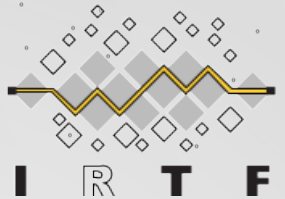
Applied Networking Research Workshop 2020



Evaluating the Impact of Path Brokenness on TCP

Korian Edeline, Benoit Donnet

INTERNET ARCHITECTURAL GUIDELINES

"… there is no architecture, but only a tradition, which was not written down for the first 25 years …"

UNIFORM OPERATIONS

"…in an ideal situation there should be **one**, and only one, **protocol at the Internet level**[…]but there can be a need for gradual transition from one version of IP to another…"

"…the community believes that the goal is connectivity, the tool is the Internet Protocol, and **the intelligence is end to end** rather than hidden in the network."

END-TO-END

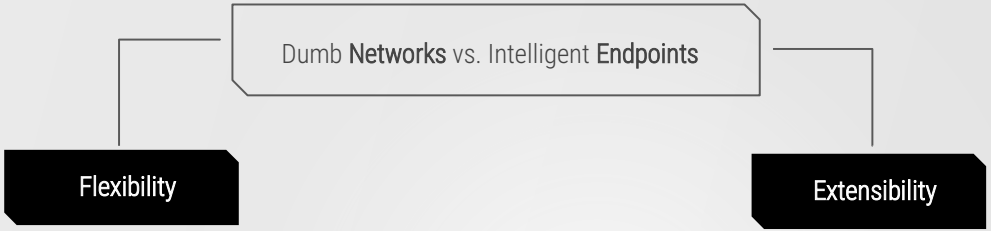
NO GLOBAL CONTROL

"…nobody owns the Internet, **there is no centralized control**[…]. Its evolution depends on rough consensus about technical proposals, and on running code."

