Oh Flow, Are Thou Happy? TCP Sendbuffer Advertising for Make Benefit of Clouds and Tenants

Alexandru Agache and Costin Raiciu

University Politehnica of Bucharest

Problem statement

- There is only so much we can find about about a connection by looking at in flight packets (losses, retransmissions, RTT, etc.)

Problem statement

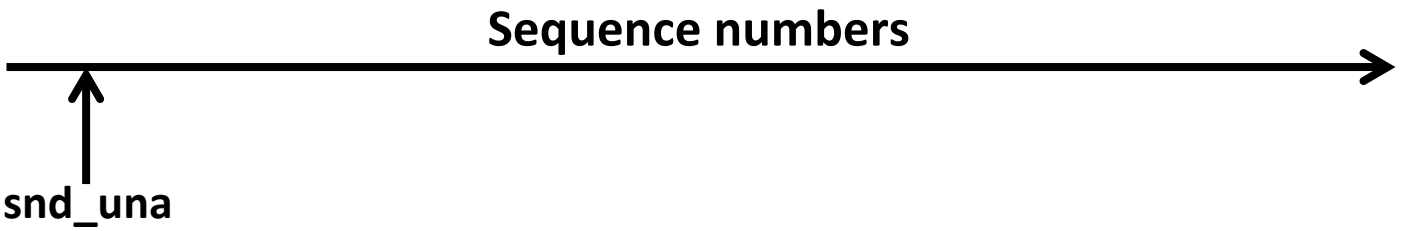
- There is only so much we can find about about a connection by looking at in flight packets (losses, retransmissions, RTT, etc.)

- Other information is more elusive: ***is the connection limited by the network ?***

What if we advertised send buffer occupancy inside TCP segments ?

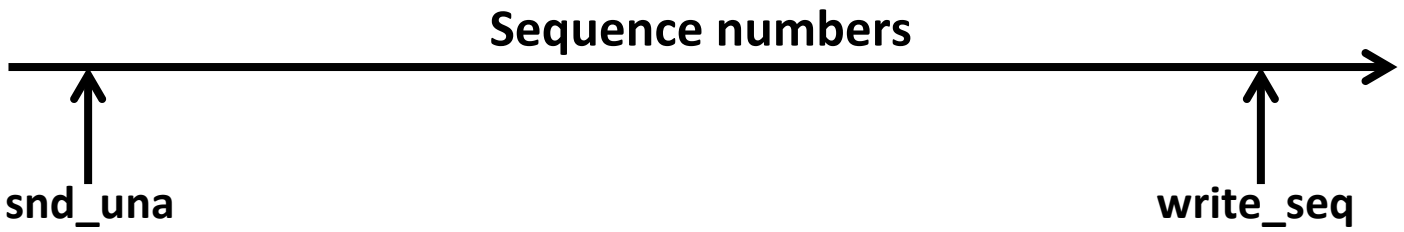
What exactly do we advertise ?

What exactly do we advertise ?



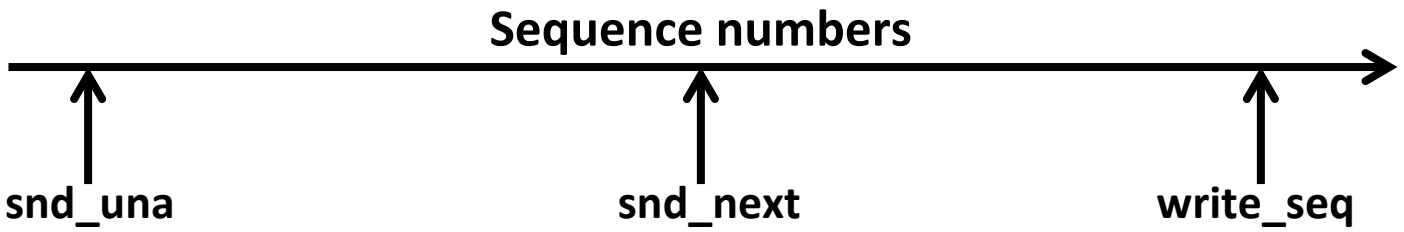
first unacknowledged
sequence number

What exactly do we advertise ?



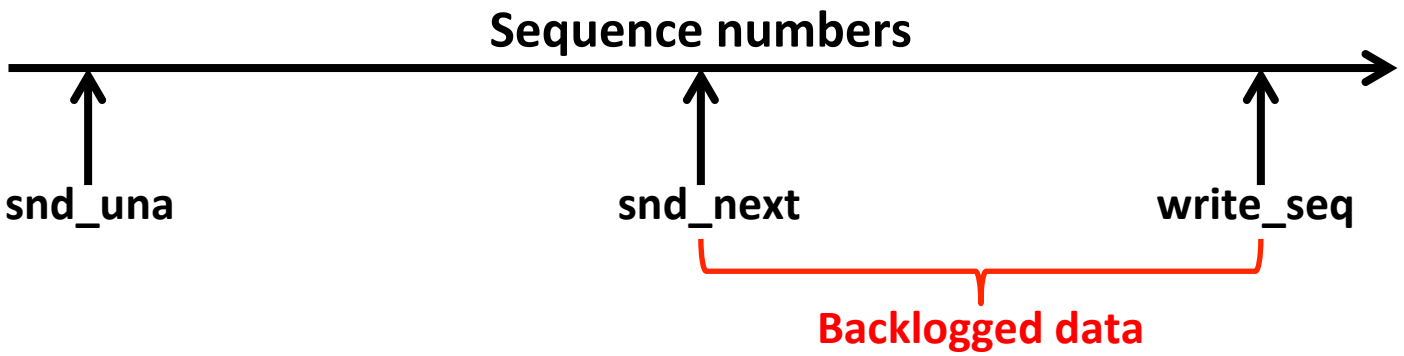
sequence number of the last written byte

What exactly do we advertise ?



