School of Computing Science

60 YEARS OF COMPUTING AT GLASGOW

UNIVERSITY of STIRLING

Evolving the Transport-Layer

Stephen McQuistin and Colin Perkins
University of Glasgow

Marwan Fayed
University of Stirling

UKNOF36
January 2017, London

We want ..

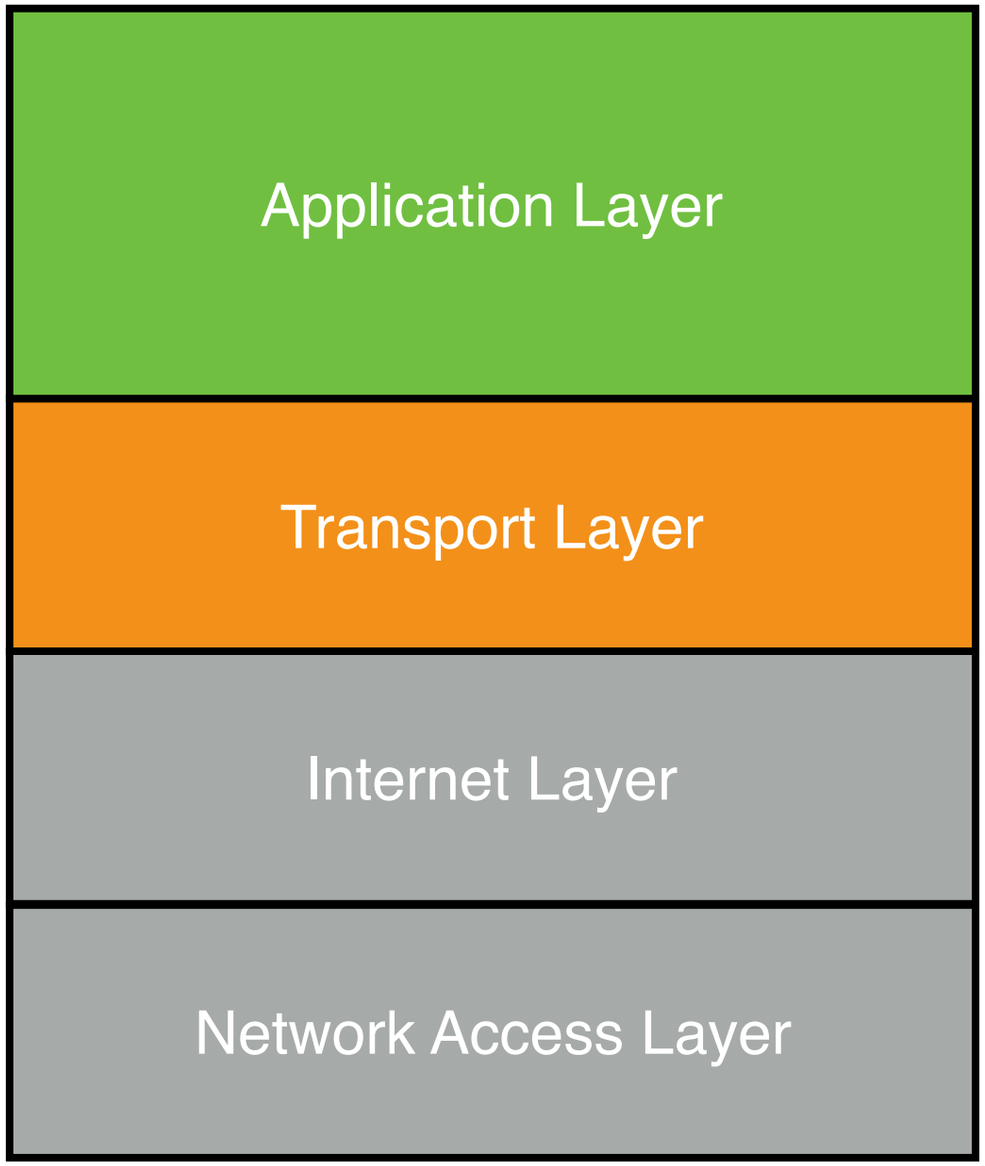
- .. to build new, *deployable* transport-layer protocols, and middleboxes make this a challenge
- .. to learn more about the middleboxes deployed in the wild
- .. your help - as operators, you *know* what middleboxes are in your networks, and what they are doing

Transport constrained to TCP/UDP

- New transport-layer protocols *should* be deployable if the IP layer is unchanged

- In practice, middleboxes — firewalls, NATs — limit change

- Transport layer ossified around TCP and UDP

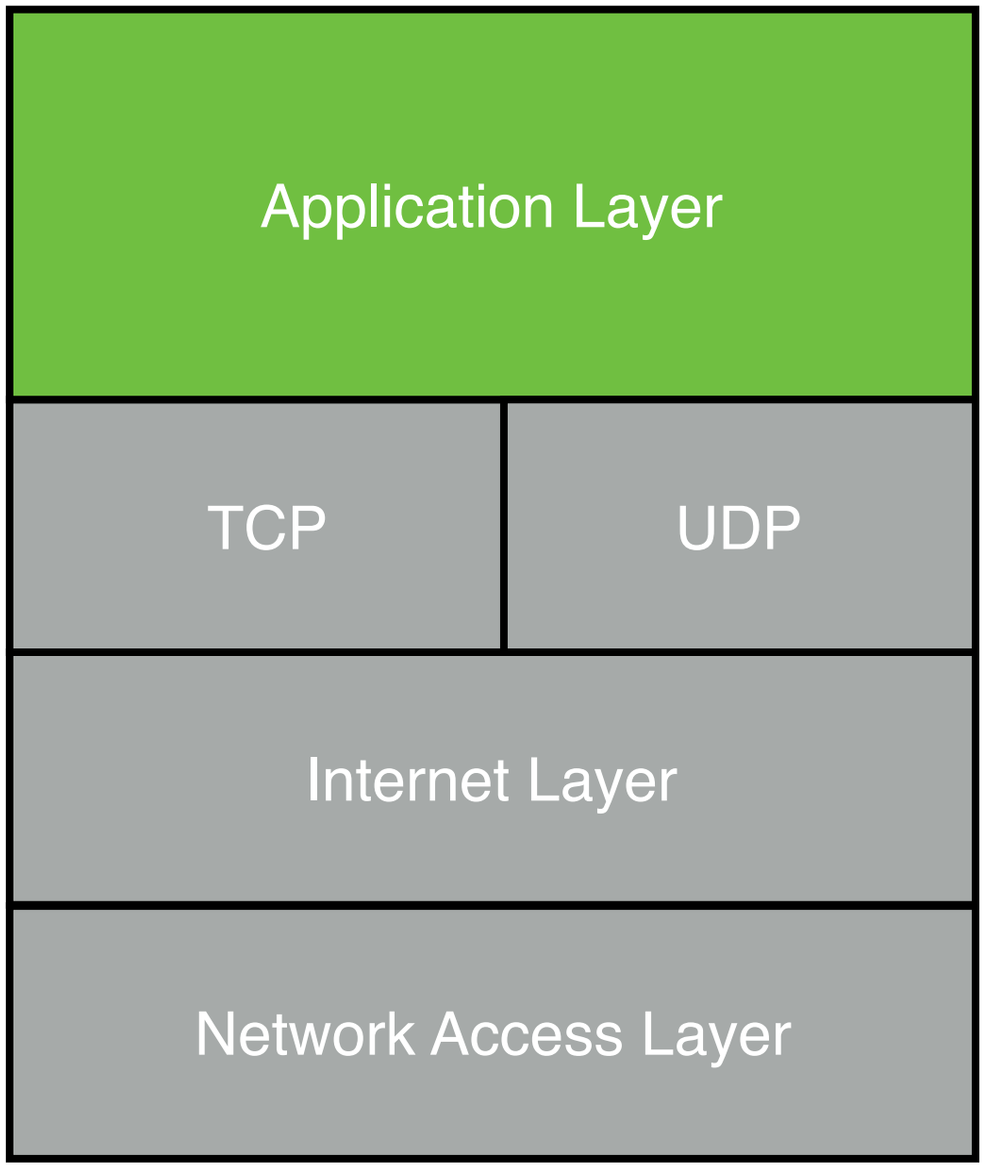


Transport constrained to TCP/UDP

- New transport-layer protocols *should* be deployable if the IP layer is unchanged

