

Evolving the Transport-Layer

Stephen McQuistin and Colin Perkins
University of Glasgow

Marwan Fayed
University of Stirling

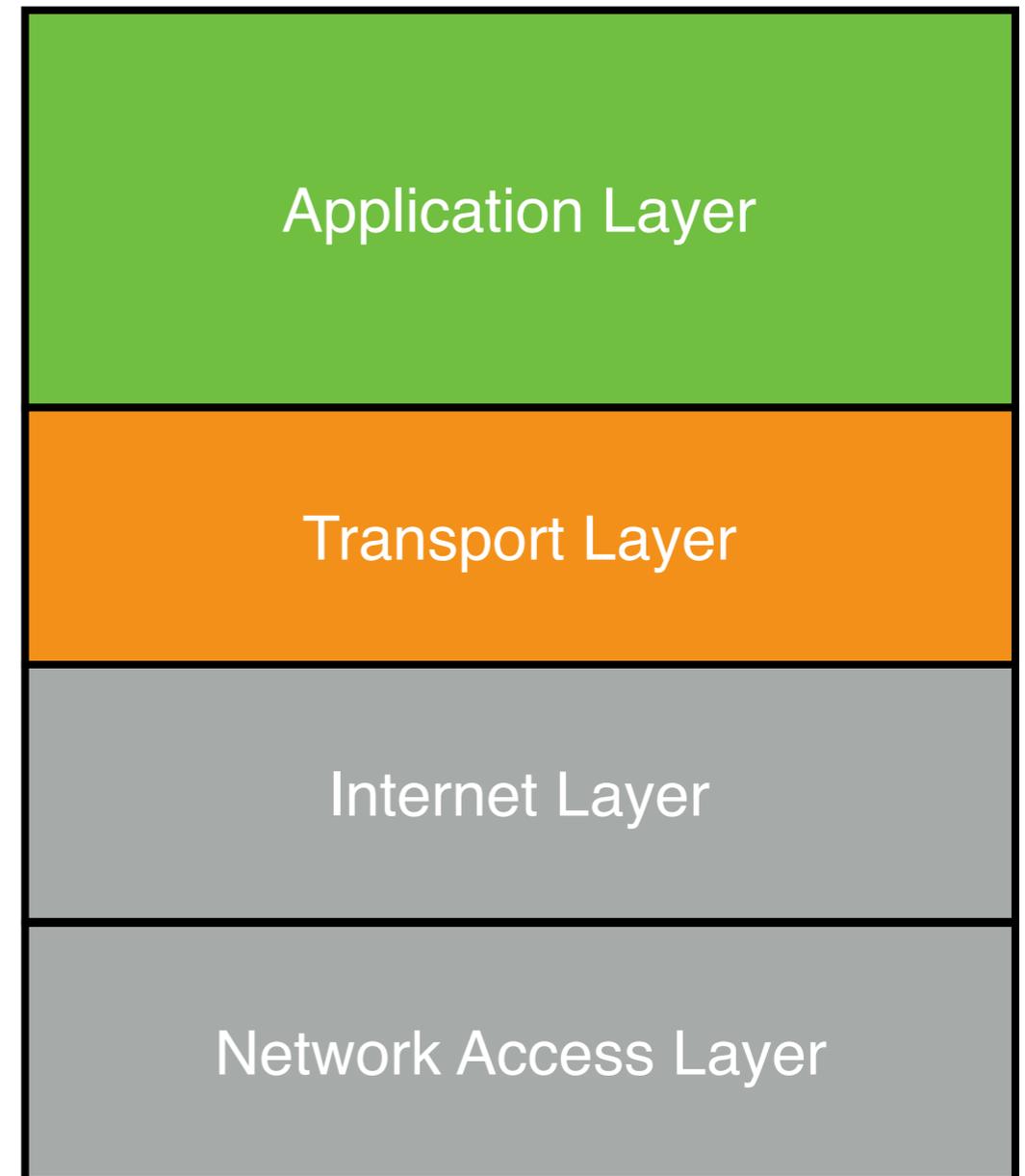
UKNOF36
January 2017, London

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

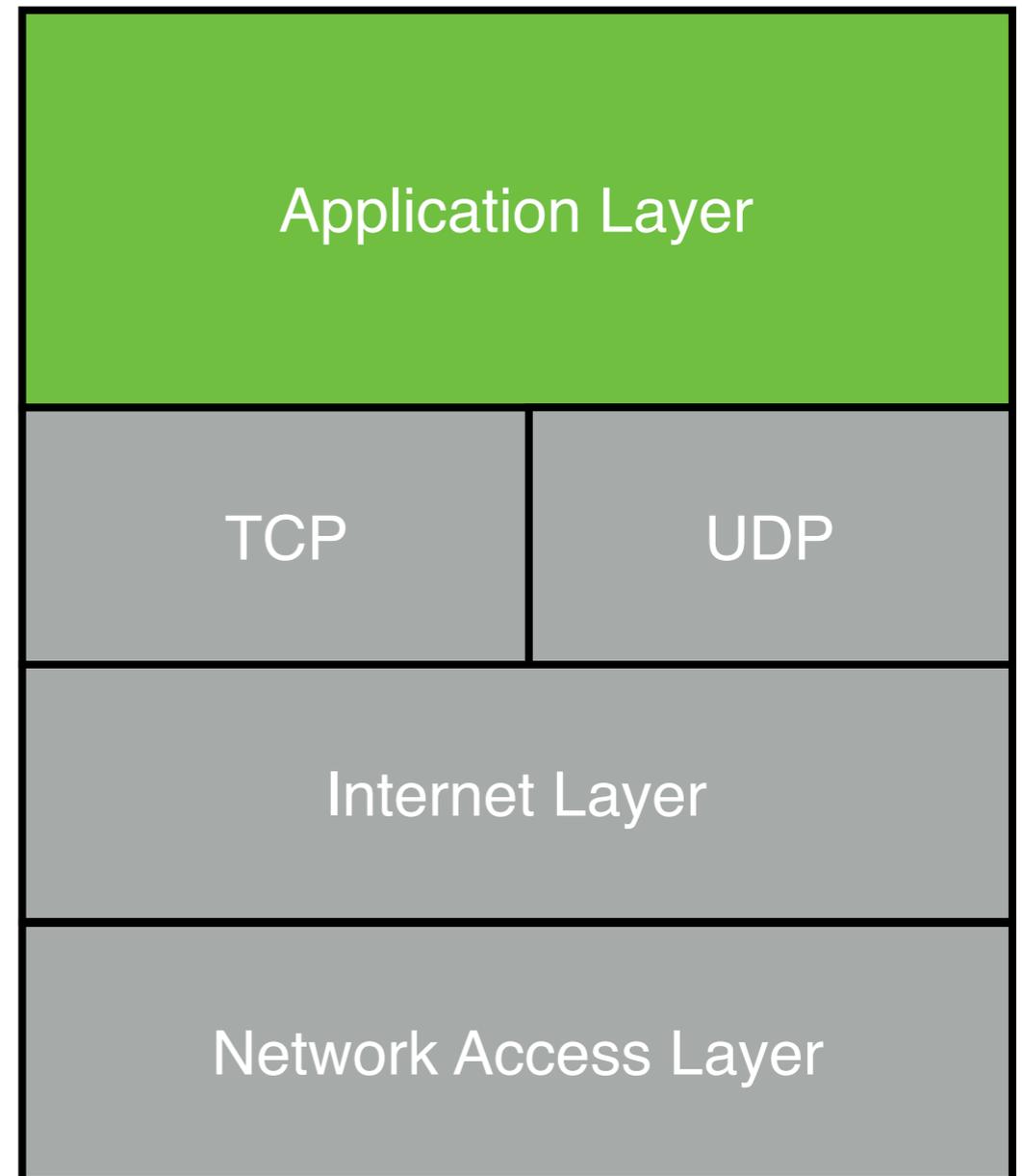
Transport constrained to TCP/UDP

- 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

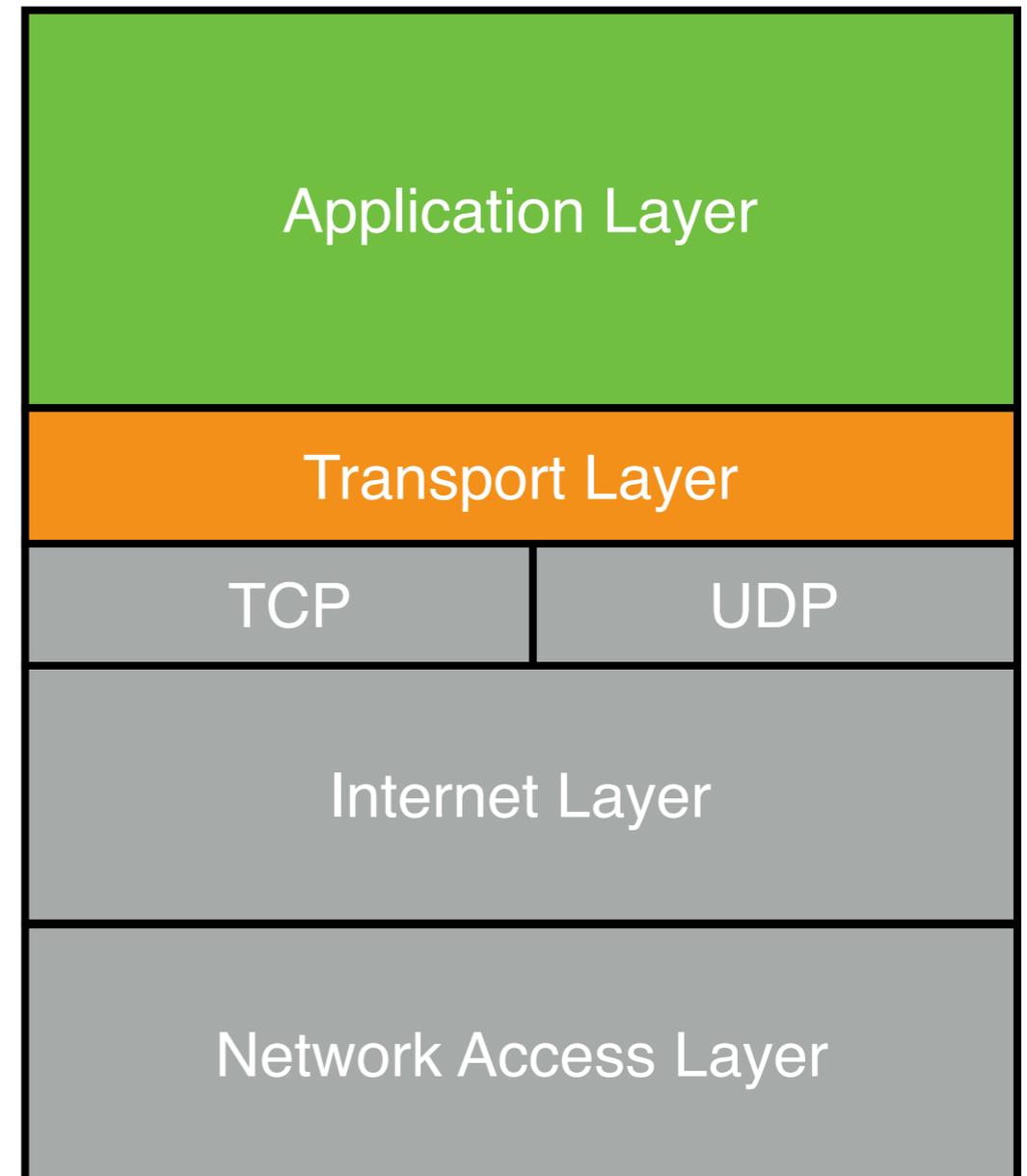


TCP and UDP aren't enough