sequence number for the
next packet to be sent

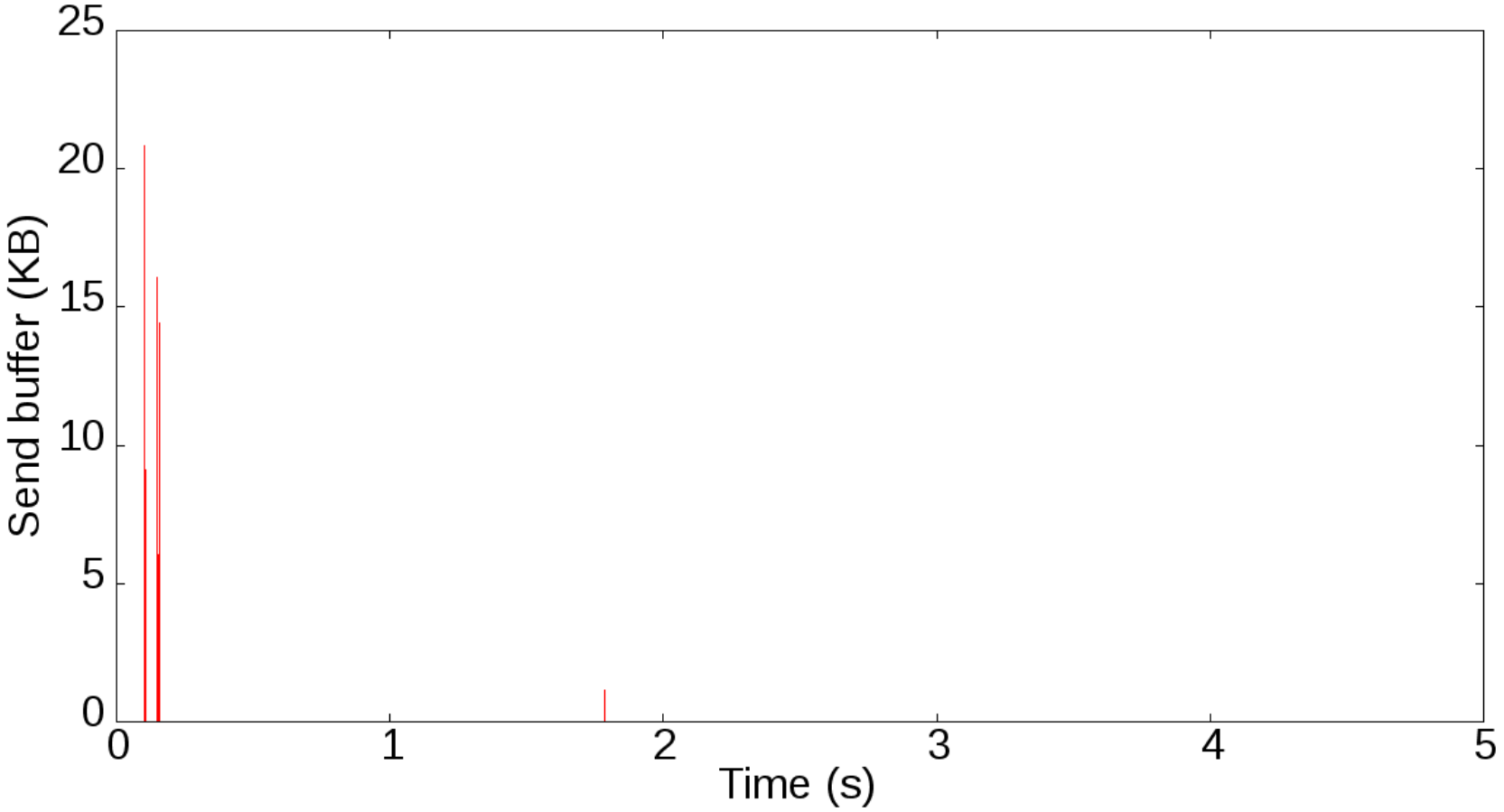
What exactly do we advertise ?



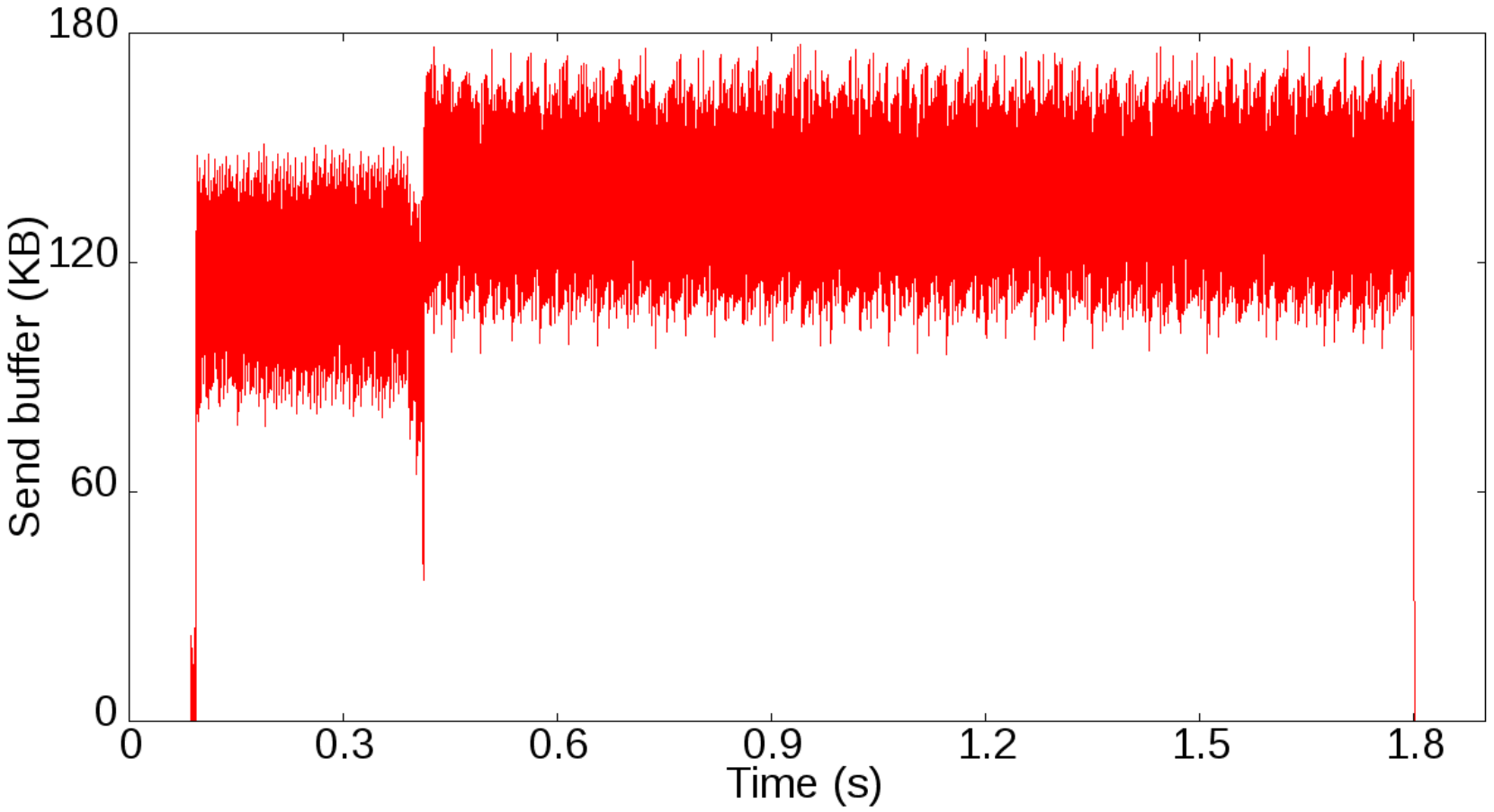
Why do we do it ?