- In practice, middleboxes — firewalls, NATs — limit change

- Transport layer ossified around TCP and UDP

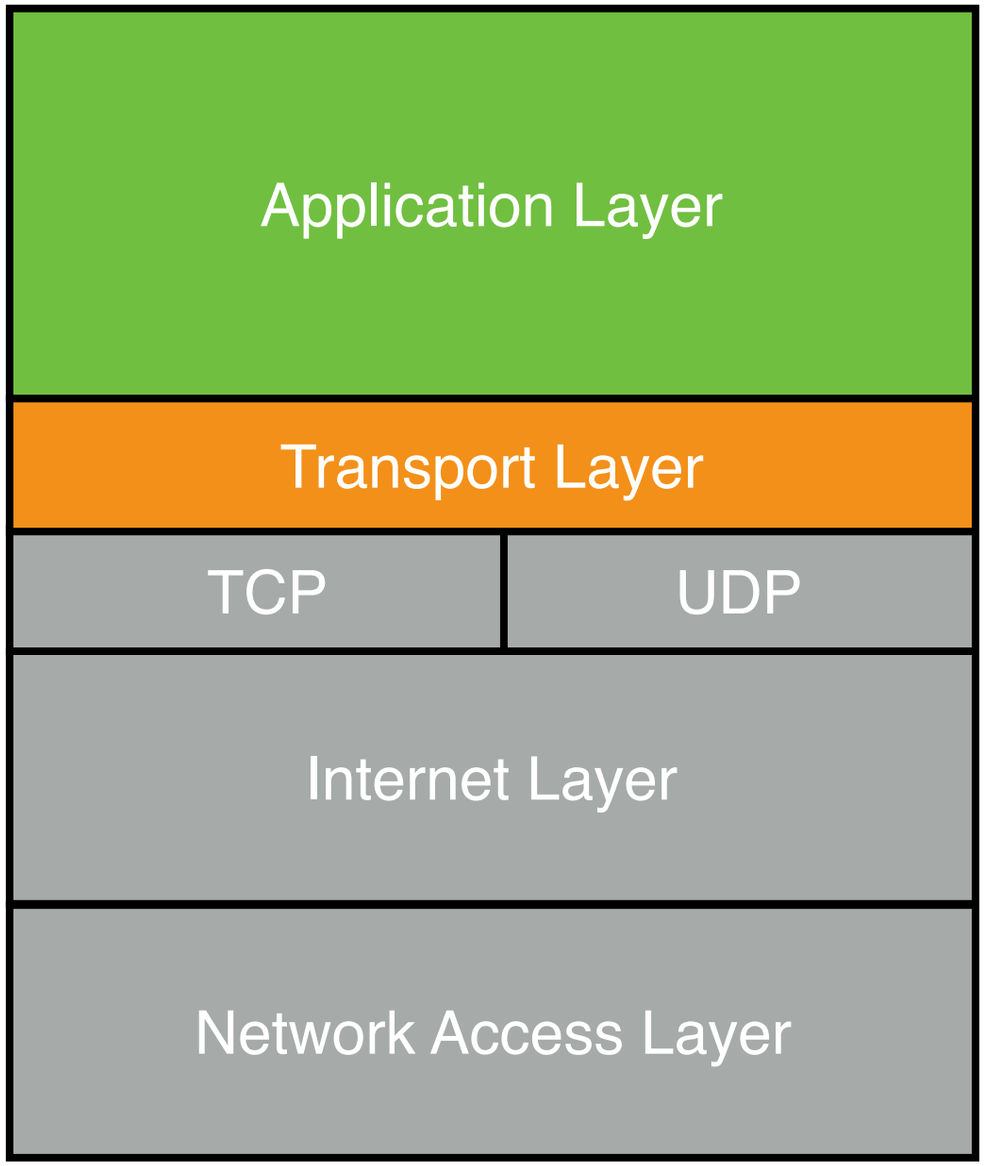


TCP and UDP aren’t enough

- TCP: ordered, reliable, congestion controlled byte stream
- UDP: unordered, unreliable datagram delivery
- Many applications don’t fit exactly to either of these service models: want a mix of features from both
- For example: real-time multimedia applications want datagram delivery, and congestion control

TCP/UDP as substrates

- Use TCP or UDP as a substrate: build new protocols within the payloads of these existing protocols
- Example: Google's QUIC (Quick UDP Internet Connections)
- Focus on low latency
- UDP isn't available to all hosts; often blocked on enterprise firewalls



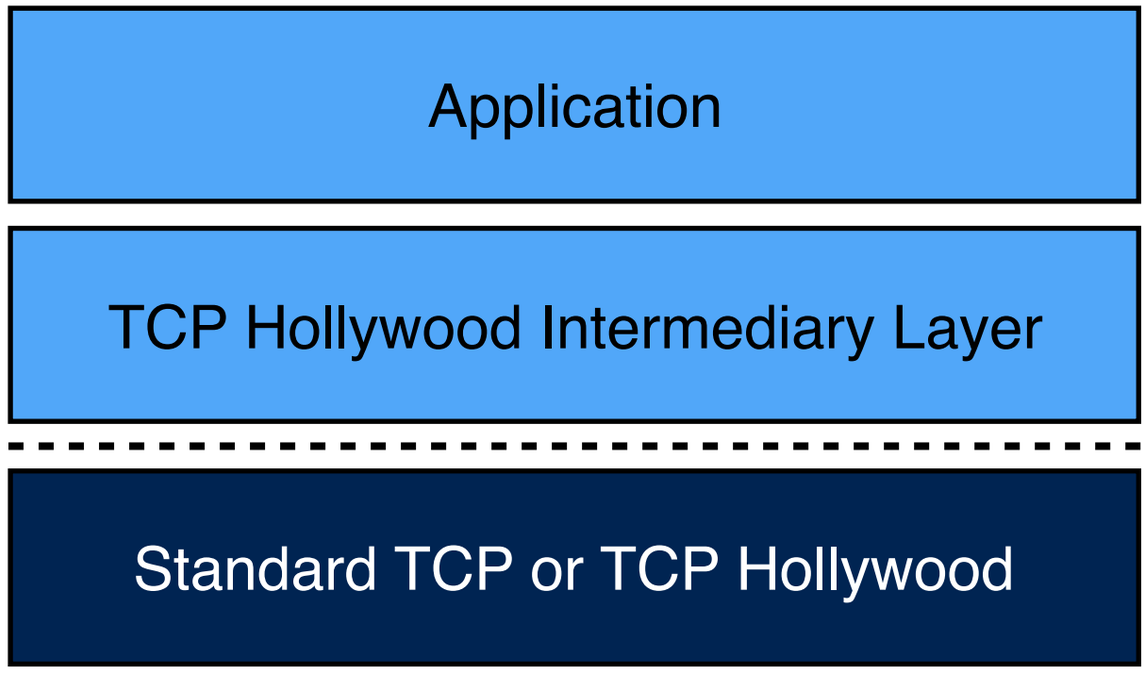
Example: TCP Hollywood

- Uses TCP as a substrate, but modified to reduce latency
- Message-oriented, to allow application data units to be sent
- Unordered, given messages are independently useful
- Partially reliable, based on metadata provided by the application

"Hollywood Sign", Gnaphron - CC BY-SA 2.0
flickr.com/photos/gnaphron/8485145044

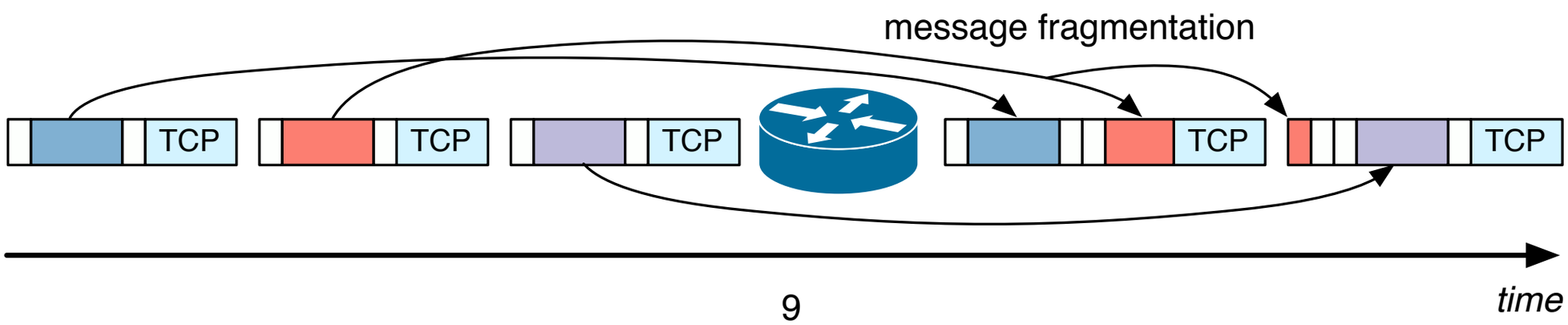
TCP Hollywood: Architecture

- Functionality split between user-space intermediary layer, and kernel extensions
- Intermediary layer works over either standard TCP or TCP Hollywood
- Supports partial deployments



TCP Hollywood: Messaging

- Nagle's algorithm disabled
- Applications pass messages with optional metadata: deadlines and dependencies
- Message data encoded to escape zero bytes for use as framing markers —*transparent channel*
- Segments passed to intermediary layer as they arrive, with ACKs generated as under standard TCP