- 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

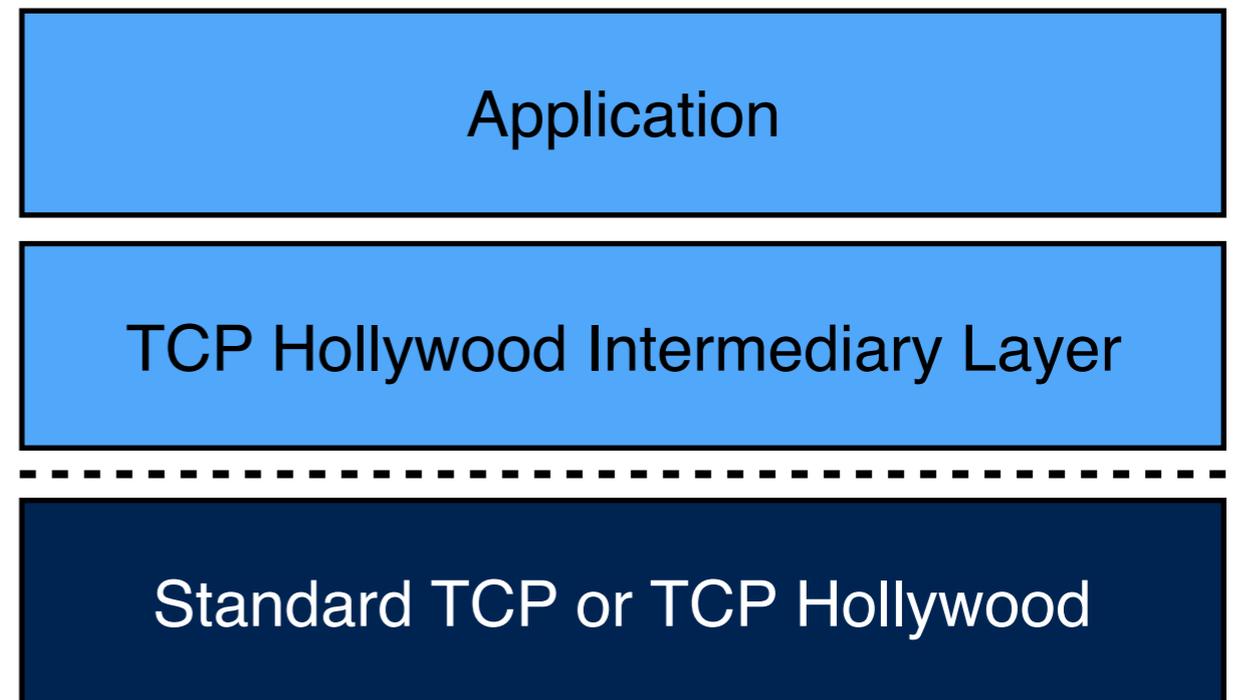
- Uses TCP as a substrate, but modified to reduce latency
- Message-oriented, to allow application data units to be sent
- Unordered, given messages are independently useful
- Partially reliable, based on metadata provided by the application



“Hollywood Sign”, Gnaphron - CC BY-SA 2.0
[flickr.com/photos/gnaphron/8485145044](https://www.flickr.com/photos/gnaphron/8485145044/)