RFC 1958

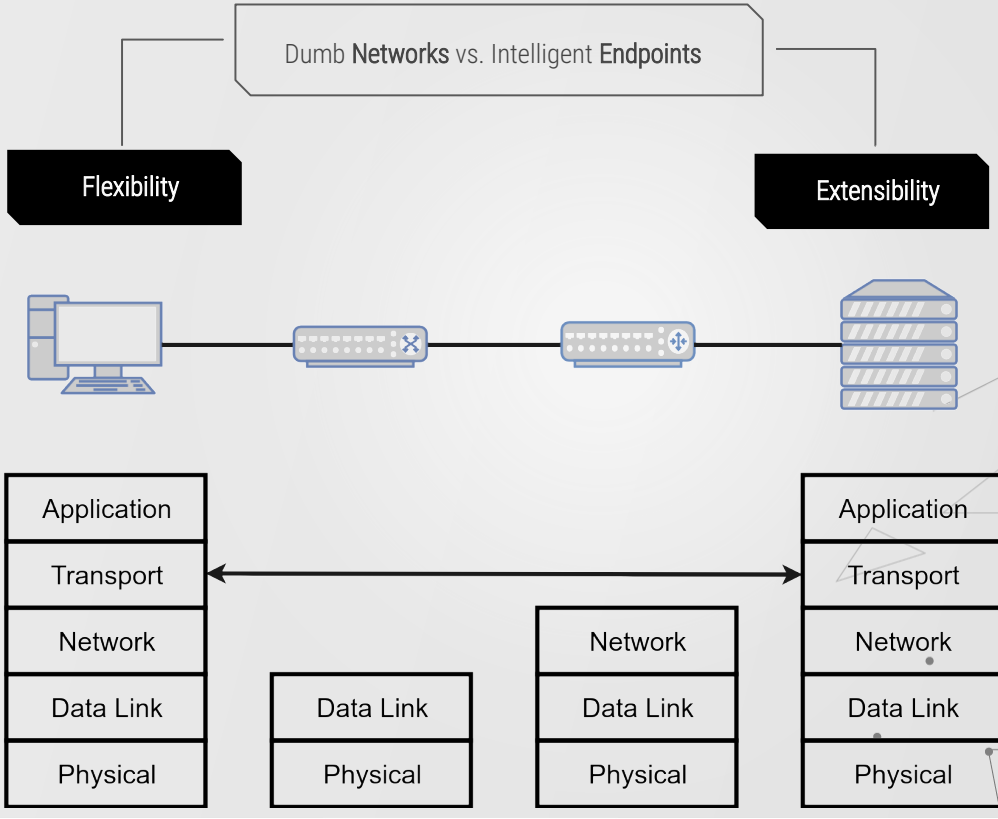
THE END-TO-END PRINCIPLE

End-to-End Arguments in System Design, Saltzer, Reed & Clark, 1981



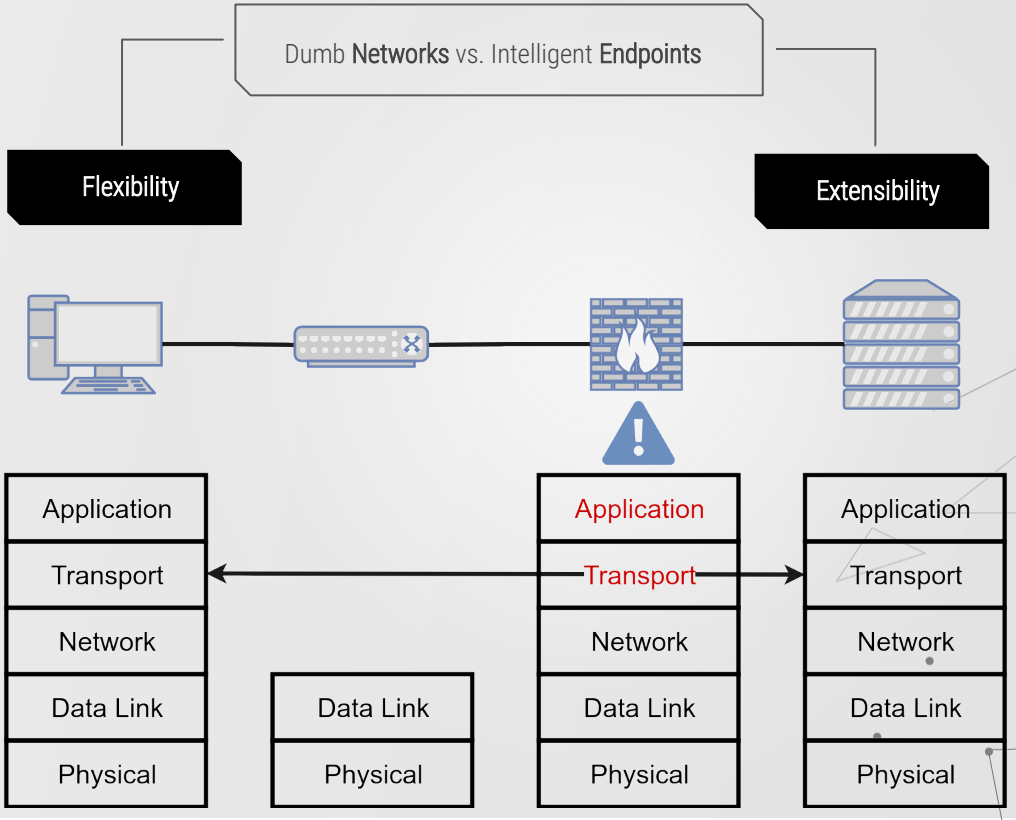
THE END-TO-END PRINCIPLE

End-to-End Arguments in System Design, Saltzer, Reed & Clark, 1981



THE END-TO-END PRINCIPLE

End-to-End Arguments in System Design, Saltzer, Reed & Clark, 1981



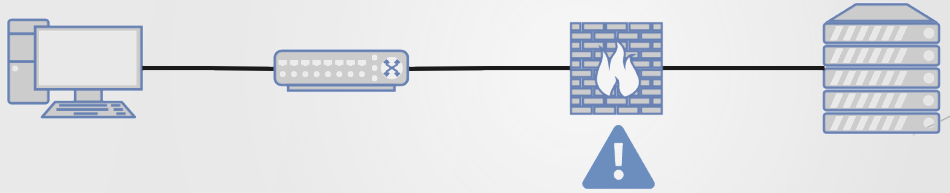
THE END-TO-END PRINCIPLE

End-to-End Arguments in System Design, Saltzer, Reed & Clark, 1981

Dumb **Networks** vs. Intelligent **Endpoints**

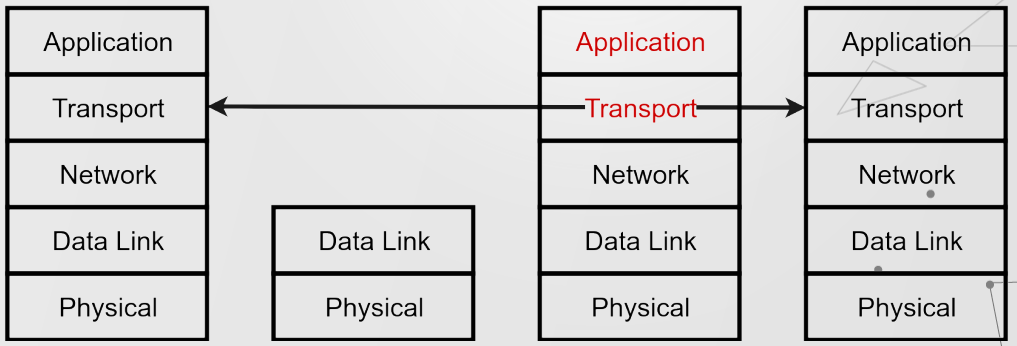
Flexibility

Extensibility



A **middlebox** is a computer networking device that transforms, inspects, filters, or otherwise manipulates traffic for purposes other than packet forwarding.

RFC 3234



THE END-TO-END PRINCIPLE

End-to-End Arguments in System Design, Saltzer, Reed & Clark, 1981

Dumb **Networks** vs. Intelligent **Endpoints**

Flexibility

Extensibility

Ossification

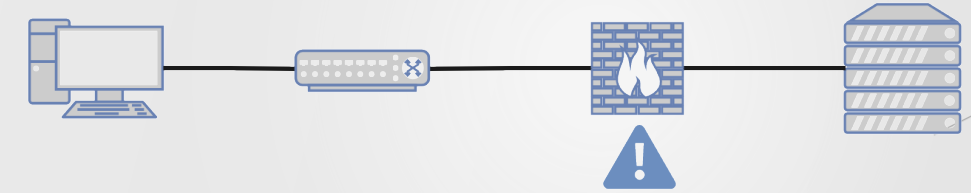
of the network infrastructure

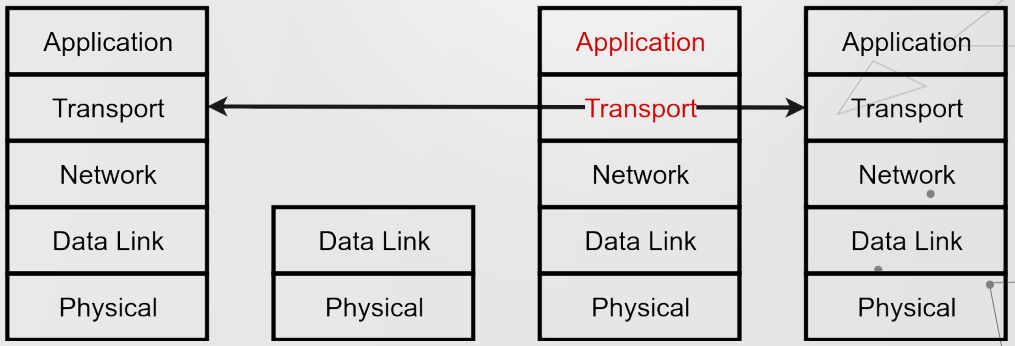
Innovation

Deployment of new TCP features and new transport protocols is crippled by middleboxes

A **middlebox** is a computer networking device that transforms, inspects, filters, or otherwise manipulates traffic for purposes other than packet forwarding.

RFC 3234





BACKGROUND

The Middlebox Problem

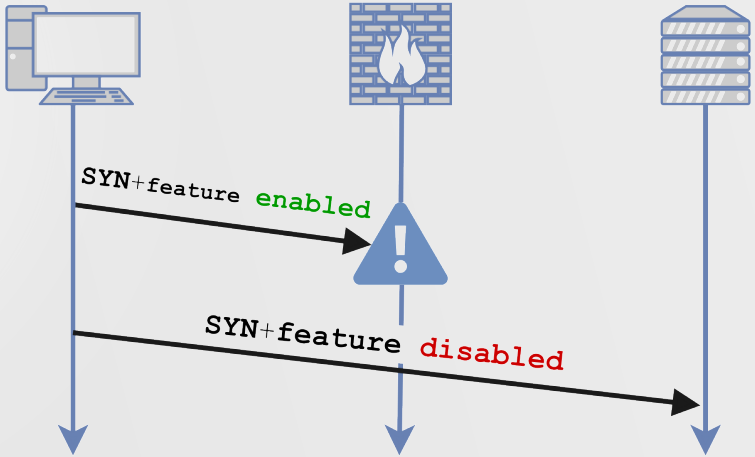
FROM PATH CONDITIONS ...

A **path condition** is a functional description of an action performed by an intermediary device on a packet, on a given path.

FROM PATH CONDITIONS ...

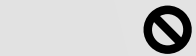
A **path condition** is a functional description of an action performed by an intermediary device on a packet, on a given path.

`feature.blocked`

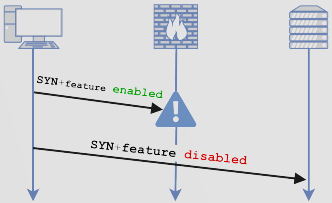


FROM PATH CONDITIONS ...

A **path condition** is a functional description of an action performed by an intermediary device on a packet, on a given path.



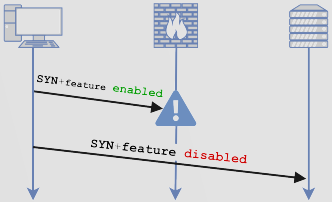
`feature.blocked`

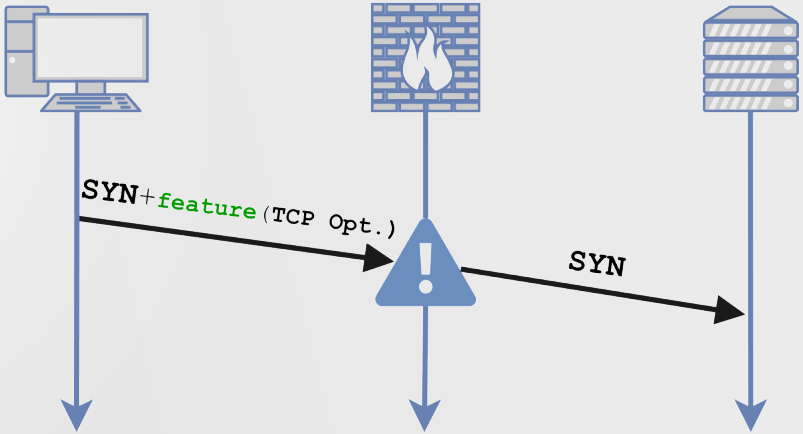


FROM PATH CONDITIONS ...

A **path condition** is a functional description of an action performed by an intermediary device on a packet, on a given path.

`feature.blocked` `feature.removed`

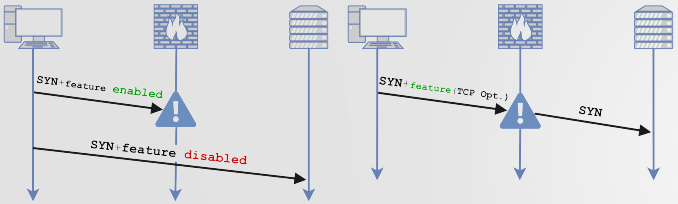




FROM PATH CONDITIONS ...

A **path condition** is a functional description of an action performed by an intermediary device on a packet, on a given path.