- Backlogged applications are usually network-limited (unless receive window limited or facing very rare issues)
- Advertising the backlog size is more informative than checking a binary threshold

Disk bound transfer



Network bound transfer



Information encoding

- Simplest way is to use a TCP option

Information encoding

- Simplest way is to use a TCP option

(but it adds overhead and may interfere with hardware offloading in NICs)

Information encoding

- Simplest way is to use a TCP option

(but it adds overhead and may interfere with hardware offloading in NICs)

- We use the Receive Window field for the value and one reserved bit for signaling

Use cases

Detecting network hotspots

- *High loss rate = congestion ?*

Detecting network hotspots

- *High loss rate = congestion ?*

Not really! Example: **incast**

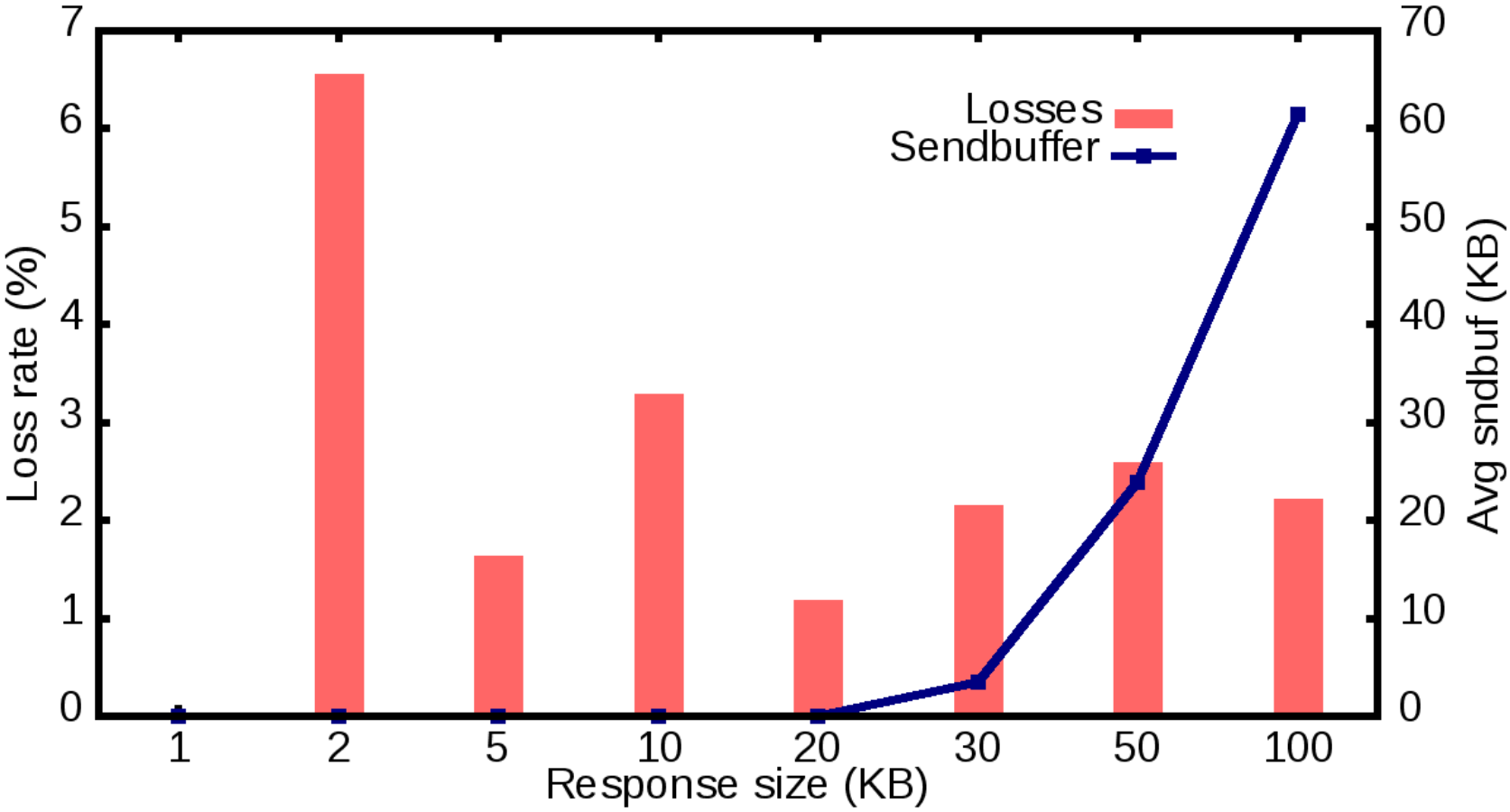
Detecting network hotspots

- *High loss rate = congestion ?*

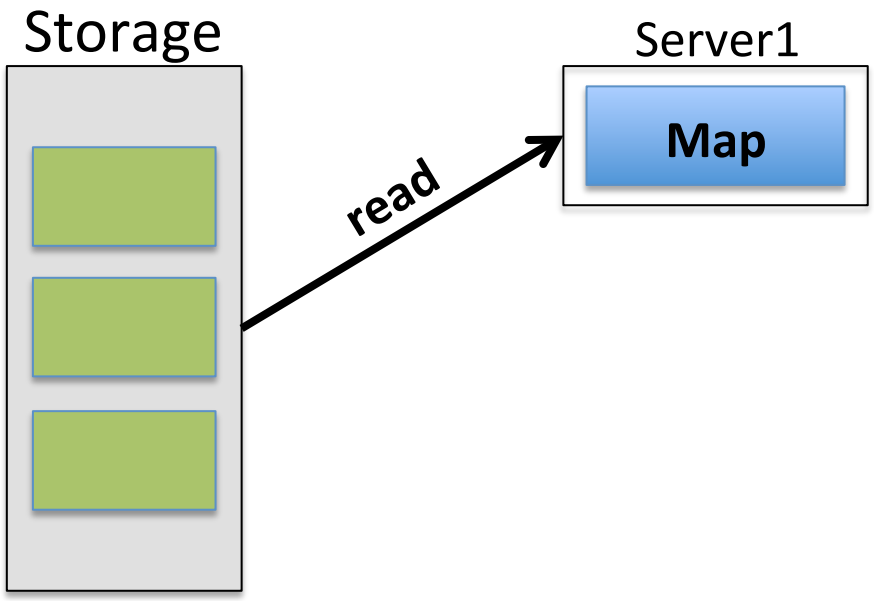
Not really! Example: **incast**

- EC2 incast scenario:
  - 99 synchronized senders and a single receiver
  - variable transfer size per round
  - average loss rate ~2.5%