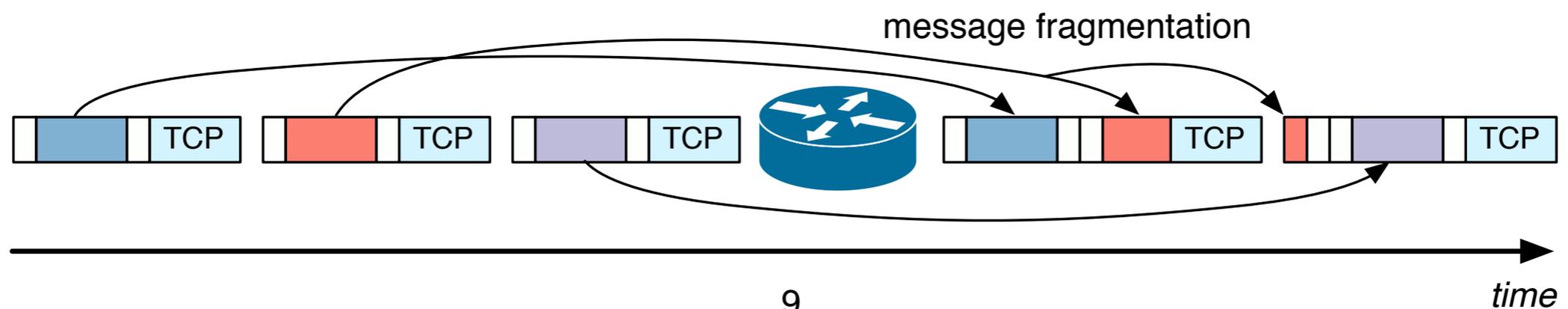
TCP Hollywood: Architecture

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



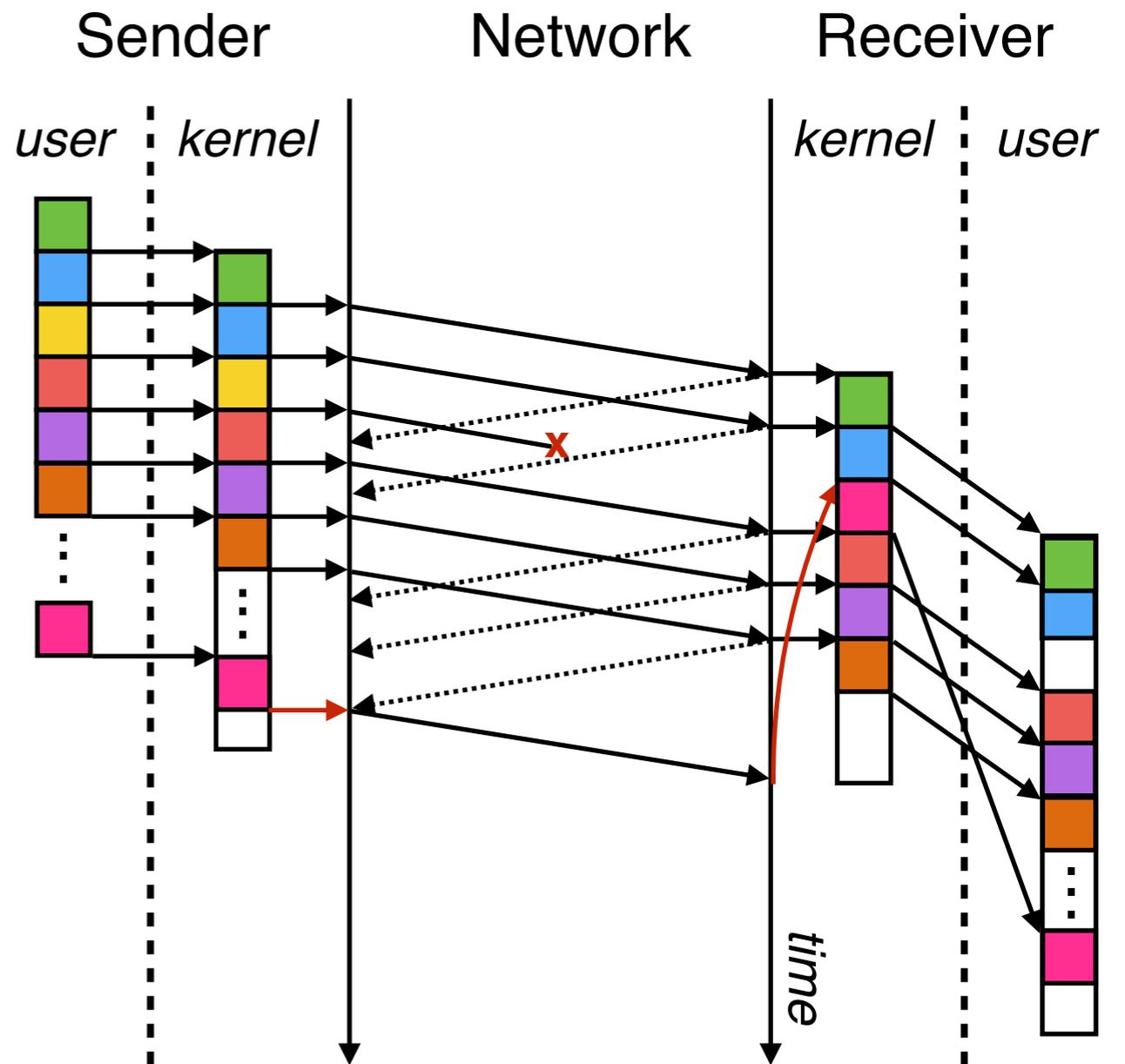
TCP Hollywood: Messaging

- 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

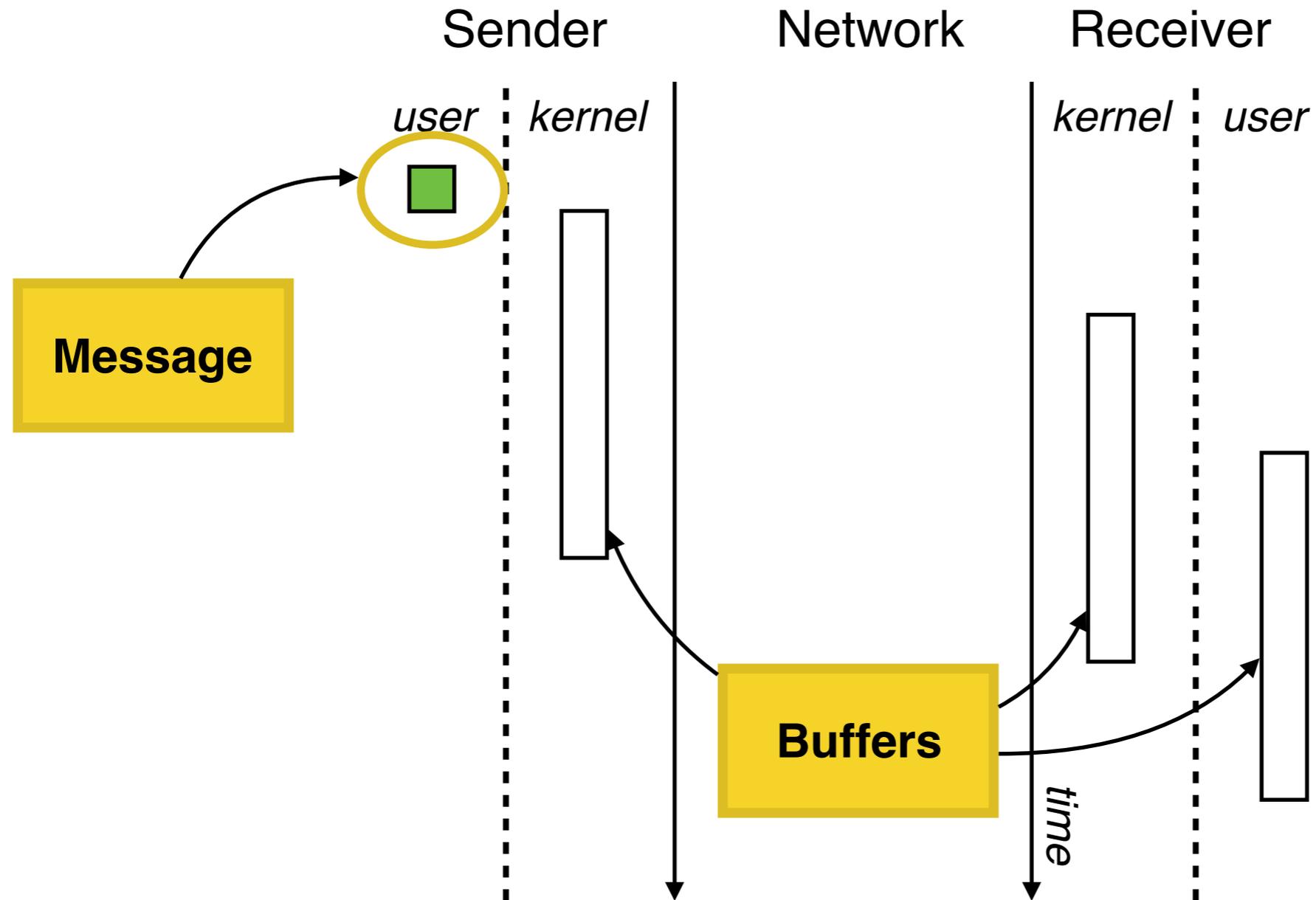


TCP Hollywood: Partial reliability

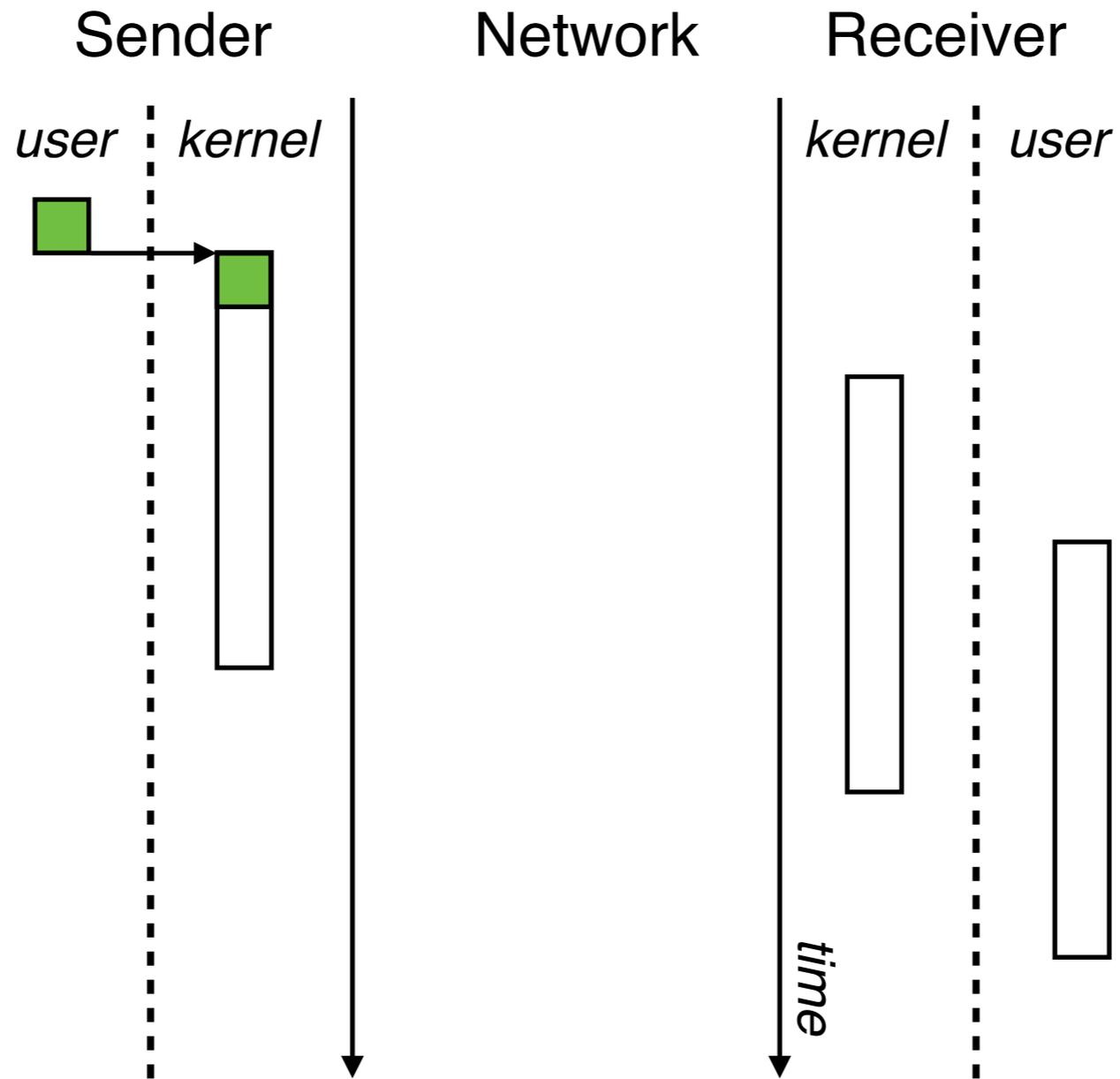
- Messages might expire: estimated to arrive too late, or depend on an undelivered message
- Expired messages aren't retransmitted - next live message sent instead as an *inconsistent retransmission*
- Same length and TCP sequence number as standard TCP — payload is different