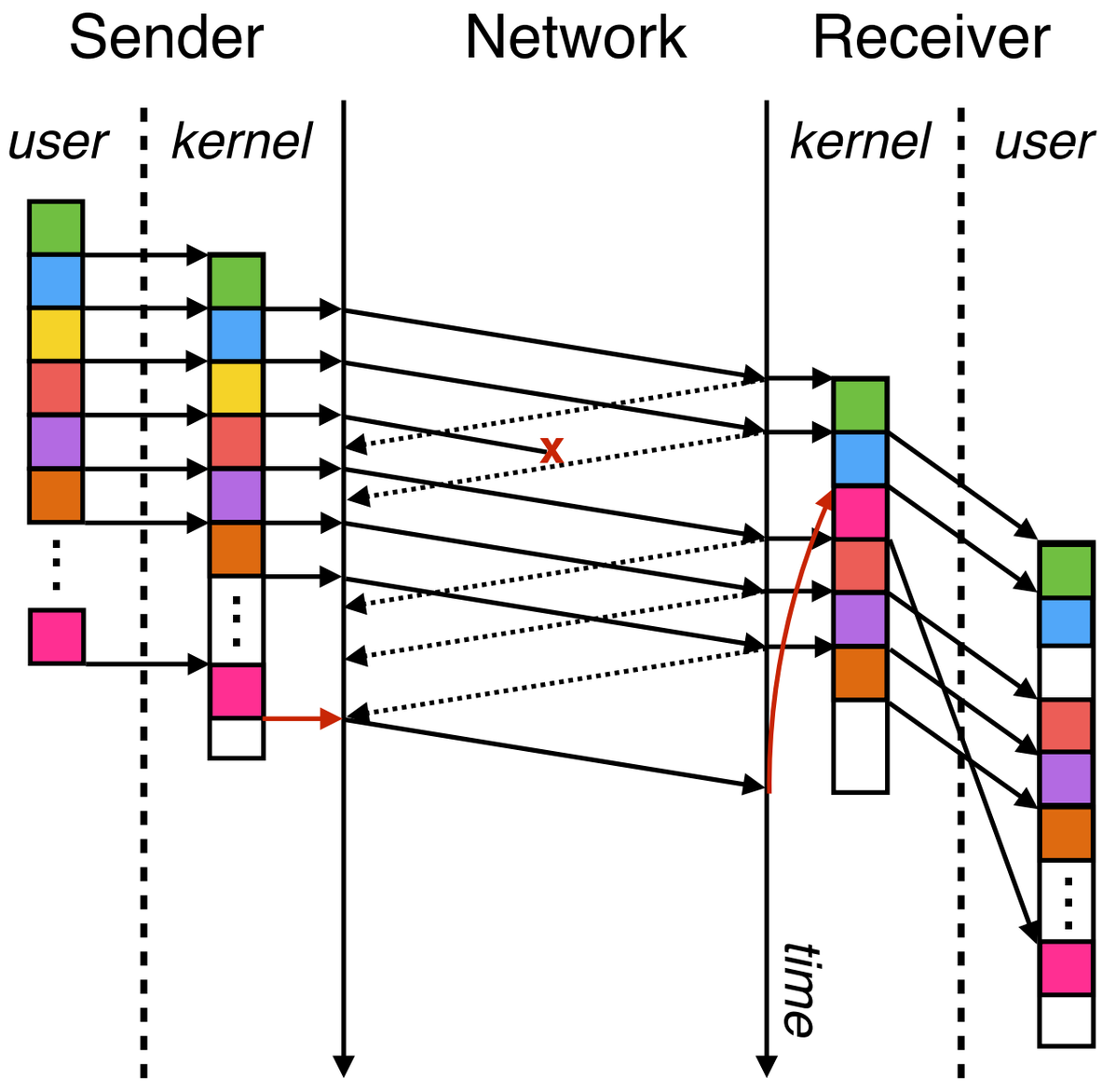


TCP Hollywood: Partial reliability

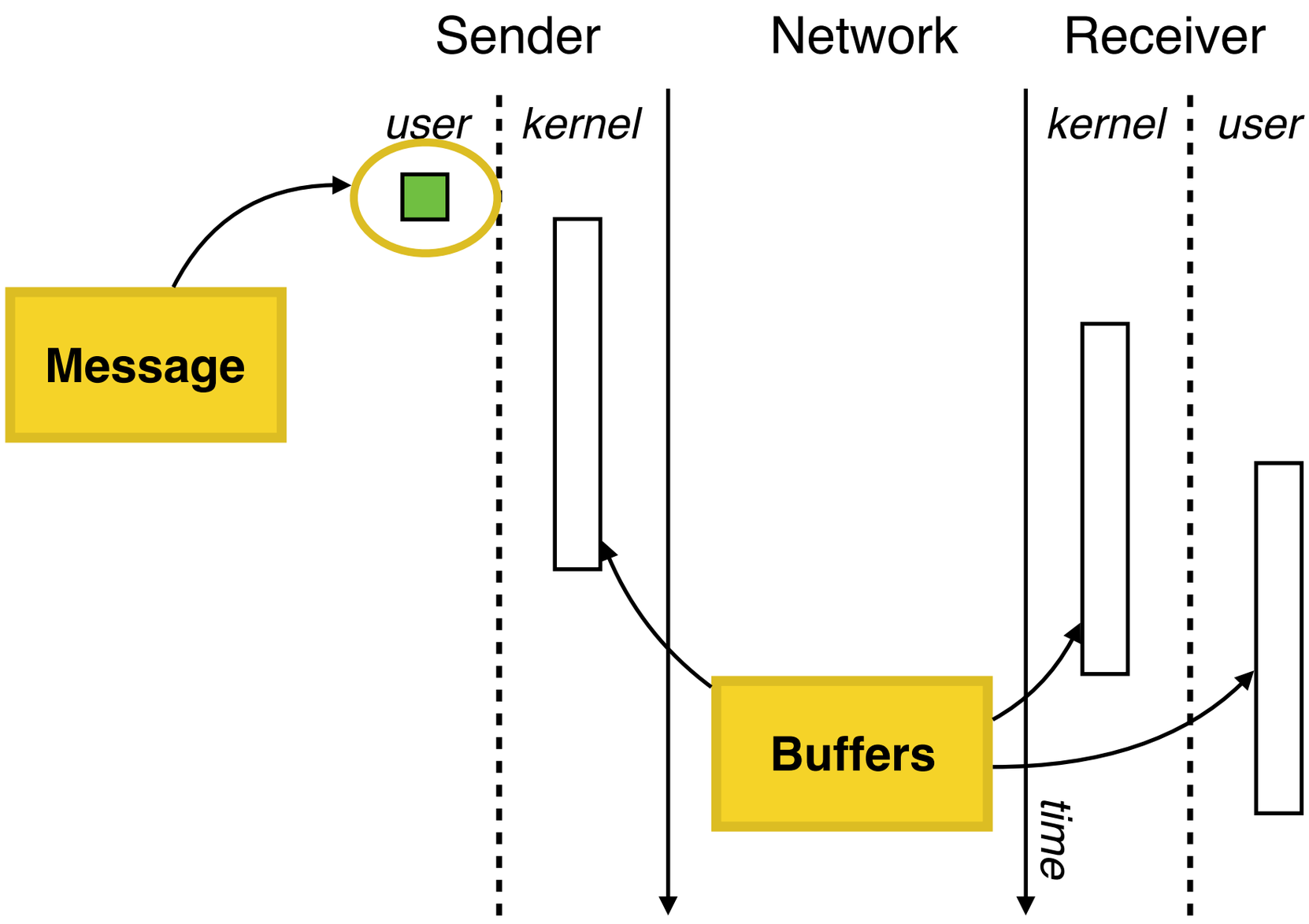
- Messages might expire: estimated to arrive too late, or depend on an undelivered message

- Expired messages aren't retransmitted - next live message sent instead as an *inconsistent retransmission*

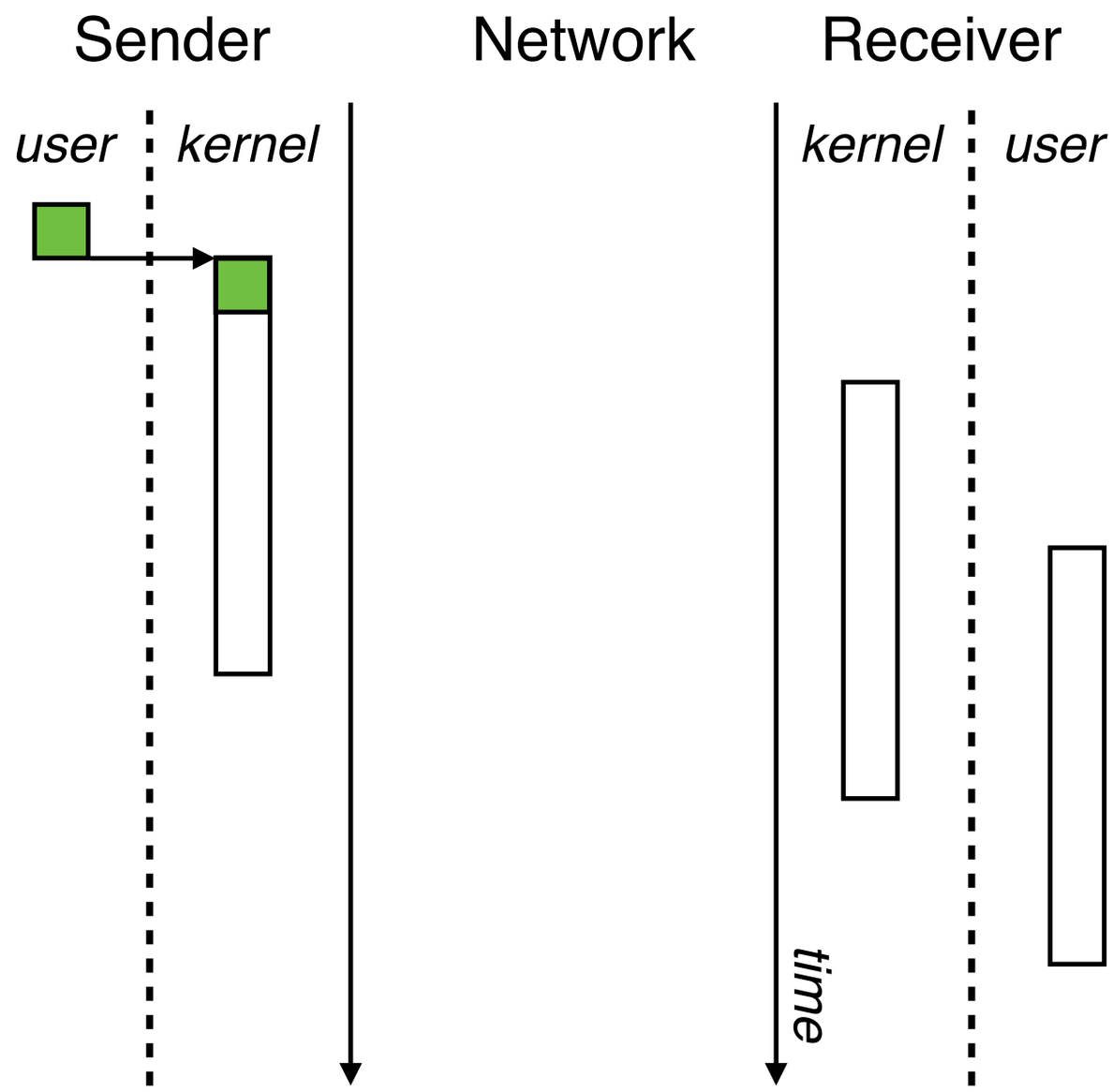
- Same length and TCP sequence number as standard TCP — payload is different