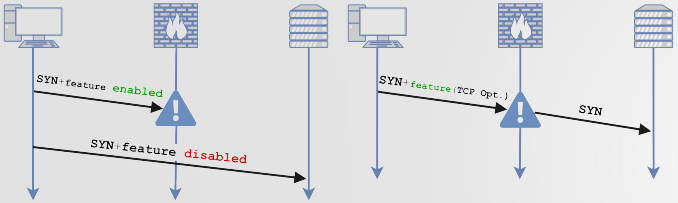
`feature.blocked` `feature.removed`

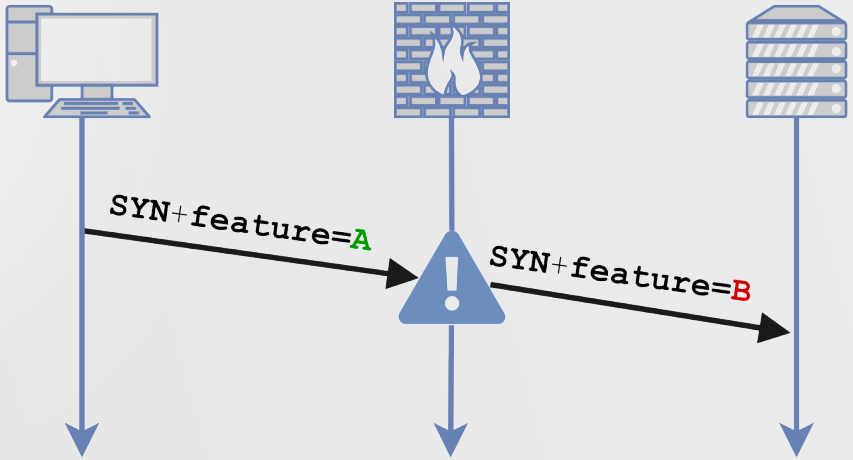


FROM PATH CONDITIONS ...

A **path condition** is a functional description of an action performed by an intermediary device on a packet, on a given path.

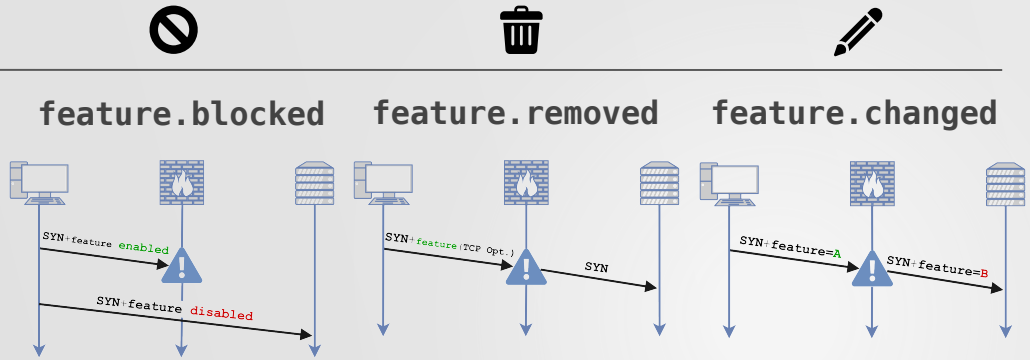
`feature.blocked` `feature.removed` `feature.changed`





FROM PATH CONDITIONS ...

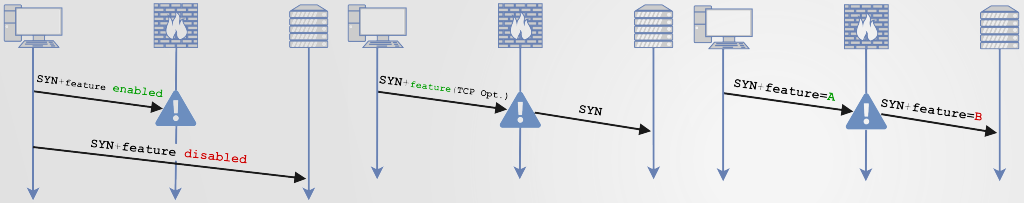
A **path condition** is a functional description of an action performed by an intermediary device on a packet, on a given path.



FROM PATH CONDITIONS ...

A **path condition** is a functional description of an action performed by an intermediary device on a packet, on a given path.

`feature.blocked` `feature.removed` `feature.changed`



WHY?

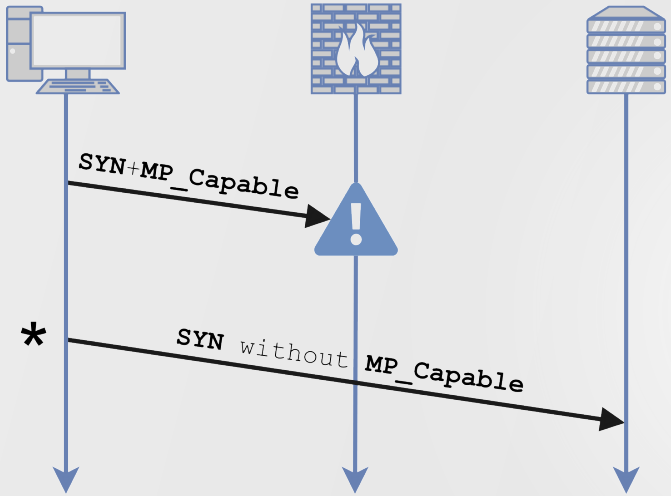
SECURITY **PERFORMANCE** **PACKET MARKING**

… TO PATH IMPAIRMENTS

A **path impairment** is a middlebox-induced connectivity issue, a decrease, or shortfall of Quality-of-Service (QoS), on a given path

... TO PATH IMPAIRMENTS

A **path impairment** is a middlebox-induced connectivity issue, a decrease, or shortfall of Quality-of-Service (QoS), on a given path

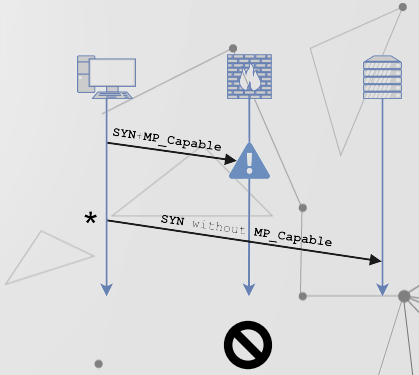


Feature not supported
TCP Normalization

BLOCKED TRAFFIC

... TO PATH IMPAIRMENTS

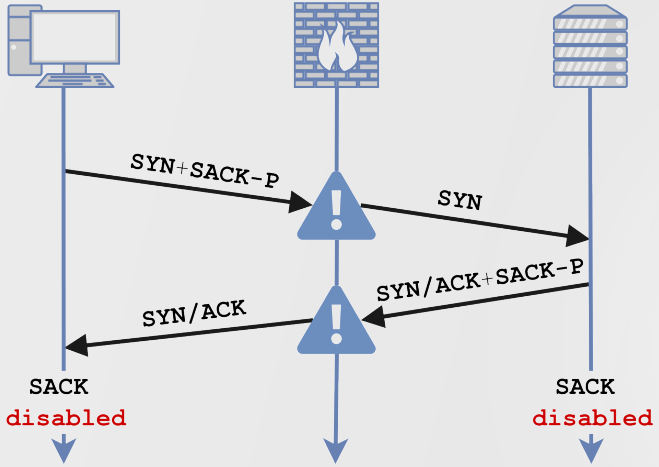
A **path impairment** is a middlebox-induced connectivity issue, a decrease, or shortfall of Quality-of-Service (QoS), on a given path



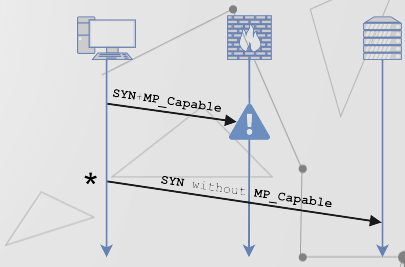
BLOCKED TRAFFIC

... TO PATH IMPAIRMENTS

A **path impairment** is a middlebox-induced connectivity issue, a decrease, or shortfall of Quality-of-Service (QoS), on a given path



TCP Normalization

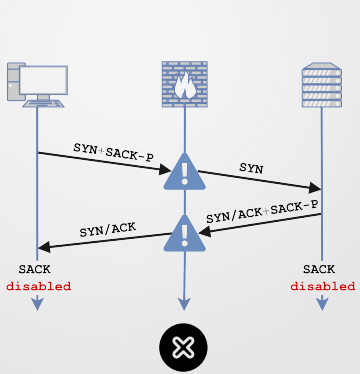


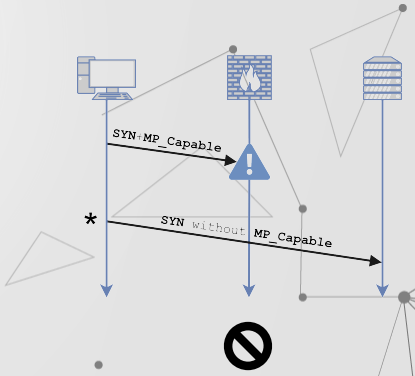
DISABLED FEATURE

BLOCKED TRAFFIC

... TO PATH IMPAIRMENTS

A **path impairment** is a middlebox-induced connectivity issue, a decrease, or shortfall of Quality-of-Service (QoS), on a given path