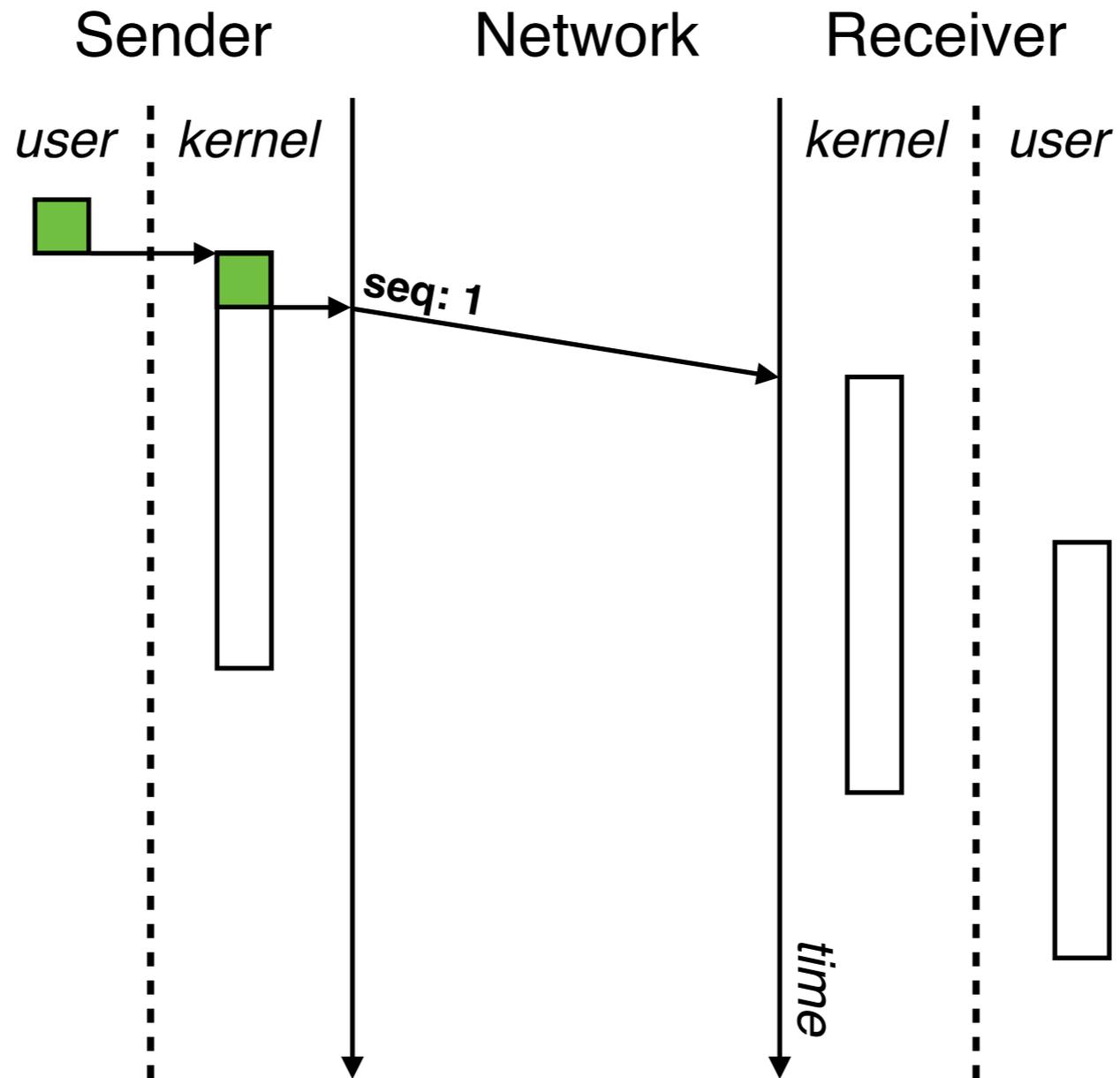
TCP Hollywood in action



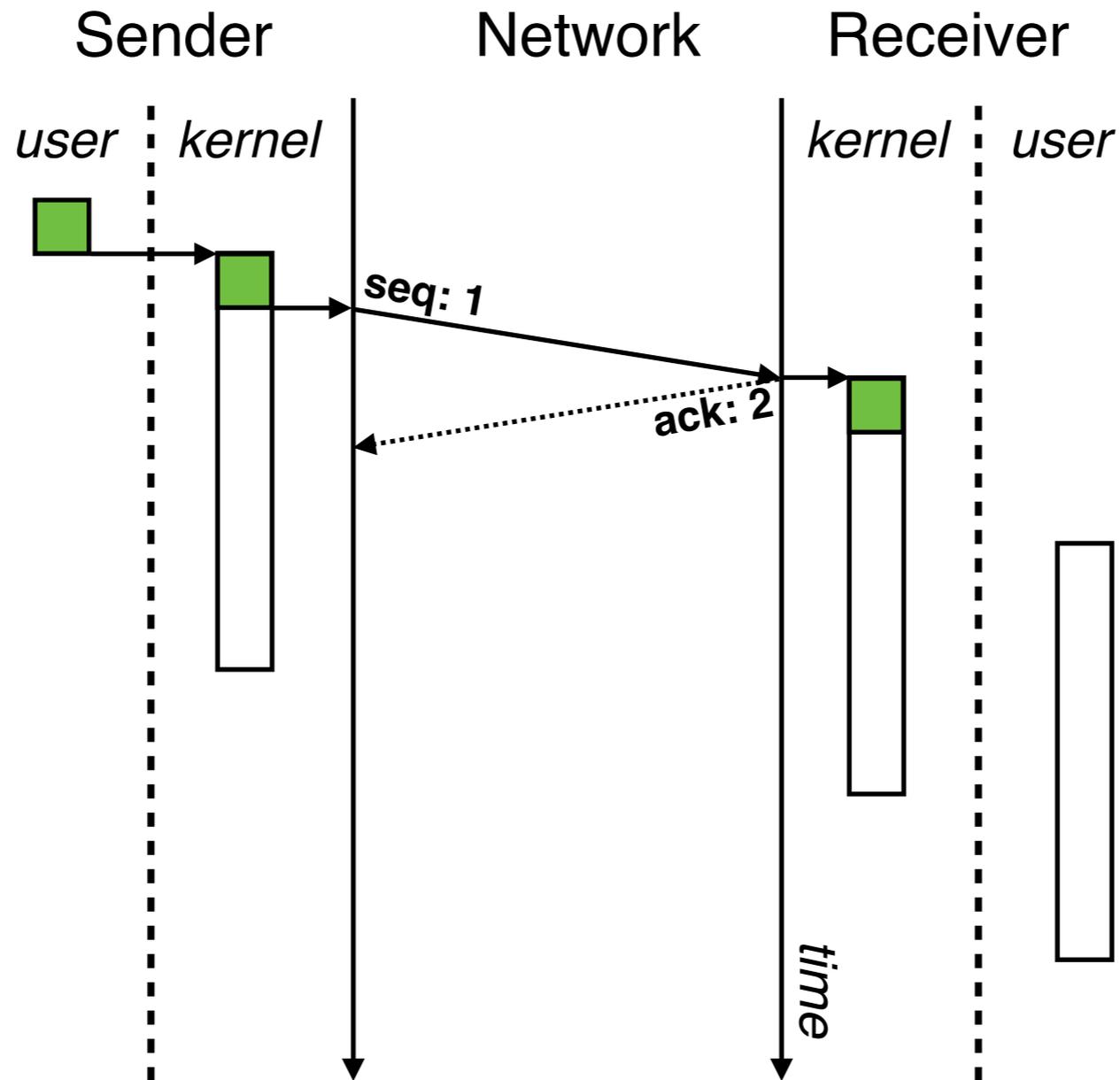
TCP Hollywood in action



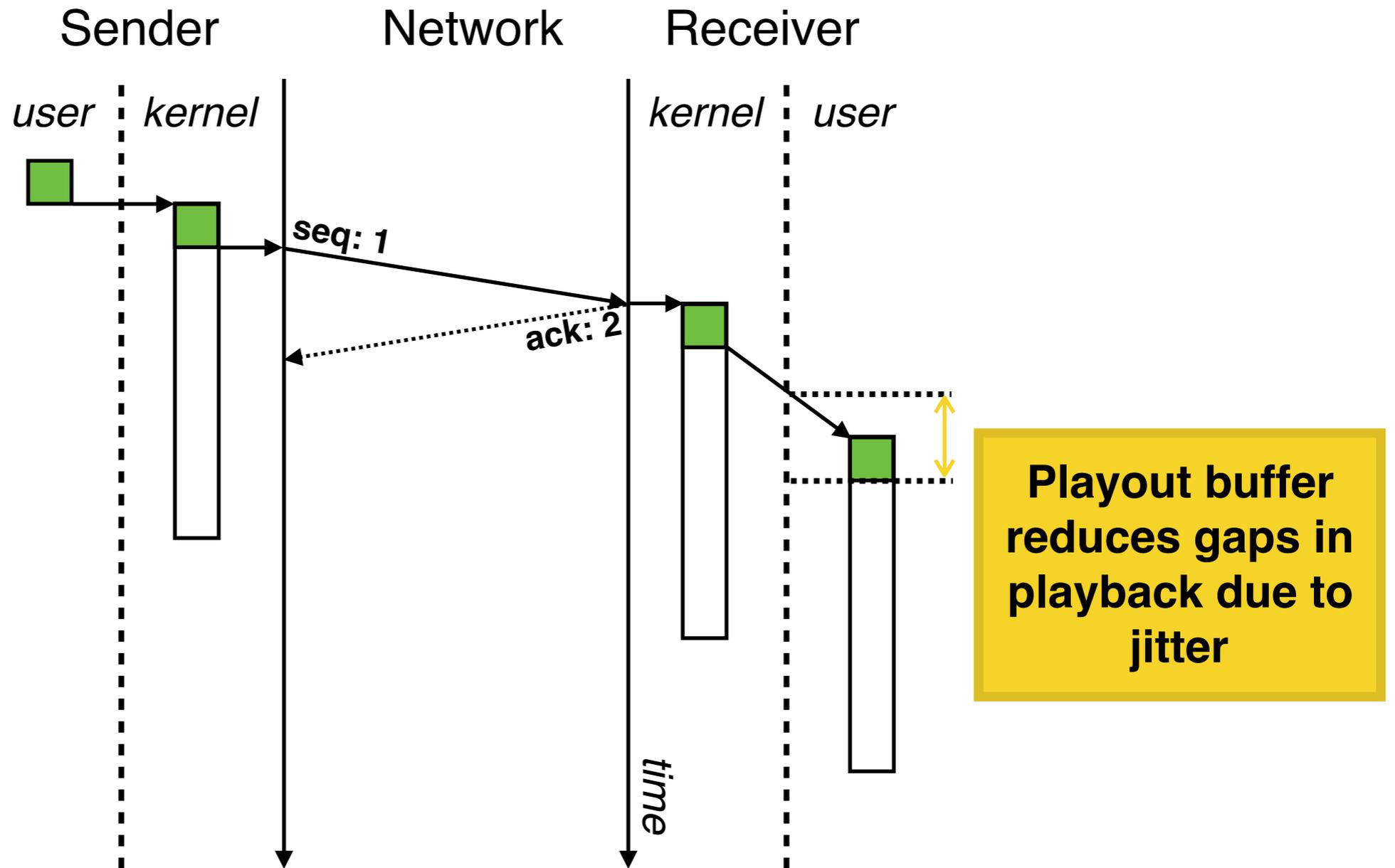
TCP Hollywood in action



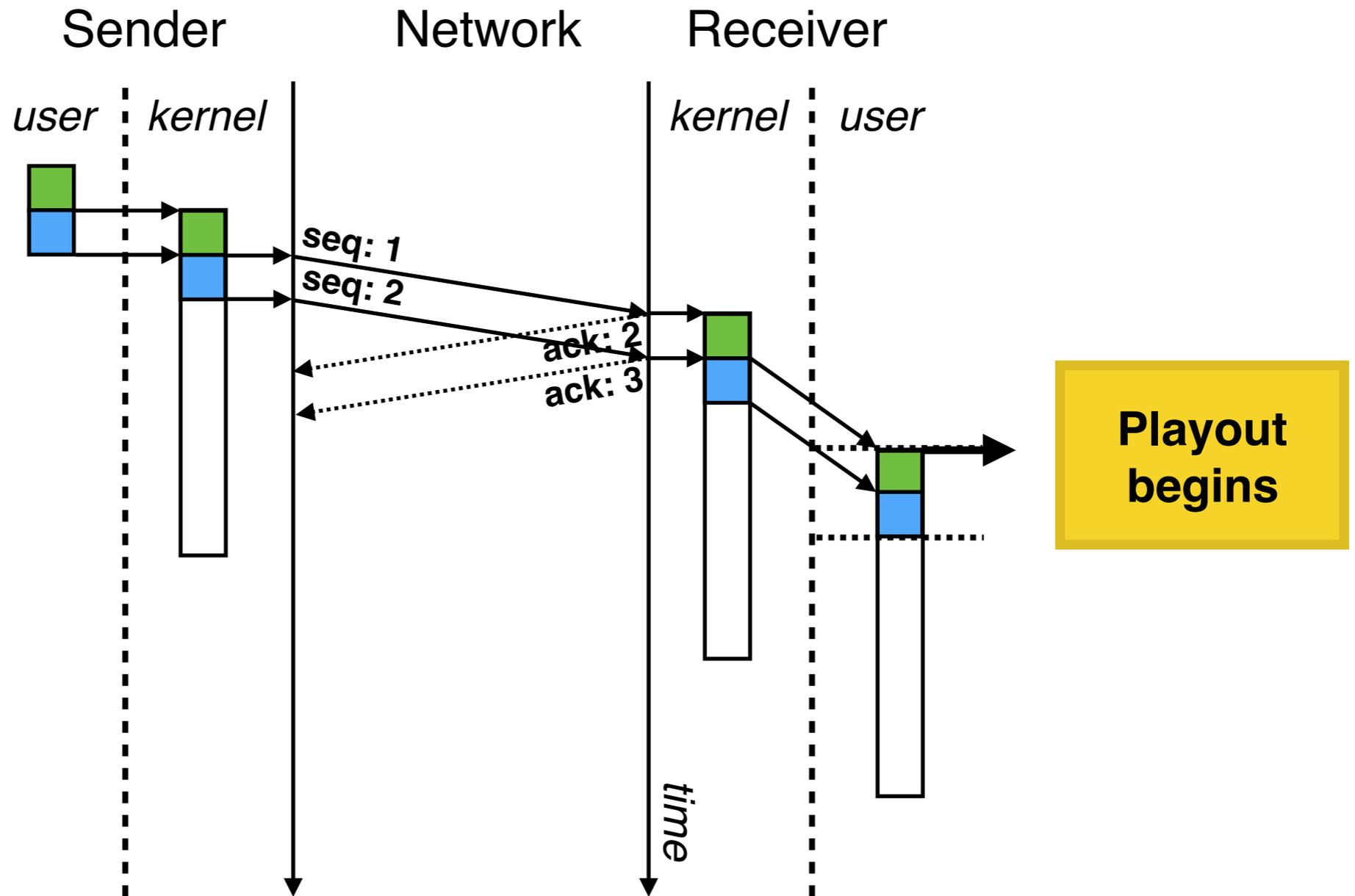
TCP Hollywood in action



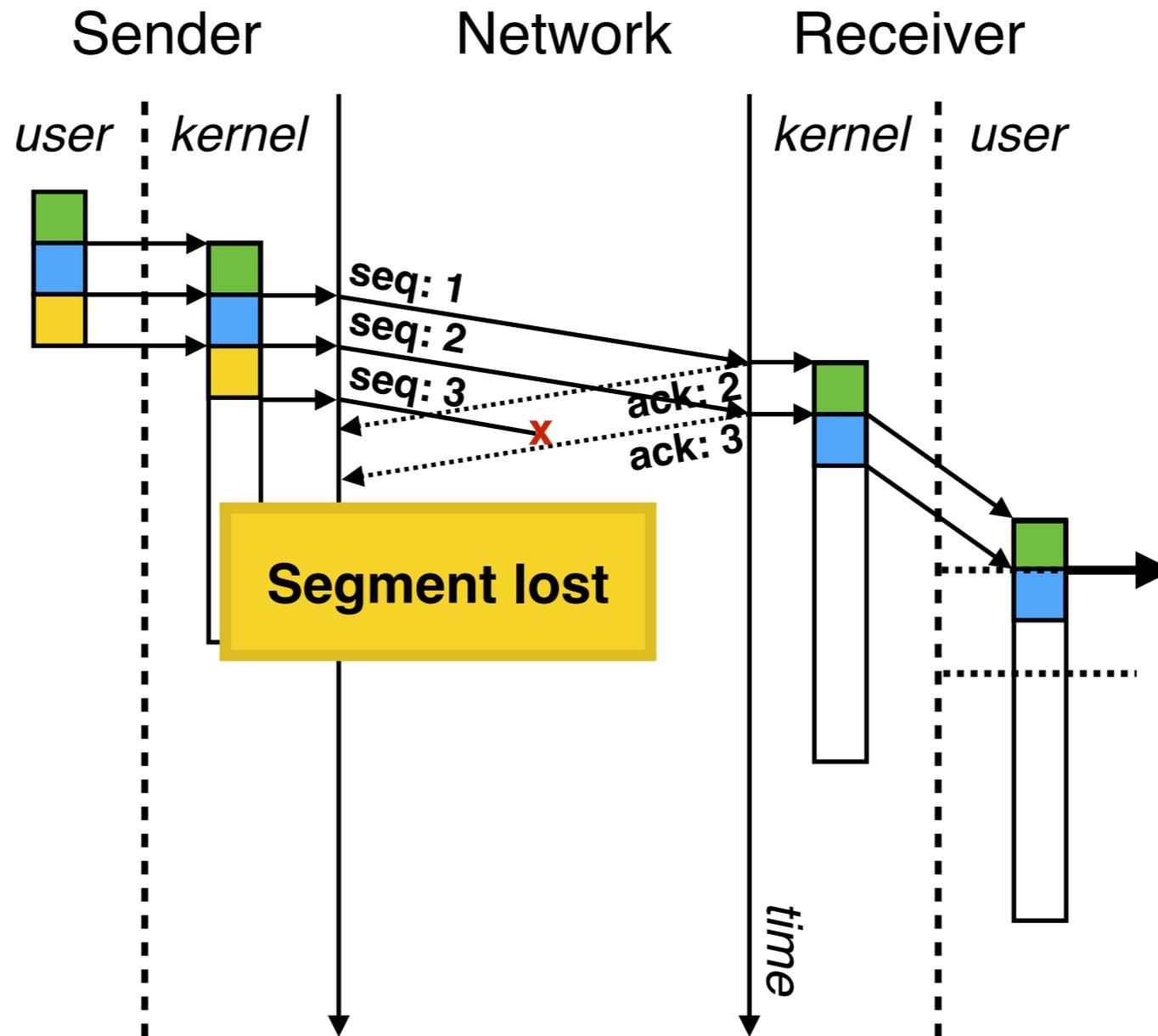
TCP Hollywood in action



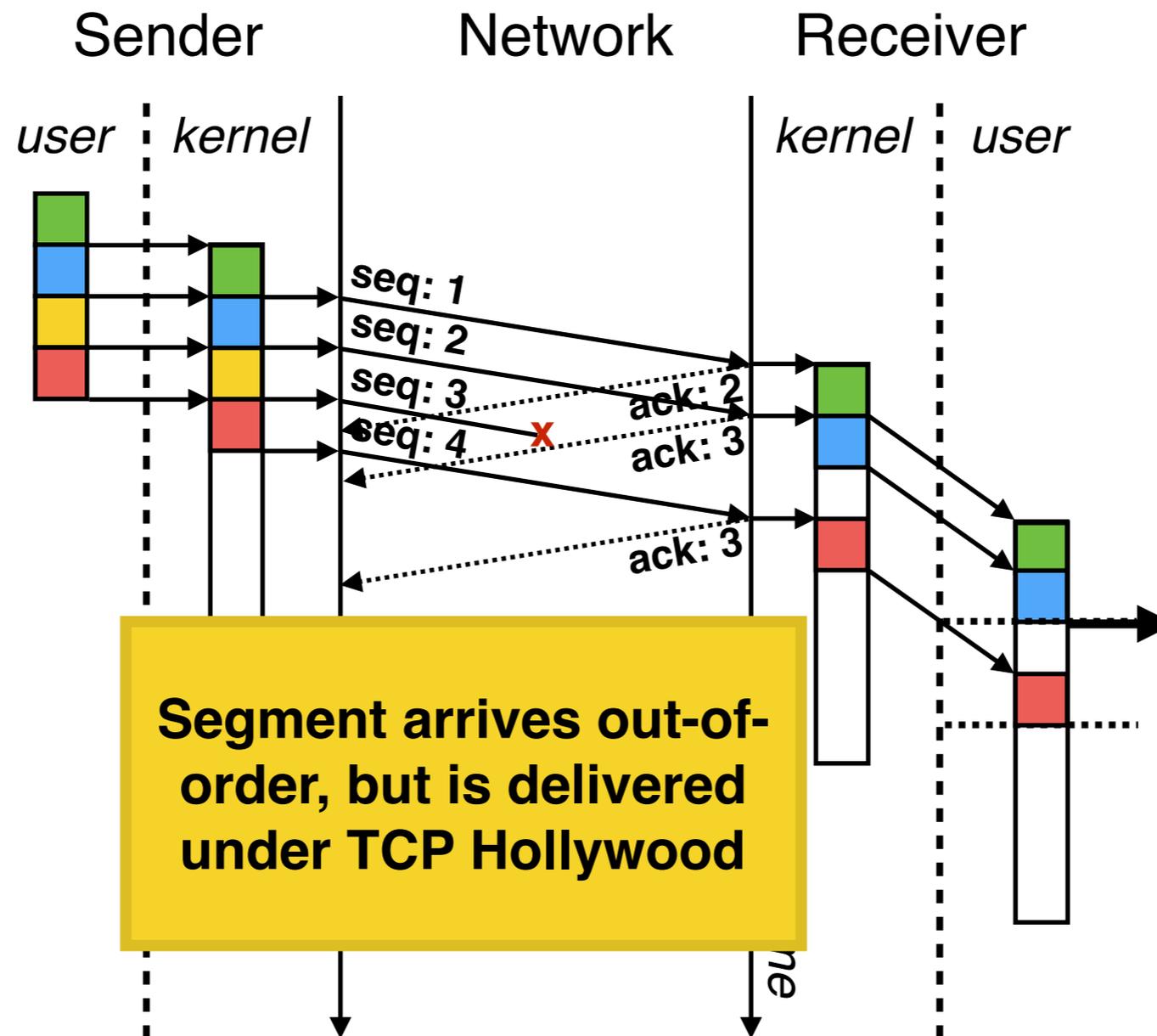
TCP Hollywood in action



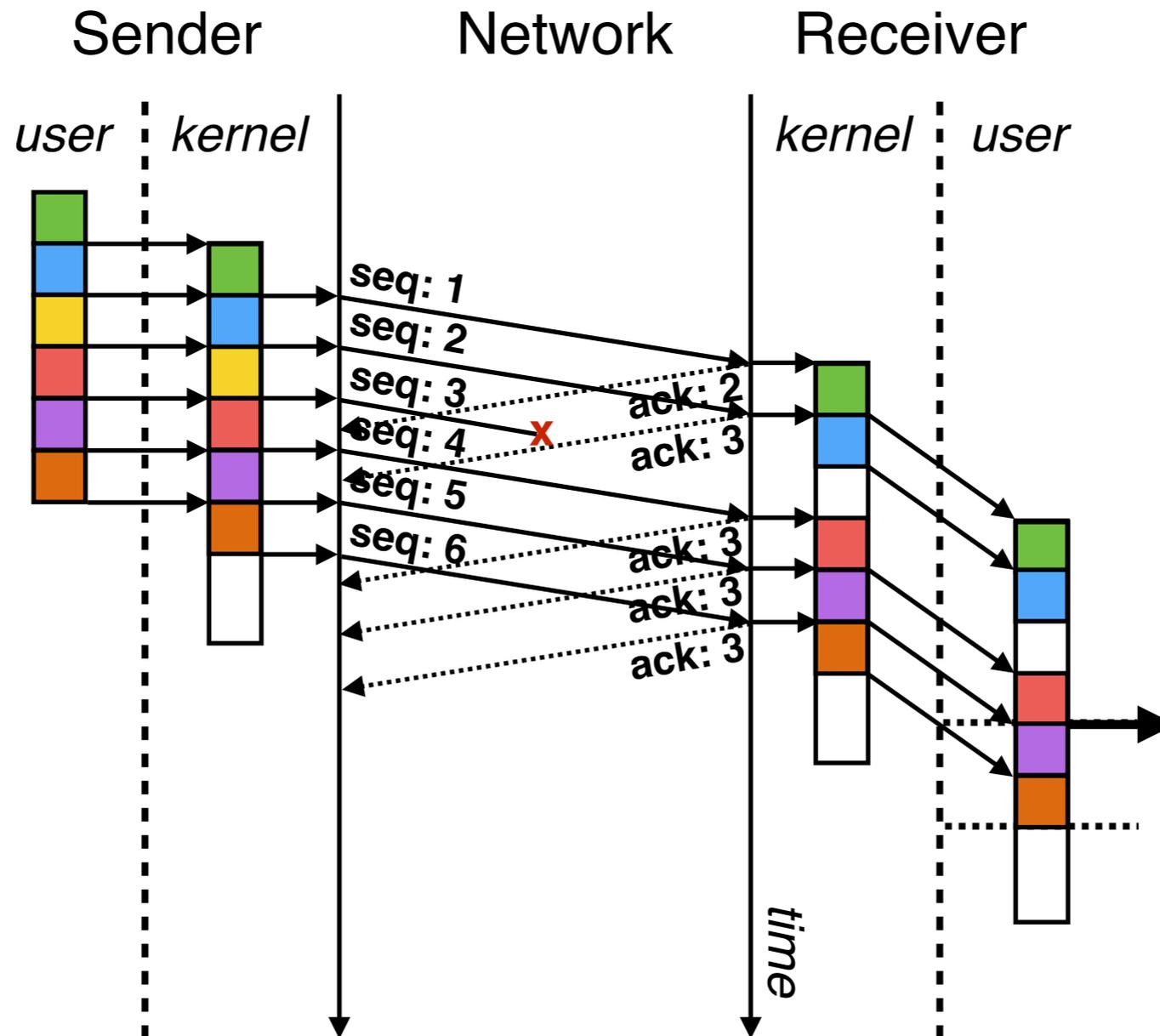
TCP Hollywood in action



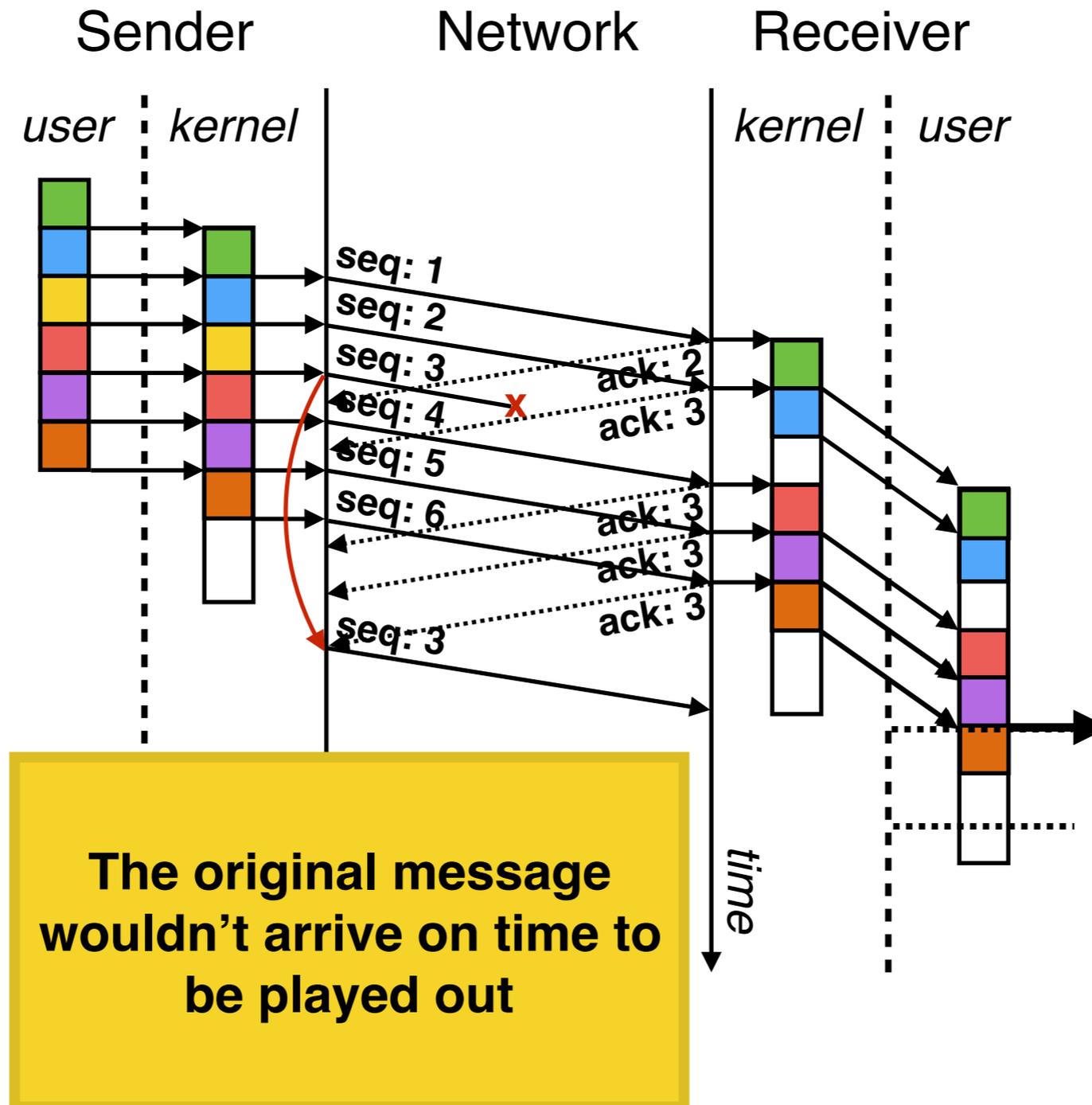
TCP Hollywood in action



