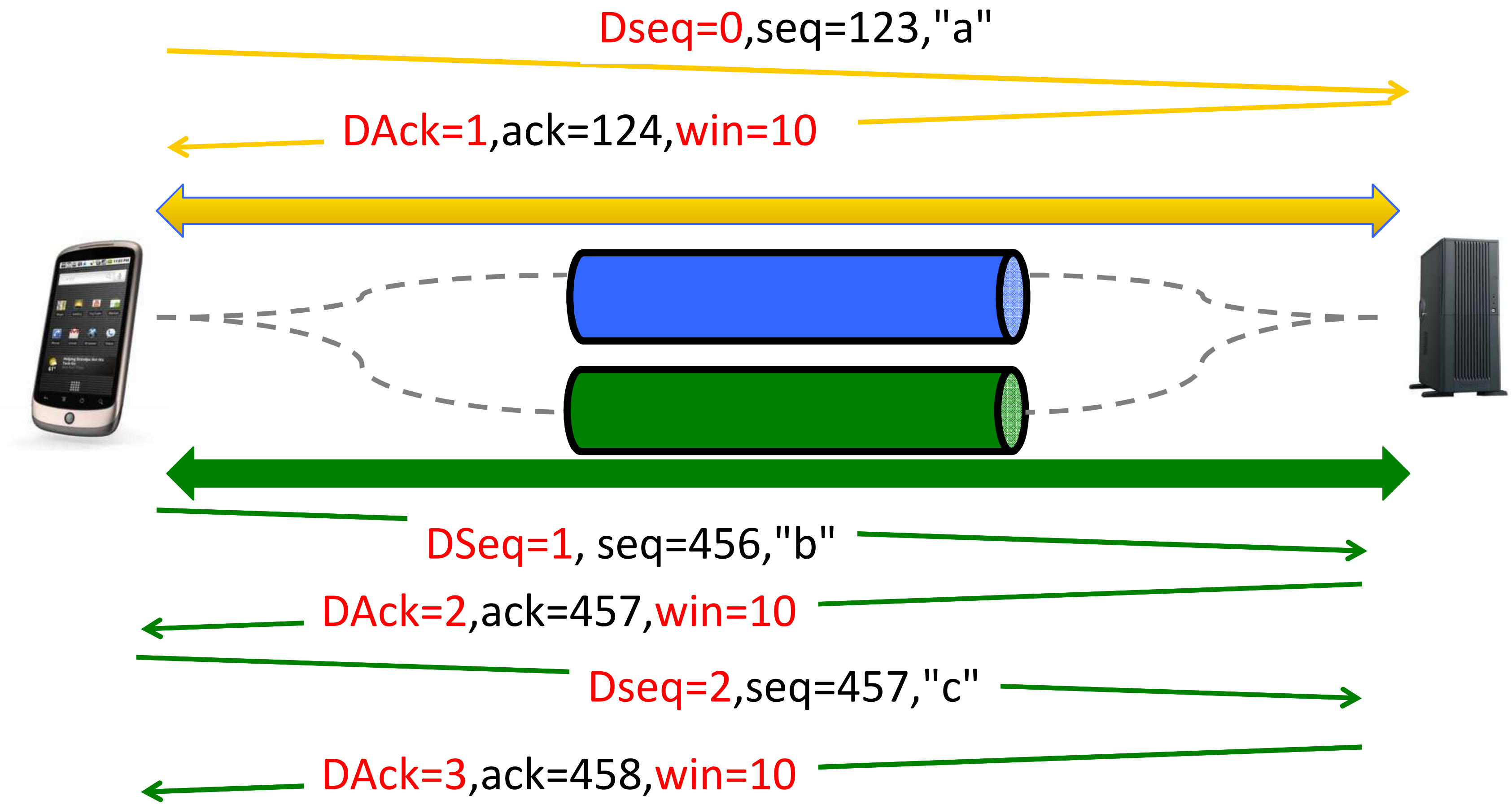
Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Independant windows - possible problem

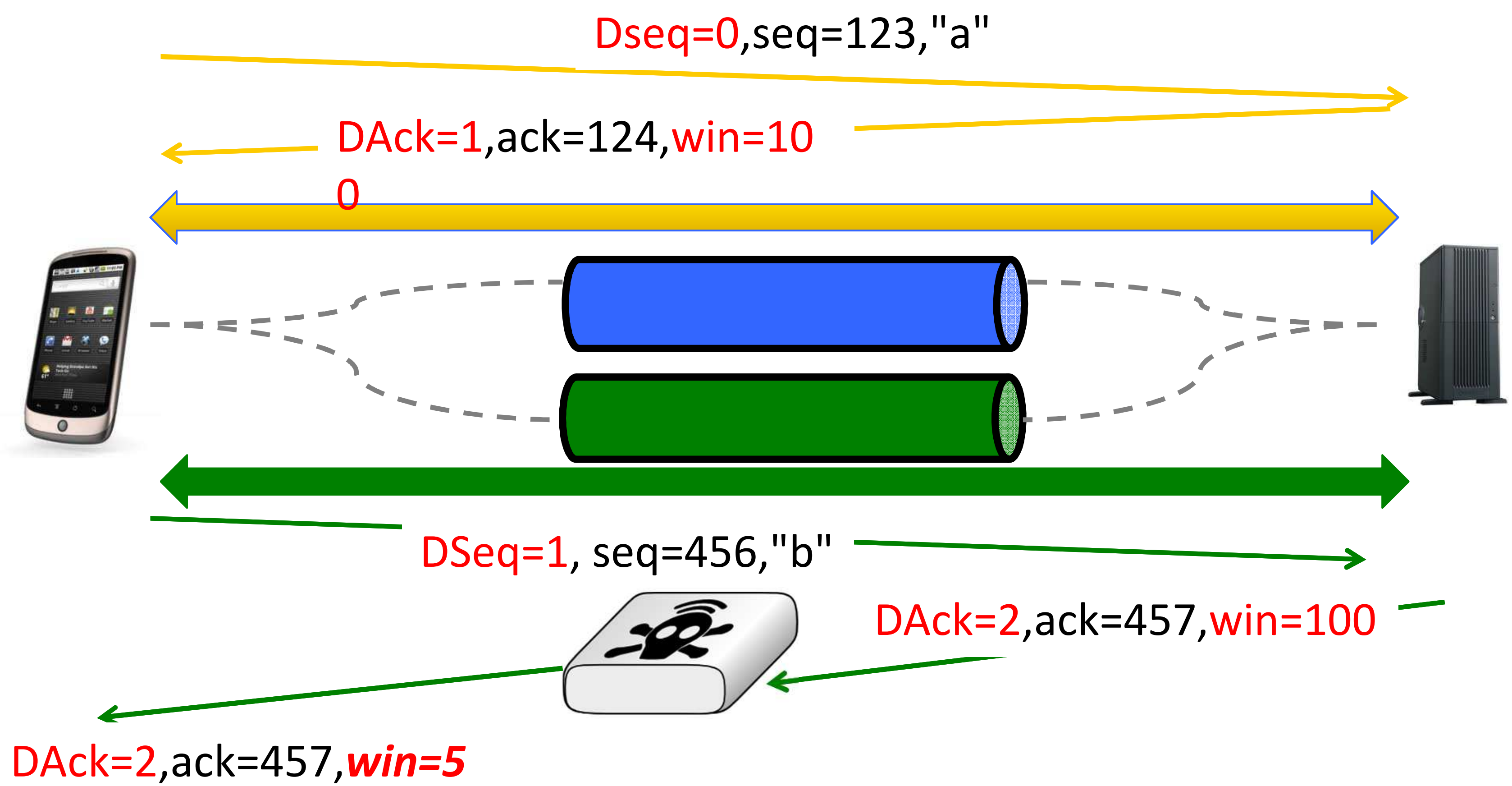


Impossible to retransmit, window is already full on green subflow

A single window shared by all subflows



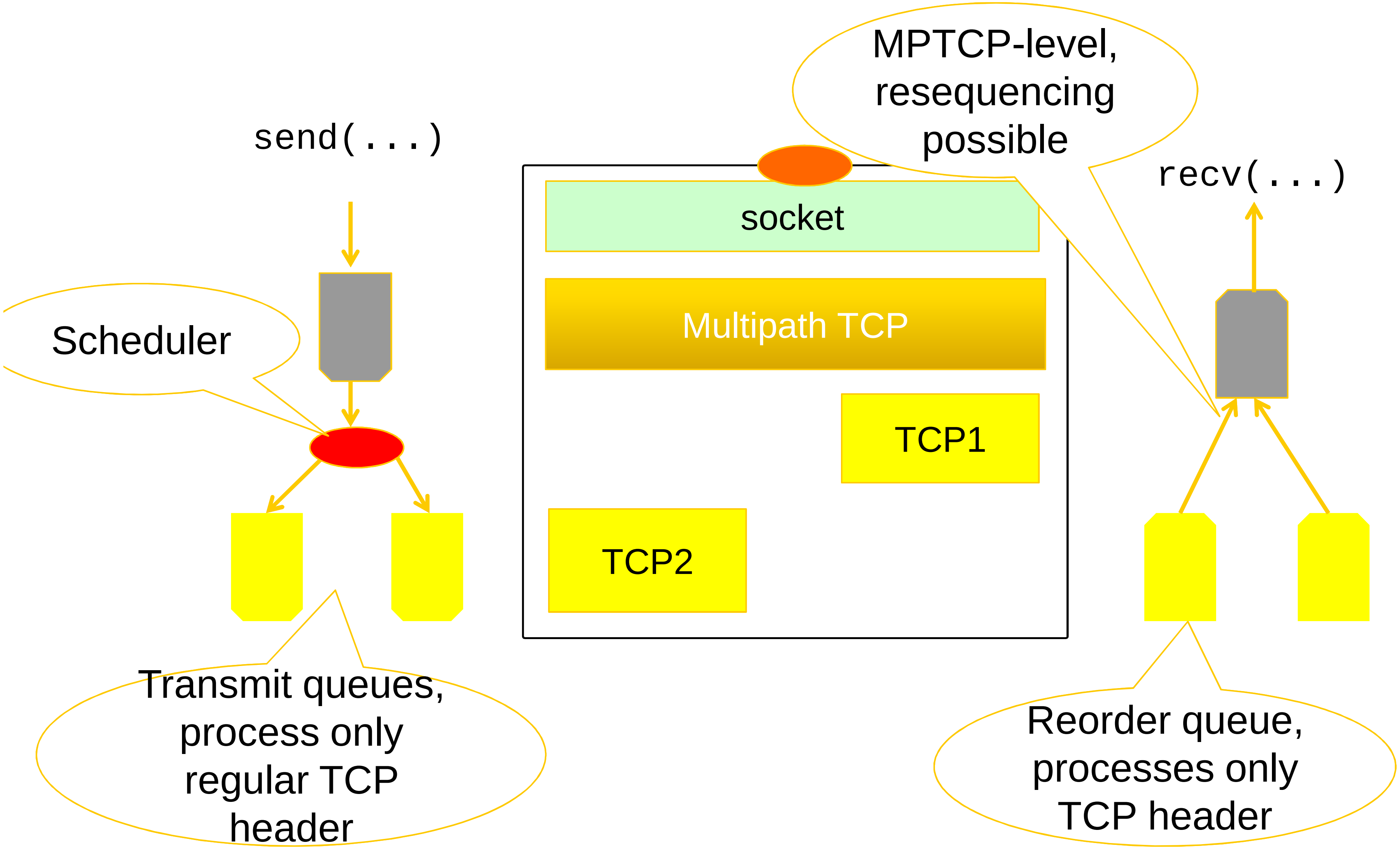
A single window shared by all subflows
Impact of middleboxes



Multipath TCP maintains one window per Multipath TCP connection

- Window is relative to the last acked data (Data Ack)
- Window is shared among all subflows
  - It's up to the implementation to decide how the window is shared
- Window is transmitted inside the `window` field of the regular TCP header
- If middleboxes change `window` field,
  - use largest `window` received at MPTCP-level
  - use received `window` over each subflow to cope with the flow control imposed by the middlebox

Multipath TCP buffers



Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Sending Multipath TCP information

How to exchange the Multipath TCP specific information between two hosts ?

Option 1

- Use TLVs to encode data and control information inside payload of subflows

Option 2

- Use TCP options to encode all Multipath TCP information

Option 1 : Michael Scharf, Thomas-Rolf Banniza , *MCTCP: A Multipath Transport Shim Layer*, GLOBECOM 2011

Advantages

- Normal way of extending TCP
- Should be able to go through middleboxes or fallback

Drawbacks

- limited size of the TCP options, notably inside SYN
- What happens when middleboxes drop TCP options in data segments

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Advantages

- Multipath TCP could start as regular TCP and move to Multipath only when needed
- Could be implemented as a library in userspace
- TLVs can be easily extended

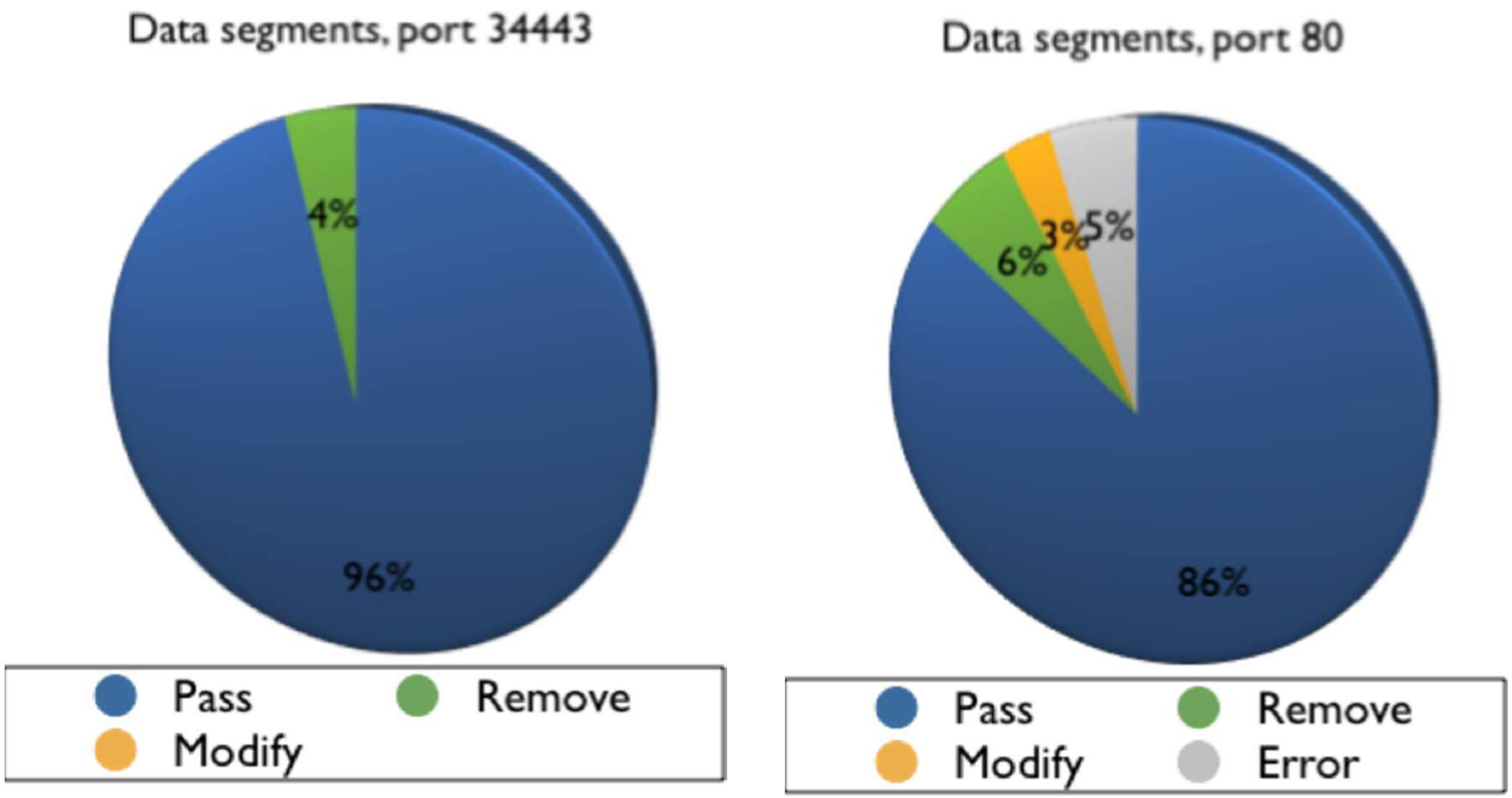
Drawbacks

- TCP segments contain TLVs including the data and not only the data
  - problem for middleboxes, DPI, ..
- Middleboxes become more difficult

Michael Scharf, Thomas-Rolf Banniza , *MCTCP: A Multipath Transport Shim Layer*, GLOBECOM 2011

TECHNISCHE UNIVERSITÄT DARMSTADT

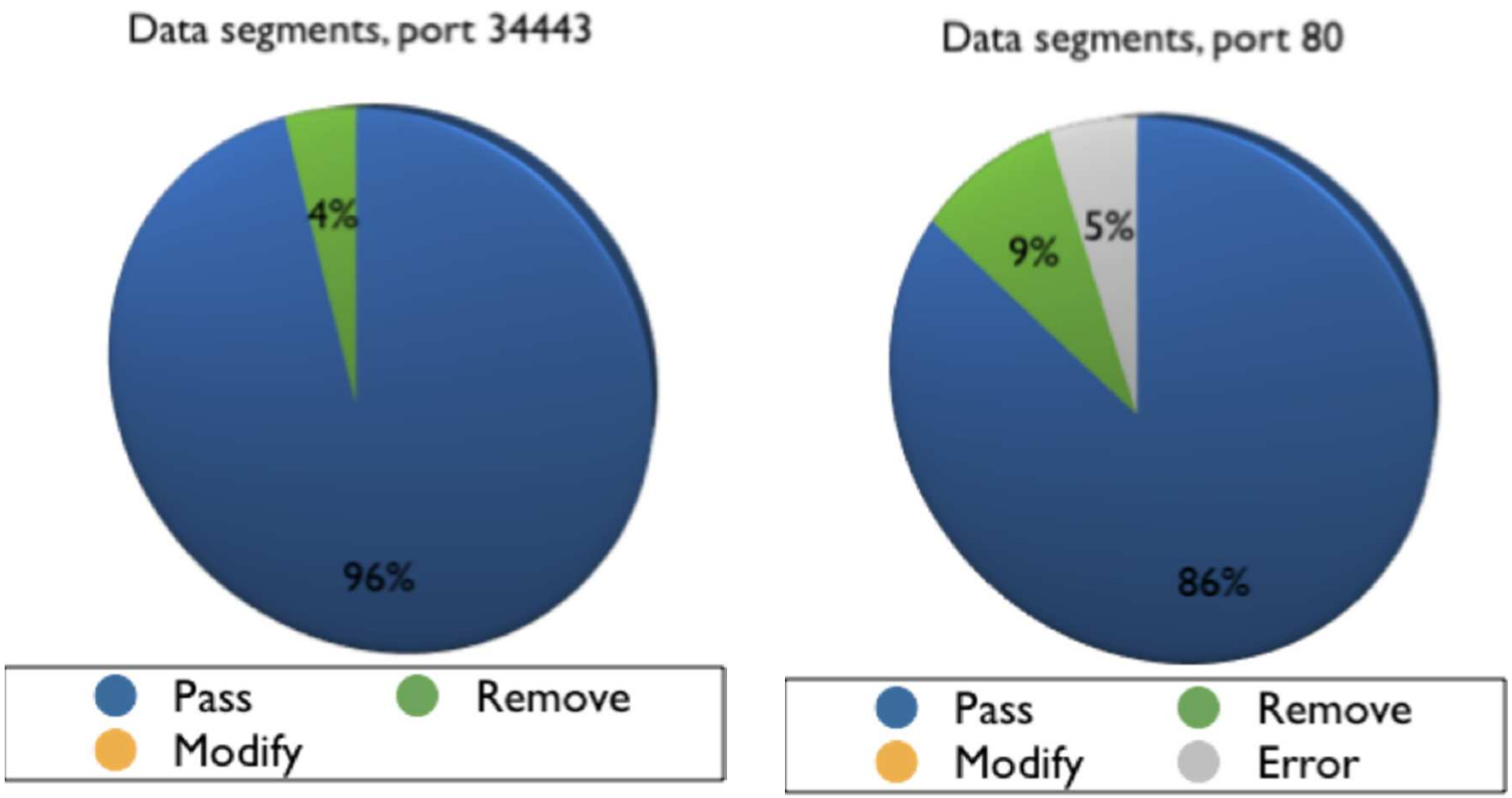
Known option (TS) in Data segments



Honda, Michio, et al. "Is it still possible to extend TCP?." Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. ACM, 2011.

Is it safe to use TCP options ?

Unknown option in Data segments



Honda, Michio, et al. "Is it still possible to extend TCP?." Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. ACM, 2011.

Multipath TCP options

TCP option format

Kind	Length	Option-specific data

Initial design

- One option kind for each purpose
  (e.g. Data Sequence number)

Final design

- A single variable-length Multipath TCP option

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

TECHNISCHE UNIVERSITÄT DARMSTADT

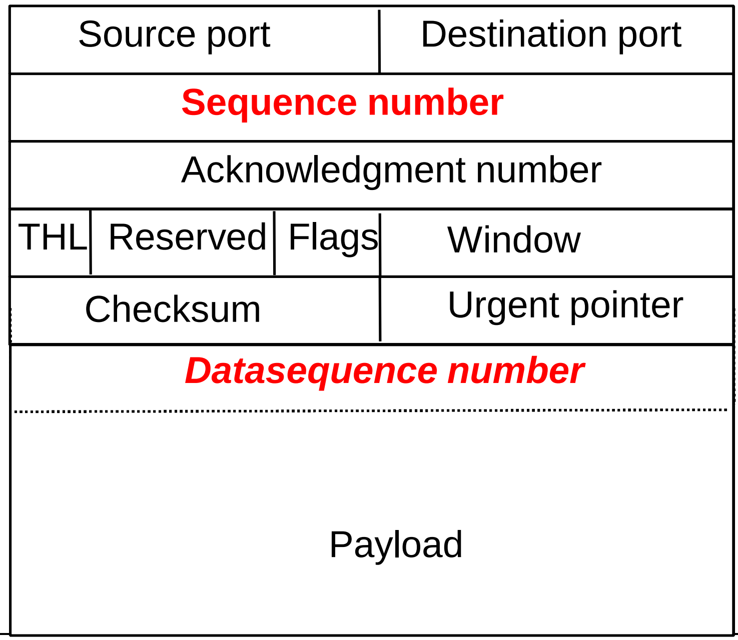
A single option type

- to minimise the risk of having one option accepted by middleboxes in SYN segments and rejected in segments carrying data

Kind	Length	Subtype	
Subtype specific data (variable length)			

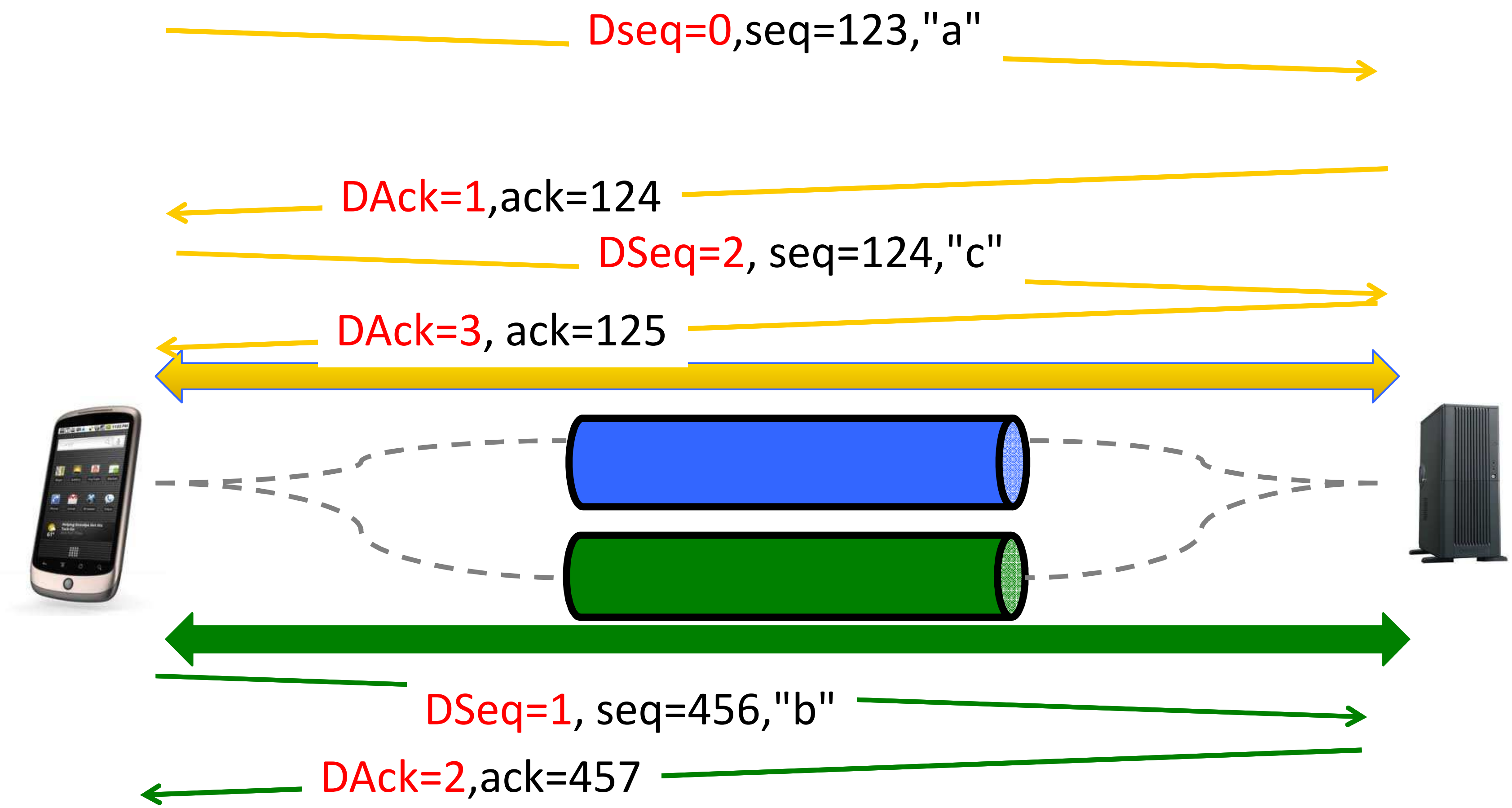
How to transport Data sequence numbers ?