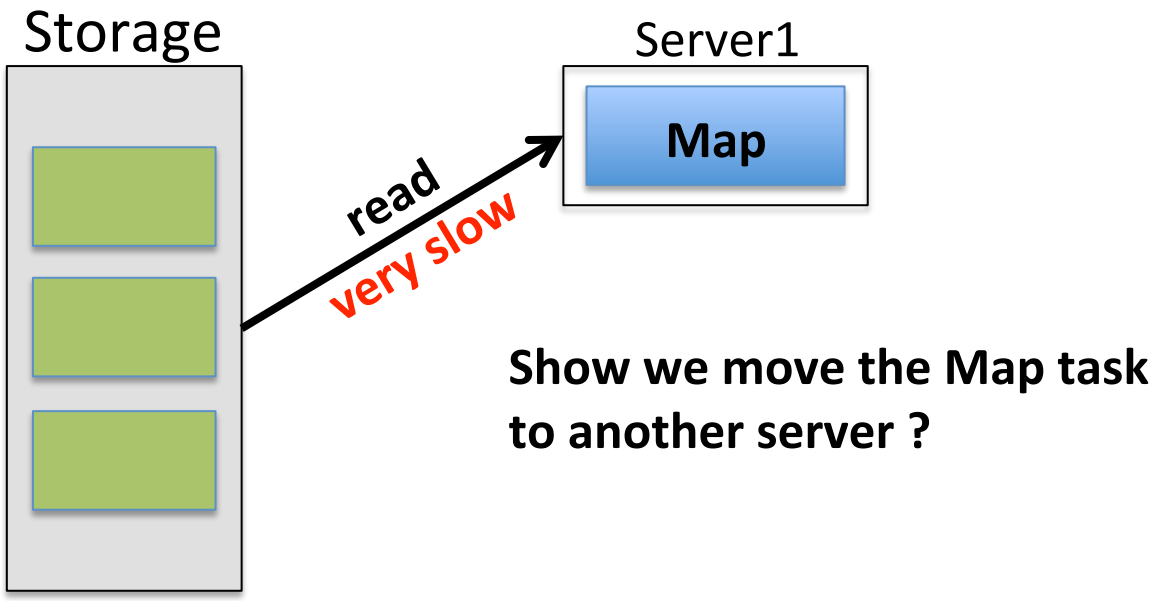
Incast results



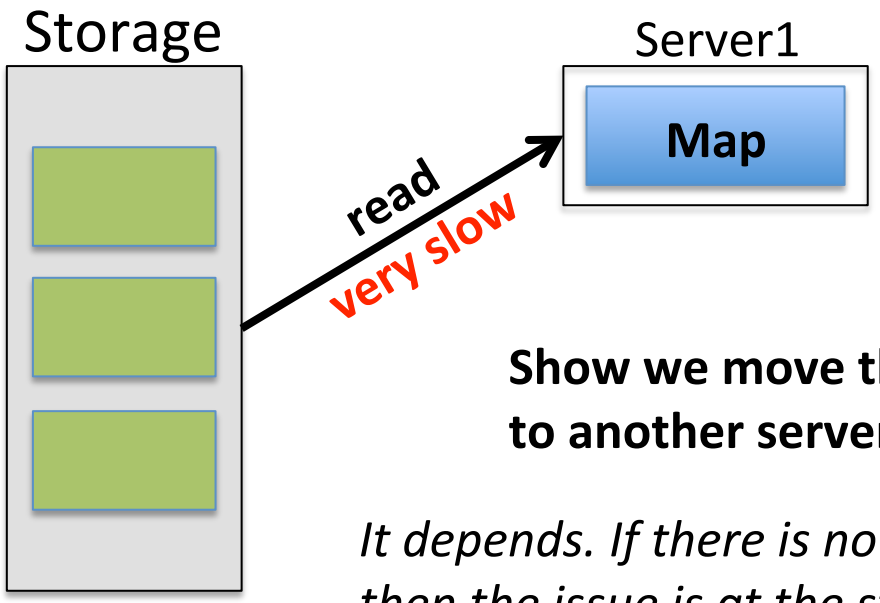
Helping applications



Helping applications



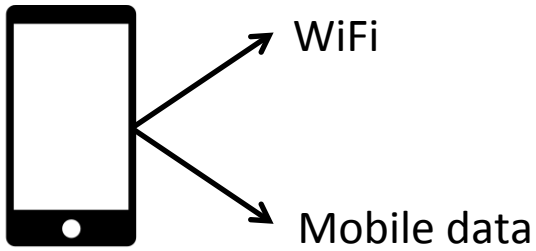
Helping applications



Show we move the Map task to another server ?

It depends. If there is no backlogged data, then the issue is at the storage node.