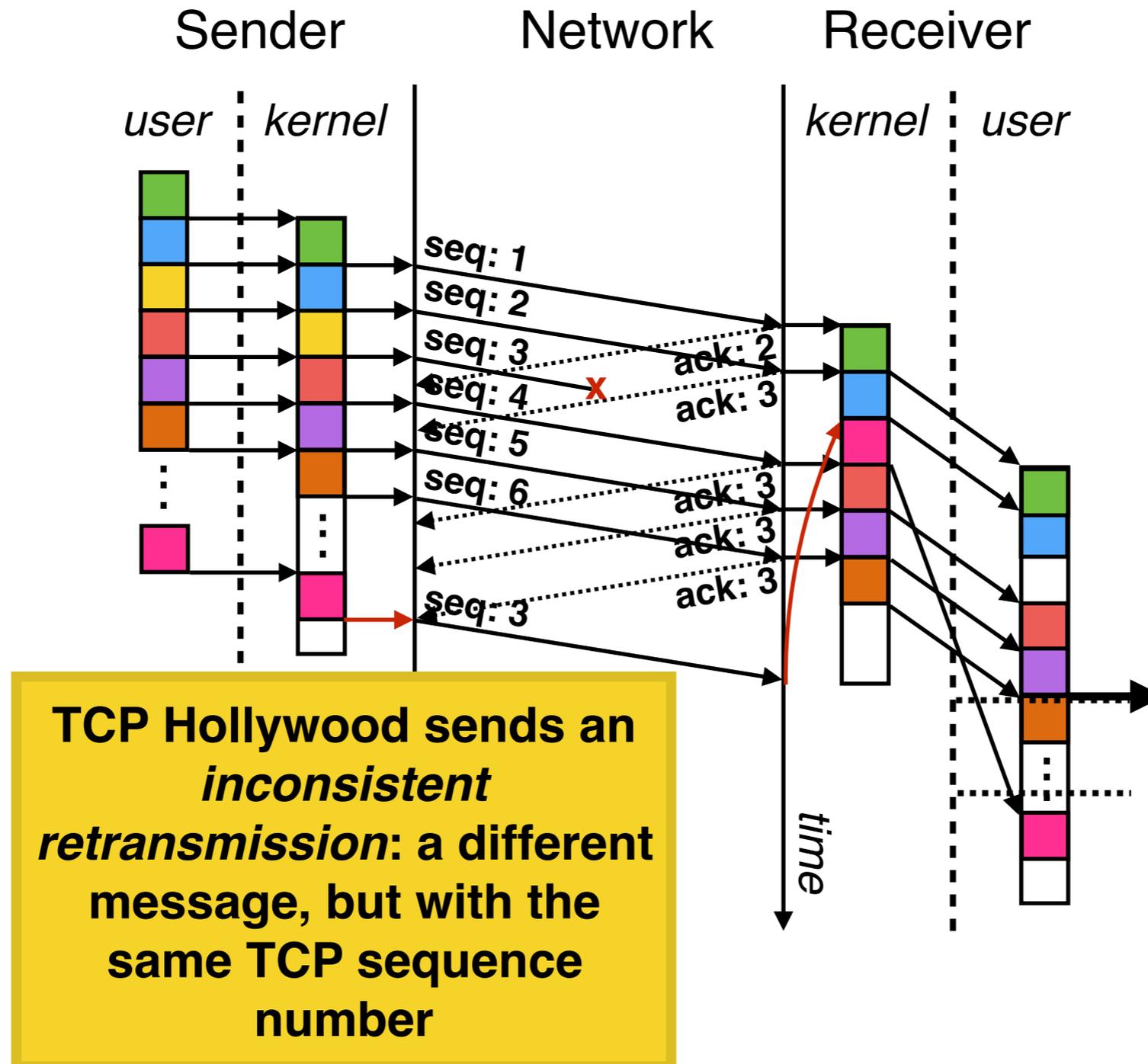
TCP Hollywood in action



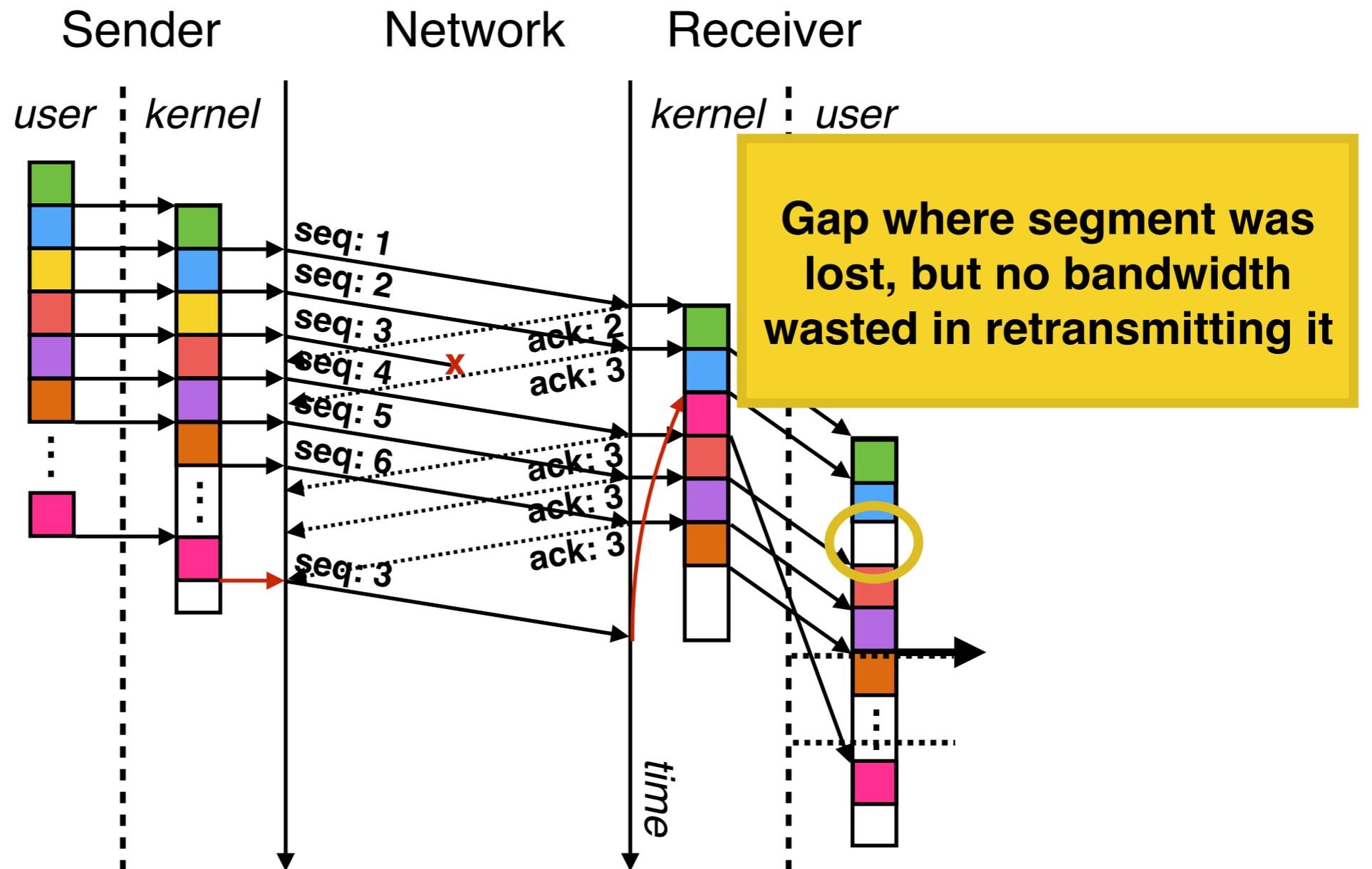
TCP Hollywood in action



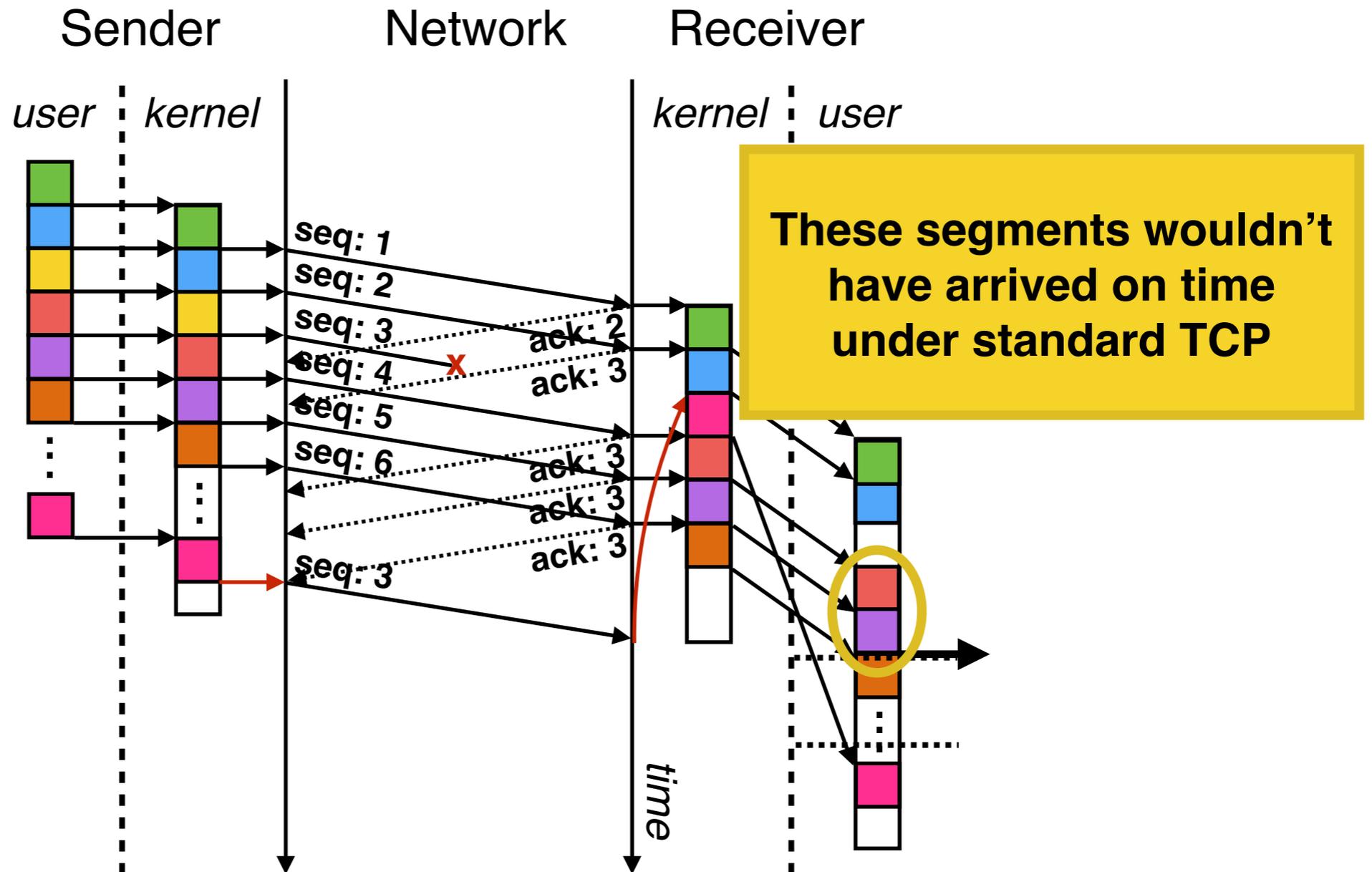
TCP Hollywood in action



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	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



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

TCP Extensibility Measurements

[About](#) - [Ethics Approval](#) - [Submit Log](#)

The aim of this experiment is to determine the extent to which Internet protocols can be changed. A range of modifications to the TCP/IP protocol, one of the infrastructure components that form the basis of the Internet, will be tested.

The experiment will evaluate how the Internet - that is, the network devices between sender and receiver - responds to these changes. It is hoped that the results of the experiments will help with future research into protocol design.

The experiment involves sending and receiving data from servers on the Internet. To do this, you will be asked to run software on your computer. The software will connect to a range of servers (operated by the experimenters), send data to these servers, and record the response it receives. The data sent by the software is designed to test how devices in the network respond to changes in the Internet protocols being used. Recording the data sent and received by the software allows us to determine if, and how, devices in the network have interacted with the data sent.

The software does not interfere with or record any other network traffic. Only the data it sends, and receives in response, is recorded.

The software will ask you to provide information about the network it is being run across. When prompted by the software, enter the country the network is located in, and the network type. The software will provide further instructions, including how to determine the network type.

[Download the measurement software](#)

If you have any questions about this experiment, please contact s.mcquistin.1@research.gla.ac.uk.

<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>