- Same solution as for TCP
  - Data sequence number in TCP option is the Data sequence number of the first byte of the segment

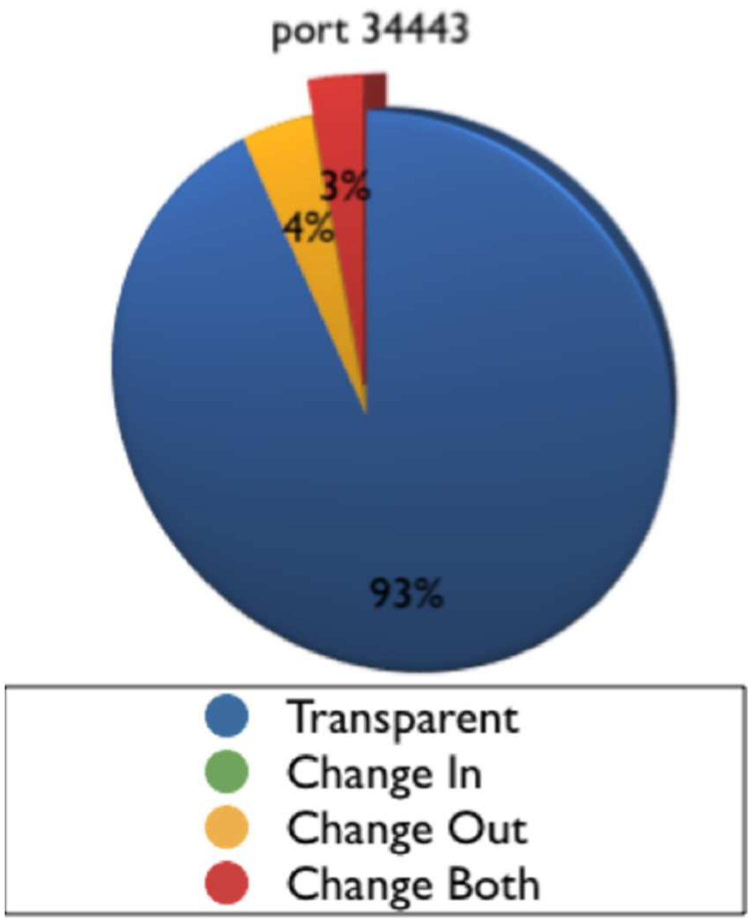


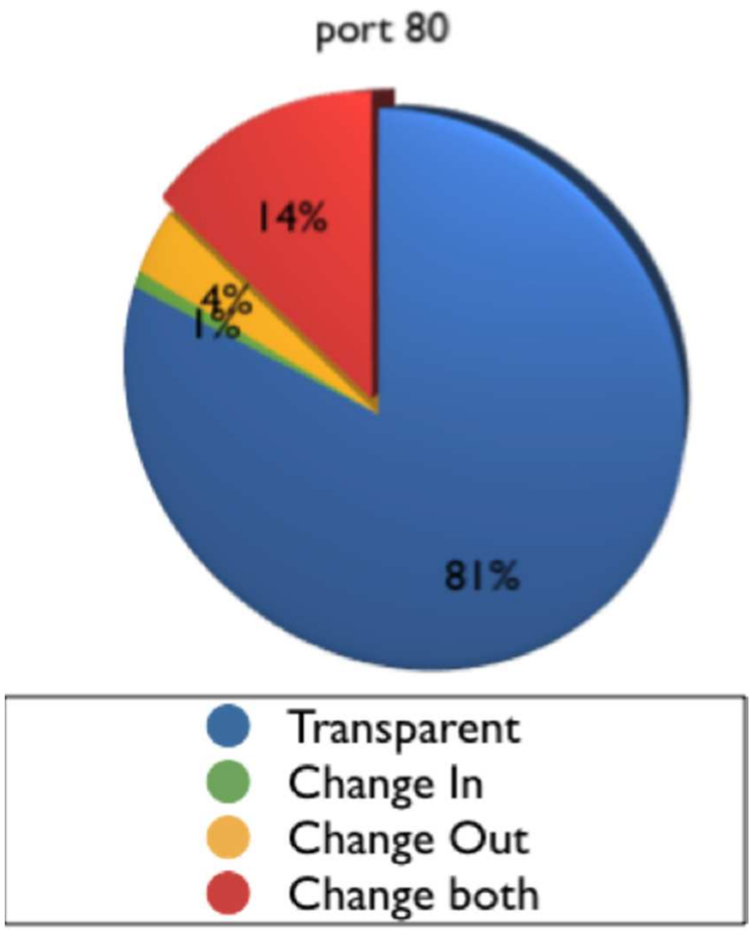
Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Multipath TCP - Data transfer



TCP sequence number and middleboxes





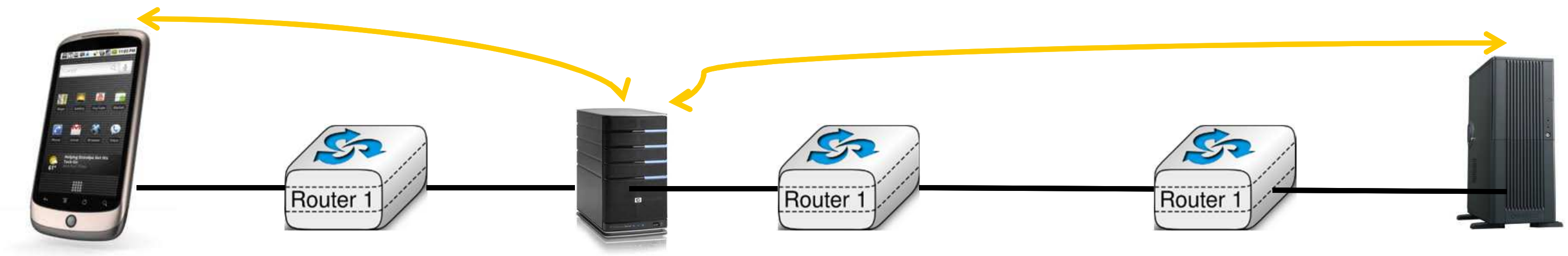
Honda, Michio, et al. "Is it still possible to extend TCP?." Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. ACM, 2011.

Which middleboxes change TCP sequence numbers ?

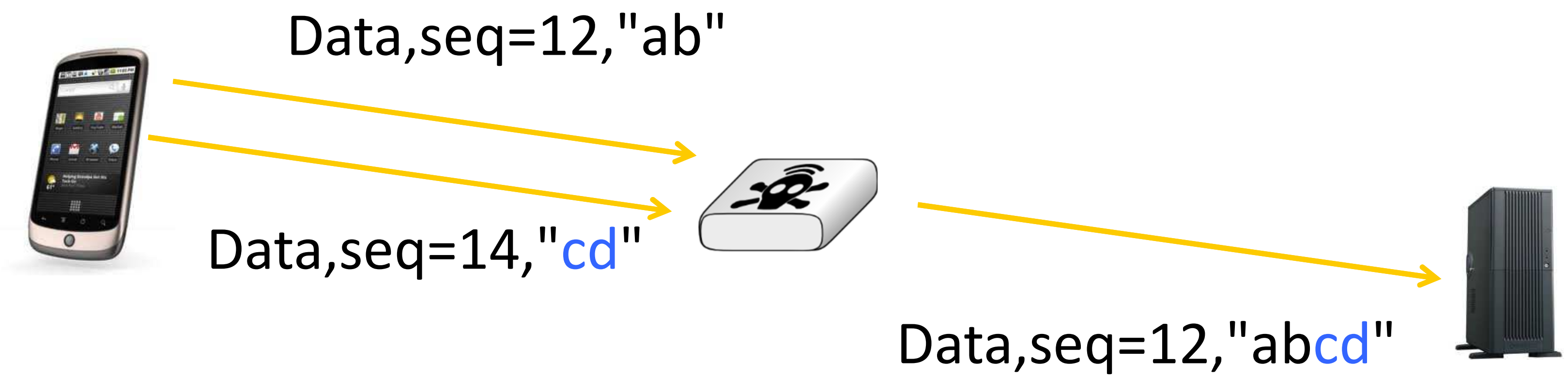
Some firewalls change TCP sequence numbers in SYN segments to ensure randomness

- fix for old windows95 bug

Transparent proxies terminate TCP connections

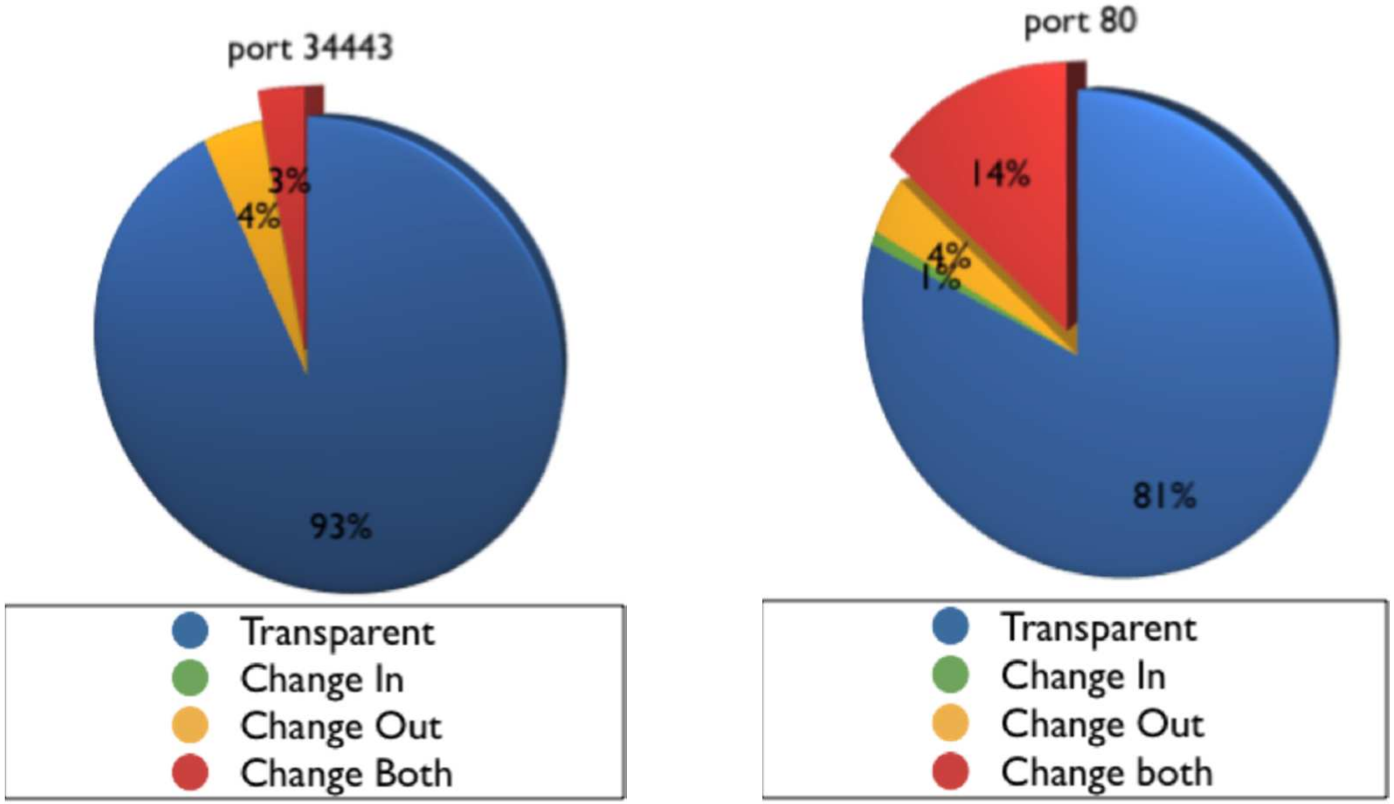


Data segments



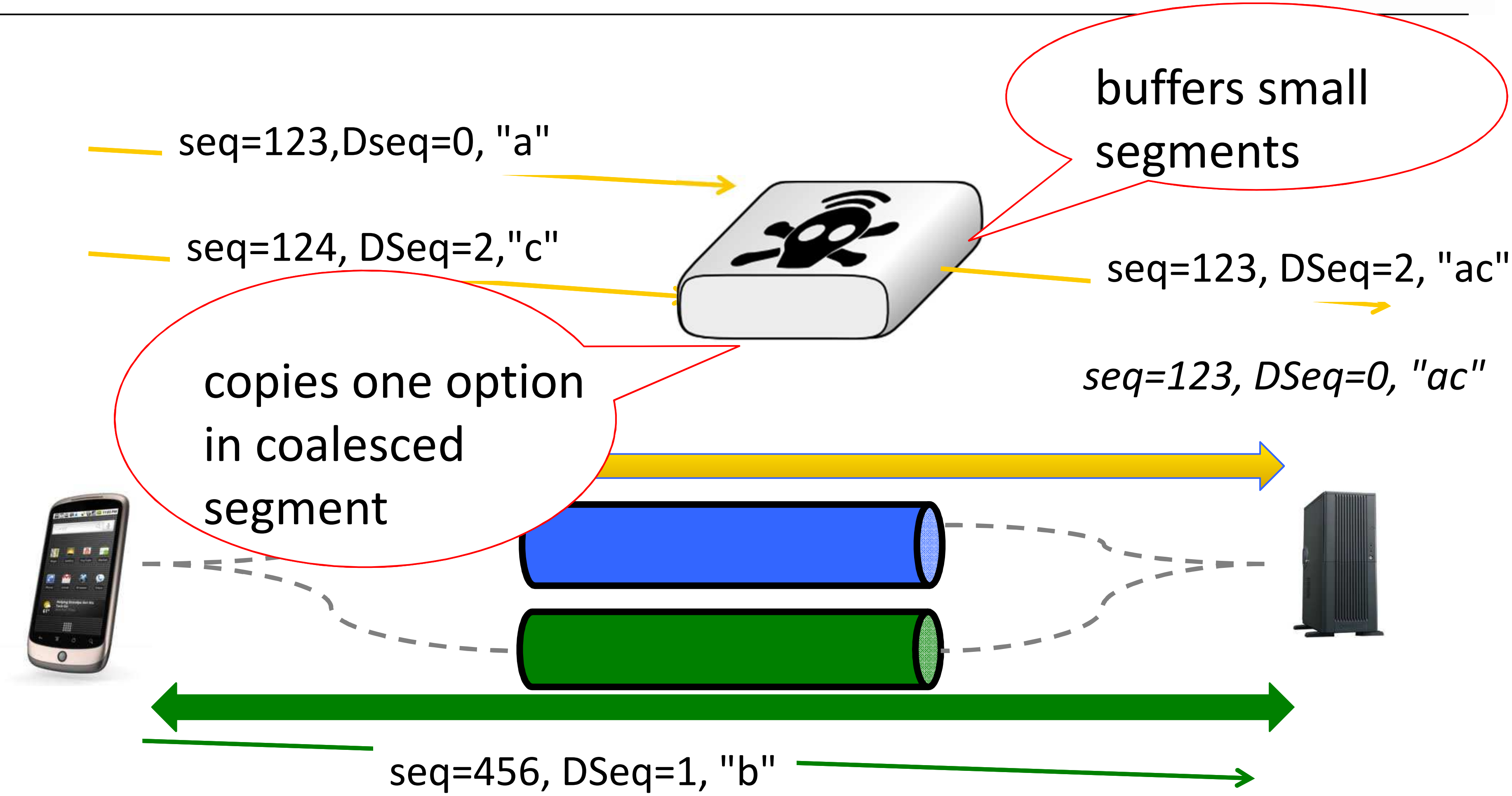
Such a middlebox could also be the network adapter of the server that uses LRO to improve performance.

Segment coalescing

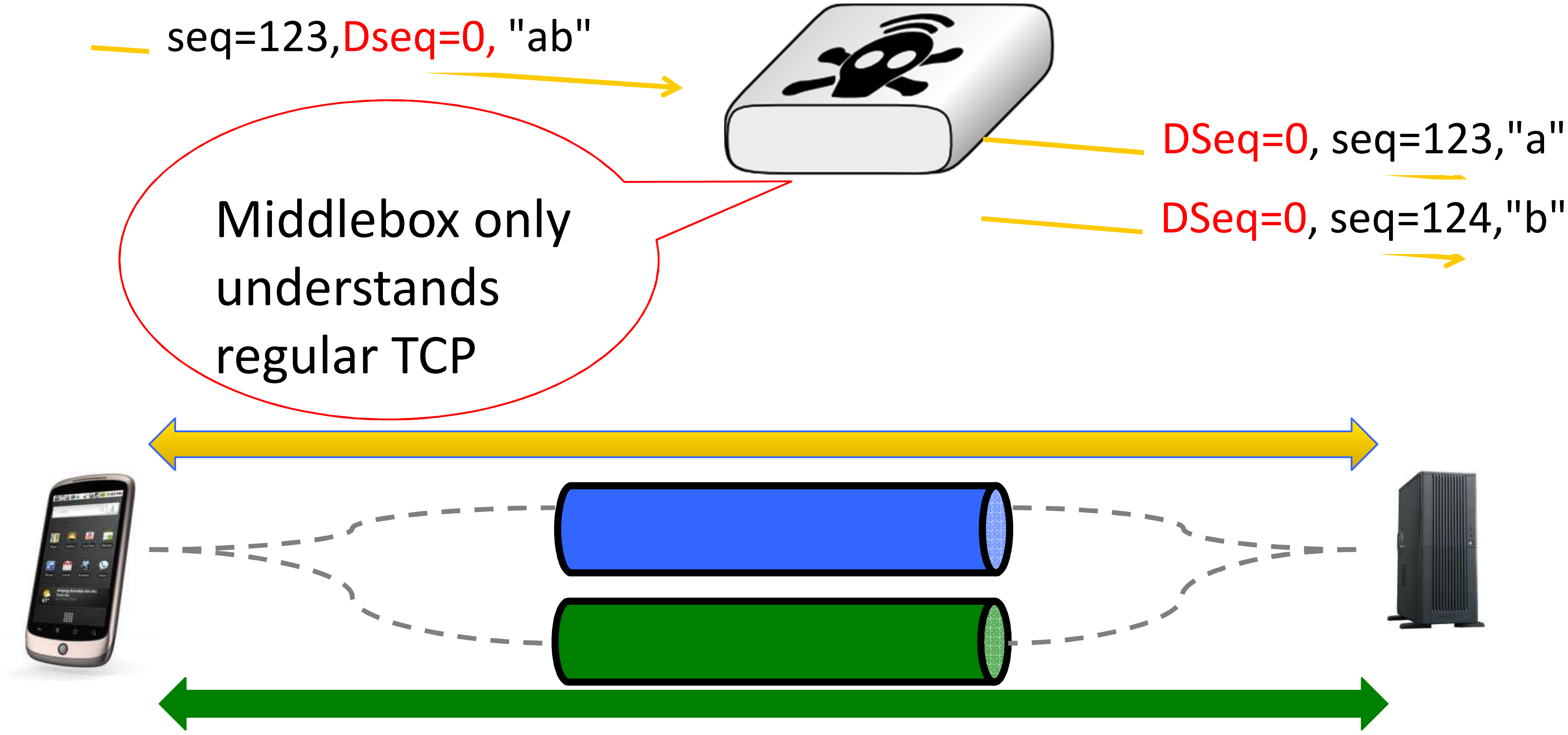


Honda, Michio, et al. "Is it still possible to extend TCP?." Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. ACM, 2011.

Data sequence numbers and middleboxes



Data sequence numbers and middleboxes



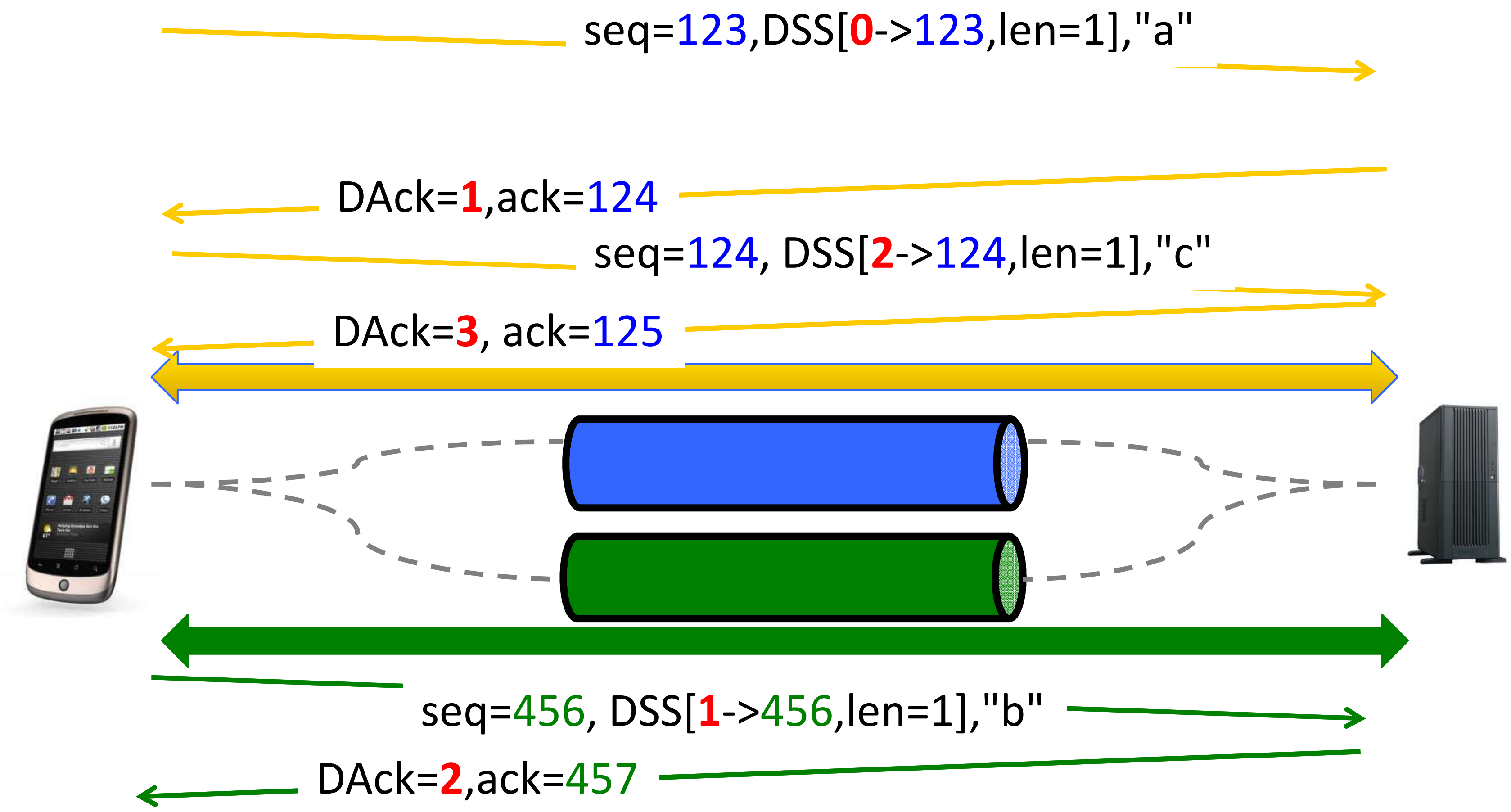
A "middlebox" that both splits and coalesces TCP segments

How to avoid desynchronisation between the bytestream and data sequence numbers ?