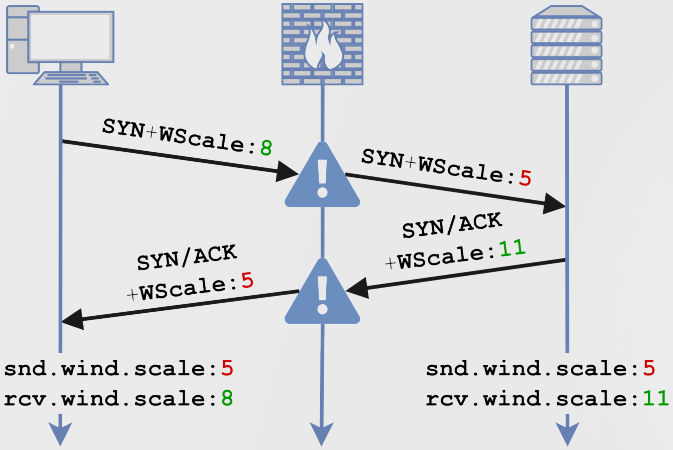


DISABLED FEATURE

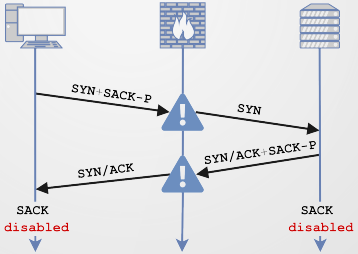
BLOCKED TRAFFIC

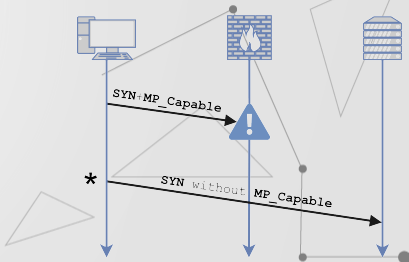
... TO PATH IMPAIRMENTS

A **path impairment** is a middlebox-induced connectivity issue, a decrease, or shortfall of Quality-of-Service (QoS), on a given path



Malconfiguration
Rate Limiting





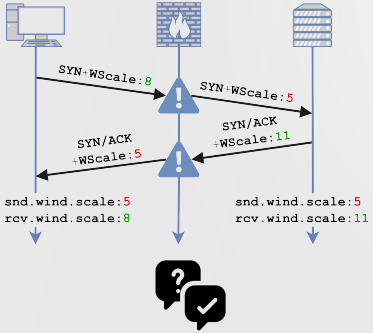
NEGOTIATION DISRUPTION

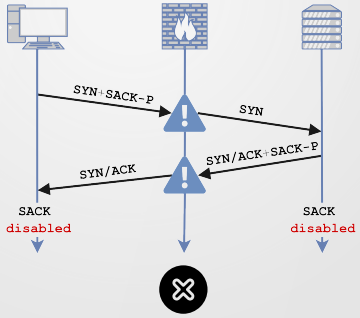
DISABLED FEATURE

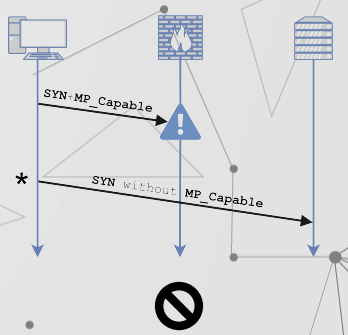
BLOCKED TRAFFIC

... TO PATH IMPAIRMENTS

A **path impairment** is a middlebox-induced connectivity issue, a decrease, or shortfall of Quality-of-Service (QoS), on a given path







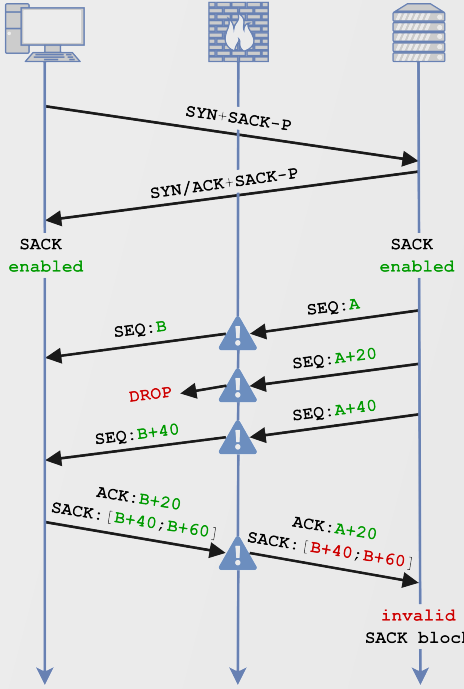
NEGOTIATION DISRUPTION

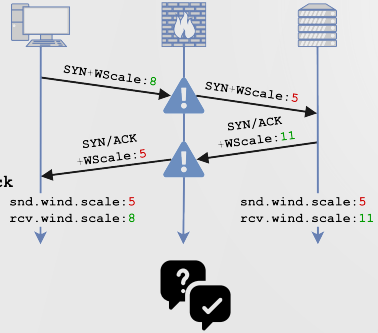
DISABLED FEATURE

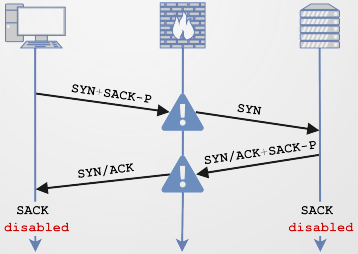
BLOCKED TRAFFIC

... TO PATH IMPAIRMENTS

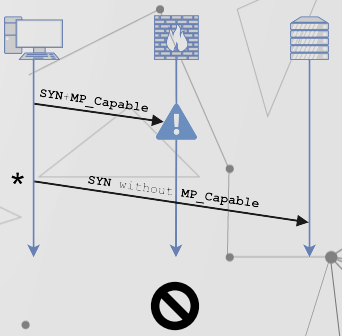
A **path impairment** is a middlebox-induced connectivity issue, a decrease, or shortfall of Quality-of-Service (QoS), on a given path







Fix TCP ISN Randomness



TRAFFIC DISRUPTION

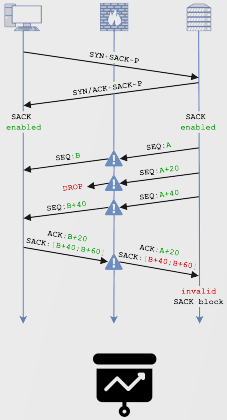
NEGOTIATION DISRUPTION

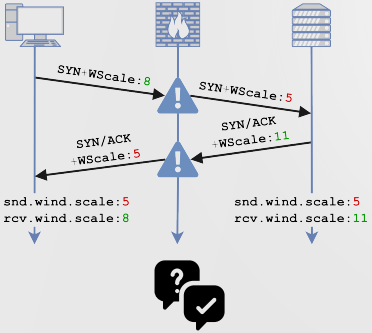
DISABLED FEATURE

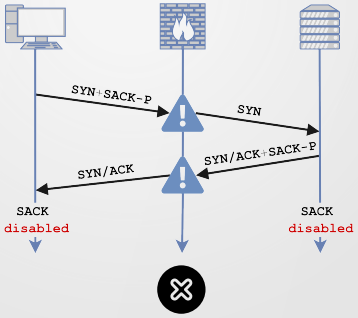
BLOCKED TRAFFIC

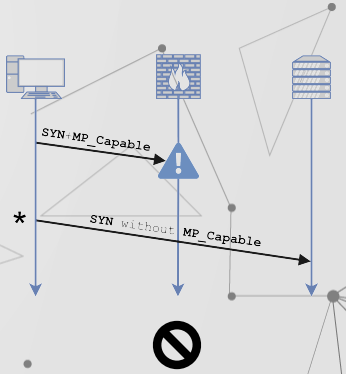
... TO PATH IMPAIRMENTS

A **path impairment** is a middlebox-induced connectivity issue, a decrease, or shortfall of Quality-of-Service (QoS), on a given path