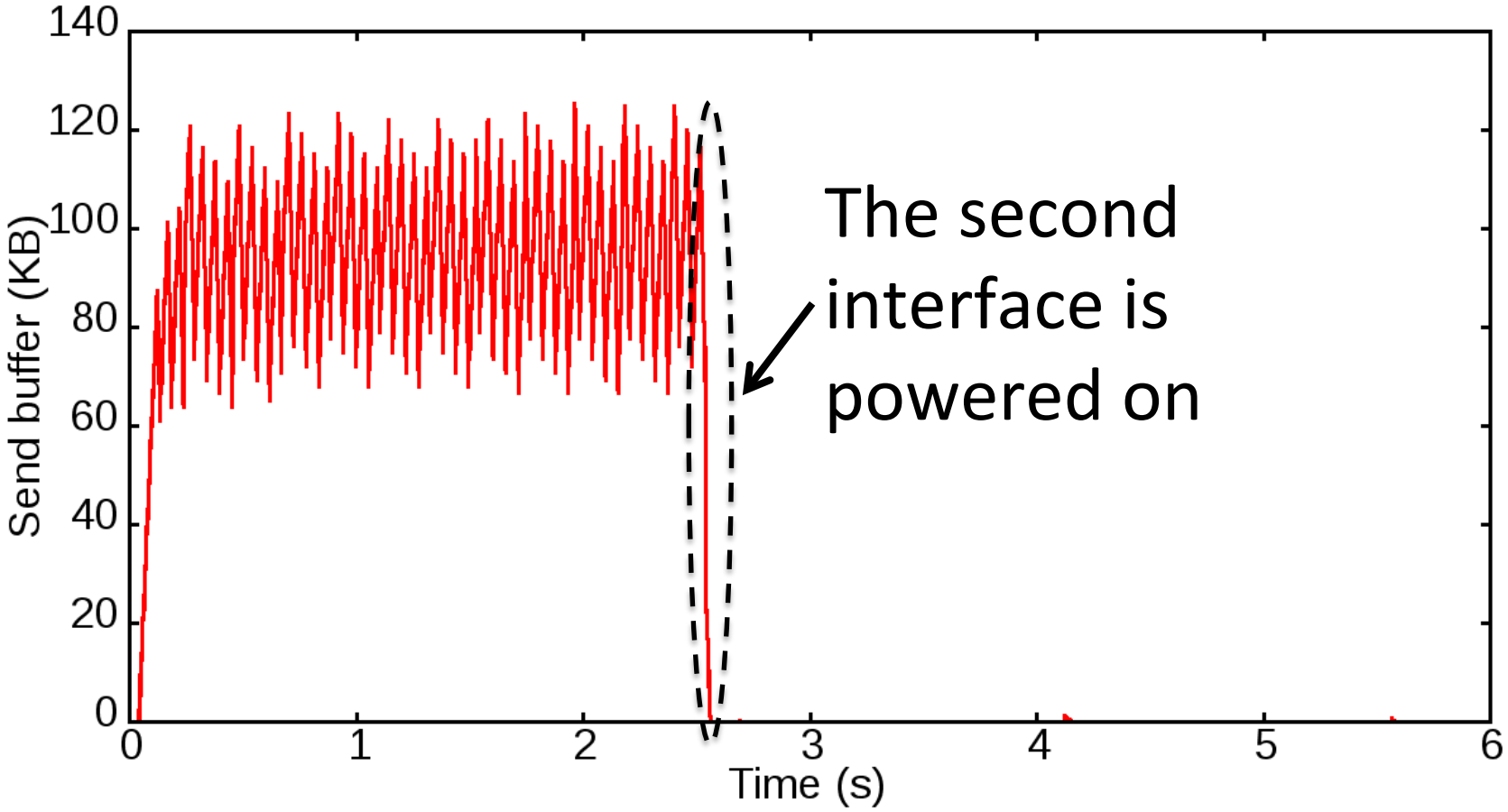
Improving mobile performance



Mobile data is generally not used if a WiFi network is available.

Some applications (video streaming for example) may benefit from also using the other interface, especially in poor network conditions.