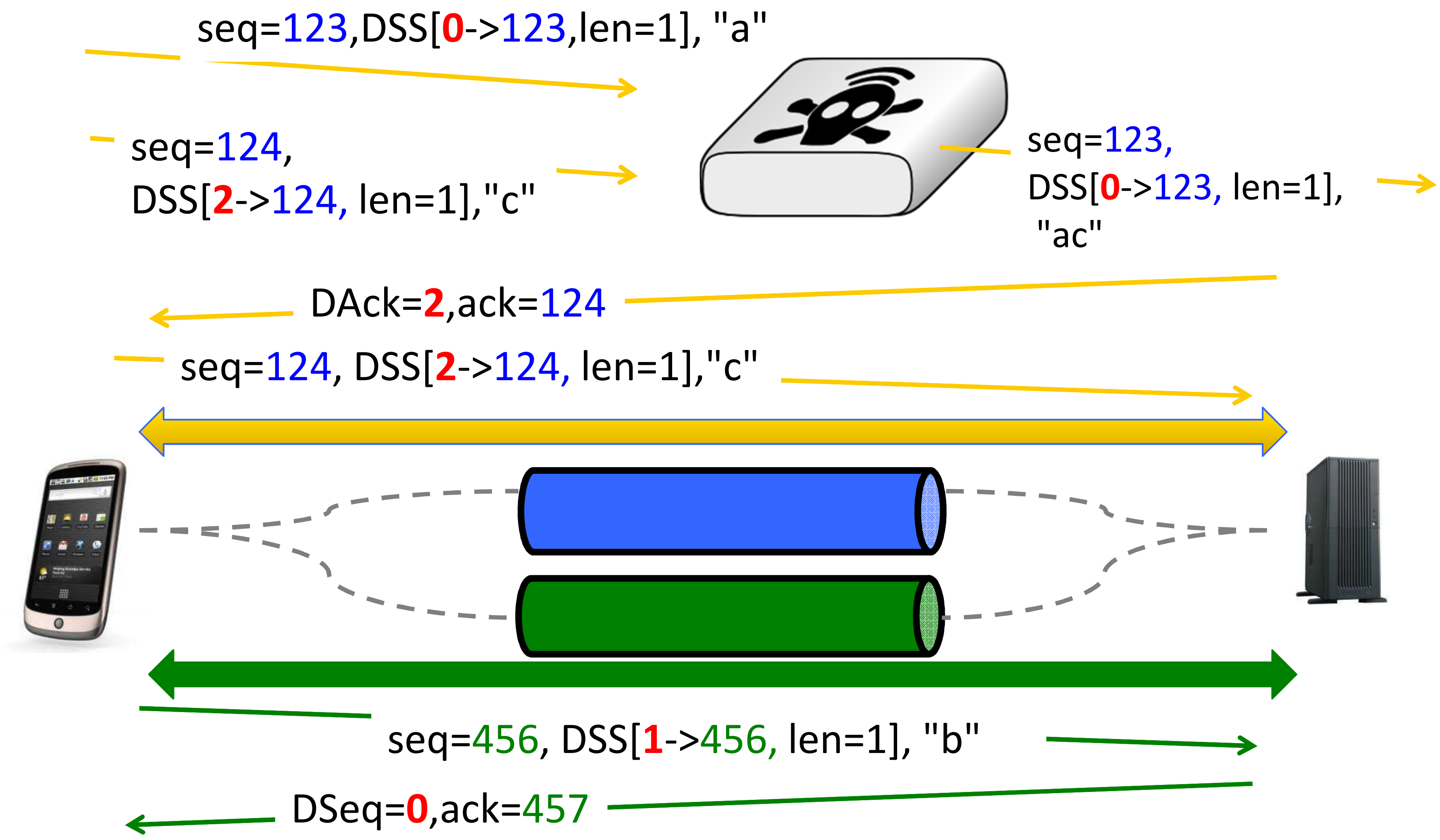
Solution

- Multipath TCP option carries mapping between Data sequence numbers and (difference between initial and current) subflow sequence numbers
  - mapping covers a part of the bytestream (length)

Multipath TCP - Data transfer



Data sequence numbers and middleboxes



Multipath TCP and middleboxes

With the DSS mapping, Multipath TCP can cope with middleboxes that

- combine segments
- split segments

Are they the most annoying middleboxes for Multipath TCP ?

– Unfortunately not

The worst middlebox

seq=123,
DSS[1->123,len=2],
"ab"

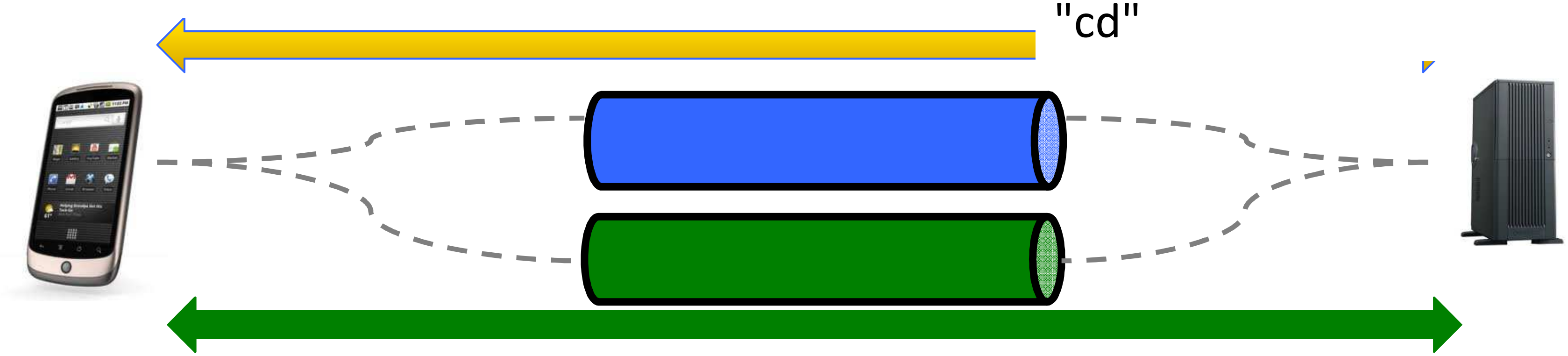
seq=123,
DSS[1->123, len=2],
"a***XXX***b"

DAck=3,ack=***125***

DAck=3,ack=128

seq=125,
DSS[3->125, len=2],
"cd"

seq=***128***,
DSS[3->125, len=2],
"cd"



Is this an academic exercise or reality ?

The worst middlebox

Is unfortunately very old...

- Any ALG for a NAT

```
220 ProFTPD 1.3.3d Server (BELNET FTPD Server) [193.190.67.15]
ftp_login: user `<null>' pass `<null>' host `ftp.belnet.be'
Name (ftp.belnet.be:obo): anonymous
---> USER anonymous
331 Anonymous login ok, send your complete email address as your password
Password:
---> PASS XXXX
---> PORT 192,168,0,7,195,120
200 PORT command successful
---> LIST
150 Opening ASCII mode data connection for file list
lrw-r--r--   1 ftp      ftp            6 Jun  1  2011 pub -> mirror
226 Transfer complete
```

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Coping with the worst middlebox

What should Multipath TCP do in the presence of such a worst middlebox ?

- Do nothing and ignore the middlebox
  - but then the bytestream and the application would be broken and this problem will be difficult to debug by network administrators
- Detect the presence of the middlebox
  - and fallback to regular TCP (i.e. use a single path and nothing fancy)

Multipath TCP **MUST** work in all networks where regular TCP works.

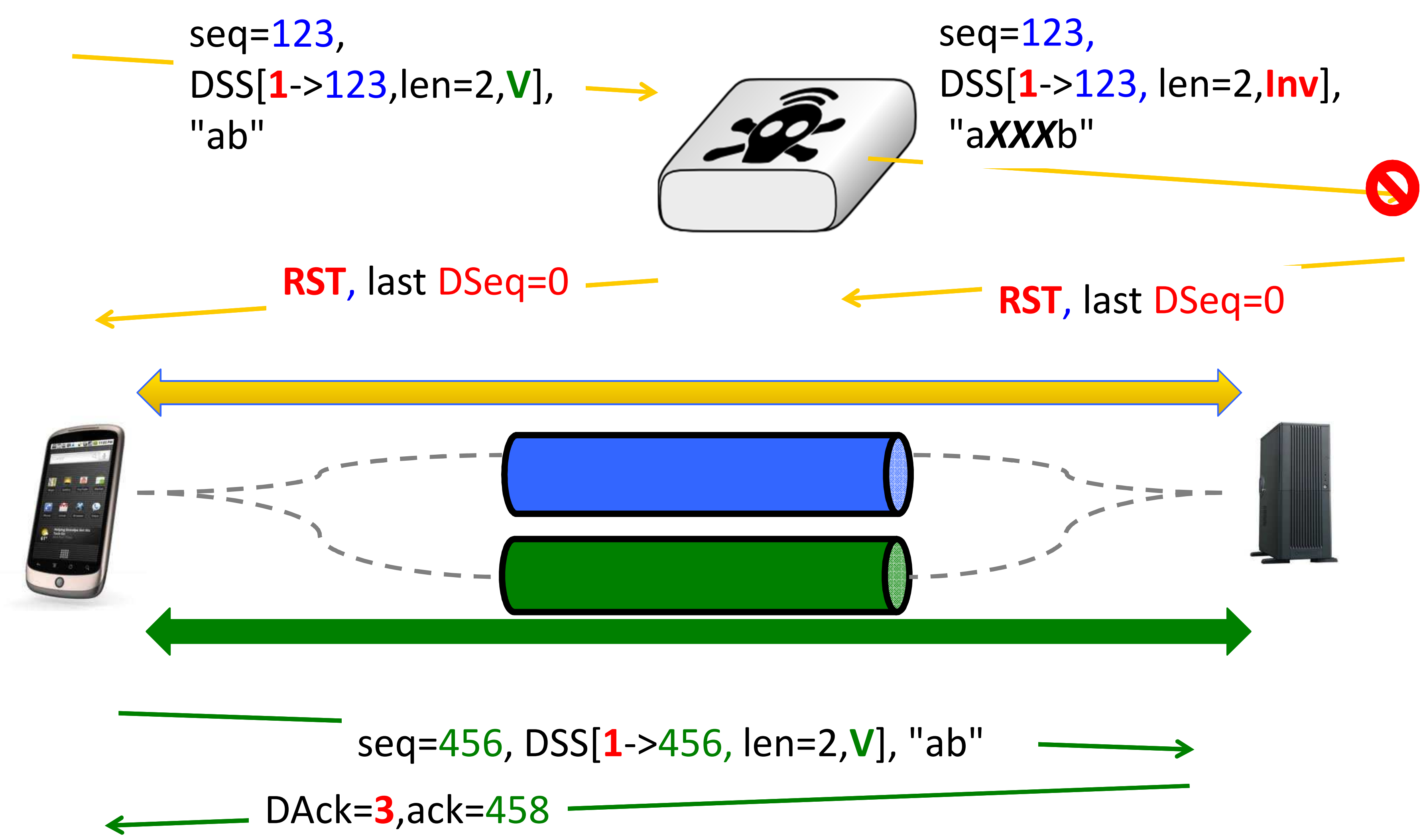
Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Detecting the worst middlebox ?

How can Multipath TCP detect a middlebox that modifies the bytestream and inserts/removes bytes ?

- Various solutions were explored
- In the end, Multipath TCP chose to include its own checksum to detect insertion/deletion of bytes

The worst middlebox



What should be the length of the data sequence numbers ?

- 32 bits
  - compact and compatible with TCP
  - wrap around problem at highspeed requires PAWS

- 64 bits
  - wrap around is not an issue for most transfers today
  - takes more space inside each segment

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Data sequence numbers and Data acknowledgements

- Maintained inside implementation as 64 bits field
- Implementations can, as an optimisation, only transmit the lower 32 bits of the data sequence and acknowledgements

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Data Sequence Signal option

A = Data ACK present
a = Data ACK is 8 octets
M = mapping present
m = DSN is 8

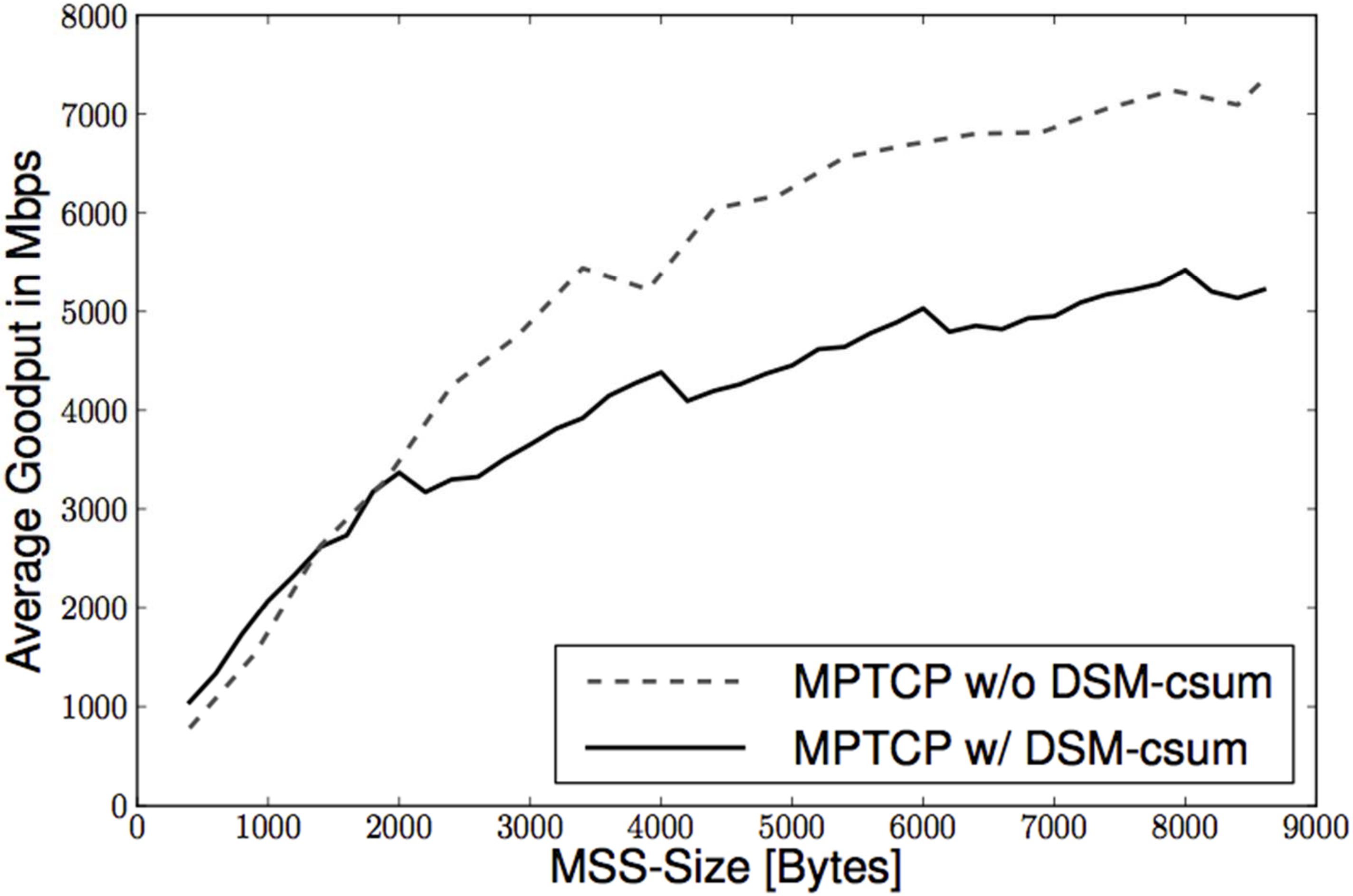
Cumulative Data ack

```
                     1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+---------------+---------------+-------+-----------------------+
|     Kind      |    Length     |Subtype| (reserved) |F|m|M|a|A|
+---------------+---------------+-------+-----------------------+
|          Data ACK (4 or 8 octets, depending on flags)         |
+---------------------------------------------------------------+
|   Data Sequence Number (4 or 8 octets, depending on flags)    |
+---------------------------------------------------------------+
|              Subflow Sequence Number (4 octets)               |
+-------------------------------+-------------------------------+
|  Data-level Length (2 octets) |      Checksum (2 octets)      |
+-------------------------------+-------------------------------+
```

Length of mapping, can extend beyond this segment

Computed over data covered by entire mapping + pseudo header

Cost of the DSN checksum



C. Raiciu, et al. "How hard can it be? designing and implementing a deployable multipath TCP," NSDI'12: Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation, 2012.

Control plane

- How to manage a Multipath TCP connection that uses several paths ?

Data plane

- How to transport data ?

Congestion control

- How to control congestion over multiple paths ?

A linear rate adaption algorithm

- $rate(t+1) = \alpha_C + \beta_C rate(t)$ when the network is congested
- $rate(t+1) = \alpha_N + \beta_N rate(t)$ when the network is *not* congested

To be fair and efficient, a linear algorithm must use additive increase and multiplicative decrease (AIMD)

```
# Additive Increase Multiplicative Decrease
if congestion :
   rate=rate*betaC    # multiplicative decrease, betaC<1
else
   rate=rate+alphaN    # additive increase, v0>0
```

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

AIMD in TCP

Congestion control mechanism

- Each host maintains a congestion window (cwnd)

No congestion

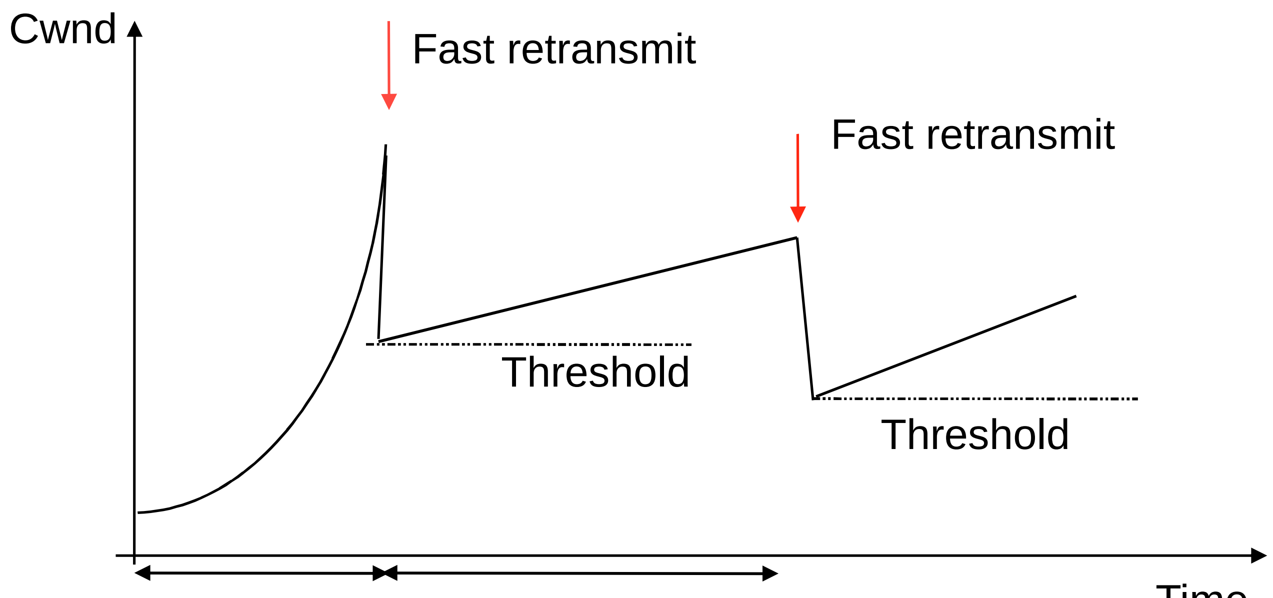
- Congestion avoidance (additive increase)
- increase cwnd by one segment every round-trip-time

Congestion

- TCP detects congestion by detecting losses
- Mild congestion (fast retransmit – multiplicative decrease)
- cwnd=cwnd/2 and restart congestion avoidance
- Severe congestion (timeout)
- cwnd=1, set slow-start-threshold and restart slow-start

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Evolution of the congestion window



Slow-start
exponential increase of cwnd

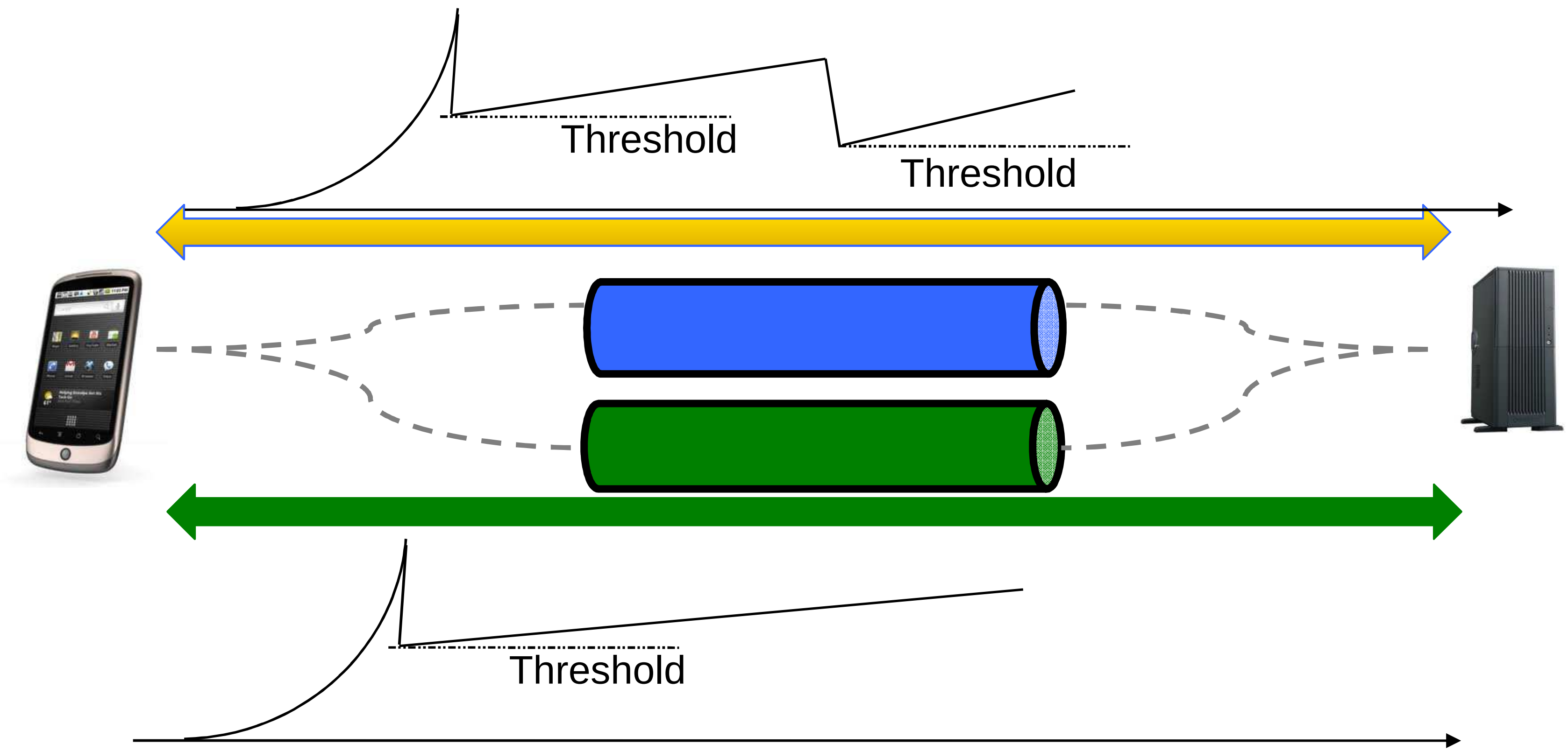
Congestion avoidance
linear increase of cwnd

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

TECHNISCHE UNIVERSITÄT DARMSTADT

Simple approach

- independant congestion windows



Control plane

- How to manage a Multipath TCP connection that uses several paths ?