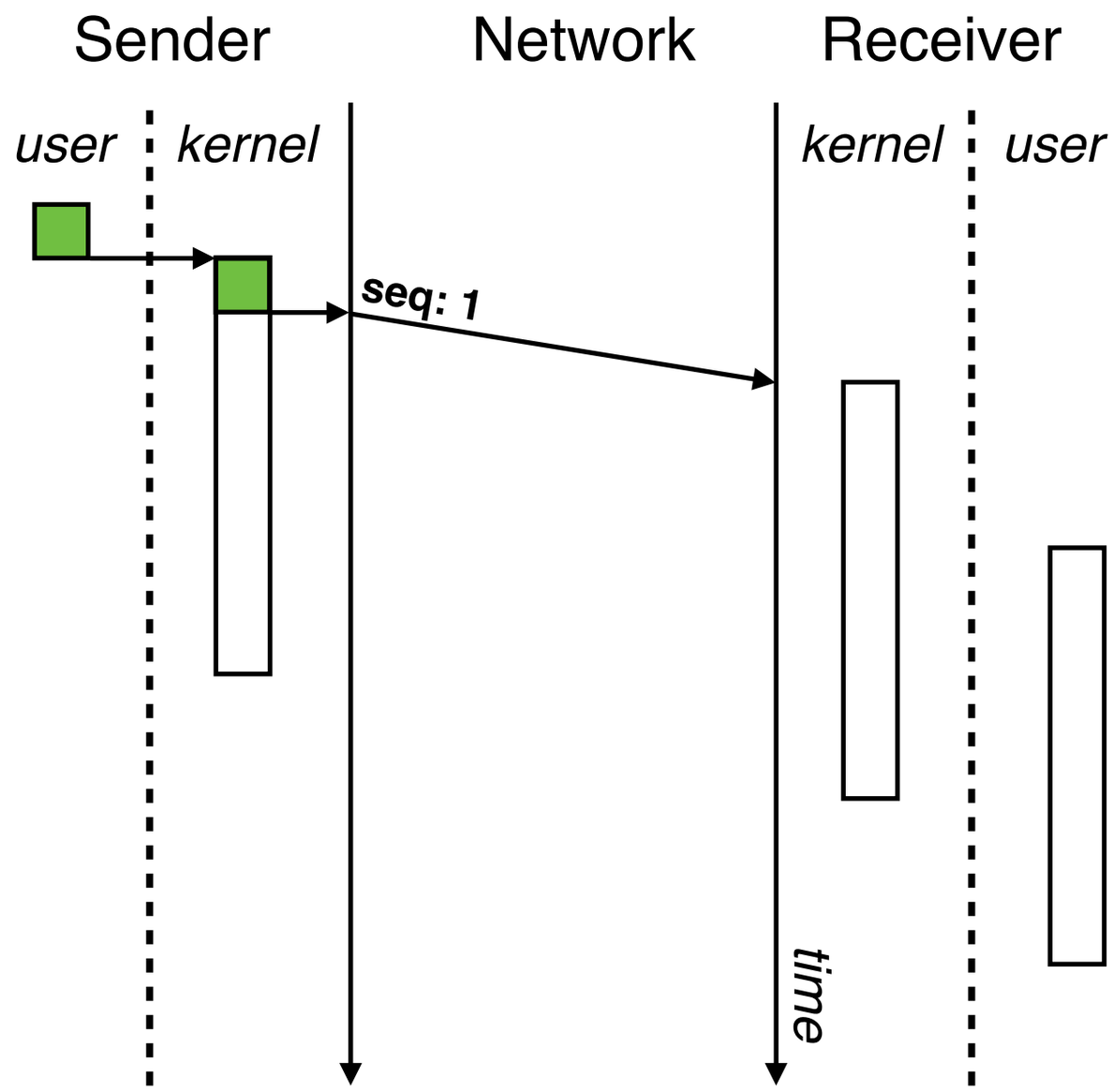
TCP Hollywood in action



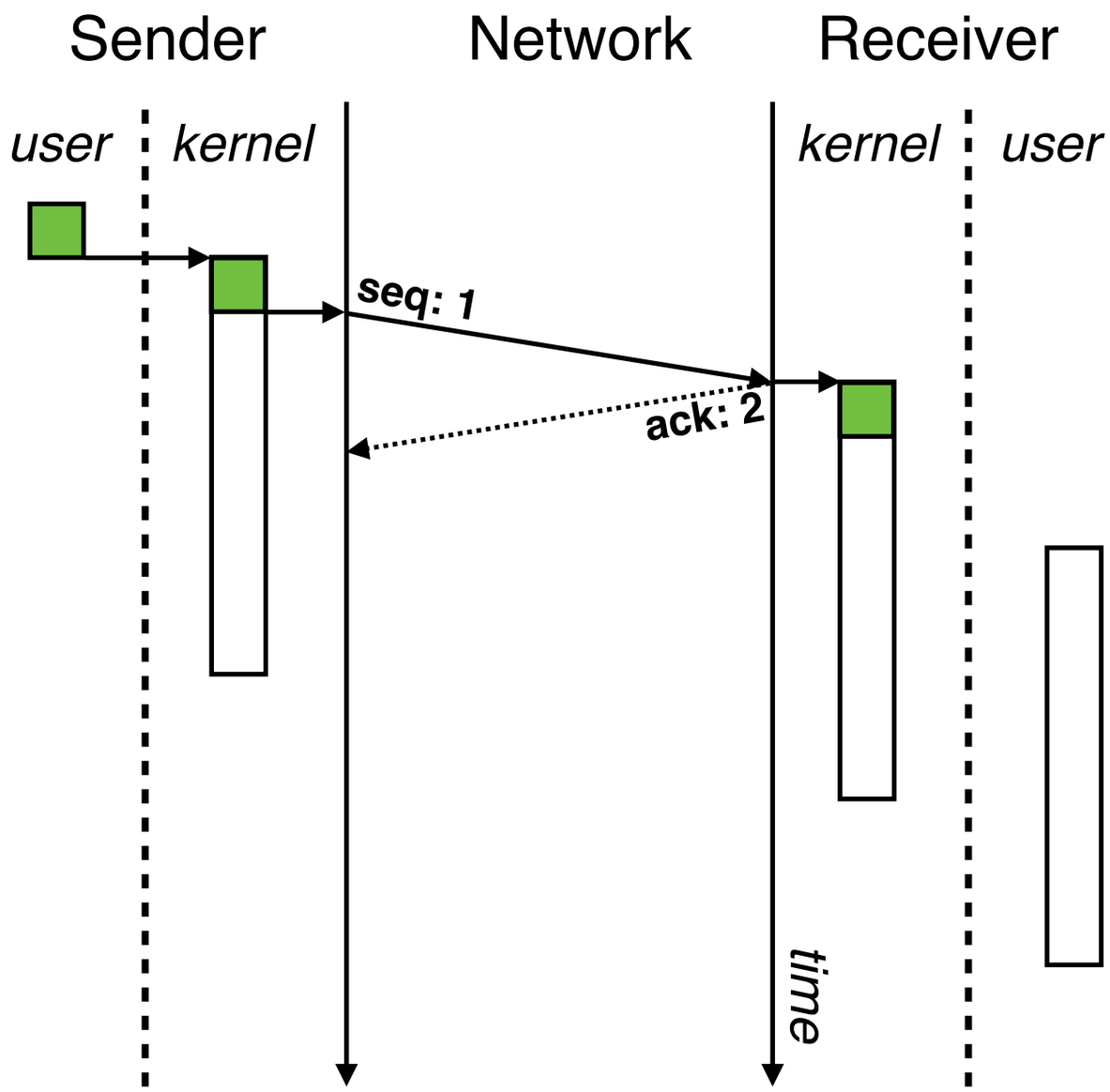
TCP Hollywood in action



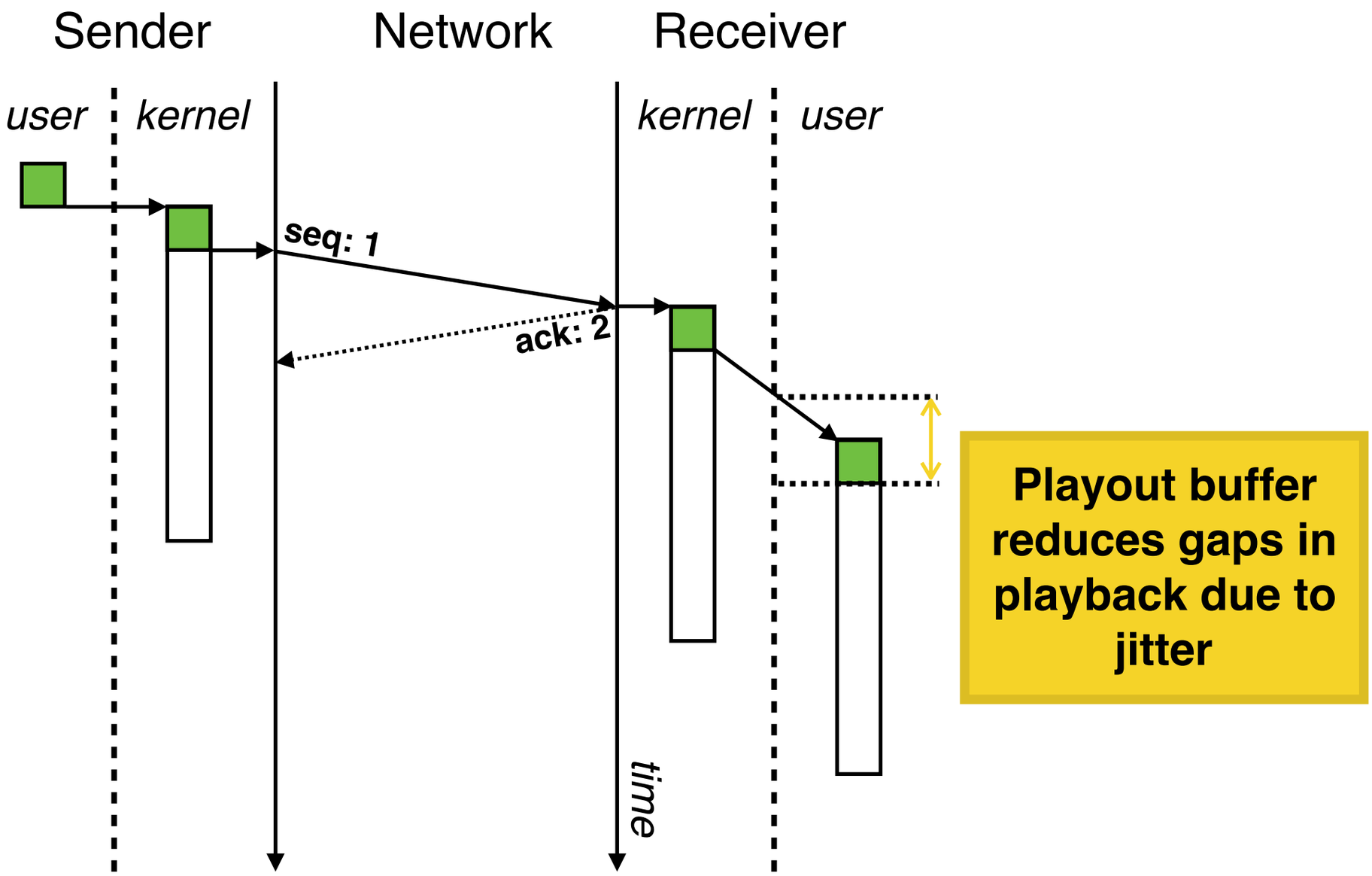