Improving mobile performance

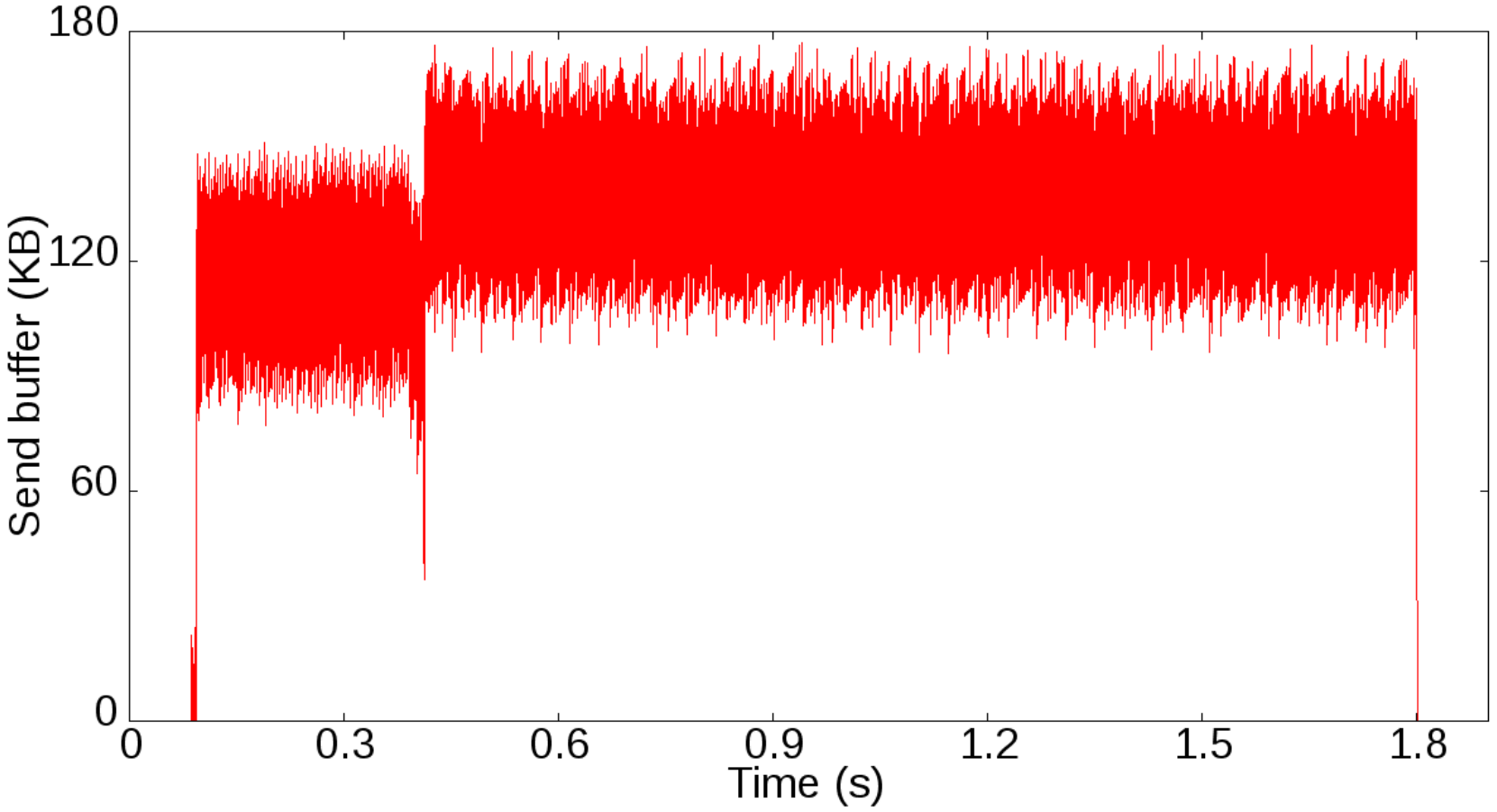


Troubleshooting flow performance

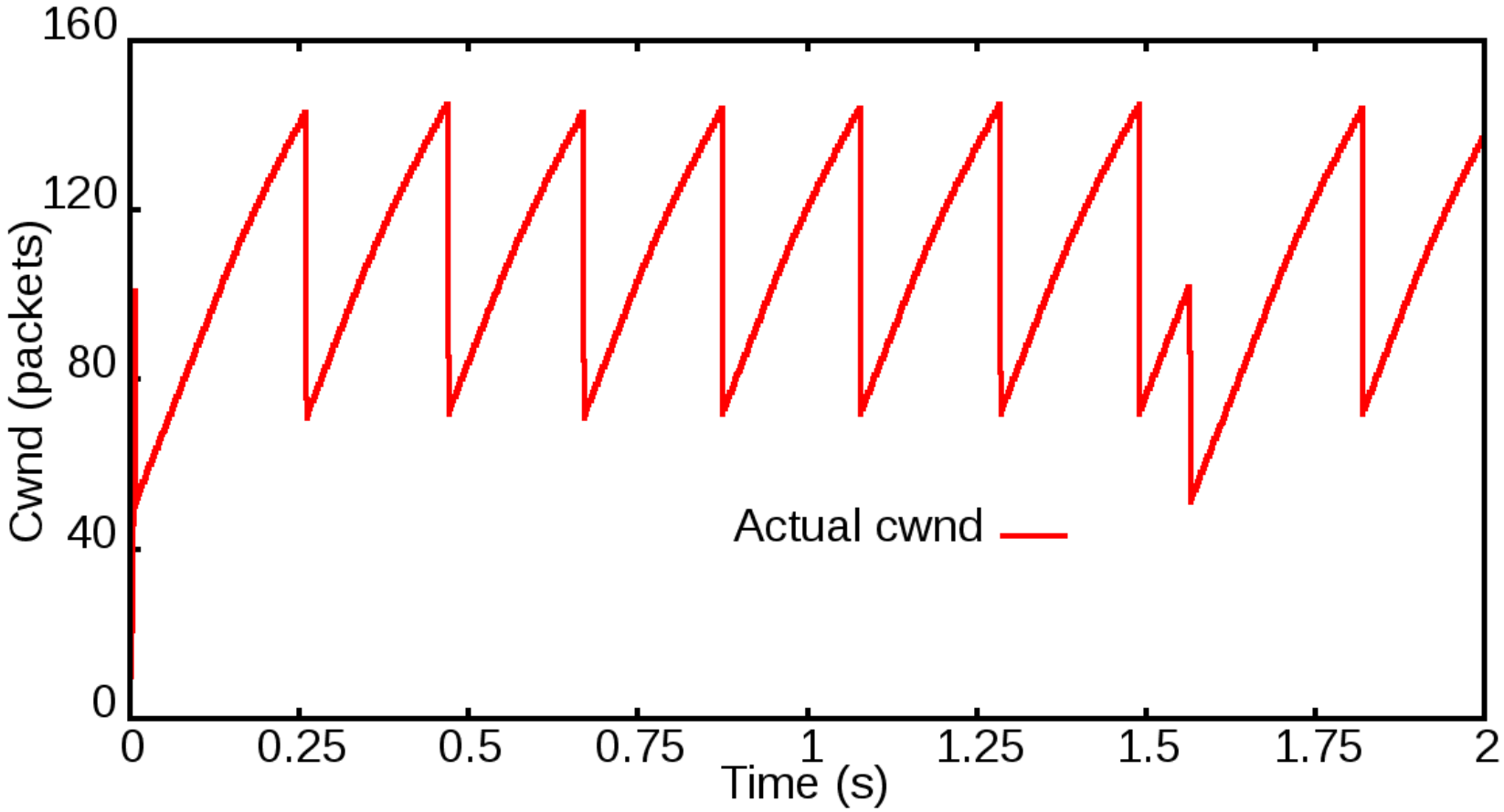
- We investigate the use of sendbuffer information to infer other flow characteristics

- For example, we try to estimate the presence of congestion events by analysing the evolution of the sendbuffer