TRAFFIC DISRUPTION

NEGOTIATION DISRUPTION

DISABLED FEATURE

BLOCKED TRAFFIC

Path impairments in the wild

2% of deployed network devices are TCP/IP MBs

38.9% of networks paths are crossing middleboxes

6.5% of network paths are affected by TCP-breaking MBs

A bottom-up investigation of the transport-layer ossification, K.Edeline and B.Donnet, in Proc. IFIP Network Traffic Measurement and Analysis Conference (TMA), 2019

EXPERIMENTATION

Path Impairments Investigation

mmb: A VPP MIDDLEBOX

https://github.com/mami-project/vpp-mb

mmb: A VPP MIDDLEBOX

VPP (Vector Packet Processing) is a high-performance kernel - bypass framework developped by Cisco.

https://github.com/mami-project/vpp-mb

mmb: A VPP MIDDLEBOX

Flexible

Intuitive

Fast

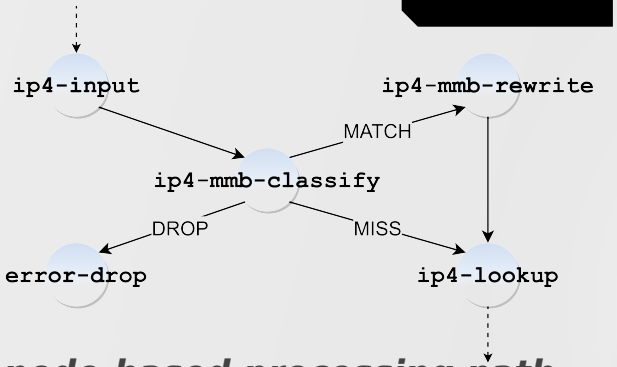
VPP (Vector Packet Processing) is a high-performance kernel - bypass framework developped by Cisco.

https://github.com/mami-project/vpp-mb

mmb: A VPP MIDDLEBOX

Flexible **Intuitive** **Fast**

VPP (Vector Packet Processing) is a high-performance kernel-bypass framework developped by Cisco.

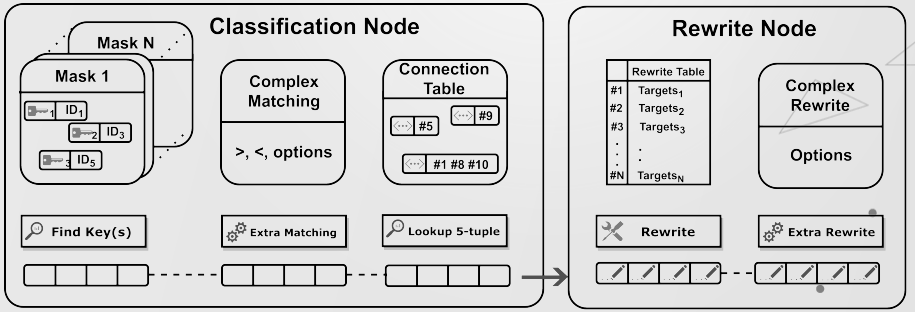


node-based processing path

```
# mmb <add-keyword> <match>  [<match>  ...  <match>]
                    <target> [<target> ... <target>]

<add-keyword> : add-stateless | add-stateful
<match>       : <field> <condition> <value>
<target>      : mod <field> <value> | add <field> <value>
                | strip [!] <field>  | map <field> <value>
                | shuffle <field>    | drop
```

CLI syntax



mmb software architecture

https://github.com/mami-project/vpp-mb

IMPACT OF PATH IMPAIRMENTS ON TCP

ECN	SACK	WScale
Explicit Congestion Notification	Selective Acknowledgment	Window Scaling Parameter

IMPACT OF PATH IMPAIRMENTS ON TCP

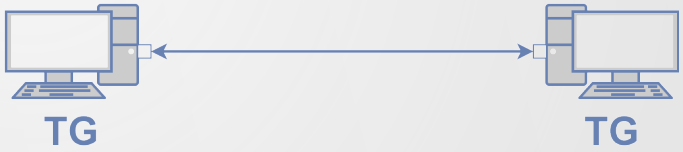
ECN	SACK	WScale
Explicit Congestion Notification	Selective Acknowledgment	Window Scaling Parameter

How do path-impairing middleboxes affect the Quality-of-Service of TCP ?

IMPACT OF PATH IMPAIRMENTS ON TCP

ECN	SACK	WScale
Explicit Congestion Notification	Selective Acknowledgment	Window Scaling Parameter

How do path-impairing middleboxes affect the Quality-of-Service of TCP ?



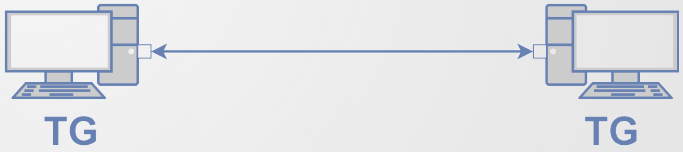
Direct

Indirect

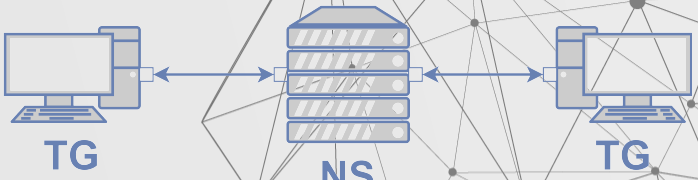
IMPACT OF PATH IMPAIRMENTS ON TCP

ECN	SACK	WScale
Explicit Congestion Notification	Selective Acknowledgment	Window Scaling Parameter

How do path-impairing middleboxes affect the Quality-of-Service of TCP ?



Direct



Indirect

LFNs ⚙ **Delay** ⚙ **Loss** ⚙ **Congestion**

IMPACT OF PATH IMPAIRMENTS ON TCP

ECN SACK WScale

Explicit Congestion Notification

Selective Acknowledgment

Window Scaling Parameter

	ECN IP bits
Not-ECT	00
ECT(1)	01
ECT(0)	10
CE	11

Path impairments scenarios without congestion:

IMPACT OF PATH IMPAIRMENTS ON TCP

	ECN IP bits
Not-ECT	00
ECT(1)	01
ECT(0)	10
CE	11

ECN

Explicit Congestion Notification

SACK

Selective Acknowledgment

WScale

Window Scaling Parameter

Path impairments scenarios without congestion:

Disabled ECN

`ip.ecn.changed.11`
with fallback mechanism

Blocked ECN

`tcp.ecn.blocked`

Broken ECN

`ip.ecn.changed.11`
fallback-proof

IMPACT OF PATH IMPAIRMENTS ON TCP

ECN

Explicit Congestion Notification

SACK

Selective Acknowledgment

WScale

Window Scaling Parameter

Path impairments scenarios without congestion:

Disabled ECN

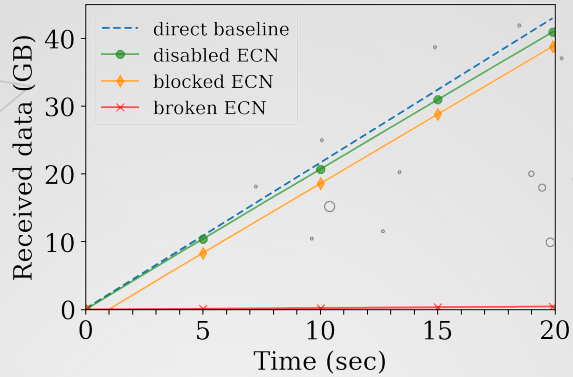
`ip.ecn.changed.11`
with fallback mechanism

Blocked ECN

`tcp.ecn.blocked`

Broken ECN

`ip.ecn.changed.11`
fallback-proof



IMPACT OF PATH IMPAIRMENTS ON TCP

ECN
Explicit Congestion Notification

SACK
Selective Acknowledgment

WScale
Window Scaling Parameter

Path impairments scenarios without congestion:

Disabled ECN
`ip.ecn.changed.11`
with fallback mechanism

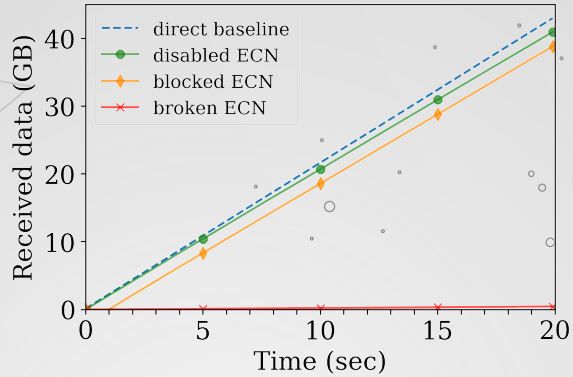
Blocked ECN
`tcp.ecn.blocked`

Broken ECN
`ip.ecn.changed.11`
fallback-proof

Path impairments scenarios with congestion:

Enabled ECN

Disabled ECN



IMPACT OF PATH IMPAIRMENTS ON TCP

ECN — Explicit Congestion Notification

SACK — Selective Acknowledgment

WScale — Window Scaling Parameter

Path impairments scenarios without congestion:

Disabled ECN
`ip.ecn.changed.11`
with fallback mechanism

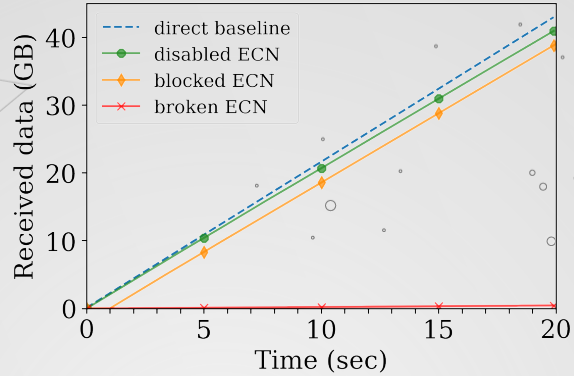
Blocked ECN
`tcp.ecn.blocked`

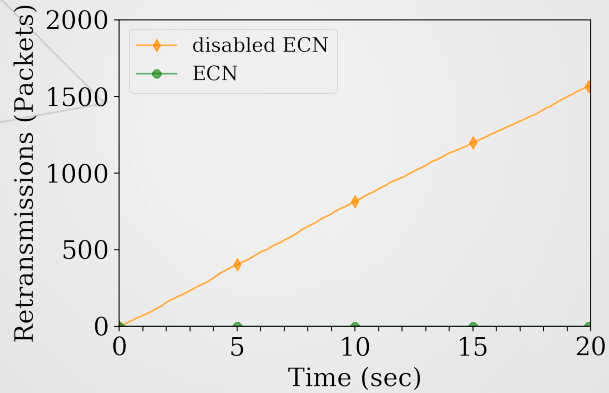
Broken ECN
`ip.ecn.changed.11`
fallback-proof

Path impairments scenarios with congestion:

Enabled ECN

Disabled ECN





IMPACT OF PATH IMPAIRMENTS ON TCP

ECN

Explicit Congestion Notification

SACK

Selective Acknowledgment

WScale

Window Scaling Parameter

Path impairments scenarios without congestion:

Disabled ECN
`ip.ecn.changed.11`
with fallback mechanism

Blocked ECN
`tcp.ecn.blocked`

Broken ECN
`ip.ecn.changed.11`
fallback-proof

Path impairments scenarios with congestion:

Enabled ECN

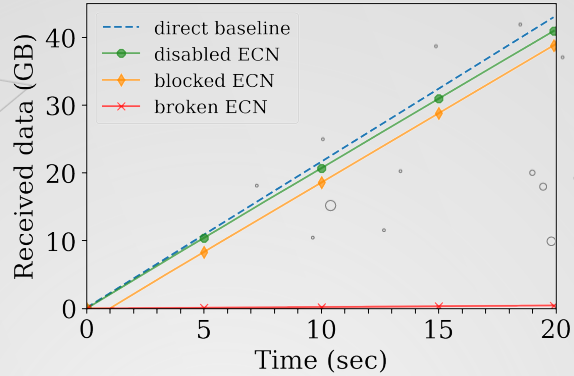
Disabled ECN

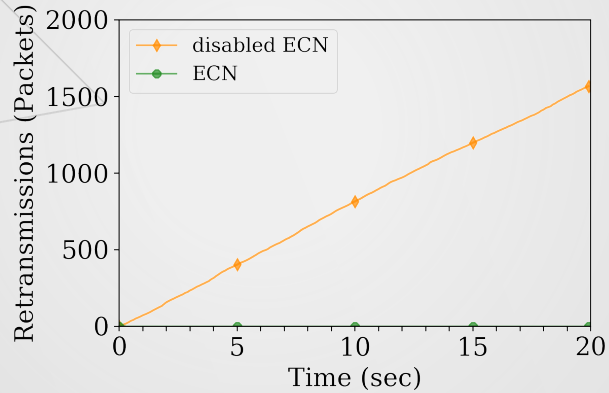
Broken ECN

Broken ECN slows down the connection to 1 MSS / RTT

Disabling ECN

Disabling ECN per default is not the solution





IMPACT OF PATH IMPAIRMENTS ON TCP

ECN

Explicit Congestion Notification

SACK

Selective Acknowledgment

WScale

Window Scaling Parameter

IMPACT OF PATH IMPAIRMENTS ON TCP

ECN — Explicit Congestion Notification

SACK — Selective Acknowledgment

WScale — Window Scaling Parameter

Path impairments scenarios with artificial loss:

Enabled SACK

SACK-enabled flows

Disabled SACK

`tcp.opt.sackok.removed`

Broken SACK

`tcp.seqnum.changed`

IMPACT OF PATH IMPAIRMENTS ON TCP

~~ECN~~ **SACK** ~~WScale~~

~~Explicit Congestion Notification~~ Selective Acknowledgment ~~Window Scaling Parameter~~

Path impairments scenarios with artificial loss:

Enabled SACK

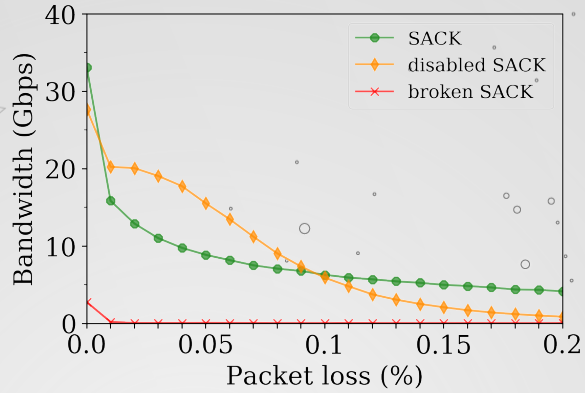
SACK-enabled flows

Disabled SACK

`tcp.opt.sackok.removed`

Broken SACK

`tcp.seqnum.changed`



IMPACT OF PATH IMPAIRMENTS ON TCP

ECN — Explicit Congestion Notification

SACK — Selective Acknowledgment

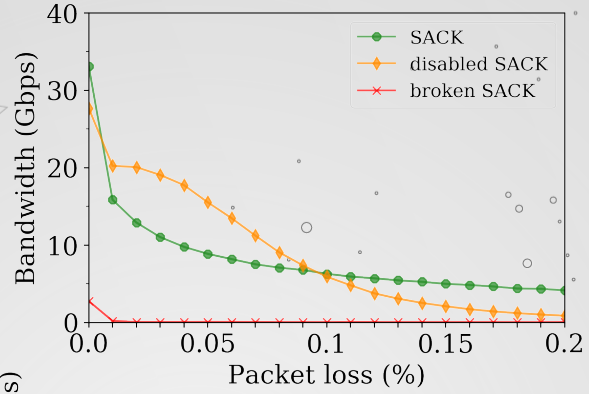
WScale — Window Scaling Parameter

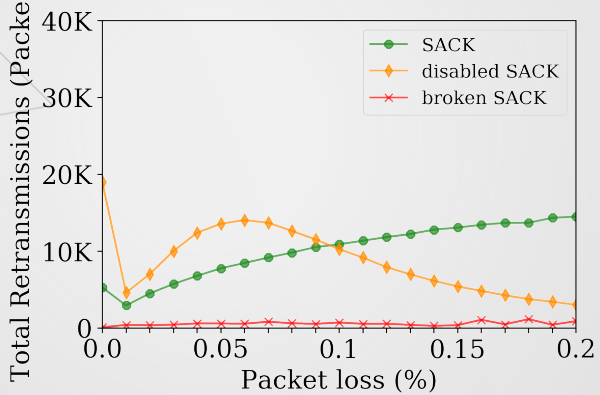
Path impairments scenarios with artificial loss:

Enabled SACK
SACK-enabled flows

Disabled SACK
`tcp.opt.sackok.removed`

Broken SACK
`tcp.seqnum.changed`





IMPACT OF PATH IMPAIRMENTS ON TCP

ECN — Explicit Congestion Notification

SACK — Selective Acknowledgment

WScale — Window Scaling Parameter

Path impairments scenarios with artificial loss:

Enabled SACK

SACK-enabled flows

Disabled SACK

`tcp.opt.sackok.removed`

Broken SACK

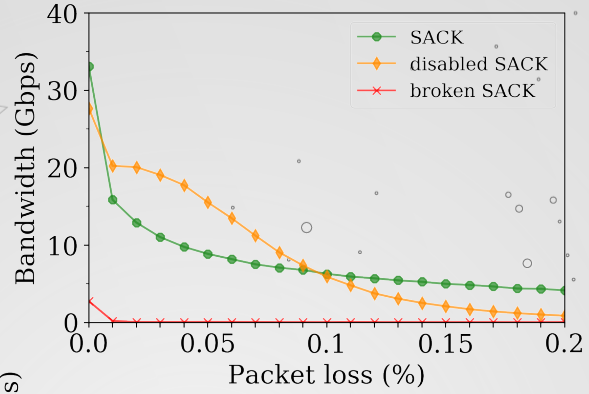
`tcp.seqnum.changed`

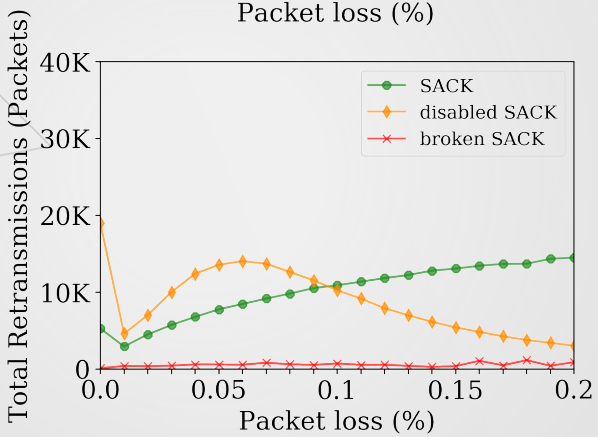
Broken SACK

Broken SACK **stalls the connection** when receiver generates a SACK block

Disabled SACK

Disabled SACK reduce TCP throughput for loss rates greater than 0.09%





IMPACT OF PATH IMPAIRMENTS ON TCP

ECN	SACK	WScale
Explicit Congestion Notification	Selective Acknowledgment	Window Scaling Parameter

TCP Receive Window : $[0 ; 2^{16}-1] * 2^{WScale}$

IMPACT OF PATH IMPAIRMENTS ON TCP

ECN	SACK	WScale
Explicit Congestion Notification	Selective Acknowledgment	Window Scaling Parameter

TCP Receive Window : $[0 ; 2^{16}-1] * 2^{WScale}$

Path impairments scenarios with artificial delay:

Clipped WScale

`tcp.opt.wscale.changed`

Stripped WScale

`tcp.opt.wscale.removed`

IMPACT OF PATH IMPAIRMENTS ON TCP

ECN — Explicit Congestion Notification

SACK — Selective Acknowledgment

WScale — Window Scaling Parameter

TCP Receive Window : $[0 ; 2^{16}-1] * 2^{WScale}$

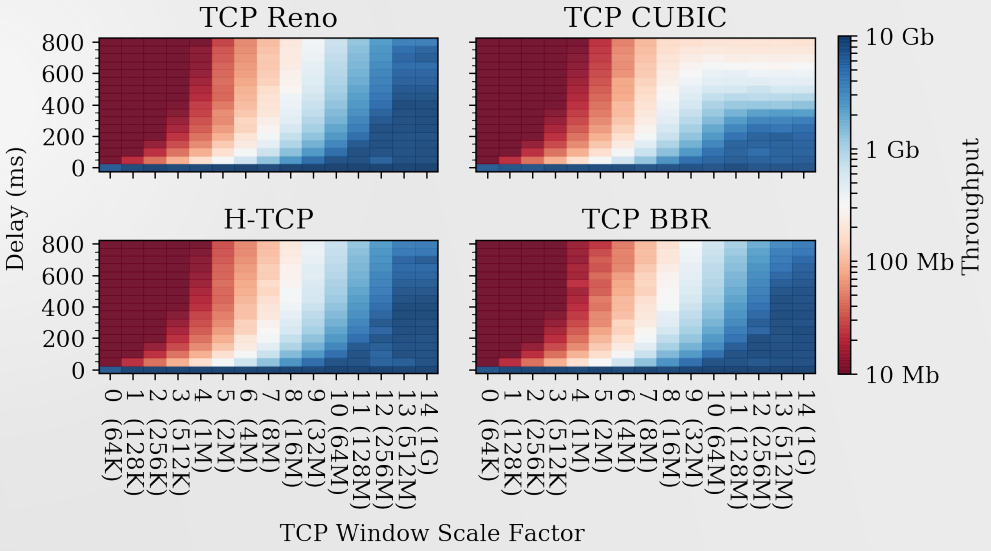
Path impairments scenarios with artificial delay:

Clipped WScale

`tcp.opt.wscale.changed`

Stripped WScale

`tcp.opt.wscale.removed`



IMPACT OF PATH IMPAIRMENTS ON TCP

ECN — Explicit Congestion Notification

SACK — Selective Acknowledgment

WScale — Window Scaling Parameter

TCP Receive Window : $[0 ; 2^{16}-1] * 2^{WScale}$

Path impairments scenarios with artificial delay:

Clipped WScale

`tcp.opt.wscale.changed`

Stripped WScale

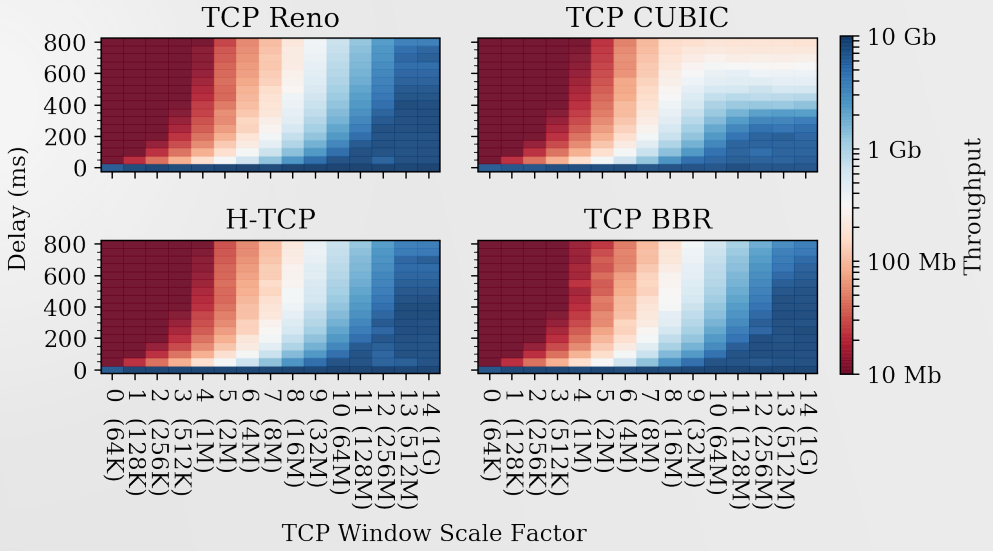
`tcp.opt.wscale.removed`

Wscale Impairment

Clipped and Stripped Wscale have a direct impact on **maximum achievable throughput**

In the future

Higher throughputs might become a problem as Higher BDPs are not supported by TCP



CONCLUSION

Lessons Learned

CONCLUSION

Lessons Learned

Middleboxes are **prevalents** in today's Internet

CONCLUSION

Lessons Learned

Middleboxes are **prevalents** in today's Internet

Middleboxes are **problematic** to existing TCP features

CONCLUSION

Measuring the Evolution of Transport Protocols in the Internet, A.Medina et al., in ACM CCR 2005

Is it still possible to extend TCP ?, M.Honda et al., in IMC11

Lessons Learned

Middleboxes are **prevalents** in today's Internet

Middleboxes are **problematic** to existing TCP features

Middleboxes are **problematic** to transport evolution

CONCLUSION

Measuring the Evolution of Transport Protocols in the Internet, A.Medina et al., in ACM CCR 2005