TCP Hollywood in action



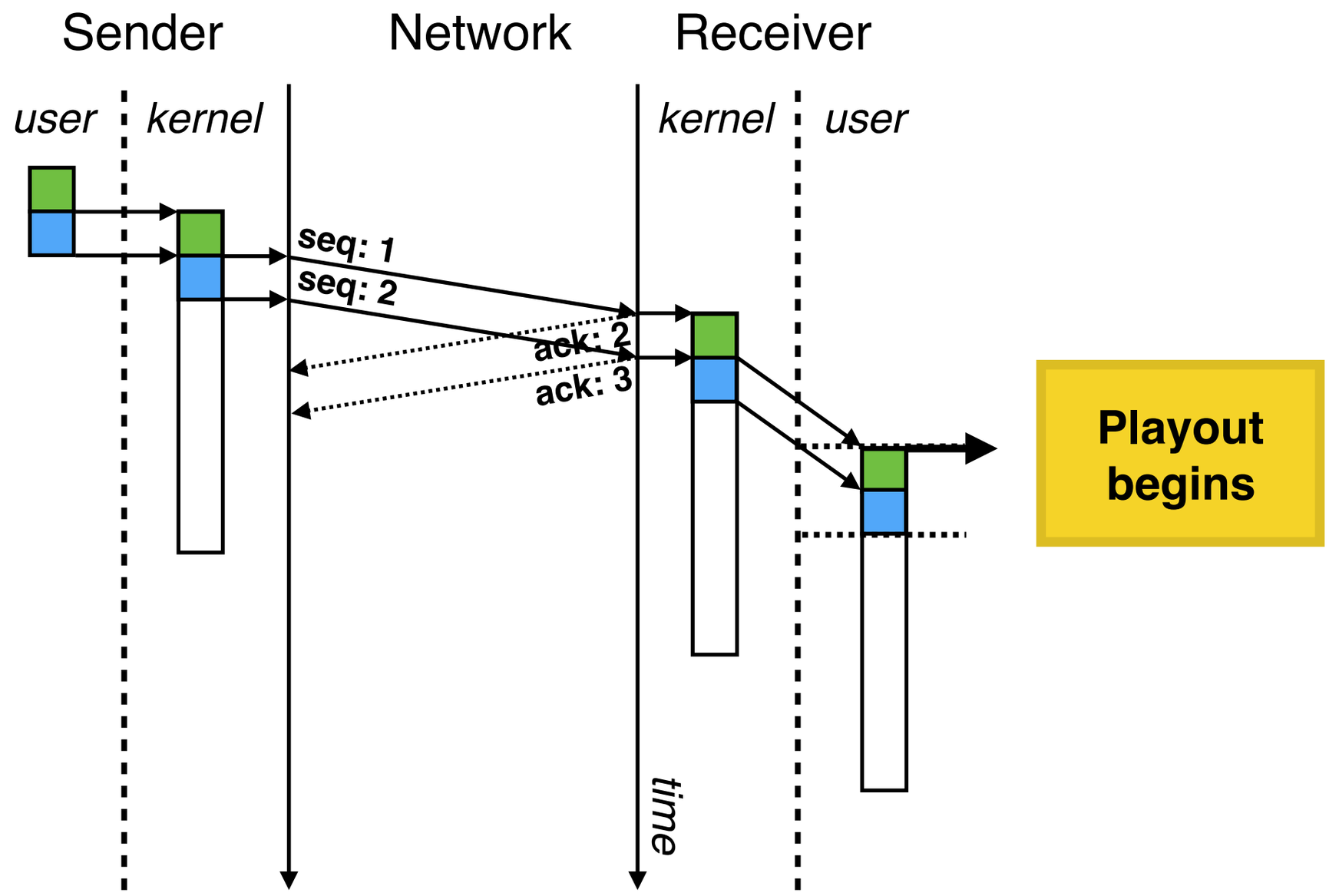
TCP Hollywood in action



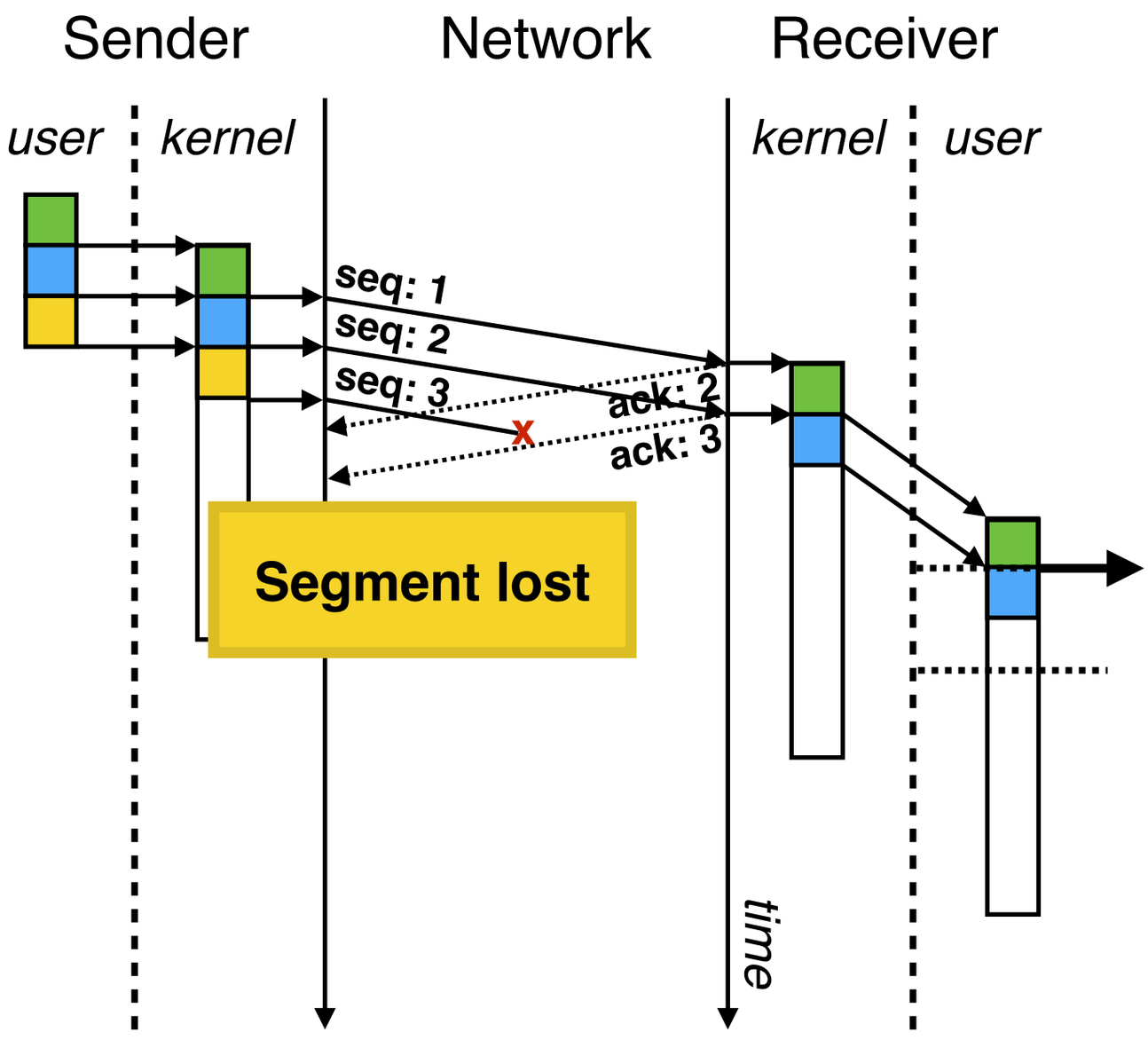
TCP Hollywood in action



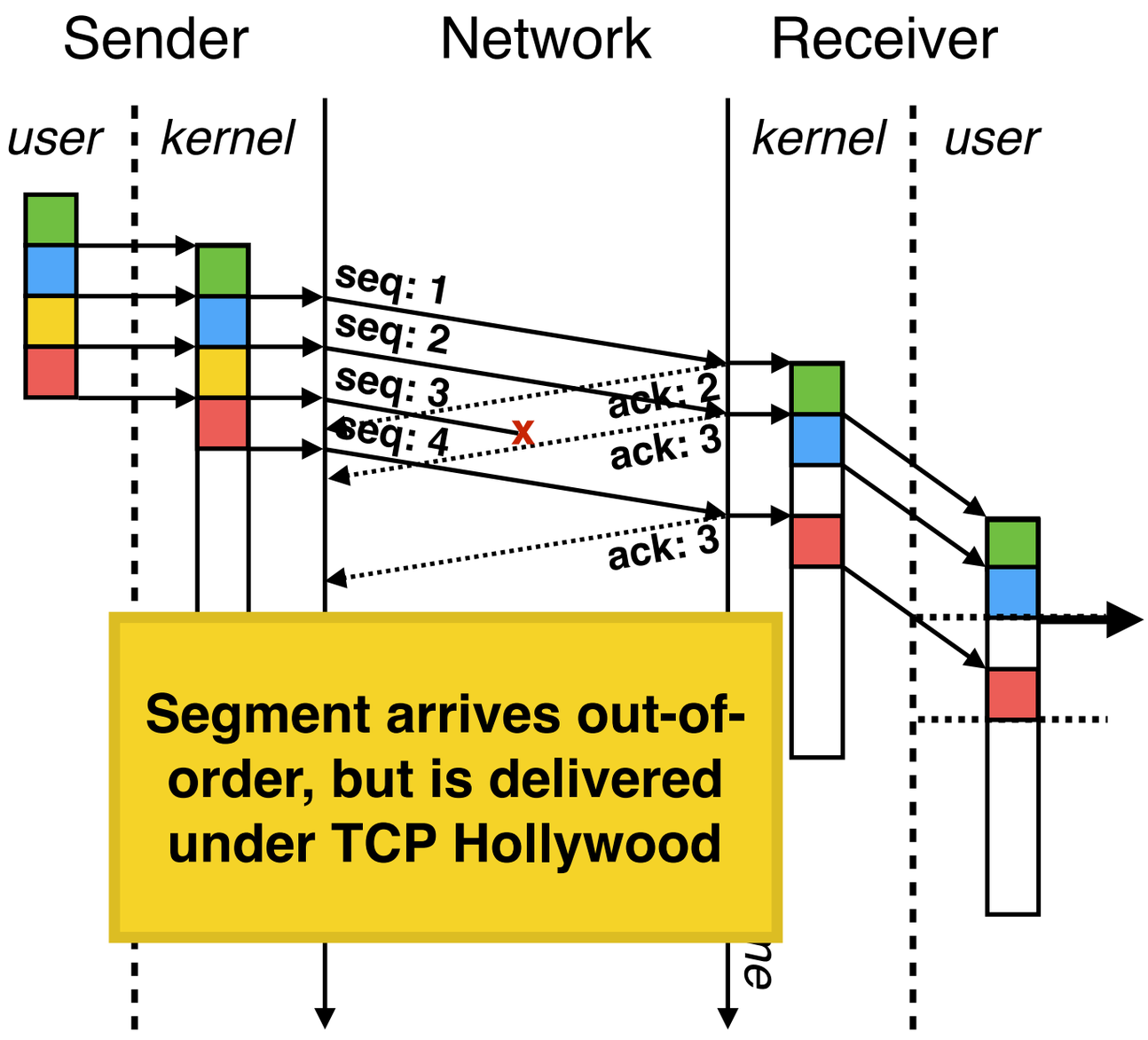