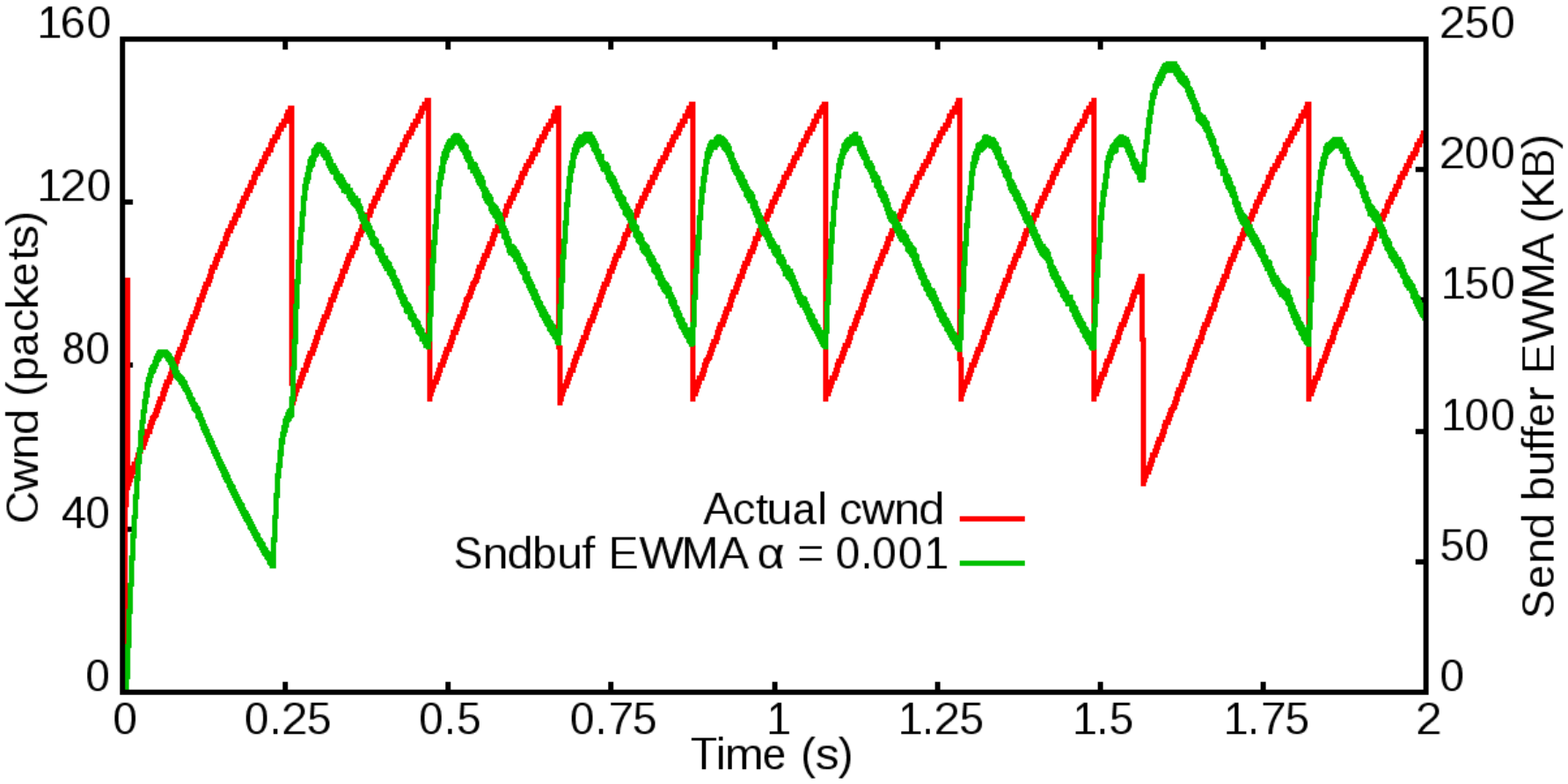
Troubleshooting flow performance



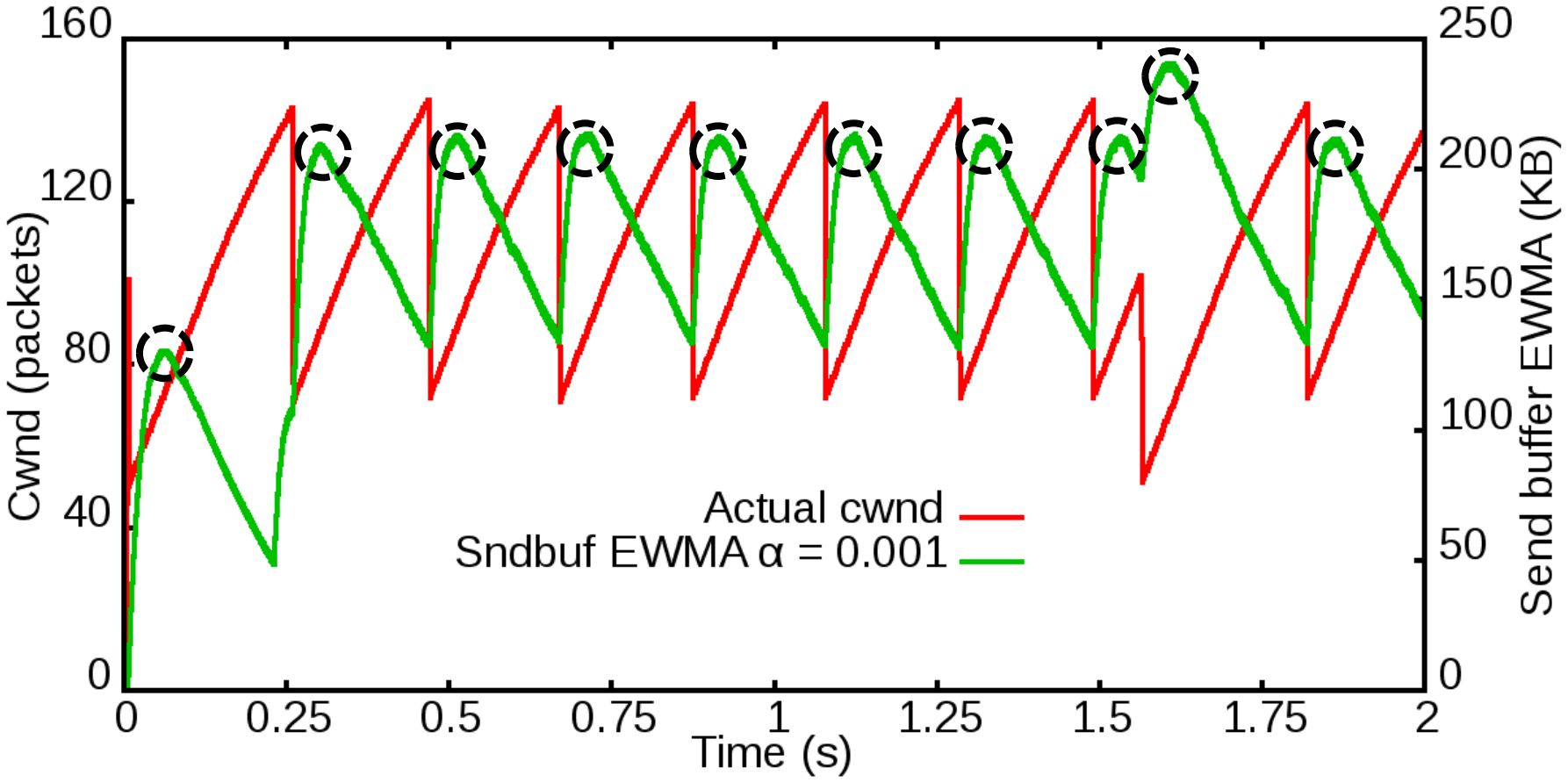
Inferring congestion events



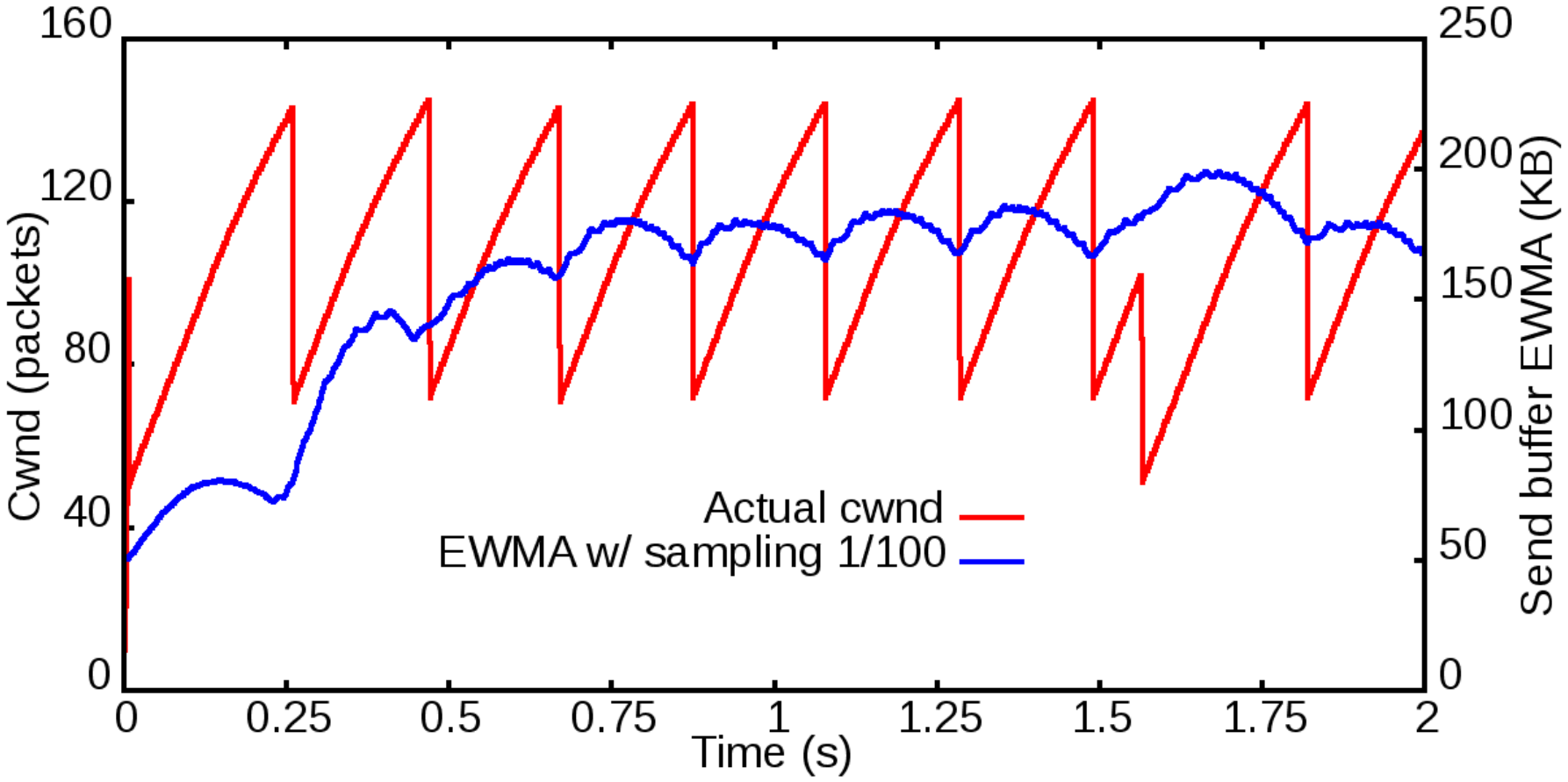
Inferring congestion events



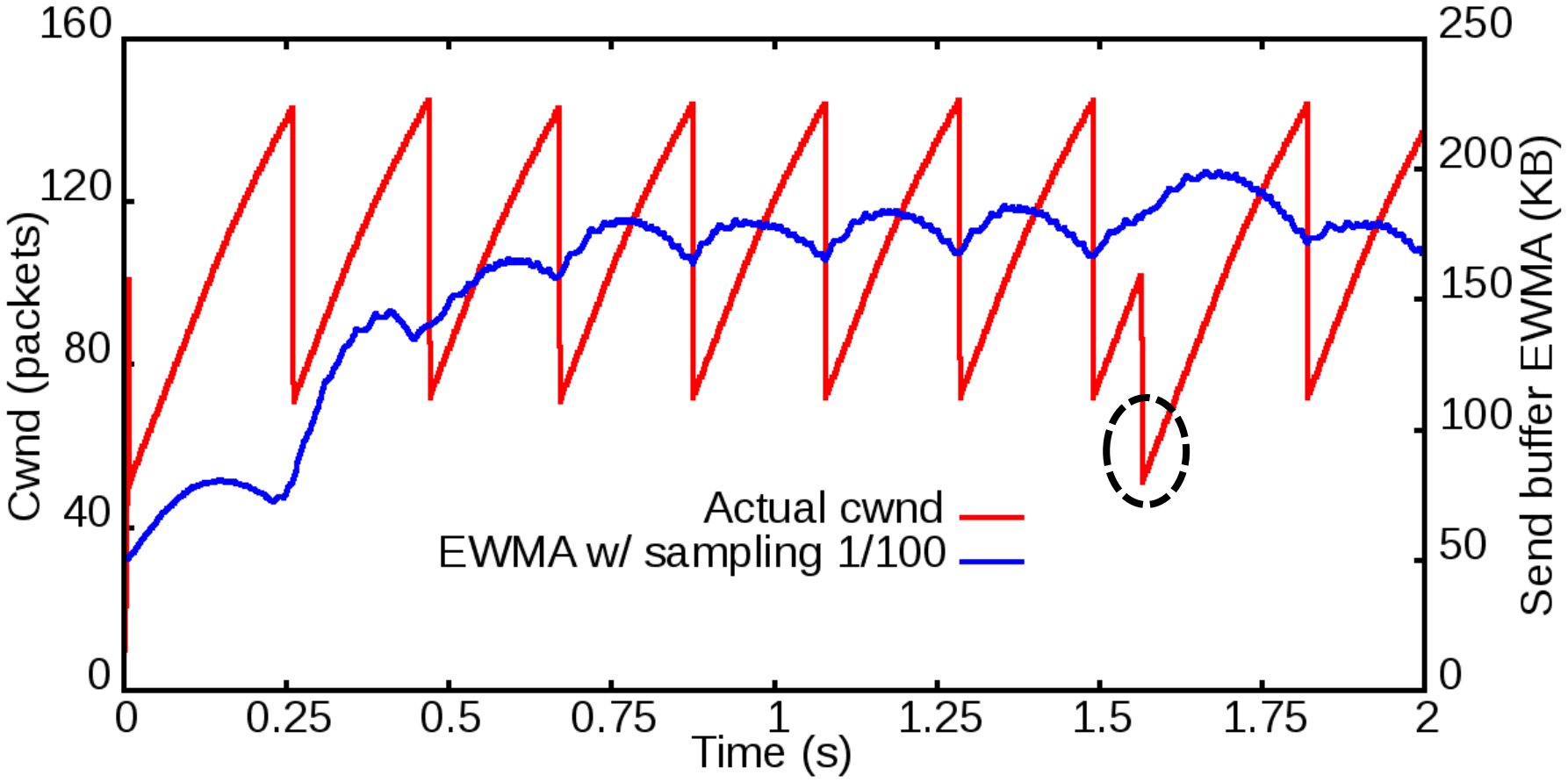
Inferring congestion events



Inferring congestion events



Inferring congestion events



Related work

- Mahout
- HONE
- XCP, SWAN, etc.

Conclusions

- Having sendbuffer information in TCP segments can prove useful in many situations
- It can be encoded in every segment without any overhead in terms of space
- Doesn’t require modified applications, but we could build some improvements on top of it

Discussion

Discussion

- Preventing cheating

Discussion

- Preventing cheating

- Tenant incentives for deployment

Discussion

- Preventing cheating
- Tenant incentives for deployment
- Minimizing flow completion times