Data plane

- How to transport data ?

Congestion control

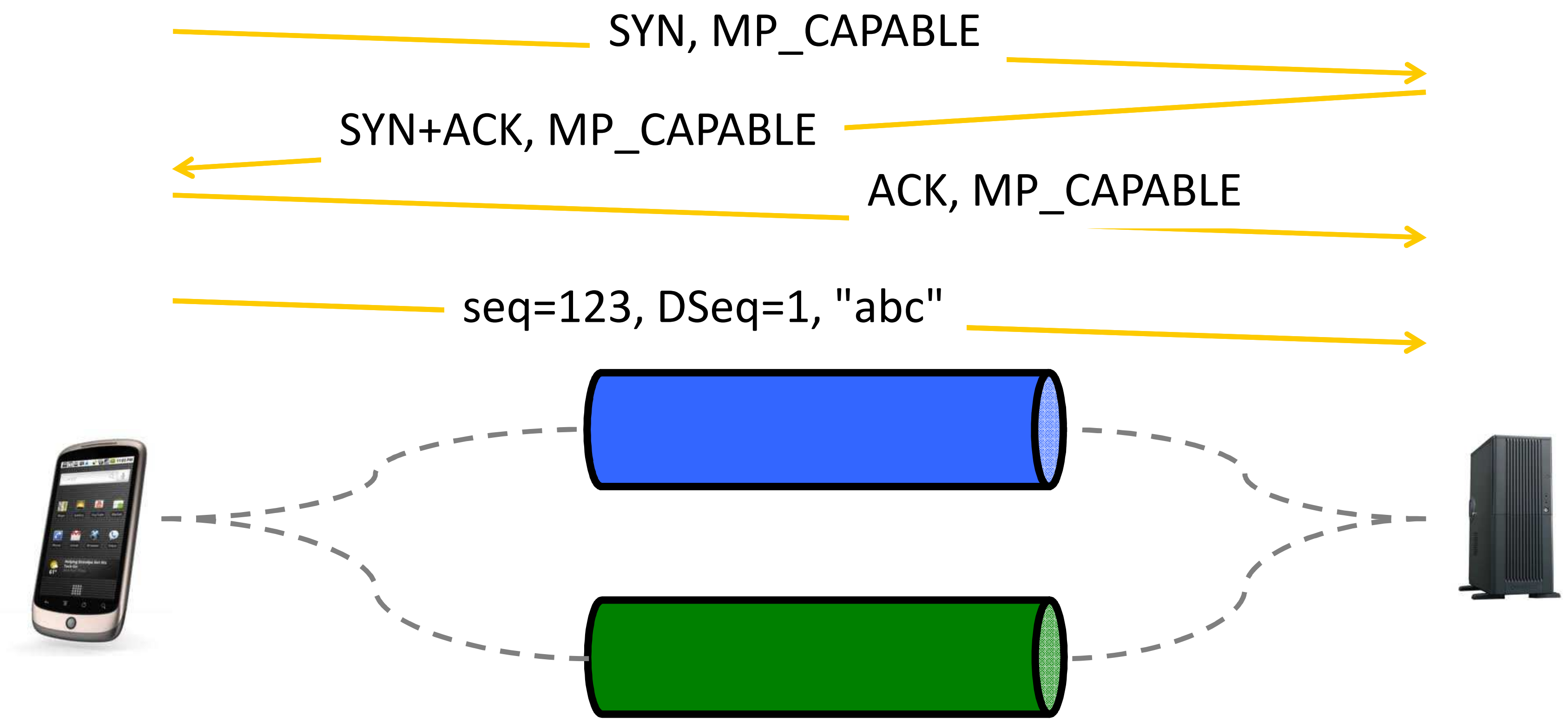
- How to control congestion over multiple paths ?

Connection establishment in details

Closing a Multipath TCP connection

Address dynamics

Principle


SYN, MP_CAPABLE
SYN+ACK, MP_CAPABLE
ACK, MP_CAPABLE
seq=123, DSeq=1, "abc"

Roles of the initial TCP handshake

Check willingness to open TCP connection

- Propose initial sequence number
- Negotiate Maximum Segment Size

TCP options

- negotiate Timestamps, SACK, Window scale

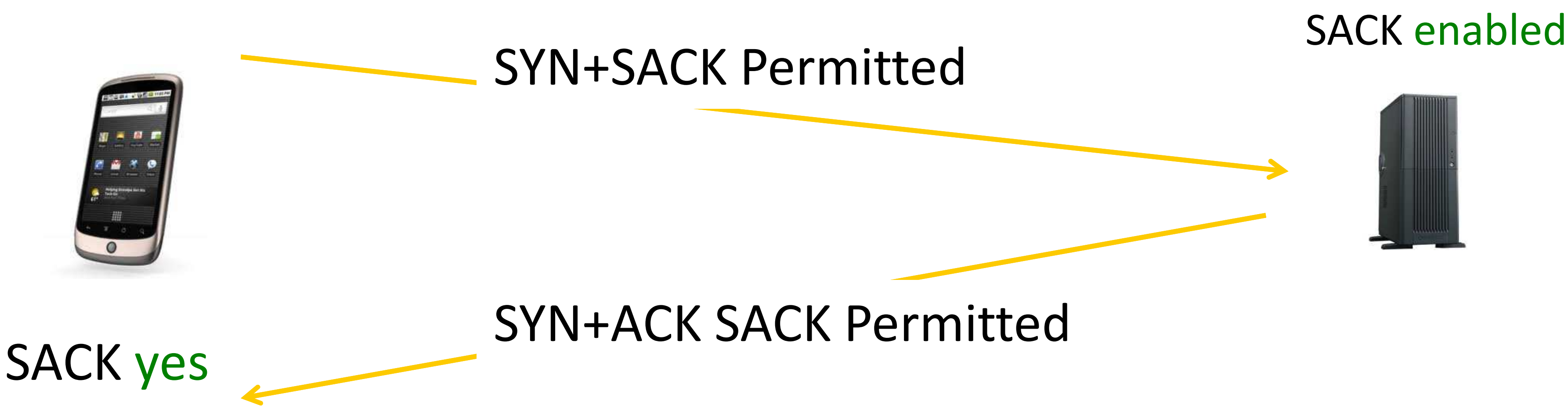
Multipath TCP

- check that server supports Multipath TCP
- propose Token in each direction
- propose initial Data sequence number in each direction
- Exchange keys to authenticate subflows

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

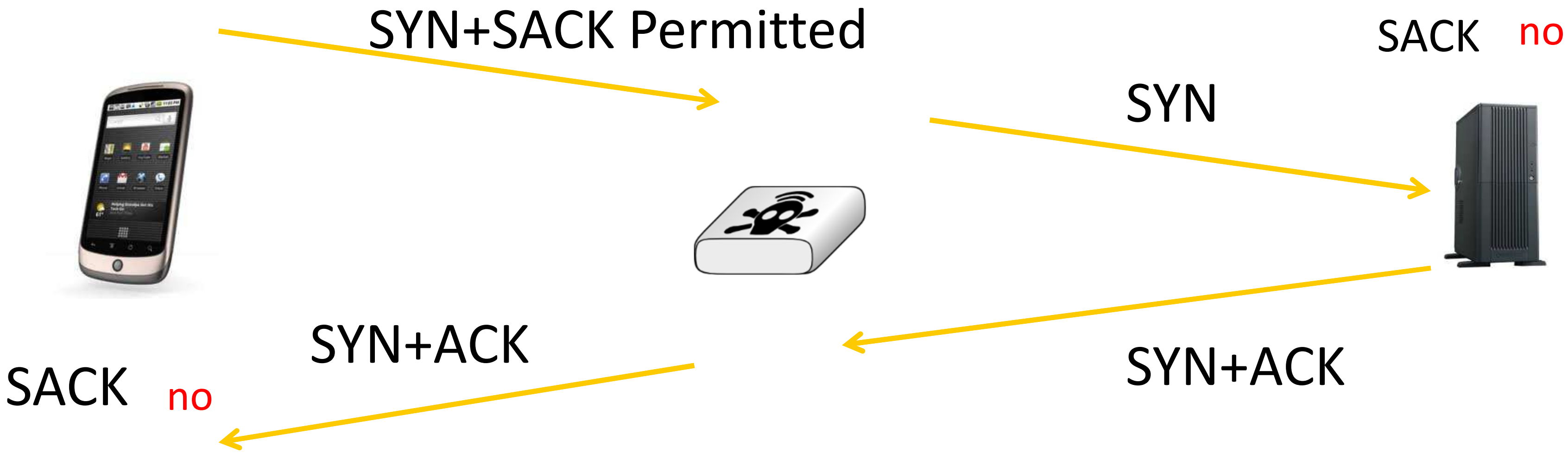
TCP options were invented for this purpose

- Exemple SACK



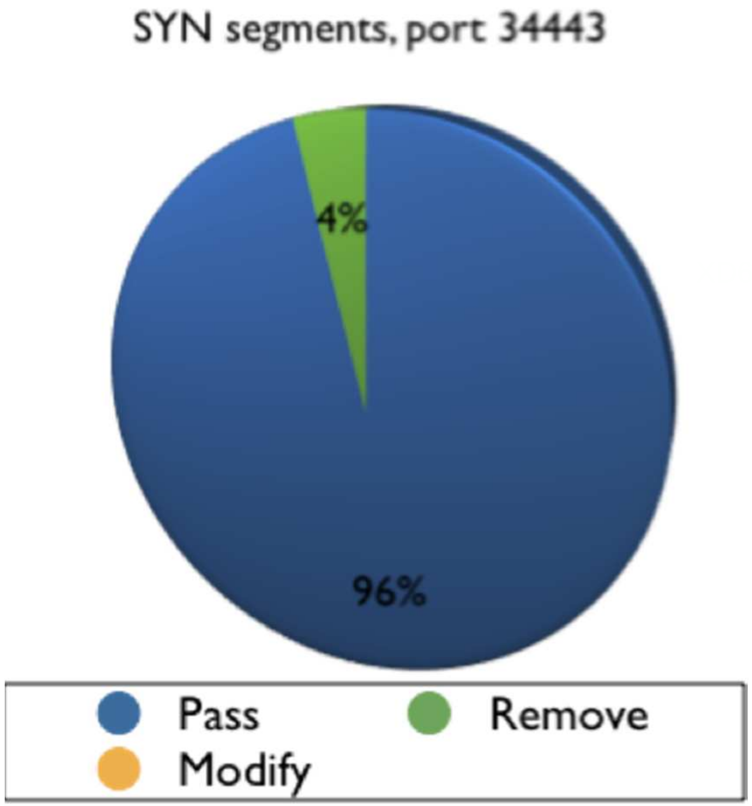
TECHNISCHE UNIVERSITÄT DARMSTADT

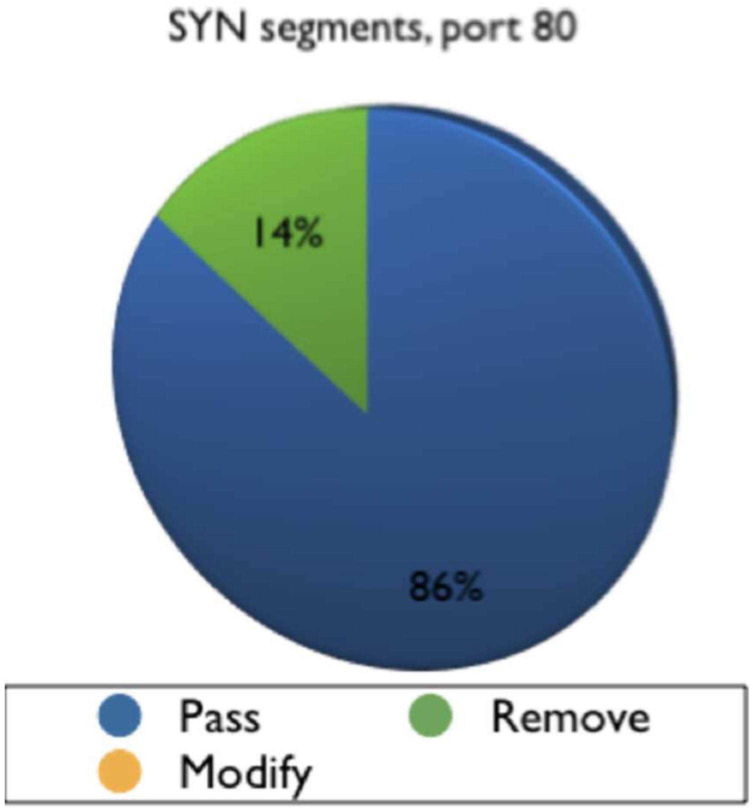
What happens when there are middleboxes on the path ?



TCP options

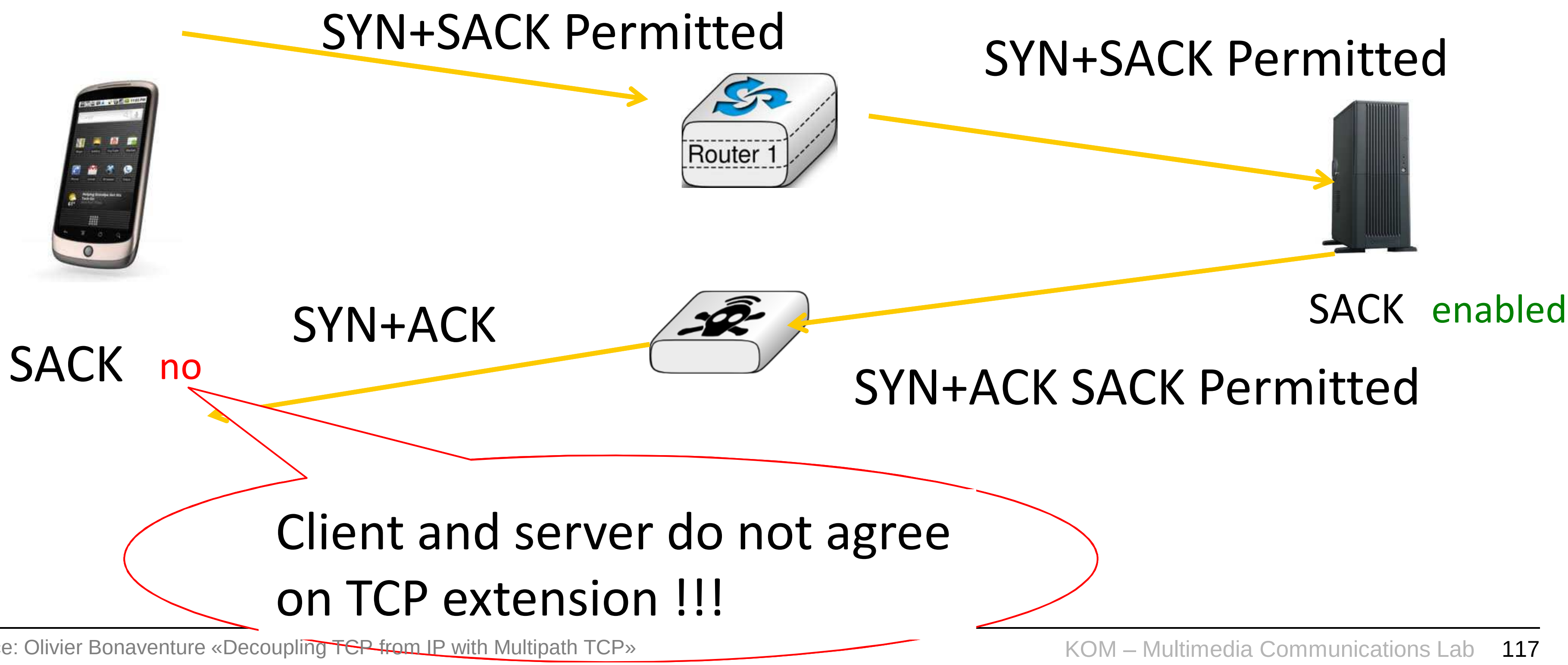
In SYN segments



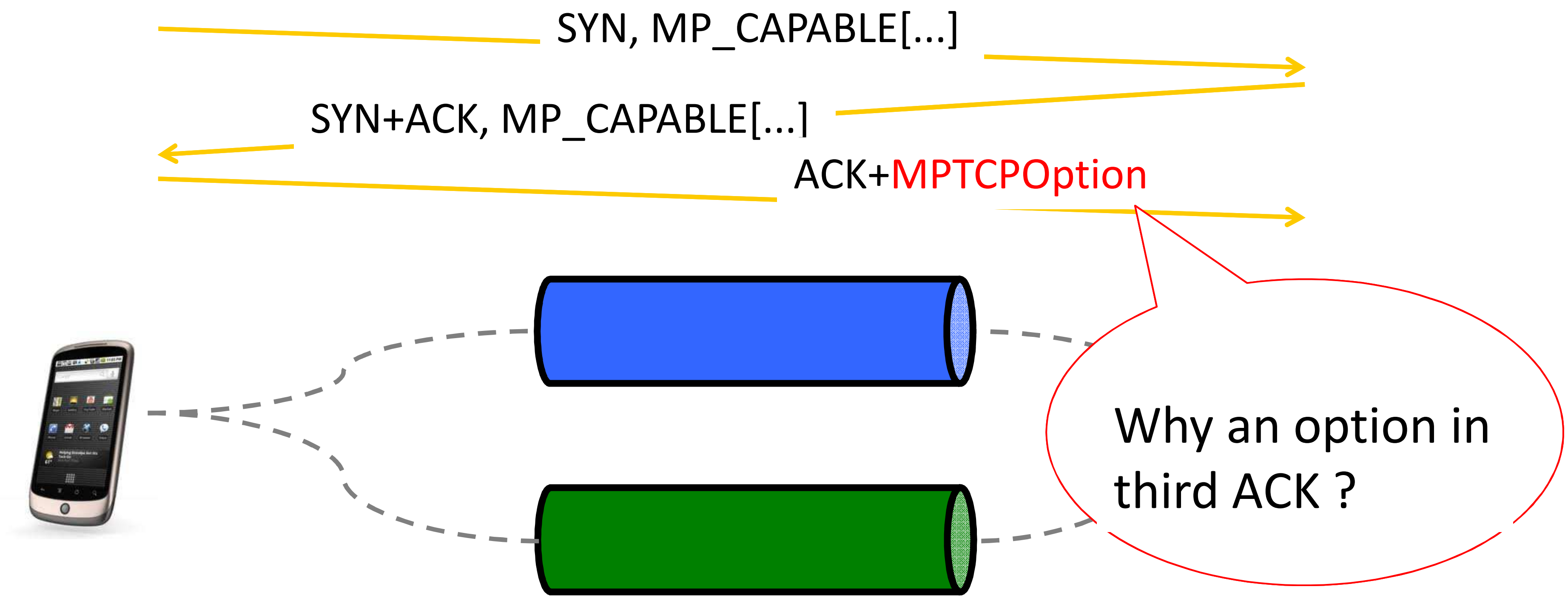


Honda, Michio, et al. "Is it still possible to extend TCP?." Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. ACM, 2011.

What happens when there are middleboxes on the path ?

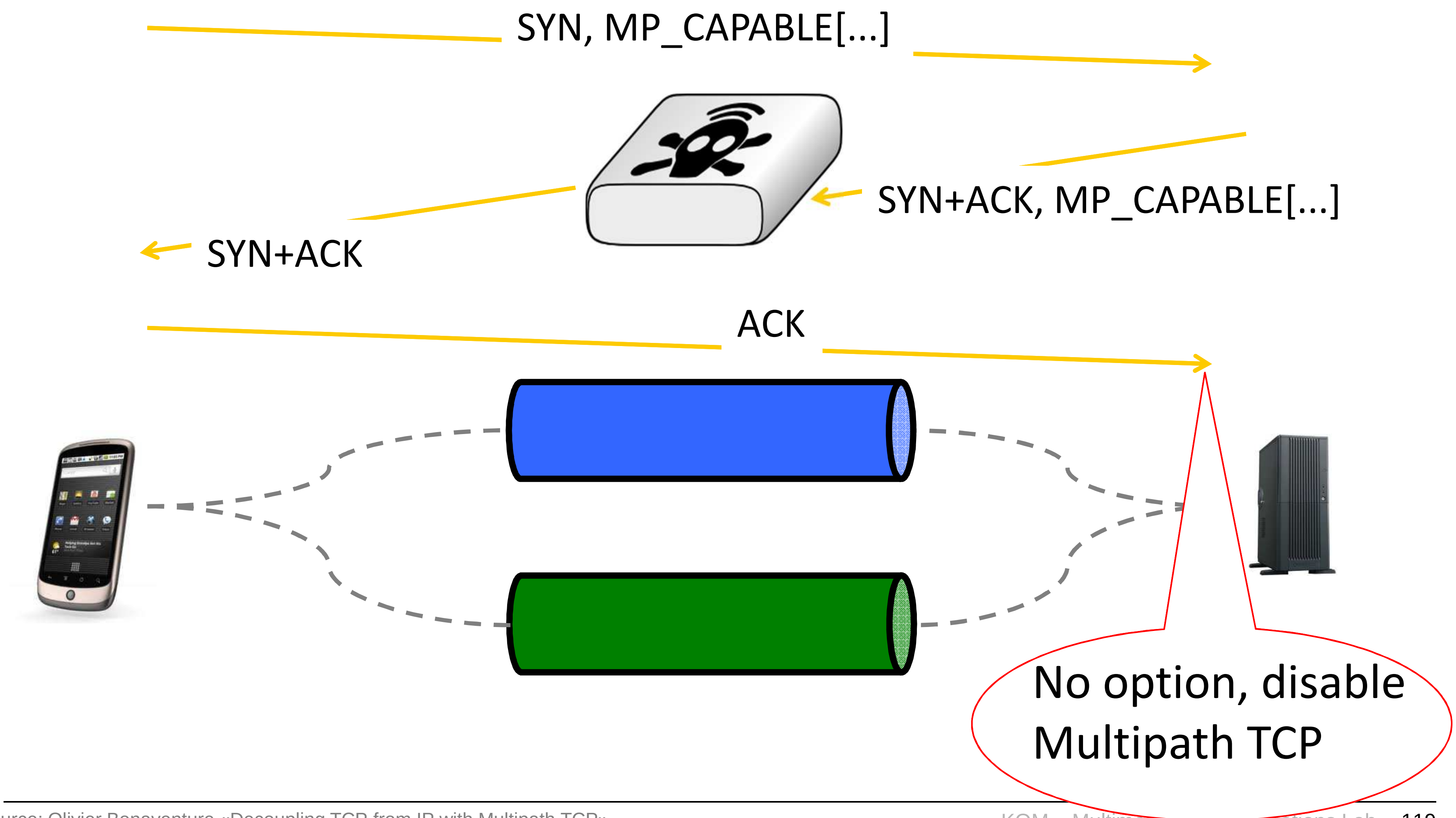


Multipath TCP handshake



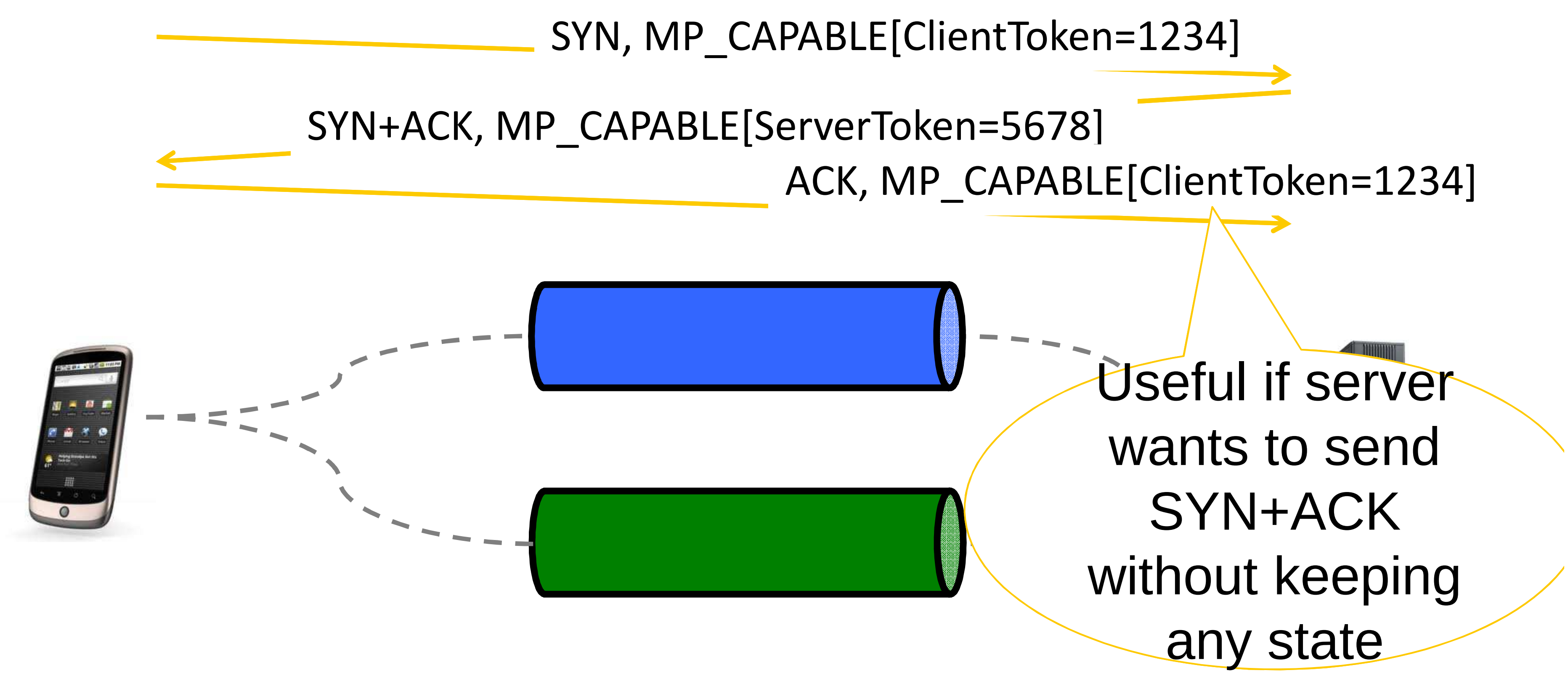
Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Multipath TCP option in third ACK



Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

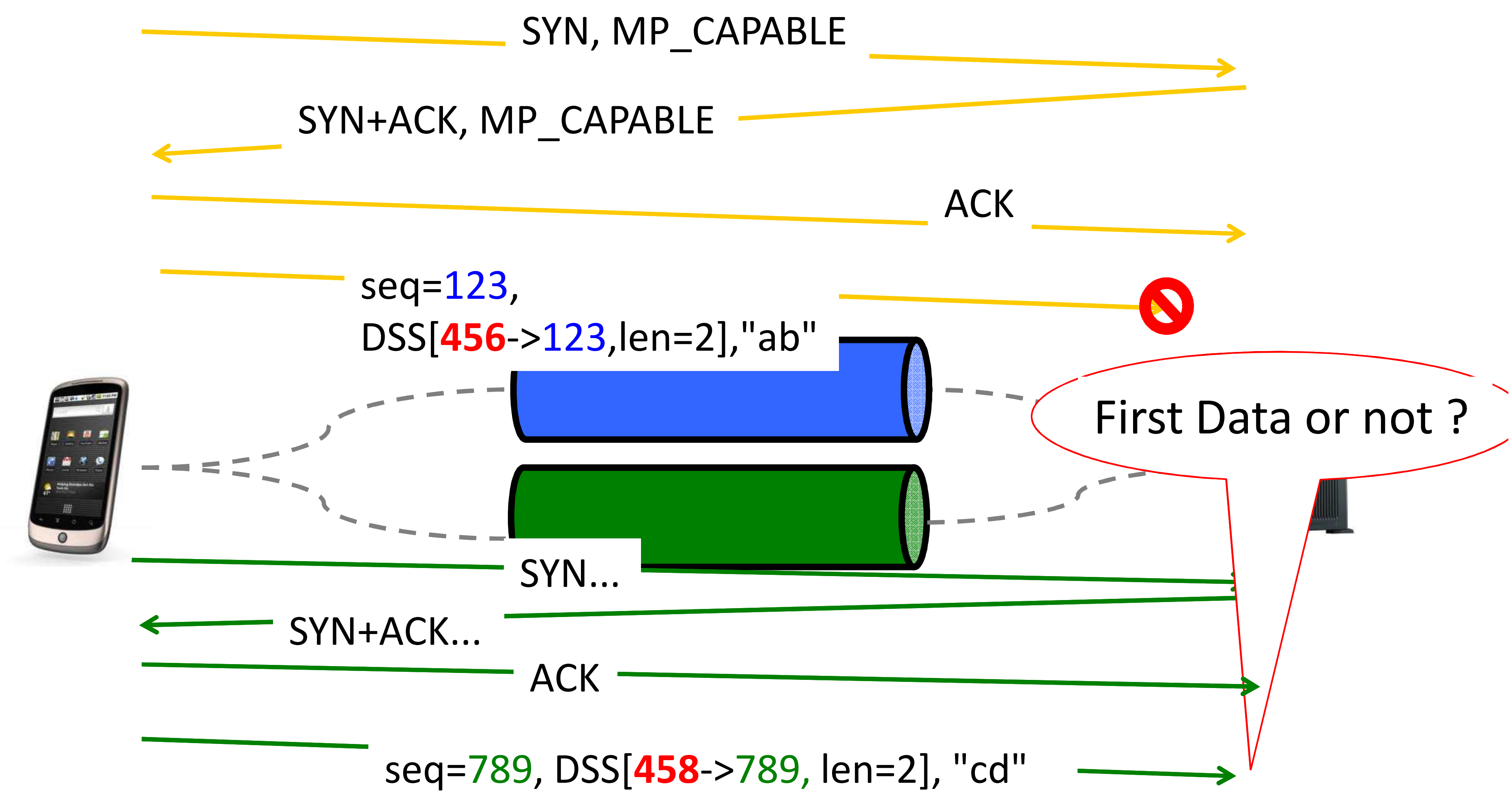
Multipath TCP handshake Token exchange



Why do we need an initial Data Sequence number ?

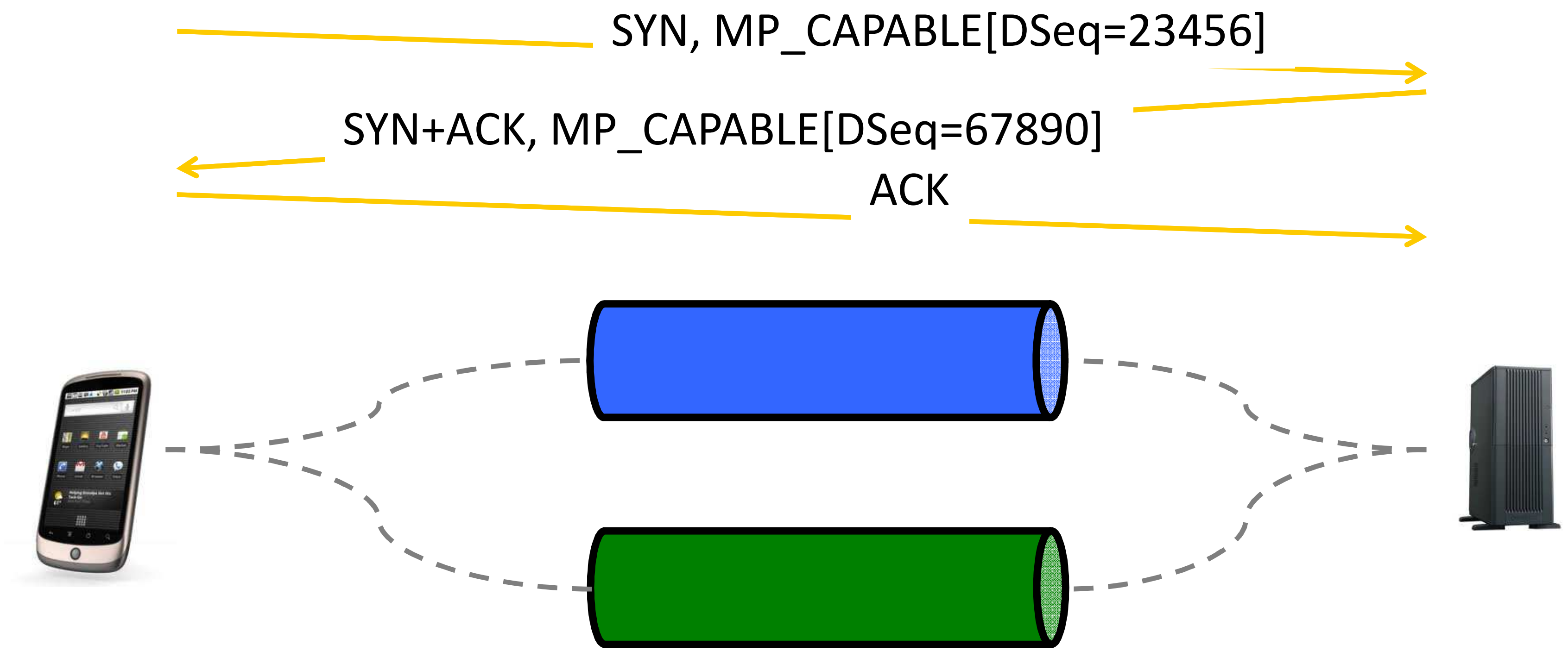
- Setting IDSN to a random value improves security
- Hosts must know IDSN to avoid losing data in some special cases

Initial Data Sequence number



Initial Data Sequence number

How to negotiate the IDSN ?



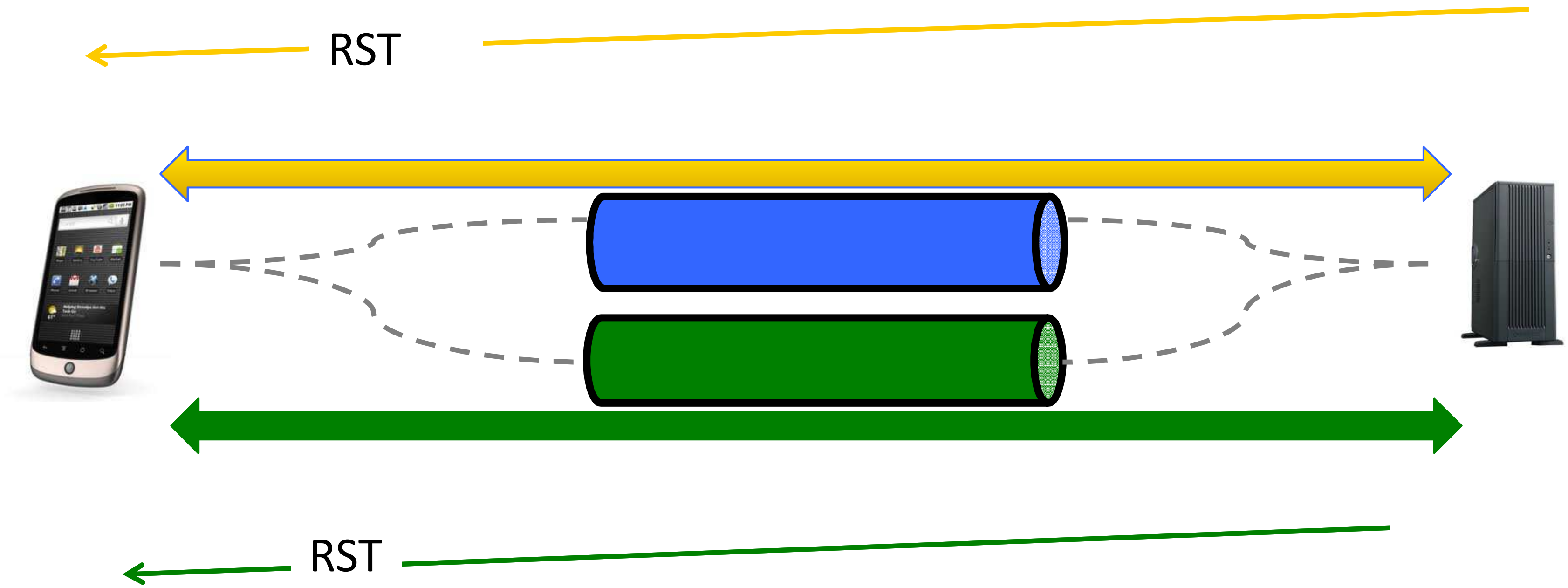
Connection establishment in details

Closing a Multipath TCP connection

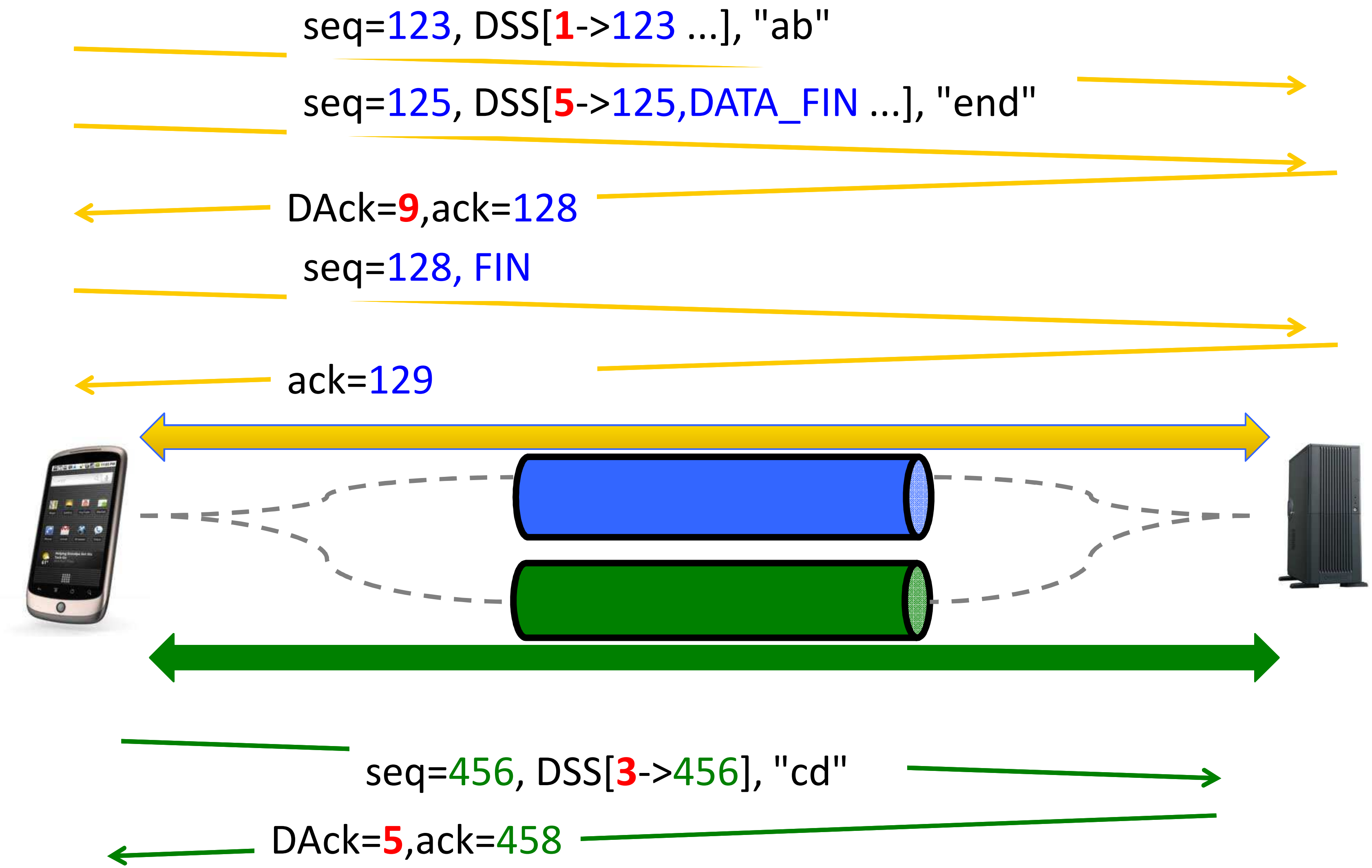
Address dynamics

How to close a Multipath TCP connection ?

- By closing all subflows ?



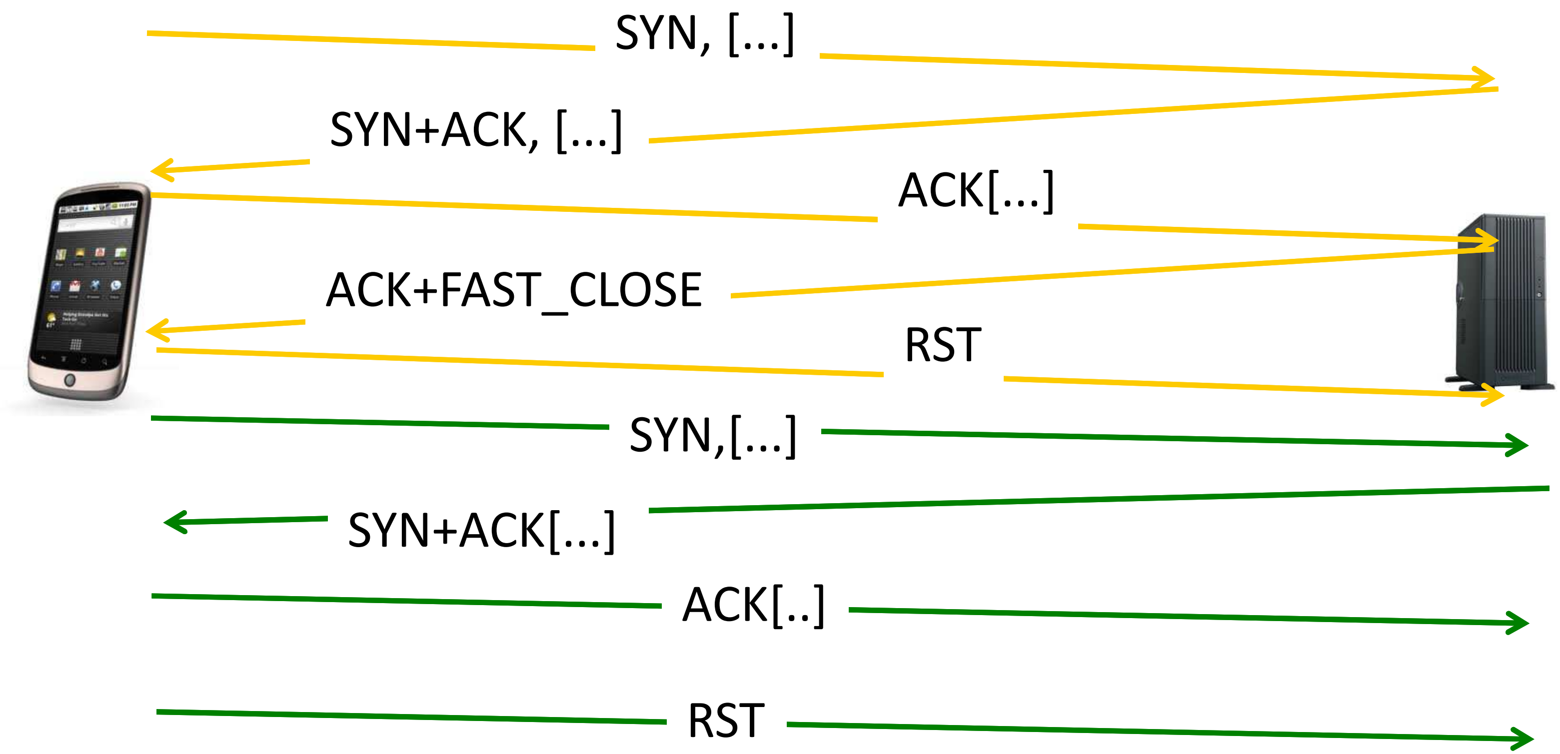
Closing a Multipath TCP connection



Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Closing a Multipath TCP connection

FAST Close



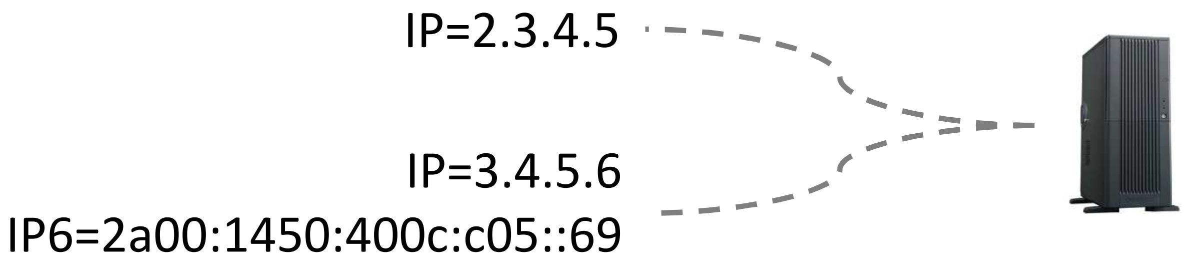
Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Connection establishment in details

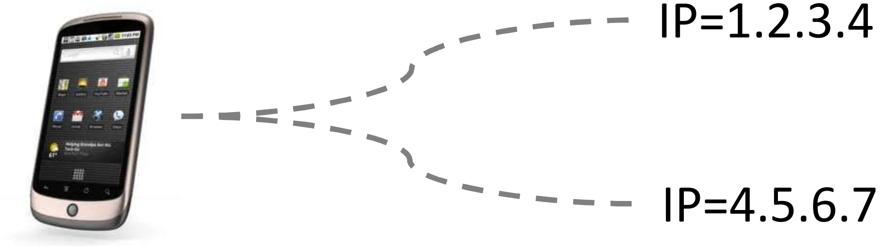
Closing a Multipath TCP connection

Address dynamics

How to learn the addresses of a host ?