TCP Hollywood in action



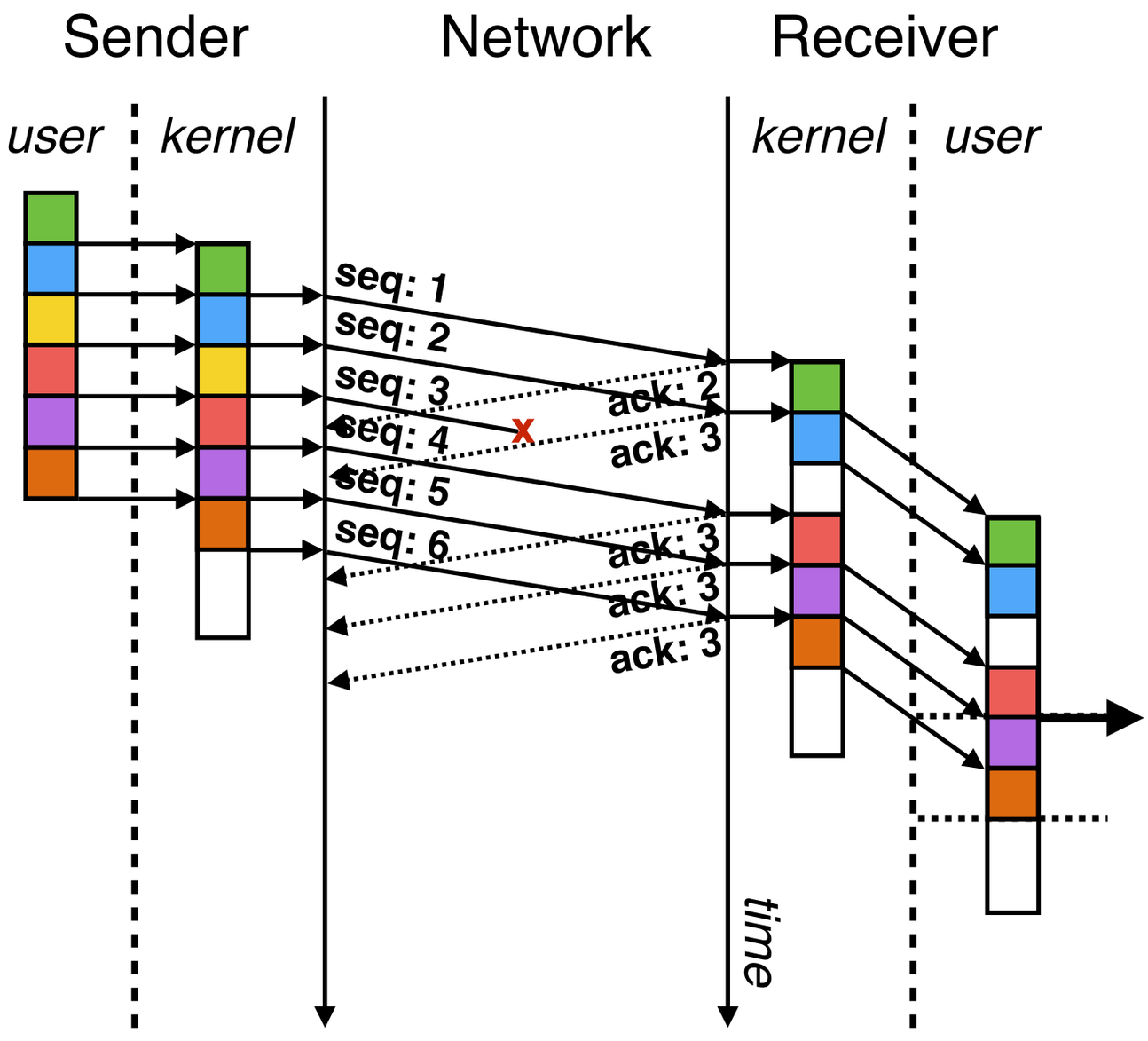
TCP Hollywood in action



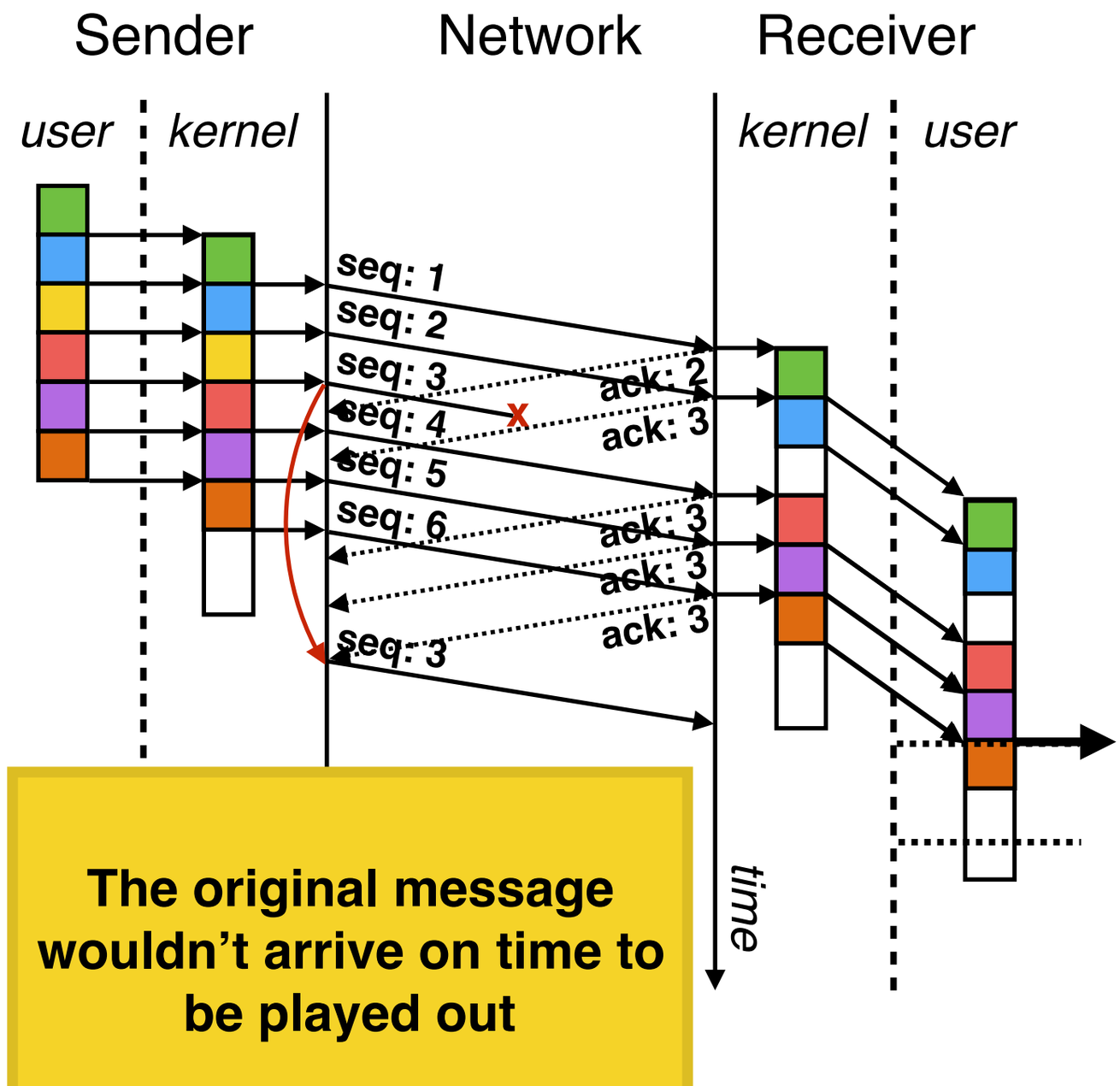
TCP Hollywood in action



TCP Hollywood in action