Is it still possible to extend TCP ?, M.Honda et al., in IMC11

Lessons Learned

Middleboxes are **prevalents** in today's Internet

Middleboxes are **problematic** to existing TCP features

Middleboxes are **problematic** to transport evolution

CONCLUSION

Lessons Learned

Middleboxes are **prevalents** in today's Internet

Middleboxes are **problematic** to existing TCP features

Middleboxes are **problematic** to transport evolution

Measuring the Evolution of Transport Protocols in the Internet, A.Medina et al., in ACM CCR 2005

Is it still possible to extend TCP ?, M.Honda et al., in IMC11

Future of path impairments

CONCLUSION

Measuring the Evolution of Transport Protocols in the Internet, A.Medina et al., in ACM CCR 2005

Is it still possible to extend TCP ?, M.Honda et al., in IMC11

Lessons Learned

Middleboxes are **prevalents** in today's Internet

Middleboxes are **problematic** to existing TCP features

Middleboxes are **problematic** to transport evolution

Future of path impairments

Encryption by Default

Transport-layer Encryption

CONCLUSION

Measuring the Evolution of Transport Protocols in the Internet, A.Medina et al., in ACM CCR 2005

Is it still possible to extend TCP ?, M.Honda et al., in IMC11

Lessons Learned

Middleboxes are **prevalents** in today's Internet

Middleboxes are **problematic** to existing TCP features

Middleboxes are **problematic** to transport evolution

Future of path impairments

Middlebox-proof TCP

e.g., MPTCP

Encryption by Default

Transport-layer Encryption

CONCLUSION

Measuring the Evolution of Transport Protocols in the Internet, A.Medina et al., in ACM CCR 2005

Is it still possible to extend TCP ?, M.Honda et al., in IMC11

Lessons Learned

Middleboxes are **problematic** to existing TCP features

Middleboxes are **problematic** to transport evolution

Future of path impairments

Middlebox-proof TCP

e.g., MPTCP

Encryption by Default

Transport-layer Encryption

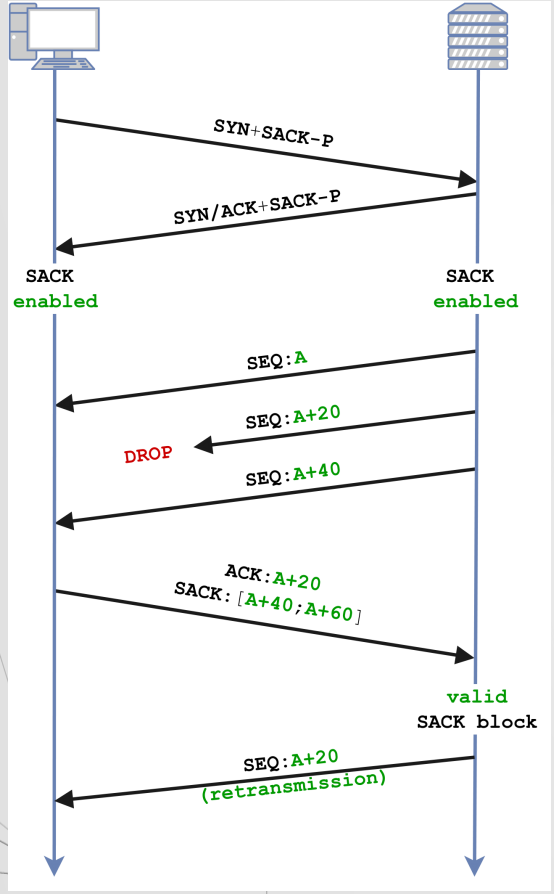
QUIC

UDP-based splitted layer transport

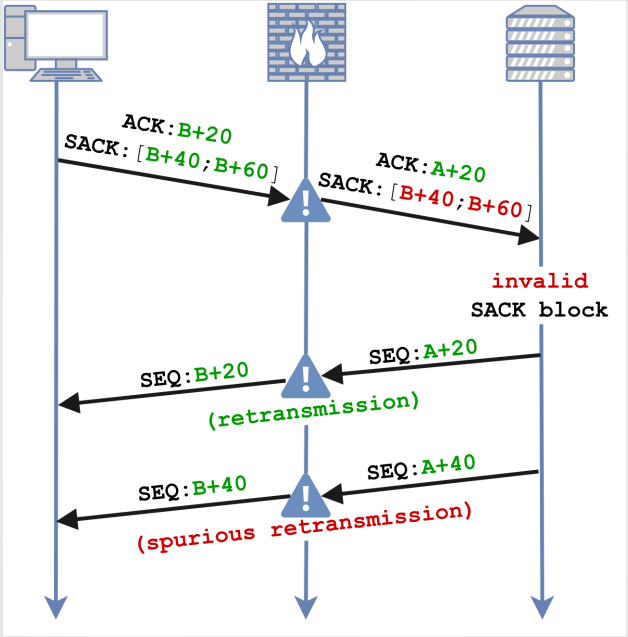
THANKS

korian.edeline@uliege.be

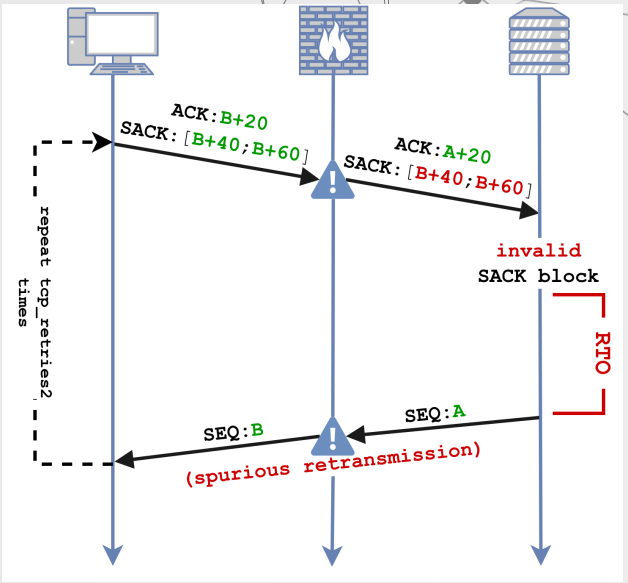
SACK



Normal SACK

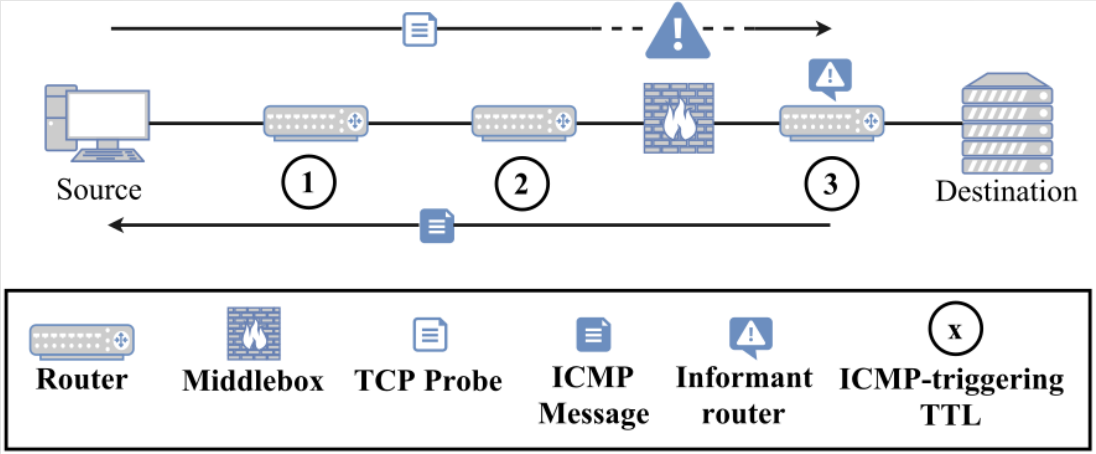


Broken SACK
(spurious retransmission)



Broken SACK (RTO)

tracebox



- **RFC 792** : “The internet header plus the first 64 bits”
- **RFC 1812** : “as much [...] as possible” (< 576 B)

Testbed specifications

- Intel Xeon E5-2620 2.1GHz, 16 Threads, 32GB RAM
- Intel XL710 2x40GB NICs
- Huawei CE6800 switch
- Debian 9