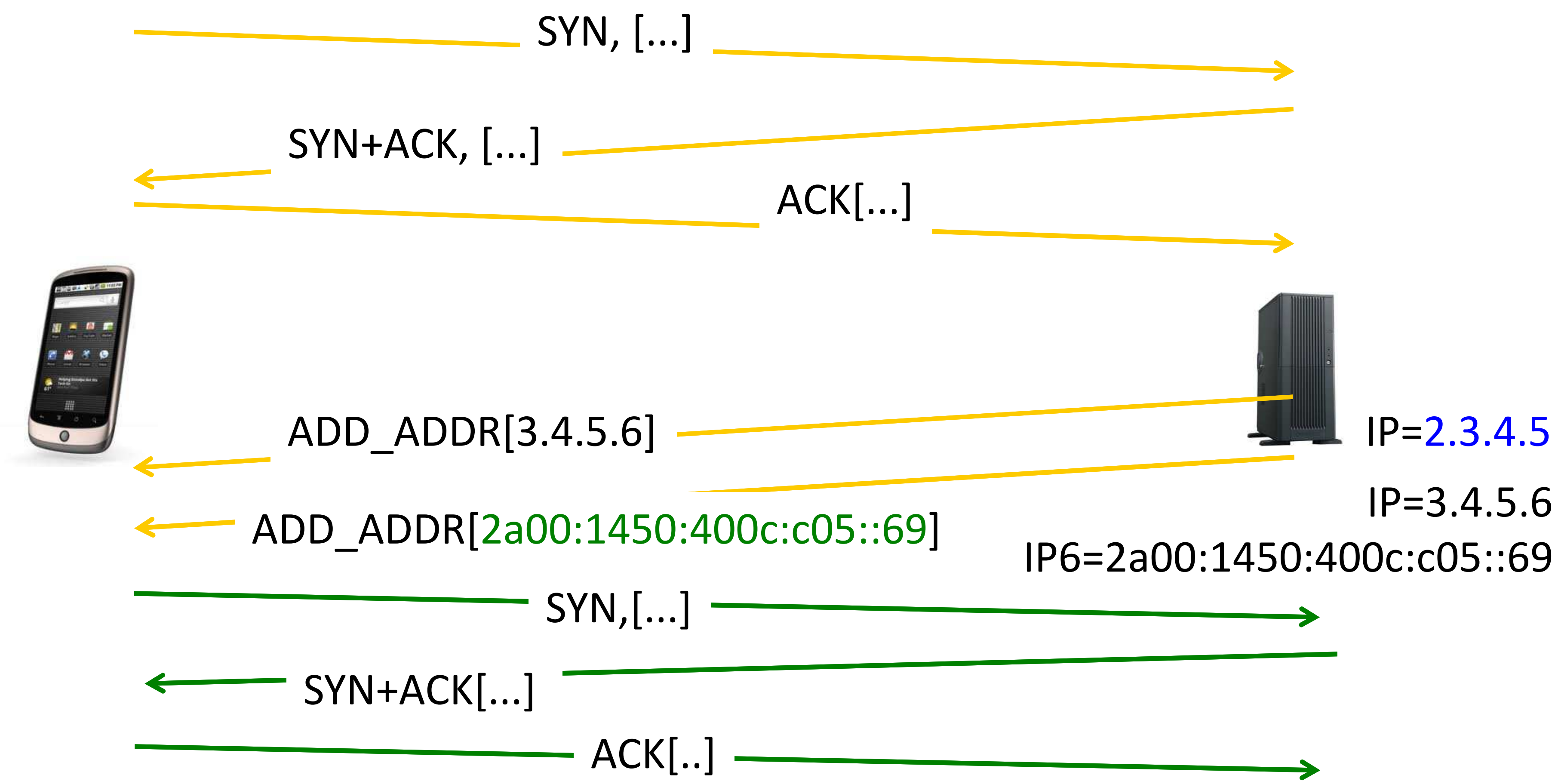
How to deal with address changes ?



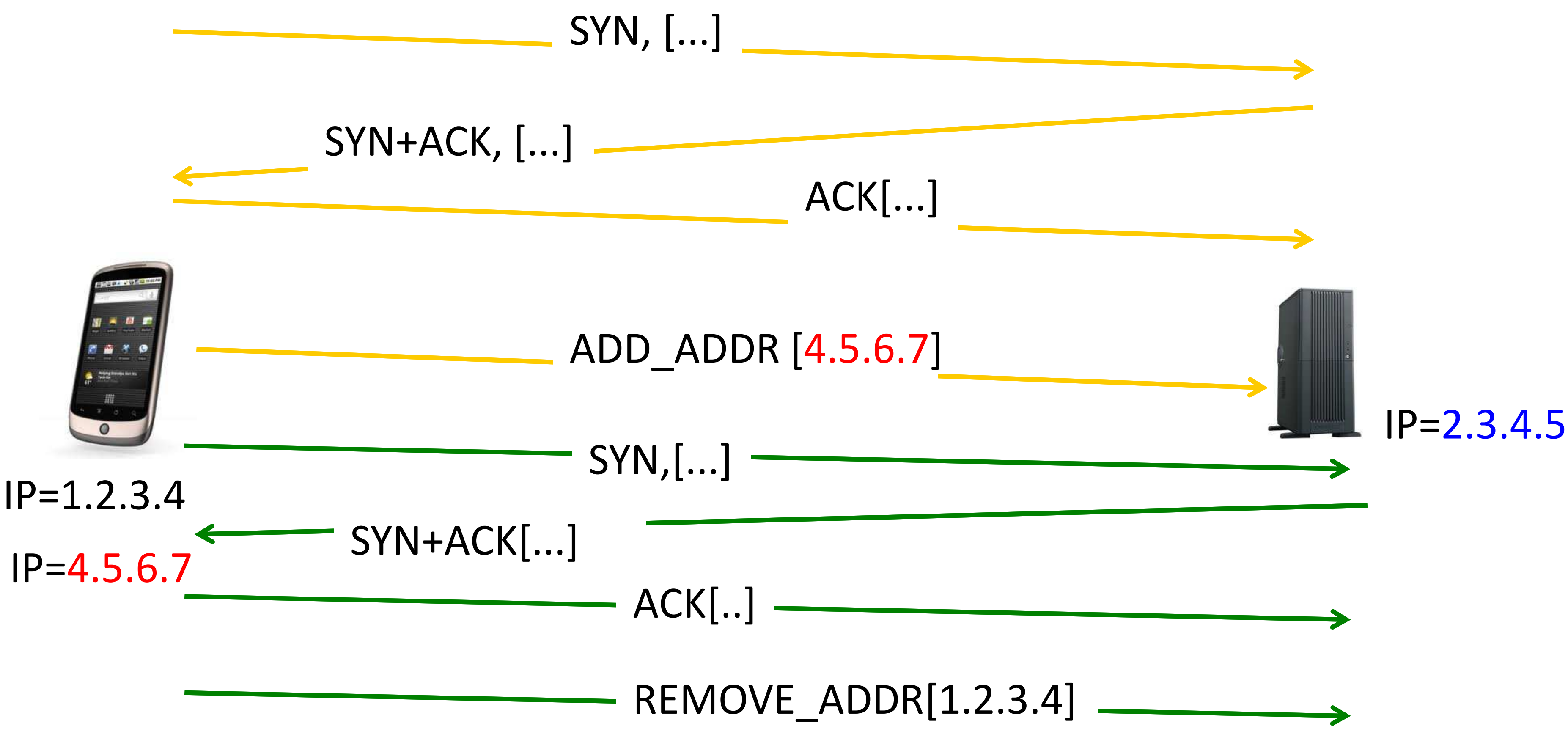
Address dynamics

TECHNISCHE UNIVERSITÄT DARMSTADT

Basic solution : multihomed server



Basic solution : mobile client



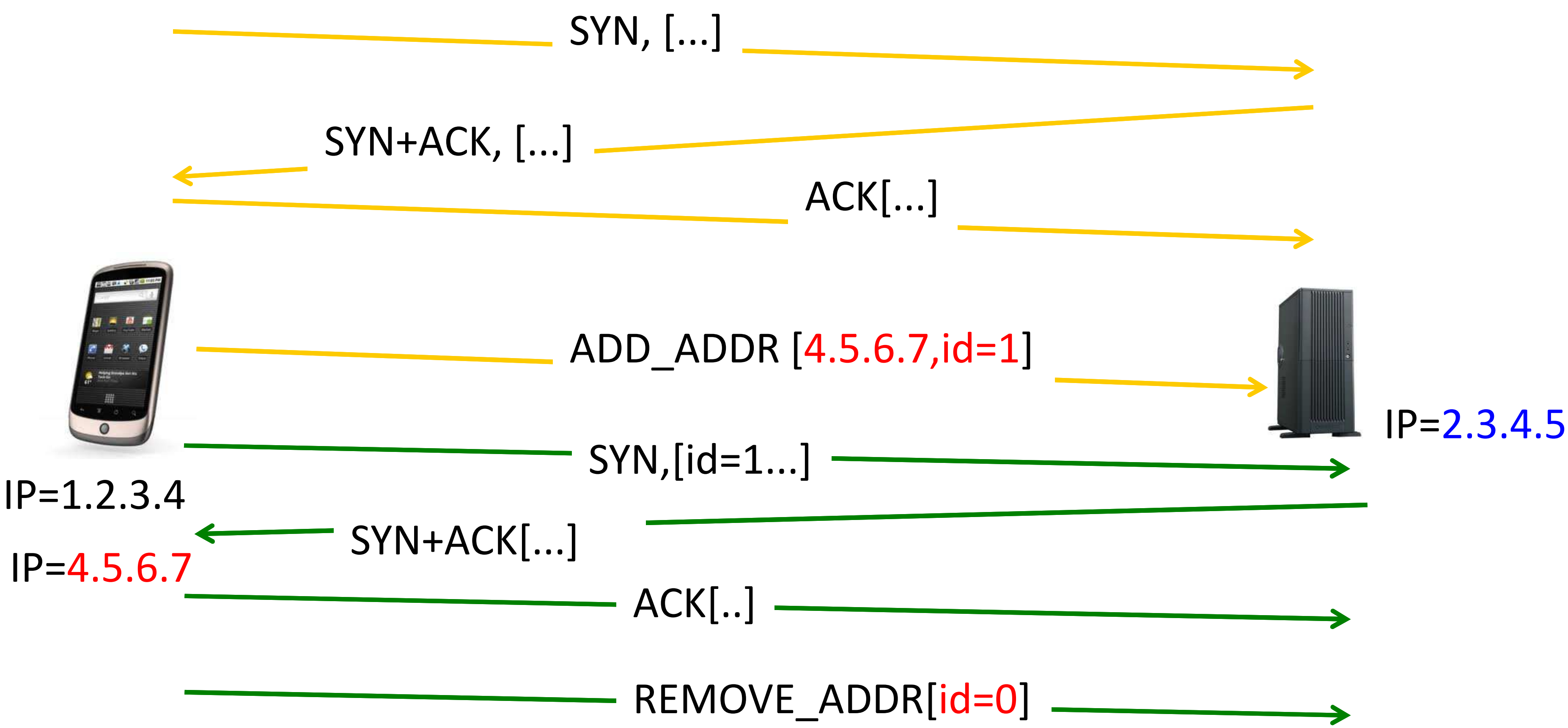
Address dynamics with NATs

Solution

- Each address has one identifier
  - Subflow is established between id=0 addresses
- Each host maintains a list of <address,id> pairs of the addresses associated to an MPTCP endpoint
- MPTCP options refer to the address identifier
  - ADD_ADDR contains <address,id>
  - REMOVE_ADDR contains <id>

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Address dynamics



Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

The motivations for Multipath TCP

The changing Internet

The Multipath TCP Protocol

Multipath TCP use cases

- Datacenters
- Smartphones

3.1 Multipath TCP use cases - Datacenters

The motivations for Multipath TCP

The changing Internet

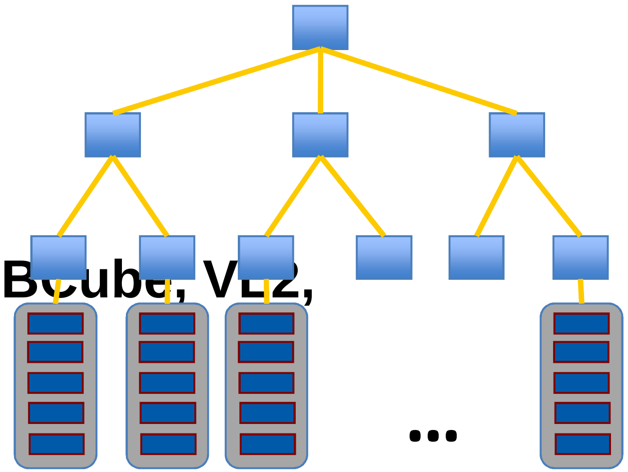
The Multipath TCP Protocol

Multipath TCP use cases

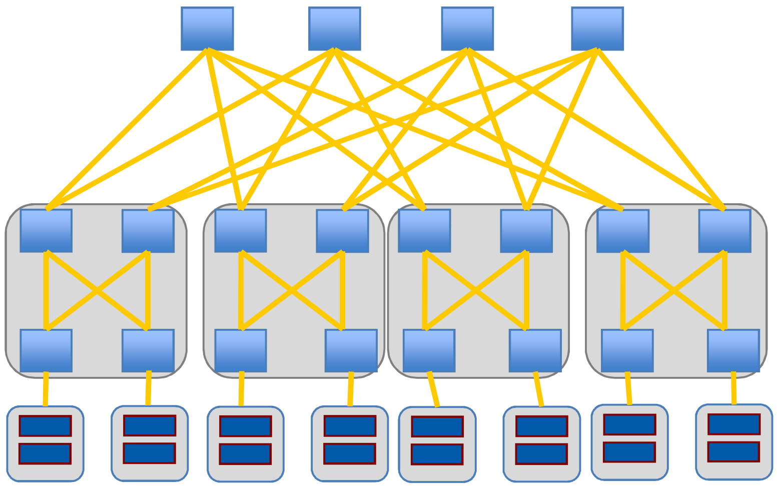
- Datacenters
- Smartphones

Traditional Topologies are tree-based

- Poor performance
- Not fault tolerant



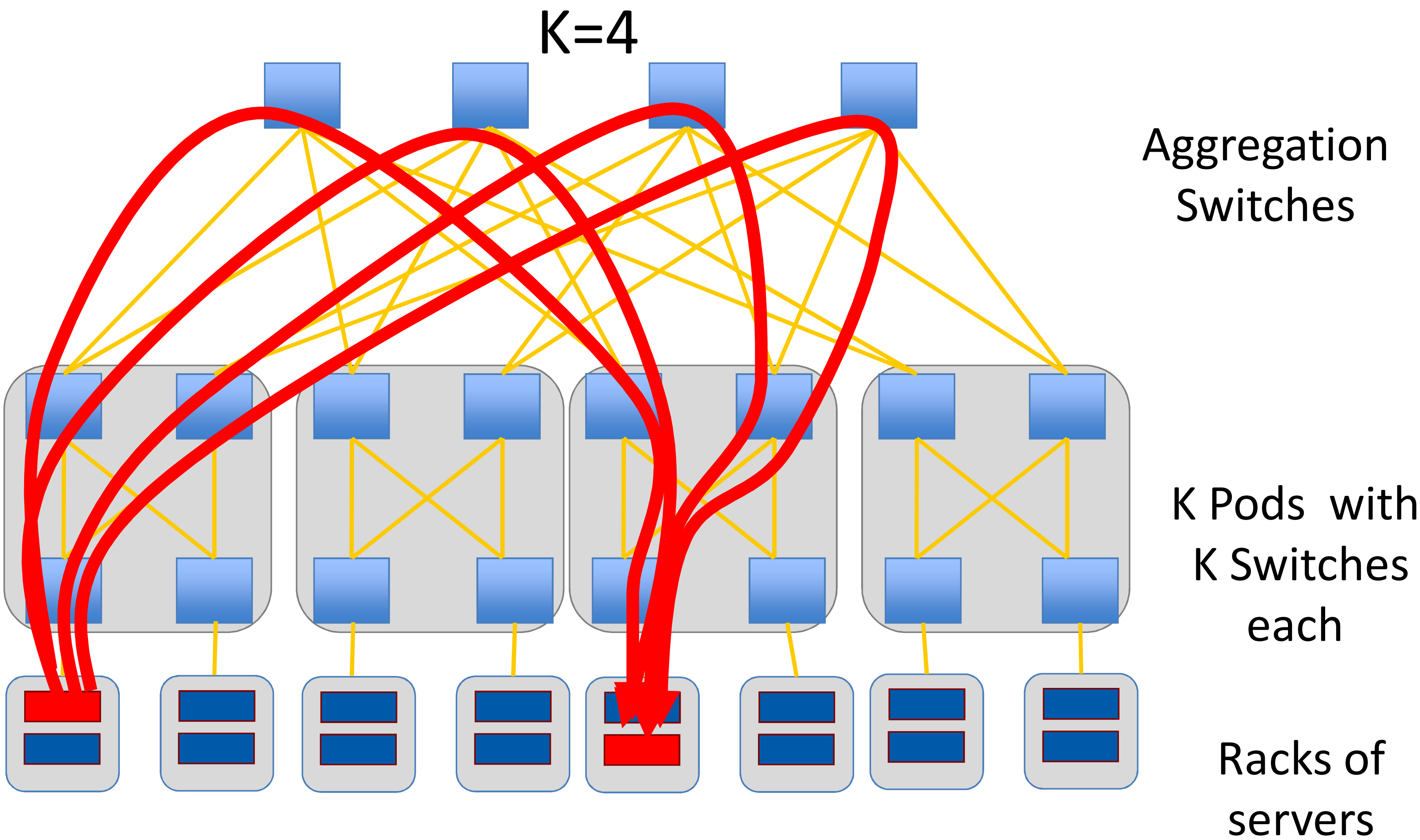
Shift towards multipath topologies: FatTree, BCube, VL2, Cisco, EC2



C. Raiciu, et al. "Improving datacenter performance and robustness with multipath TCP," *ACM SIGCOMM* 2011.

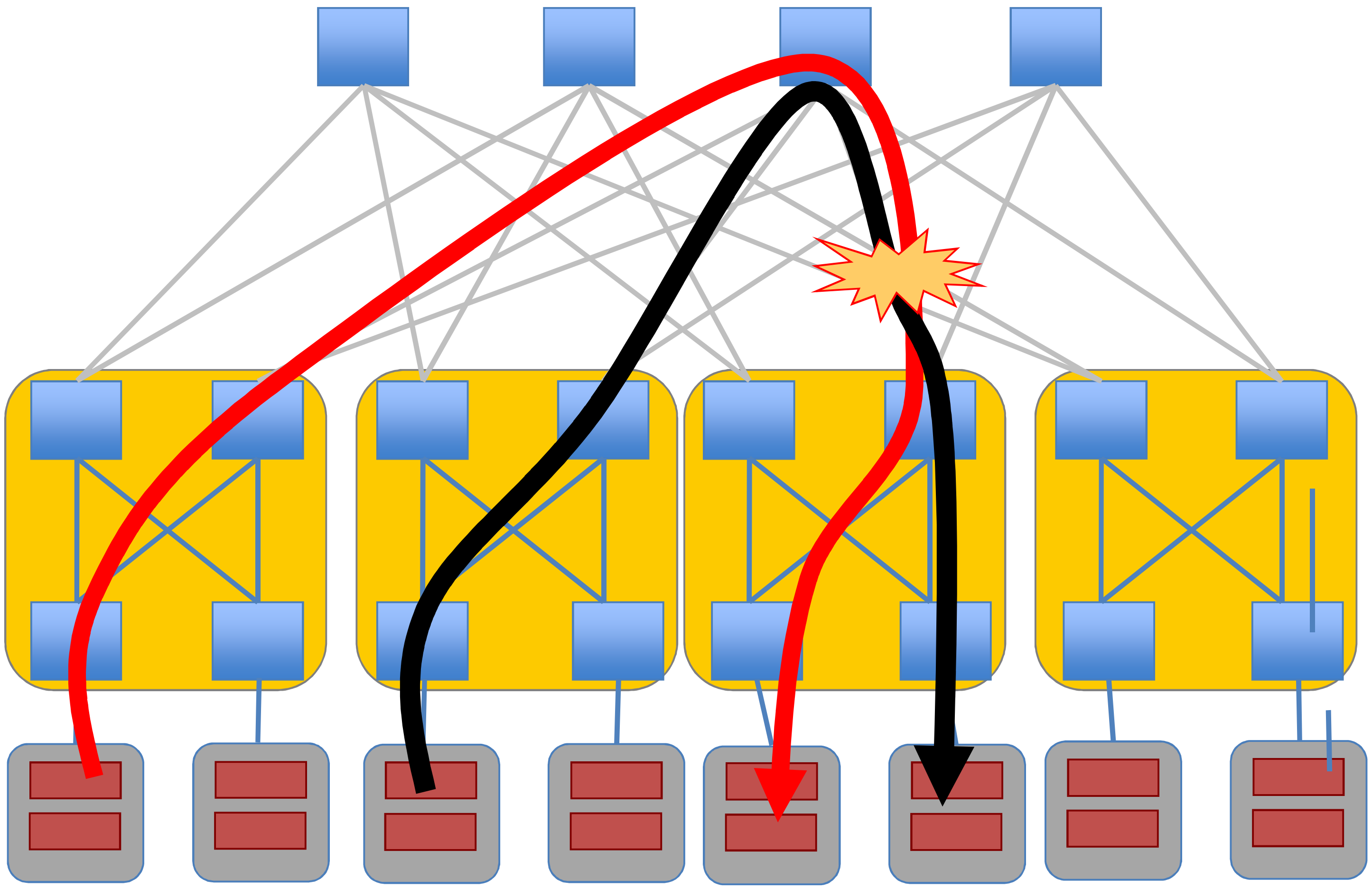
Fat Tree Topology [Fares et al., 2008; Clos, 1953]

TECHNISCHE UNIVERSITÄT DARMSTADT



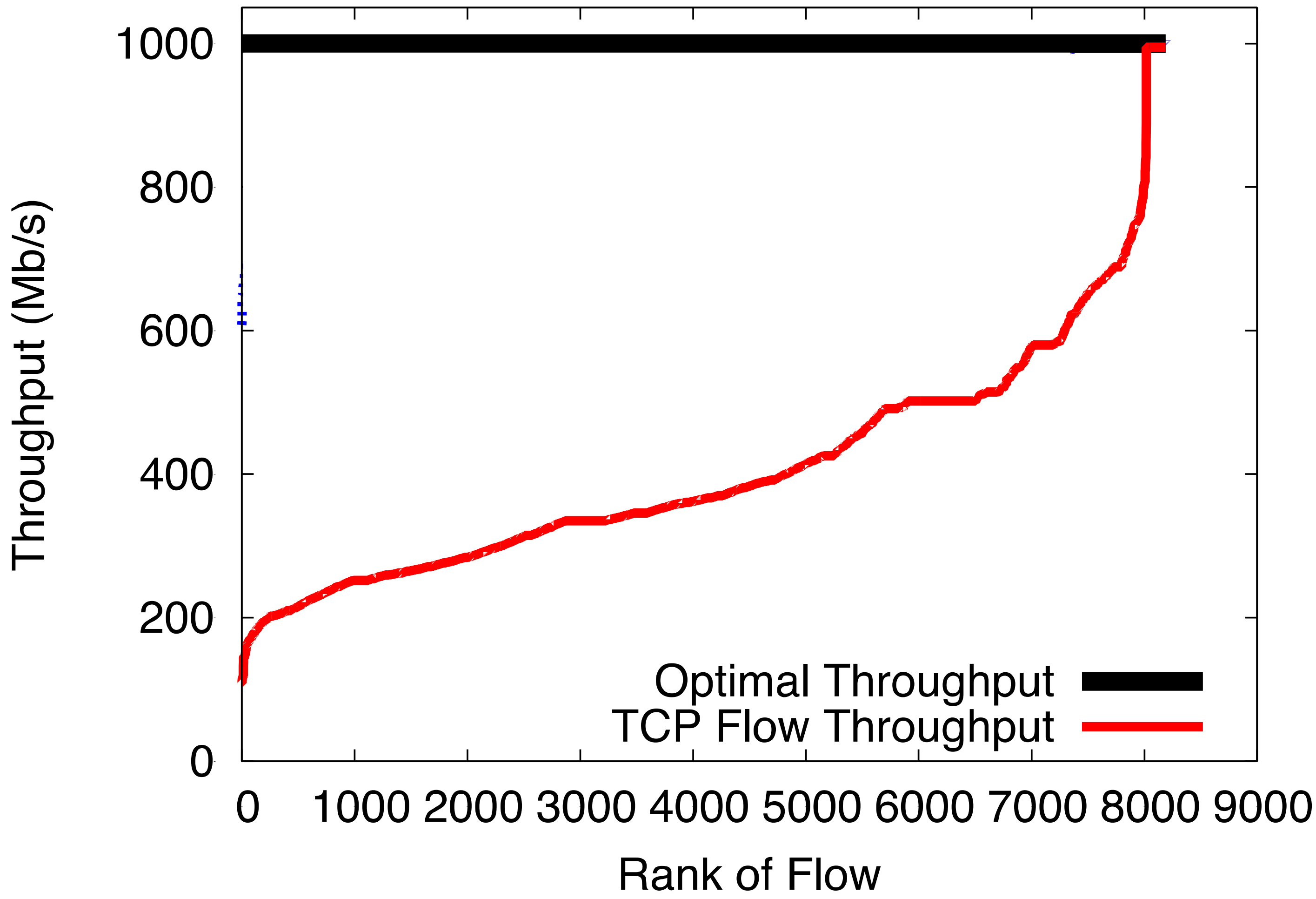
C. Raiciu, et al. "Improving datacenter performance and robustness with multipath TCP," *ACM SIGCOMM* 2011.