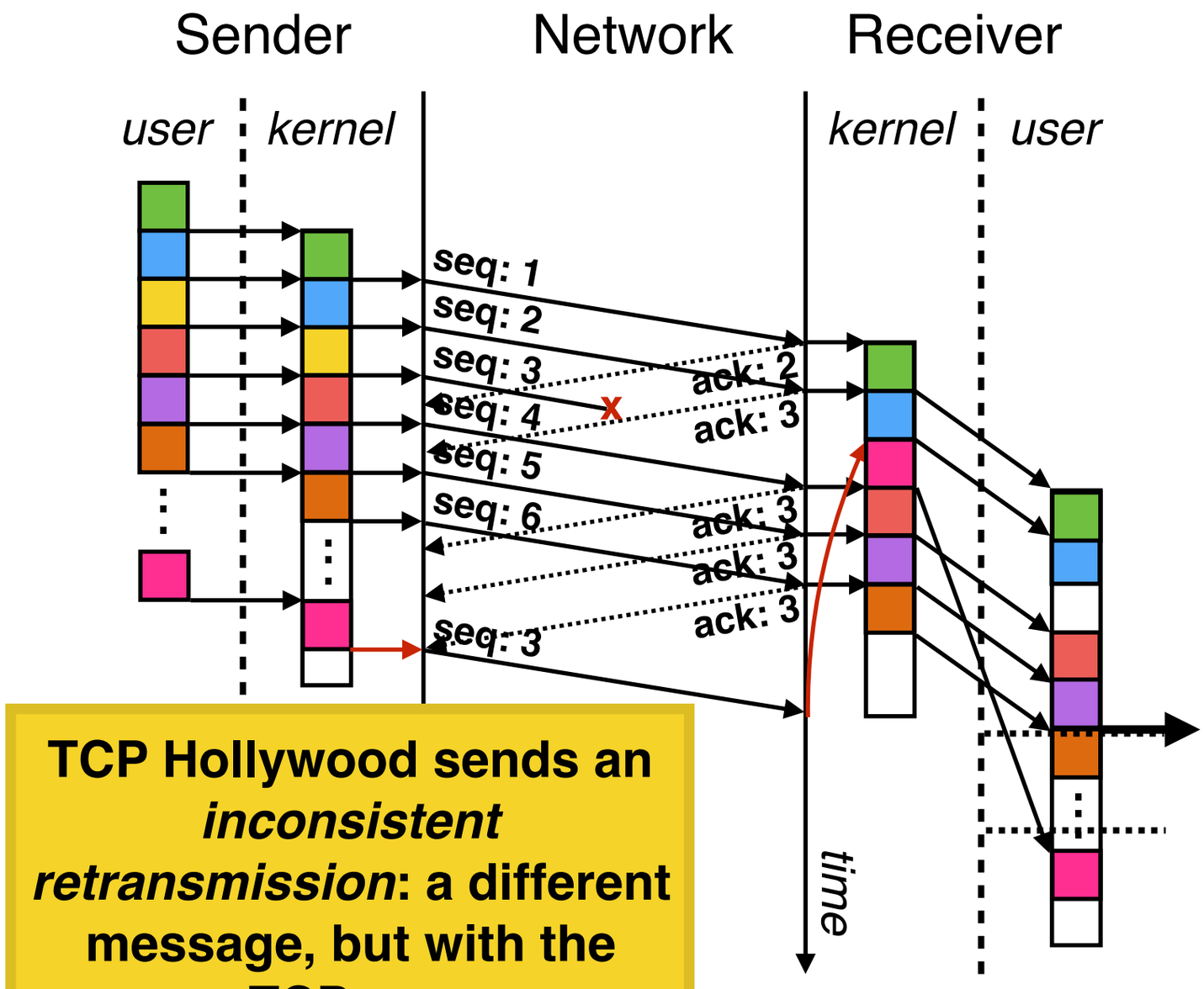


TCP Hollywood in action



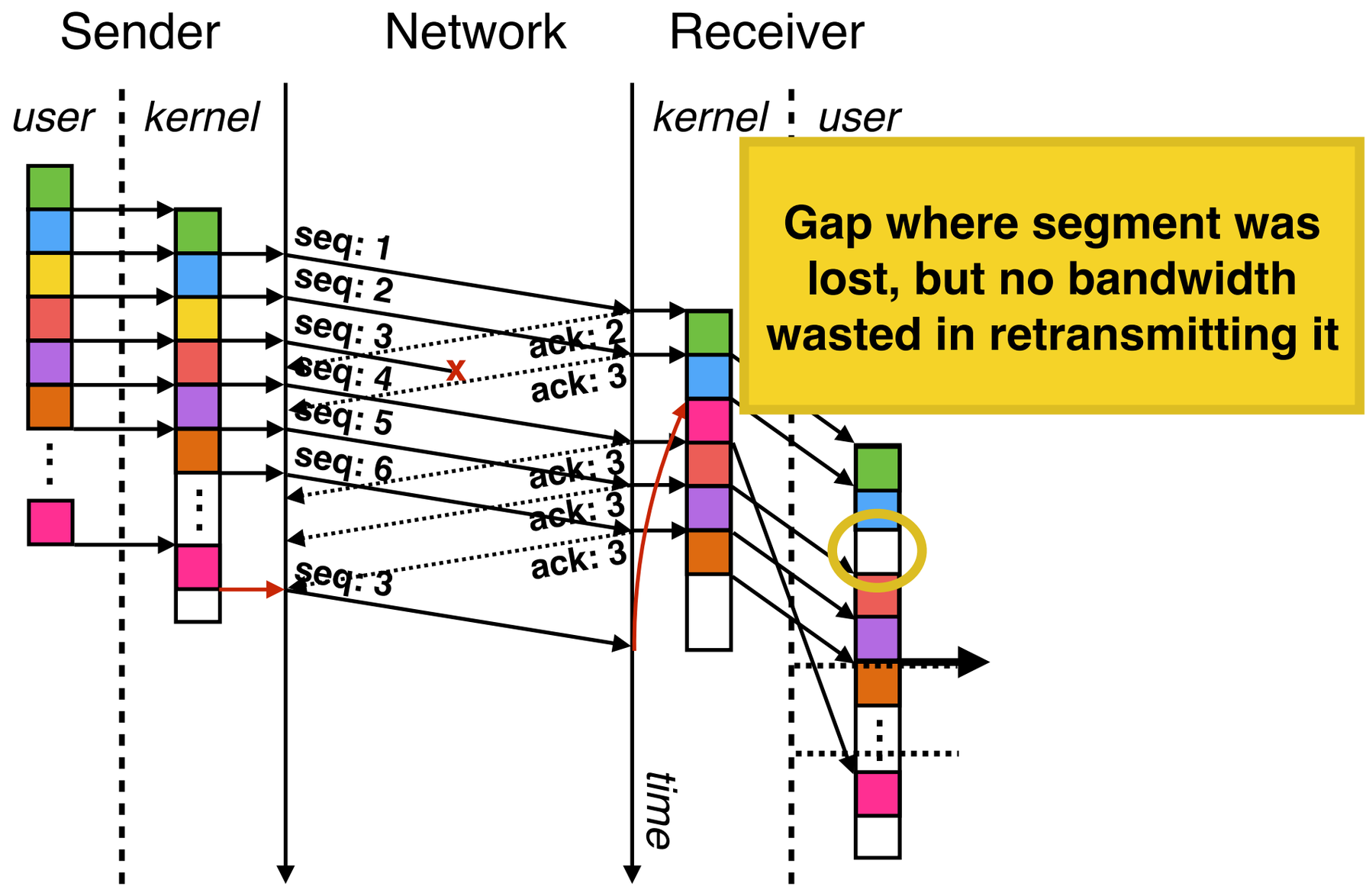
The original message wouldn't arrive on time to be played out