TCP in data centers



Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

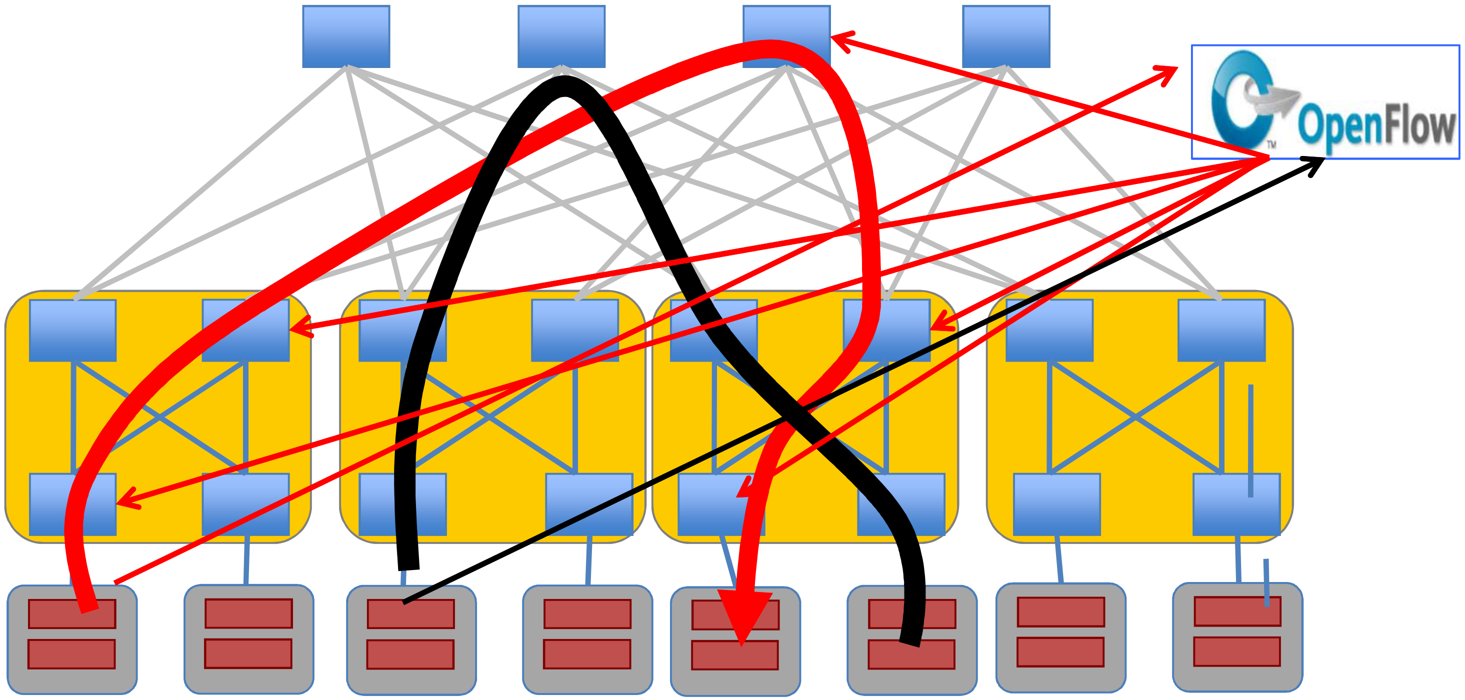
TCP in FAT tree networks
Cost of collissions



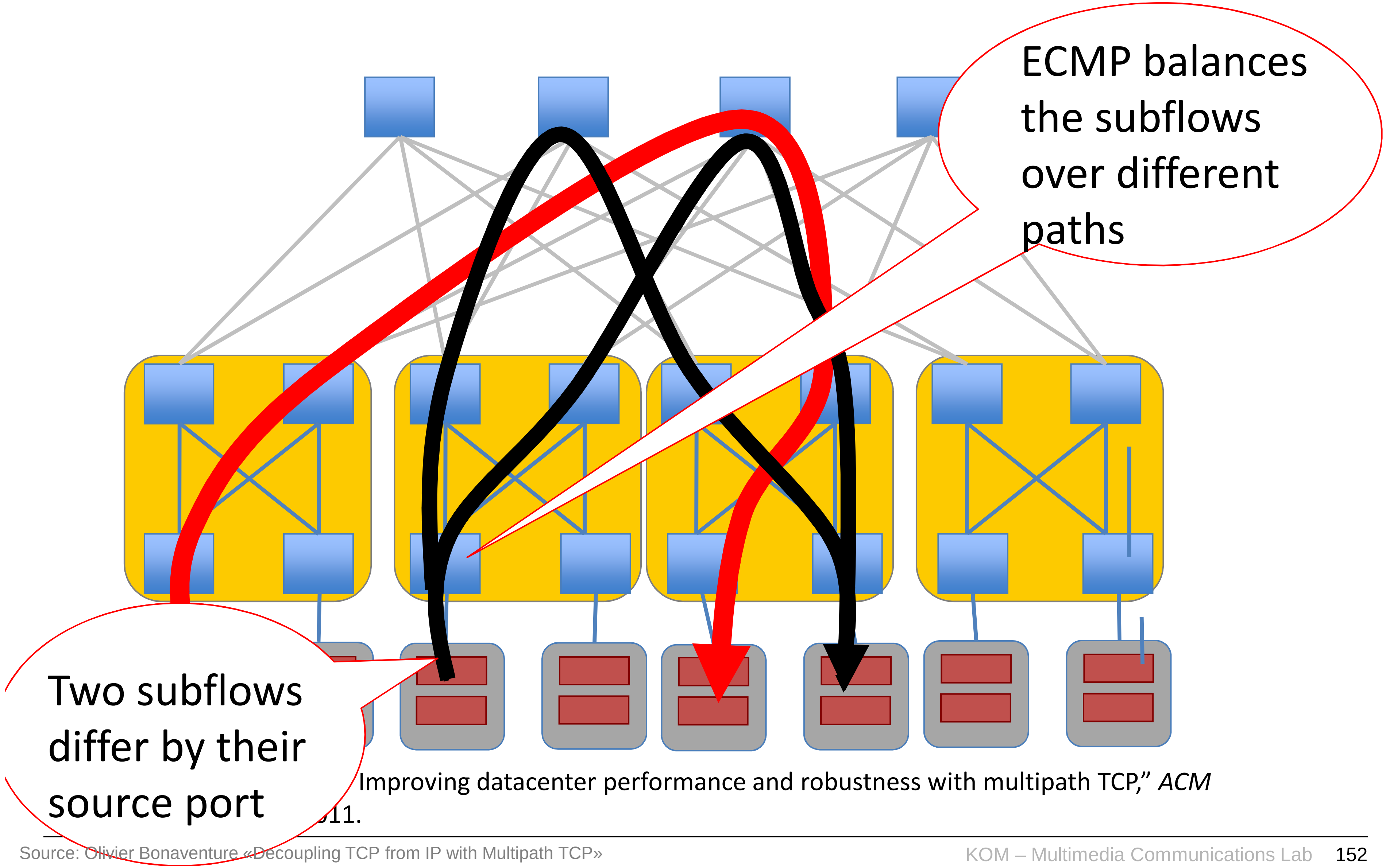
C. Raiciu, et al. “Improving datacenter performance and robustness with multipath TCP,” *ACM SIGCOMM* 2011.

How to get rid of these collisions ?

Consider TCP performance as an optimization problem

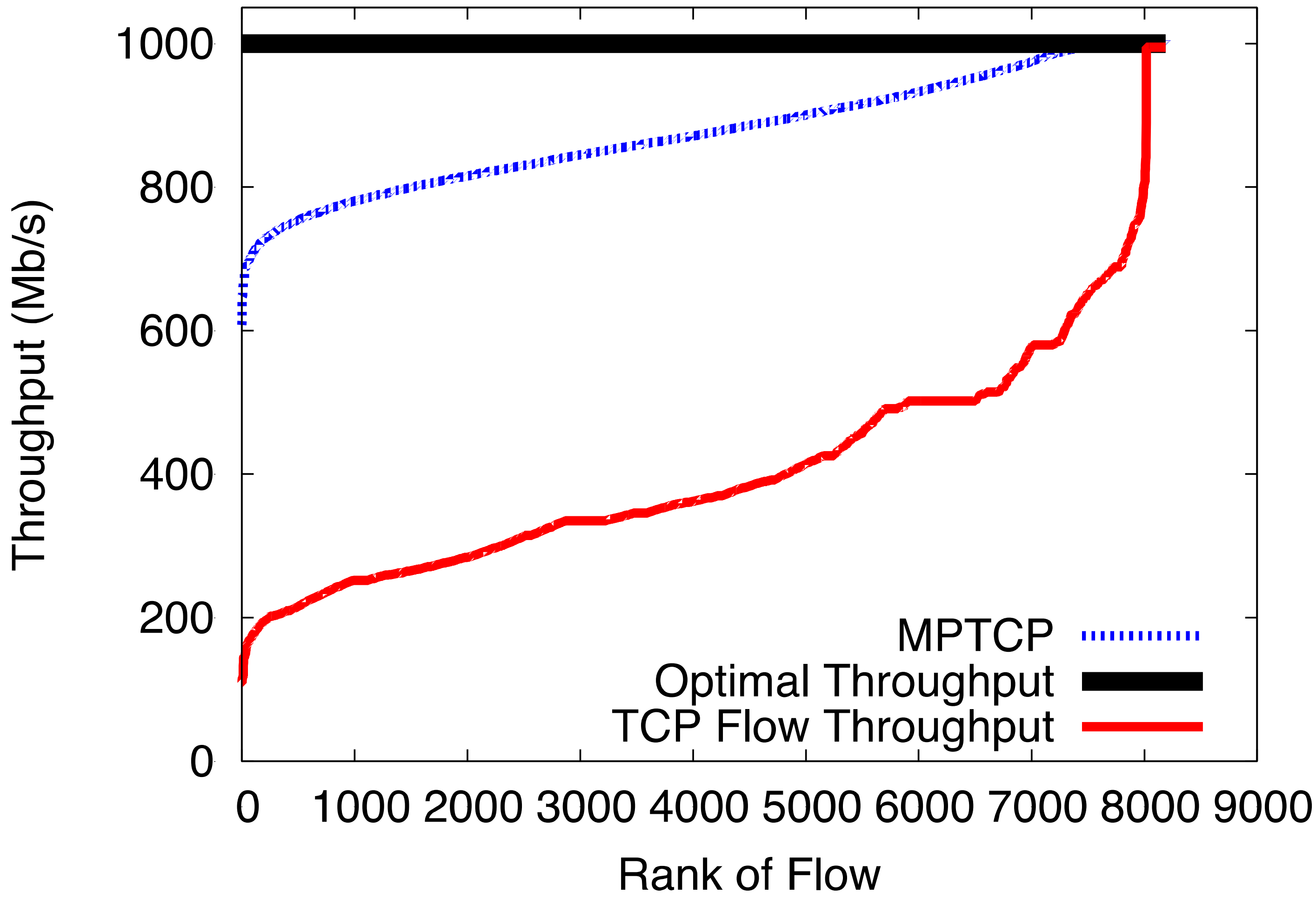


The Multipath TCP way



Improving datacenter performance and robustness with multipath TCP," *ACM*
011.

MPTCP better utilizes the FatTree network



C. Raiciu, et al. "Improving datacenter performance and robustness with multipath TCP," *ACM SIGCOMM* 2011.

The motivations for Multipath TCP

The changing Internet

The Multipath TCP Protocol

Multipath TCP use cases

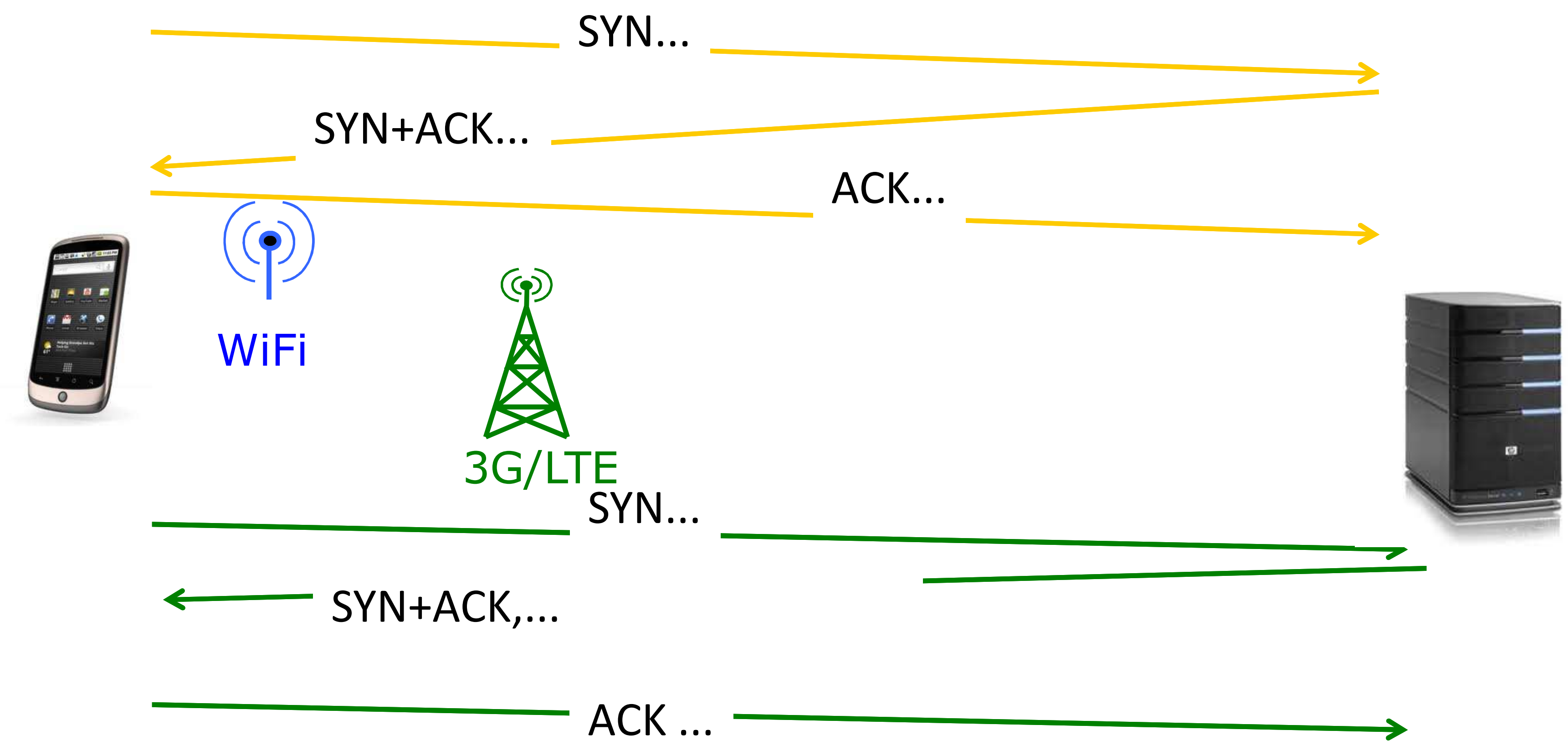
- Datacenters
- Smartphones

How should Multipath TCP use 3G and WiFi ?

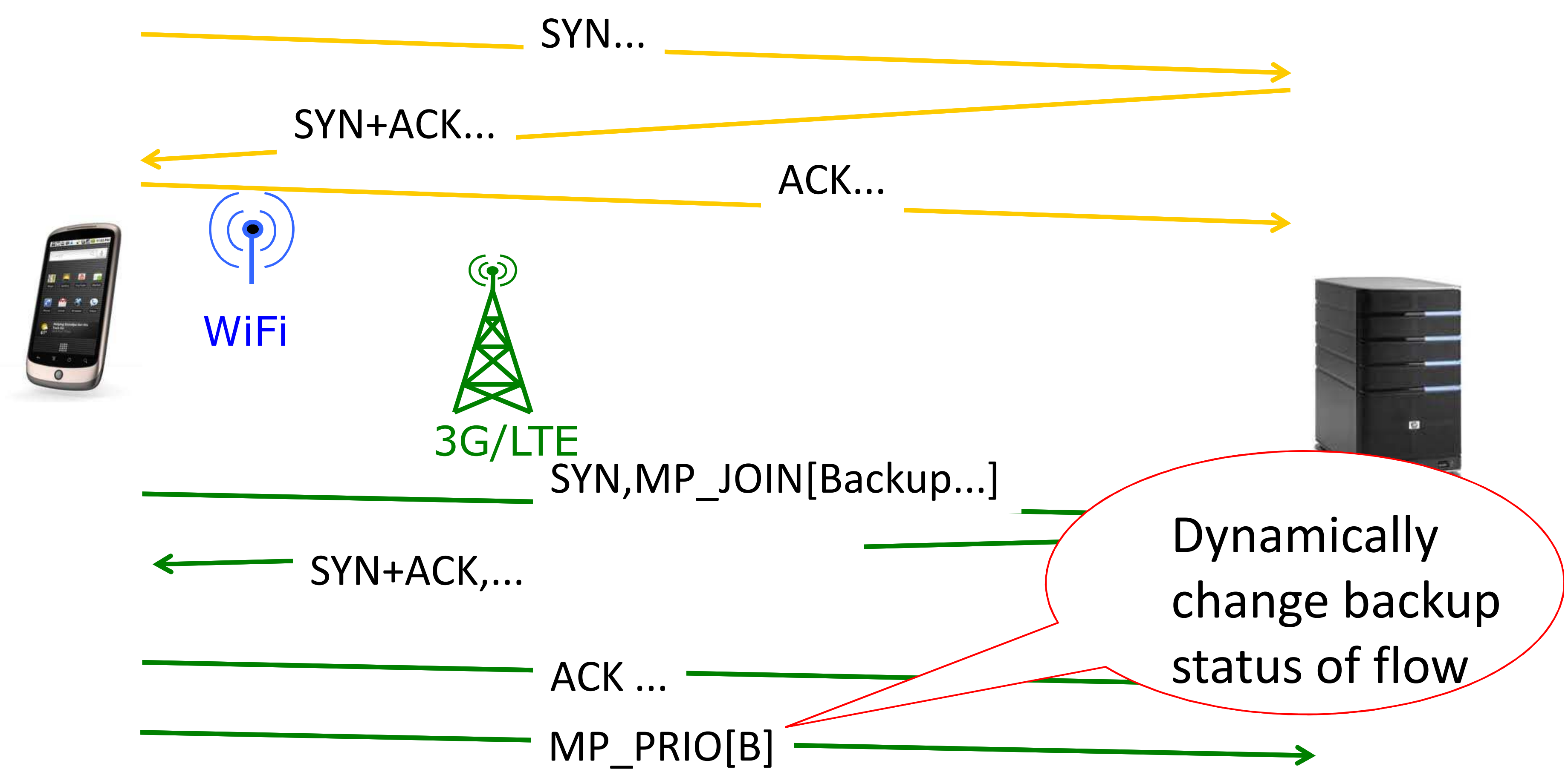
- Full mode
  - Both wireless networks are used at the same time
- Backup mode
  - Prefer WiFi when available, open subflows on 3G and use them as backup
- Single path mode
  - Only one path is used at a time, WiFi preferred over 3G