TCP Hollywood in action

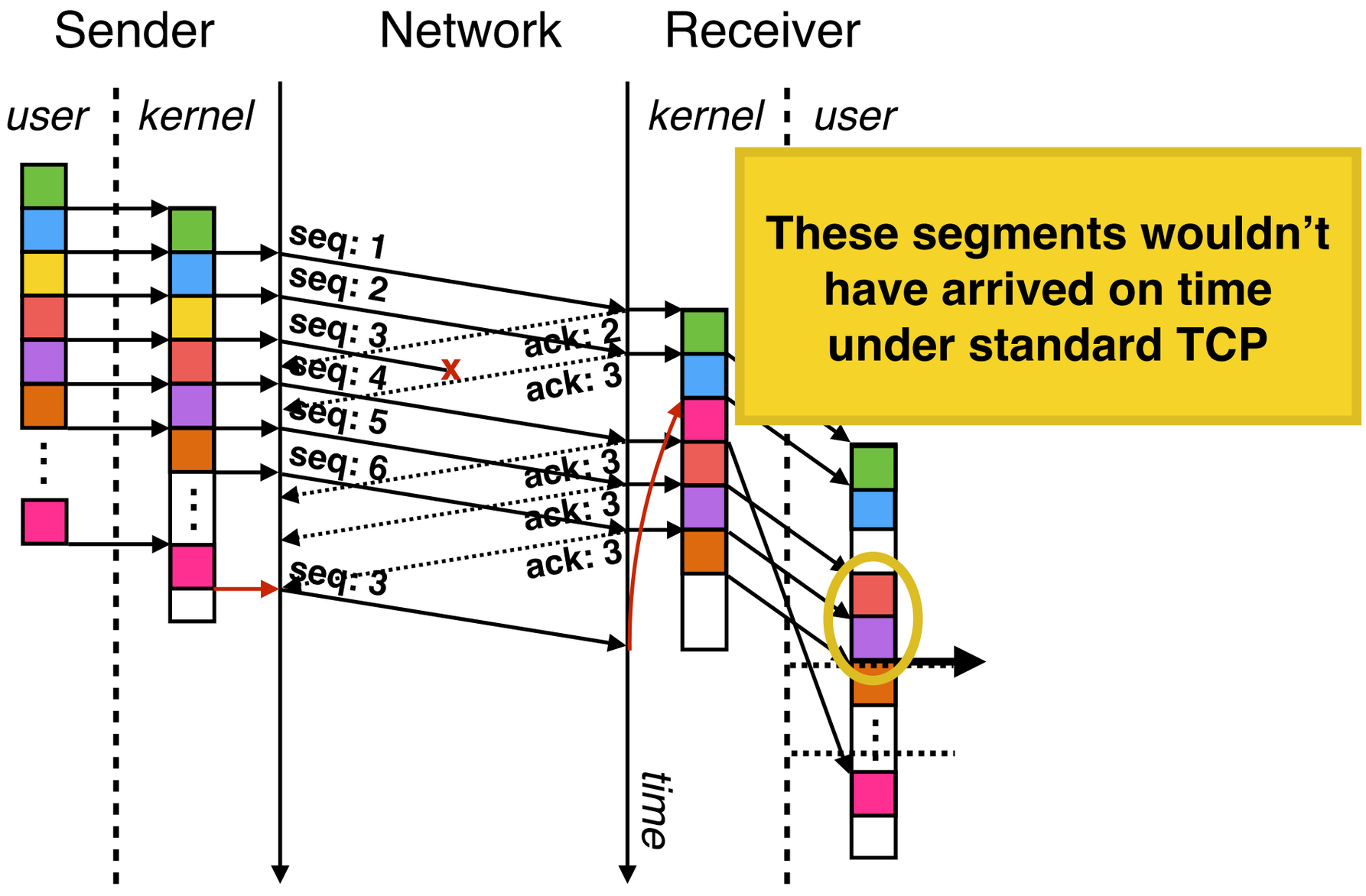


TCP Hollywood sends an *inconsistent retransmission*: a different message, but with the same TCP sequence number

TCP Hollywood in action



TCP Hollywood in action



TCP Hollywood is deployable

- Inconsistent retransmissions are visible on the wire: different payloads with the same sequence number
- Caching seen on the majority of mobile networks
- Safe failure mode for TCP Hollywood

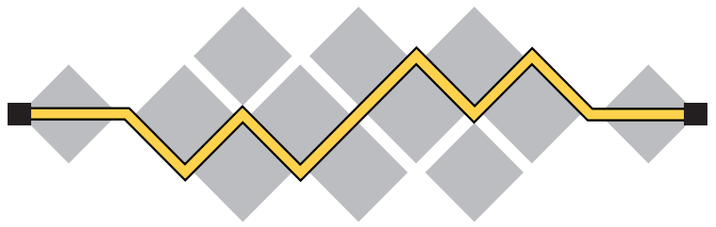
ISP	Port 80	Port 4001
Fixed-line		
Andrews & Arnold	●	●
BT	●	●
Demon	●	●
EE	●	●
Eclipse	●	●
Sky	●	●
TalkTalk	●	●
Virgin Media	●	●
Mobile		
EE	▲	▲
O2	▲	▲
Three	●	●
Vodafone	▲	●

TCP Hollywood is deployable, but..

- “Black box” measurements are limited
- Measuring a large enough set of paths is challenging — both endpoints need to be controlled
- Many different protocol variations — each needs to be measured separately
- Results provide a snapshot; could change significantly over time

Other examples

- Explicit Congestion Notification
- Differentiated Services
- Multipath TCP
- Path Layer UDP Substrate
- All require some level of middlebox support or cooperation



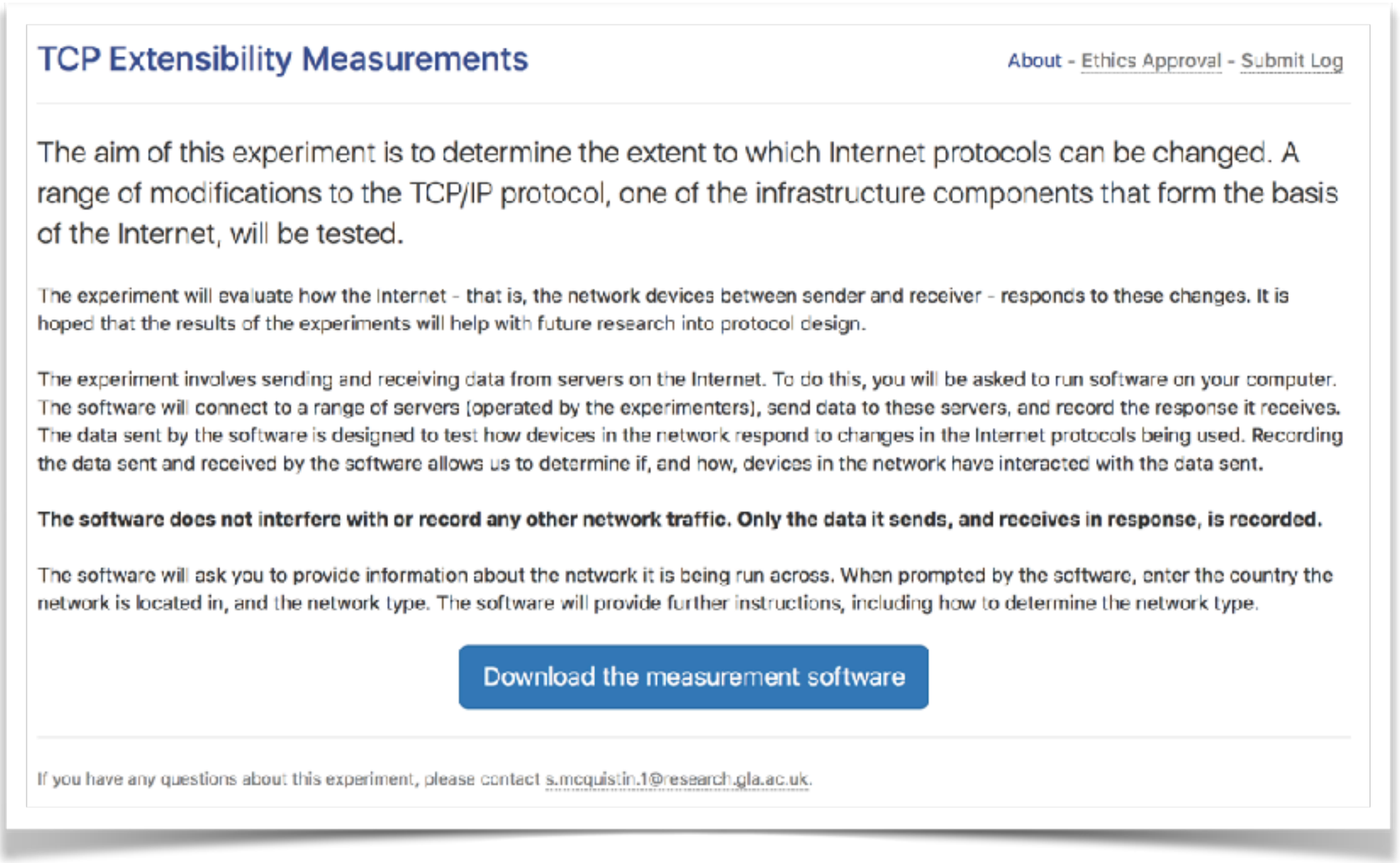
IETF®

Network Operator ‘Secrets’

- What middleboxes are deployed in networks today?
- How do they impact innovation at the transport layer?
- How should transport protocols and middleboxes interact?

Measurement Study

- Measurements between a client and a server running at the University of Glasgow
- Client sends TCP/IP packets, server responds with the headers it received, client analyses them
- Goal: measure as many paths as possible — diverse set of ISPs and locations



https://smcquistin.uk/tcp-measurements

Summary

- Are there middleboxes deployed that aren't compatible with TCP Hollywood?
- More broadly, what innovation *is* possible at the transport-layer?
- Please contact us — we're looking to hear about middlebox deployments, and run measurements

Stephen McQuistin
sm@smcquistin.uk

https://smcquistin.uk/tcp-measurements