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Multipath TCP : Full mode

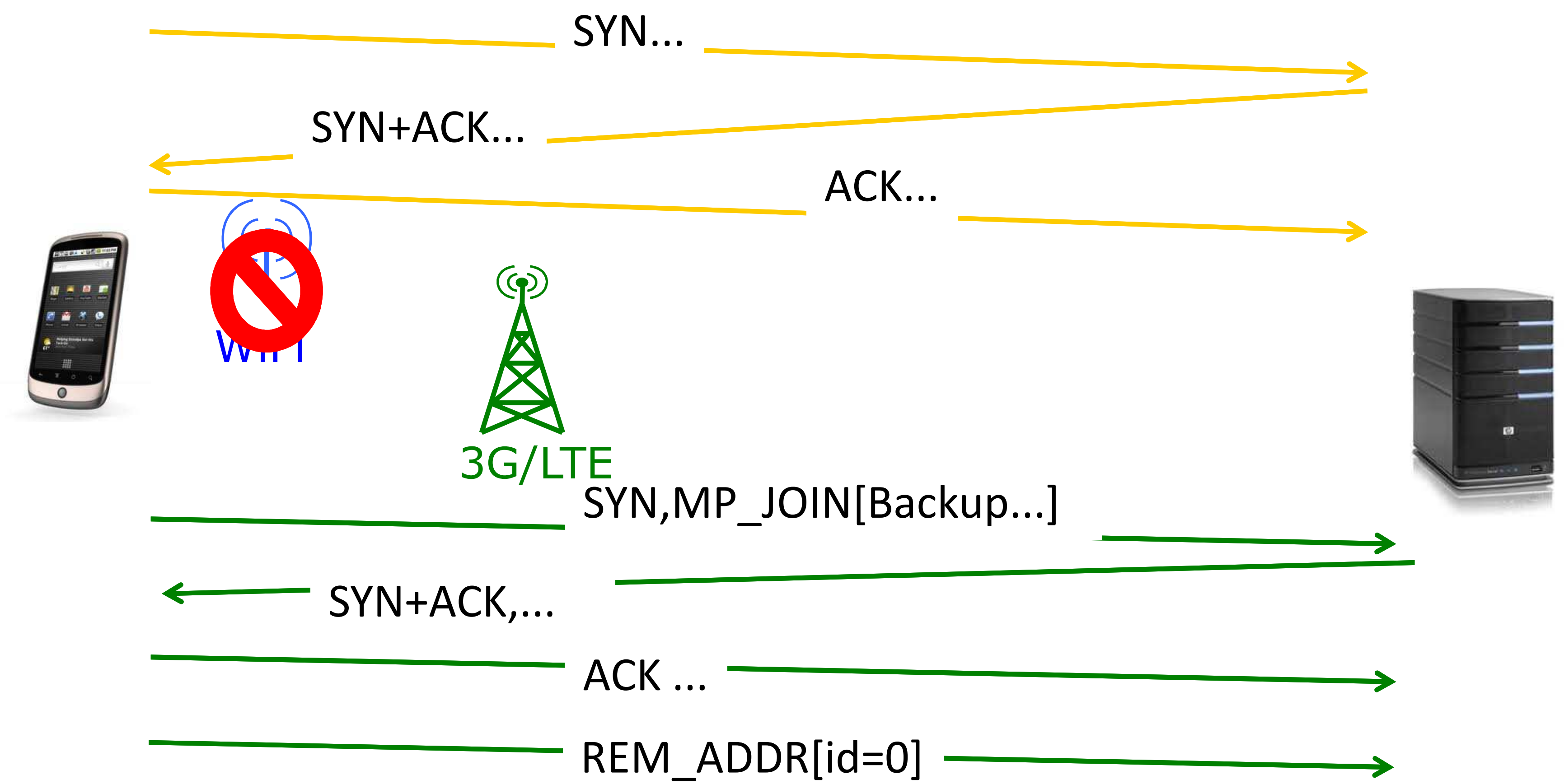


Multipath TCP : Backup mode

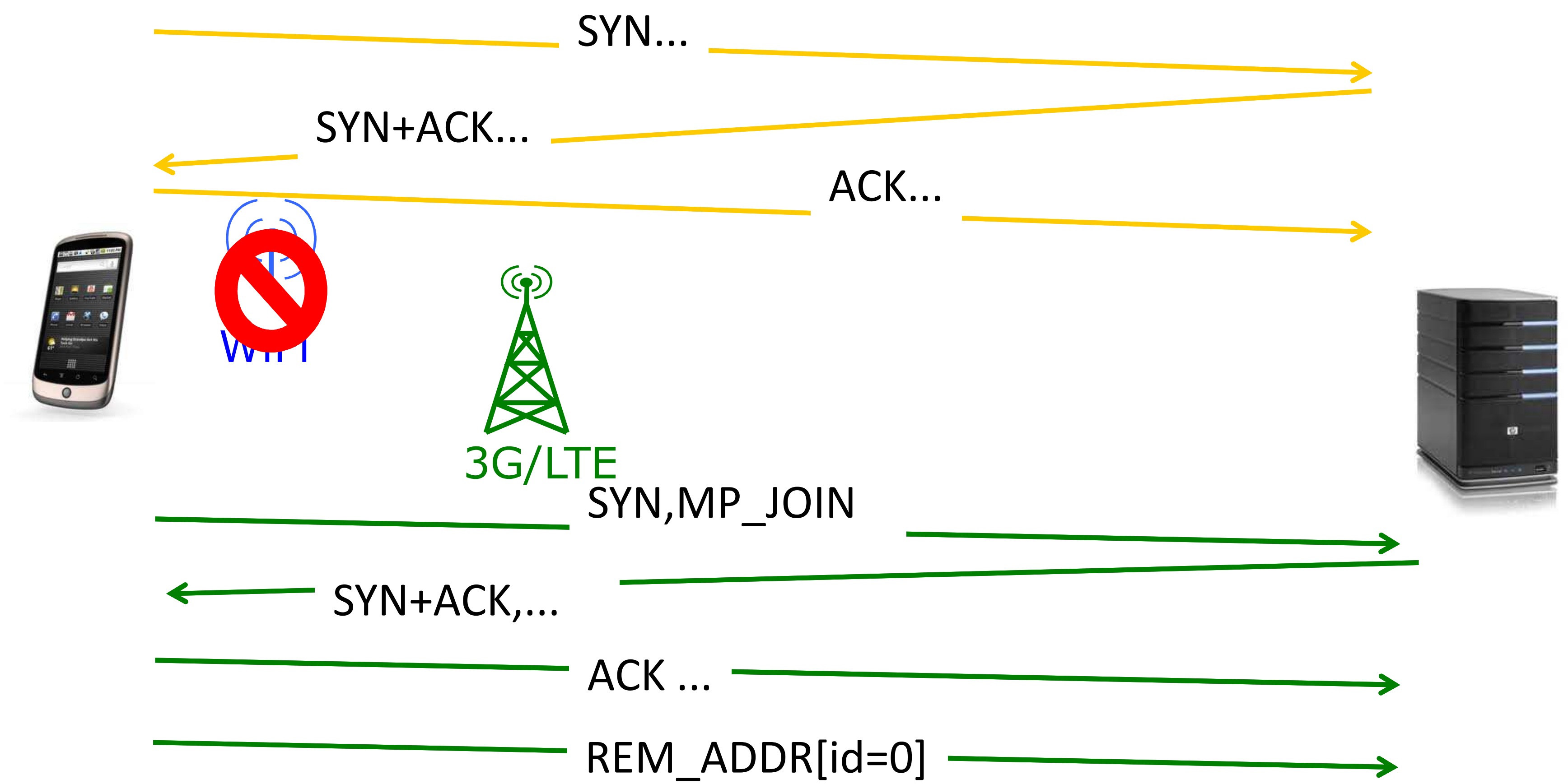


Multipath TCP : Backup mode

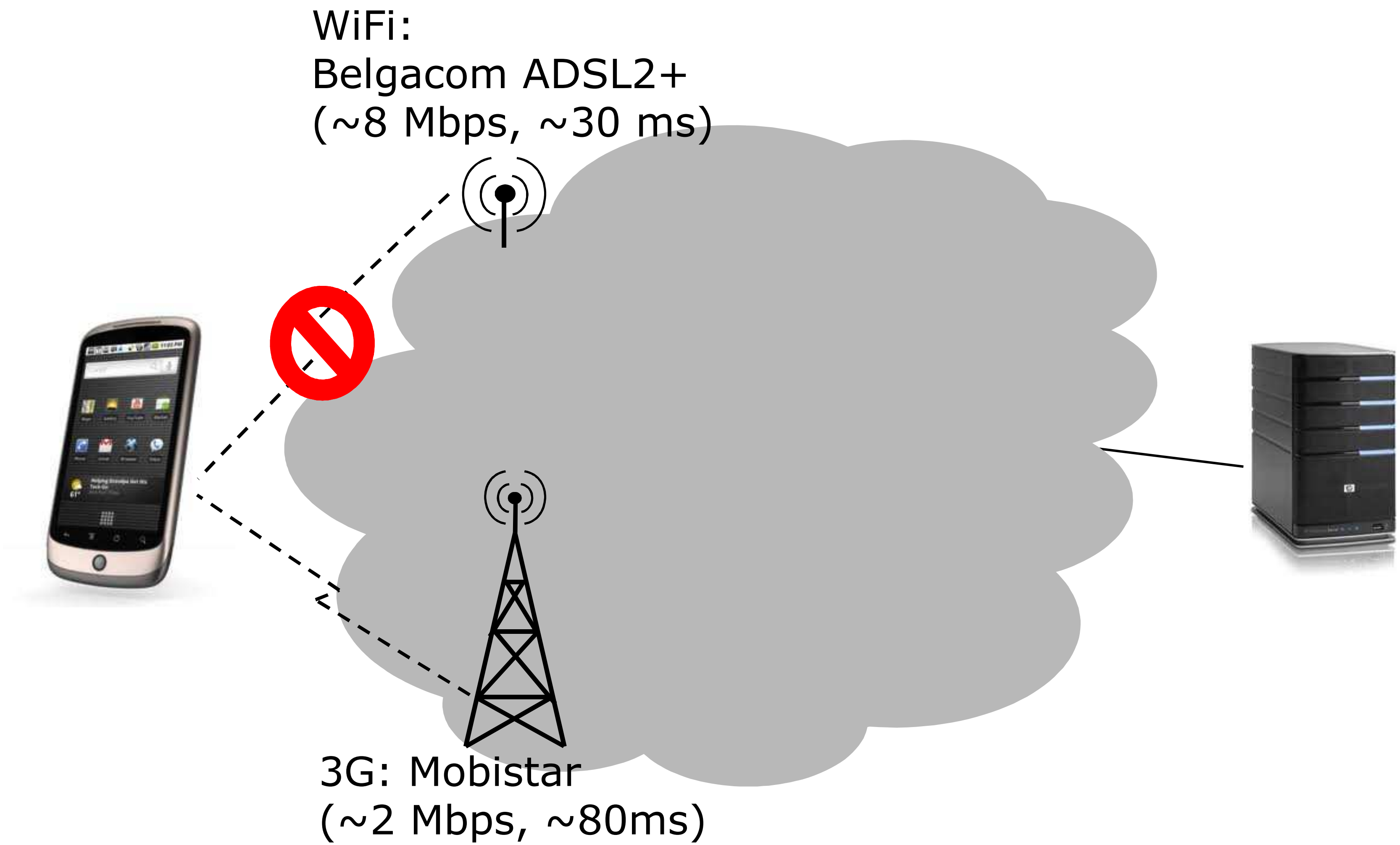
What happens when link fails ?



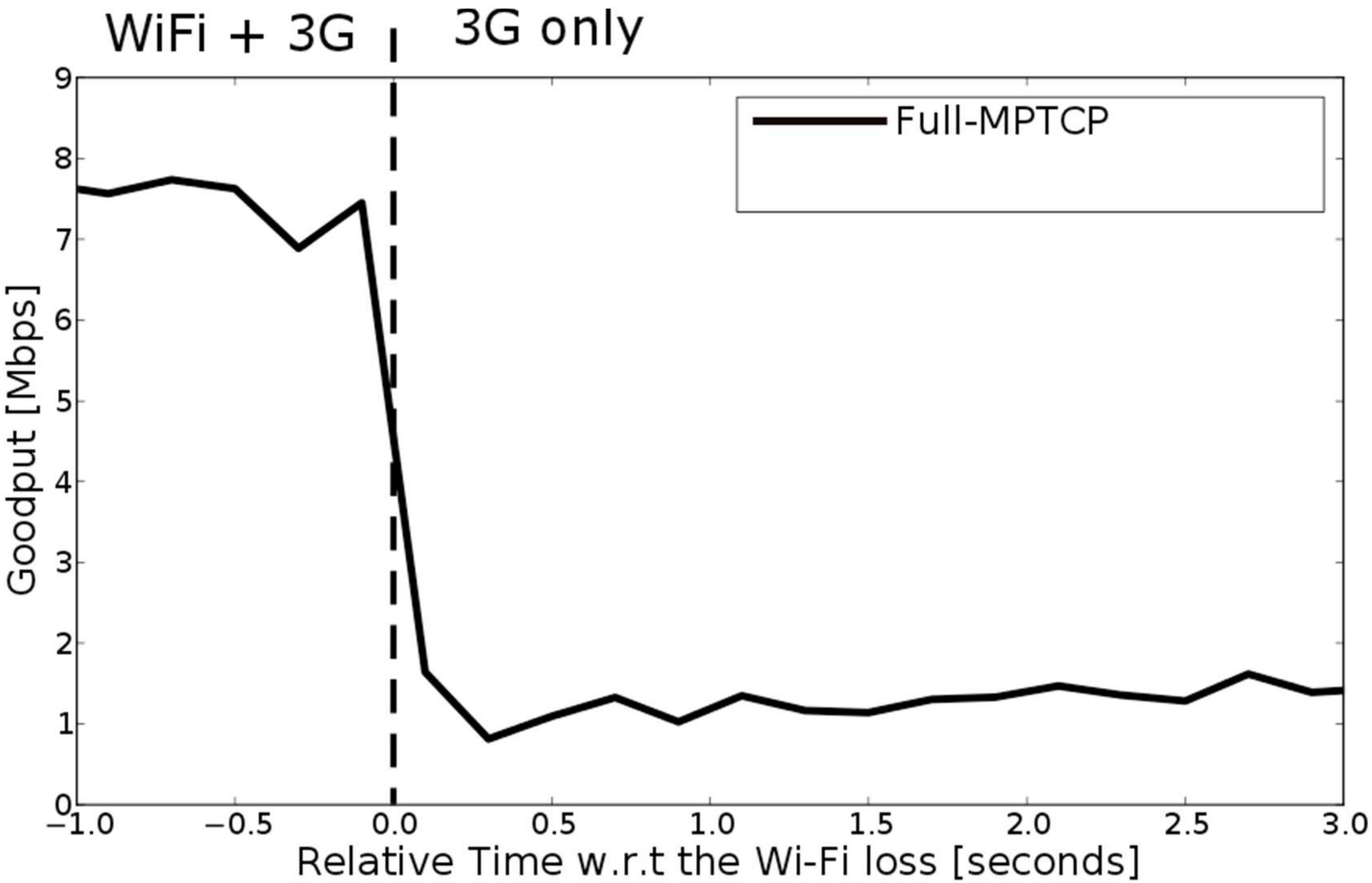
Multipath TCP supports break before make



Evaluation scenario

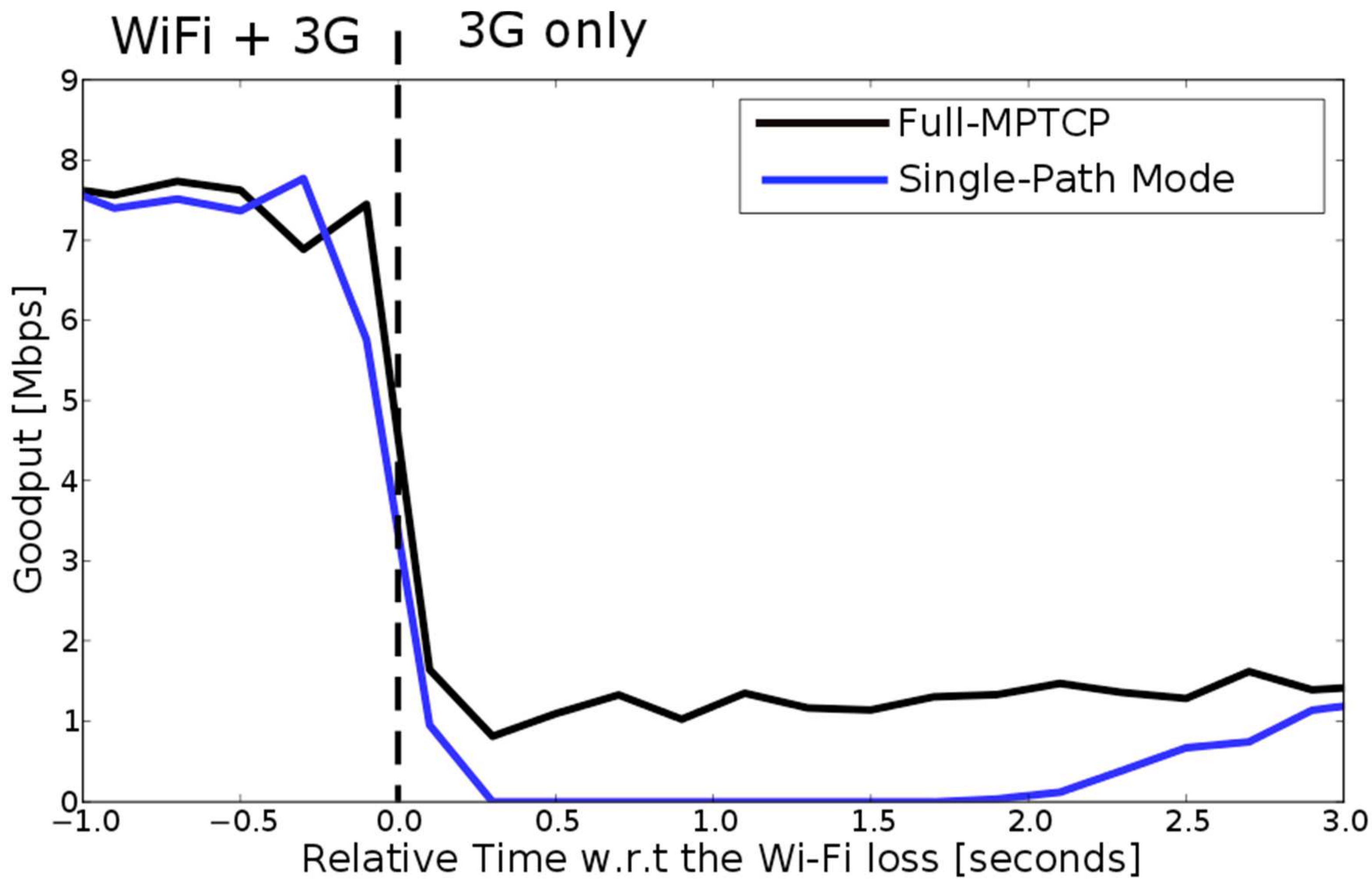


Recovery after failure



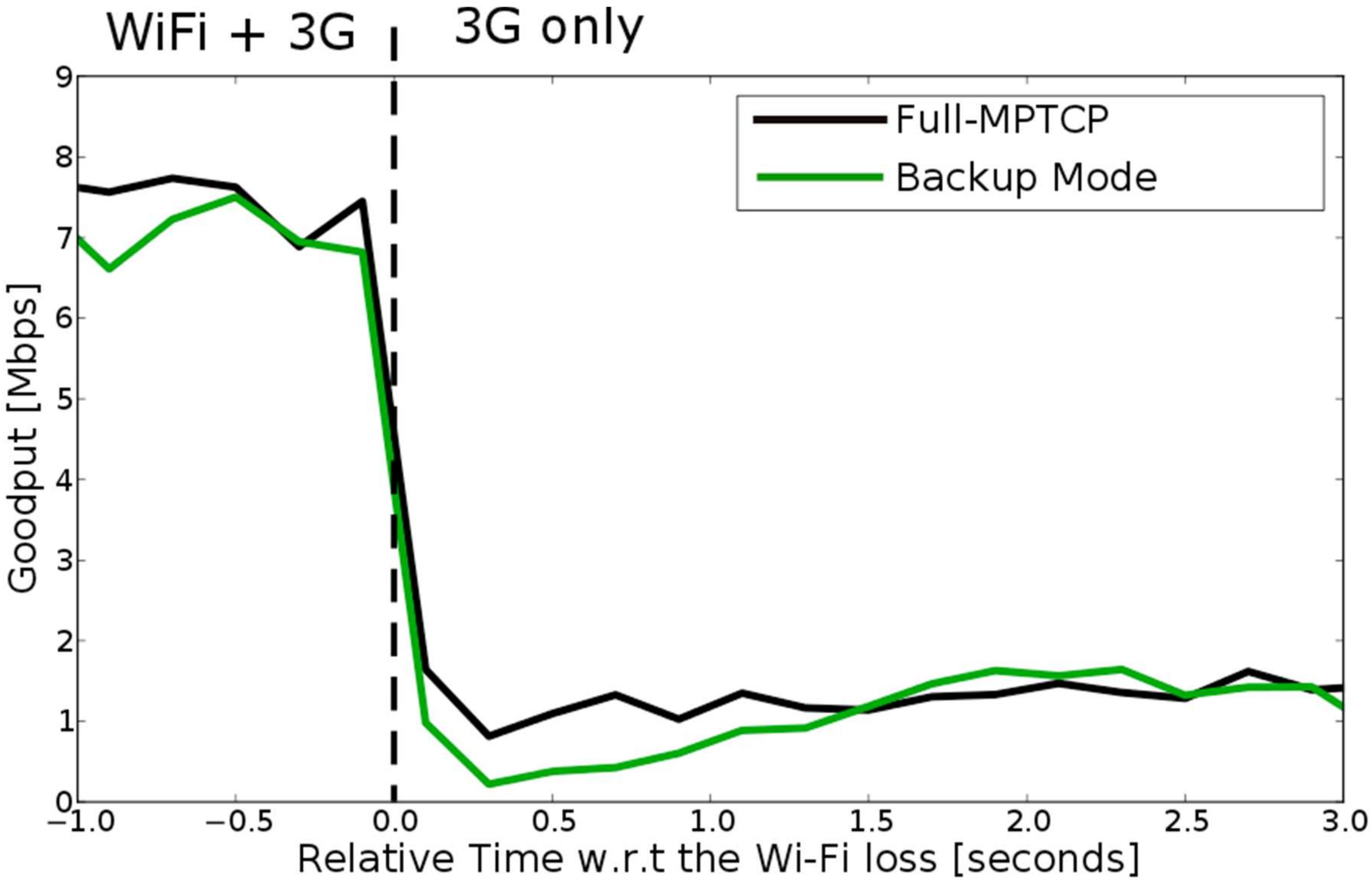
C. Paasch, et al. , "Exploring mobile/WiFi handover with multipath TCP," presented at the CellNet '12: Proceedings of the 2012 ACM SIGCOMM workshop on Cellular networks: operations, challenges, and future design, 2012.

Recovery after failure



C. Paasch, et al. , "Exploring mobile/WiFi handover with multipath TCP," presented at the CellNet '12: Proceedings of the 2012 ACM SIGCOMM workshop on Cellular networks: operations, challenges, and future design, 2012.

Recovery after failure



C. Paasch, et al. , "Exploring mobile/WiFi handover with multipath TCP," presented at the CellNet '12: Proceedings of the 2012 ACM SIGCOMM workshop on Cellular networks: operations, challenges, and future design, 2012.

4 Conclusion

Multipath TCP is becoming a reality

- Due to the middleboxes the protocol is more complex than initially expected
- RFC has been published
- there is running code !
- Multipath TCP works over today's Internet !

What's next ?

- More use cases
  - IPv4/IPv6, anycast, load balancing, deployment
- Measurements and improvements to the protocol
  - Time to revisit 20+ years of heuristics added to TCP

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

TECHNISCHE UNIVERSITÄT DARMSTADT

The Multipath TCP protocol

- http://www.multipath-tcp.org
- http://tools.ietf.org/wg/mptcp/

A. Ford, C. Raiciu, M. Handley, S. Barre, and J. Iyengar, “Architectural guidelines for multipath TCP development", RFC6182 2011.

A. Ford, C. Raiciu, M. J. Handley, and O. Bonaventure, “TCP Extensions for Multipath Operation with Multiple Addresses,” RFC6824, 2013

C. Raiciu, C. Paasch, S. Barre, A. Ford, M. Honda, F. Duchene, O. Bonaventure, and M. Handley, “How hard can it be? designing and implementing a deployable multipath TCP,” NSDI'12: Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation, 2012.

Implementations

Linux

- http://www.multipath-tcp.org

FreeBSD

- http://caia.swin.edu.au/urp/newtcp/mptcp/

Simulators

- http://nrg.cs.ucl.ac.uk/mptcp/implementation.html
- http://code.google.com/p/mptcp-ns3/

Sébastien Barré. Implementation and assessment of Modern Host-based Multipath Solutions. PhD thesis. UCL, 2011

S. Barre, C. Paasch, and O. Bonaventure, "Multipath tcp: From theory to practice," *NETWORKING 2011*, 2011.

Middleboxes

M. Honda, Y. Nishida, C. Raiciu, A. Greenhalgh, M. Handley, and H. Tokuda, “Is it still possible to extend TCP?,” IMC '11: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference, 2011.

V. Sekar, N. Egi, S. Ratnasamy, M. K. Reiter, and G. Shi, “Design and implementation of a consolidated middlebox architecture,” *USENIX NSDI*, 2012.

J. Sherry, S. Hasan, C. Scott, A. Krishnamurthy, S. Ratnasamy, and V. Sekar, “Making middleboxes someone else's problem: network processing as a cloud service,” SIGCOMM '12: Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication, 2012.

Multipath congestion control

- Background

D. Wischik, M. Handley, and M. B. Braun, “The resource pooling principle,” *ACM SIGCOMM Computer …*, vol. 38, no. 5, 2008.

- Coupled congestion control

F. Kelly and T. Voice. Stability of end-to-end algorithms for joint routing and rate control. ACM SIGCOMM CCR, 35, 2005.

P. Key, L. Massoulie, and P. D. Towsley, “Path Selection and Multipath Congestion Control,” INFOCOM 2007. 2007, pp. 143–151.

C. Raiciu, M. J. Handley, and D. Wischik, “Coupled Congestion Control for Multipath Transport Protocols,” *RFC*, vol. 6356, Oct. 2011.

D. Wischik, C. Raiciu, A. Greenhalgh, and M. Handley, “Design, implementation and evaluation of congestion control for multipath TCP,” NSDI'11: Proceedings of the 8th USENIX conference on Networked systems design and implementation, 2011.

Source: Olivier Bonaventure «Decoupling TCP from IP with Multipath TCP»

Multipath congestion control

- More

R. Khalili, N. Gast, M. Popovic, U. Upadhyay, J.-Y. Le Boudec , MPTCP is not Pareto-optimal: Performance issues and a possible solution, Proc. ACM Conext 2012

Y. Cao, X. Mingwei, and X. Fu, “Delay-based Congestion Control for Multipath TCP,” ICNP2012, 2012.

T. A. Le, C. S. Hong, and E.-N. Huh, “Coordinated TCP Westwood congestion control for multiple paths over wireless networks,” ICOIN '12: Proceedings of the The International Conference on Information Network 2012, 2012, pp. 92–96.

T. A. Le, H. Rim, and C. S. Hong, “A Multipath Cubic TCP Congestion Control with Multipath Fast Recovery over High Bandwidth-Delay Product Networks,” *IEICE Transactions*, 2012.

T. Dreibholz, M. Becke, J. Pulinthanath, and E. P. Rathgeb, “Applying TCP-Friendly Congestion Control to Concurrent Multipath Transfer,” Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on, 2010, pp. 312–319.

Use cases

- Datacenter

C. Raiciu, S. Barre, C. Pluntke, A. Greenhalgh, D. Wischik, and M. J. Handley, "Improving datacenter performance and robustness with multipath TCP," *ACM SIGCOMM* 2011.

- Mobile

G. Detal, Ch. Paasch, S. van der Linden, P. Mérindol, G. Avoine, O. Bonaventure, *Revisiting Flow-Based Load Balancing: Stateless Path Selection in Data Center Networks*, to appear in Computer Networks

C. Pluntke, L. Eggert, and N. Kiukkonen, "Saving mobile device energy with multipath TCP," *MobiArch '11: Proceedings of the sixth international workshop on MobiArch*, 2011.

C. Paasch, G. Detal, F. Duchene, C. Raiciu, and O. Bonaventure, "Exploring mobile/WiFi handover with multipath TCP," CellNet '12: Proceedings of the 2012 ACM SIGCOMM workshop on Cellular networks: operations, challenges, and future design, 2012.