

On estimating the number of flows

Bruce Spang, Nick McKeown
December 3, 2019

How big should a buffer be?

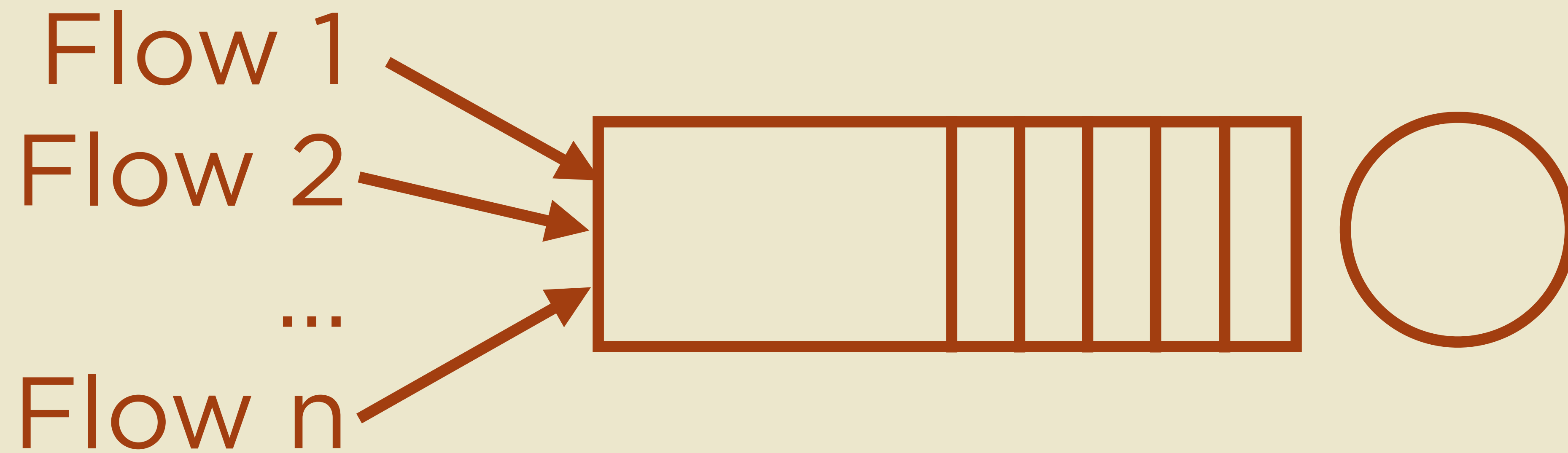
Depends on the number of flows

BDP/ \sqrt{n} : Appenzeller, McKeown, Keslassy 2004

$O(n)$: Dhamdhere, Jiang, Dovrolis 2005

$O(1)$: Enachescu, Ganjali, Goel, McKeown, Roughgarden 2006

Setting in existing work



What is the number of flows?

Ideally:

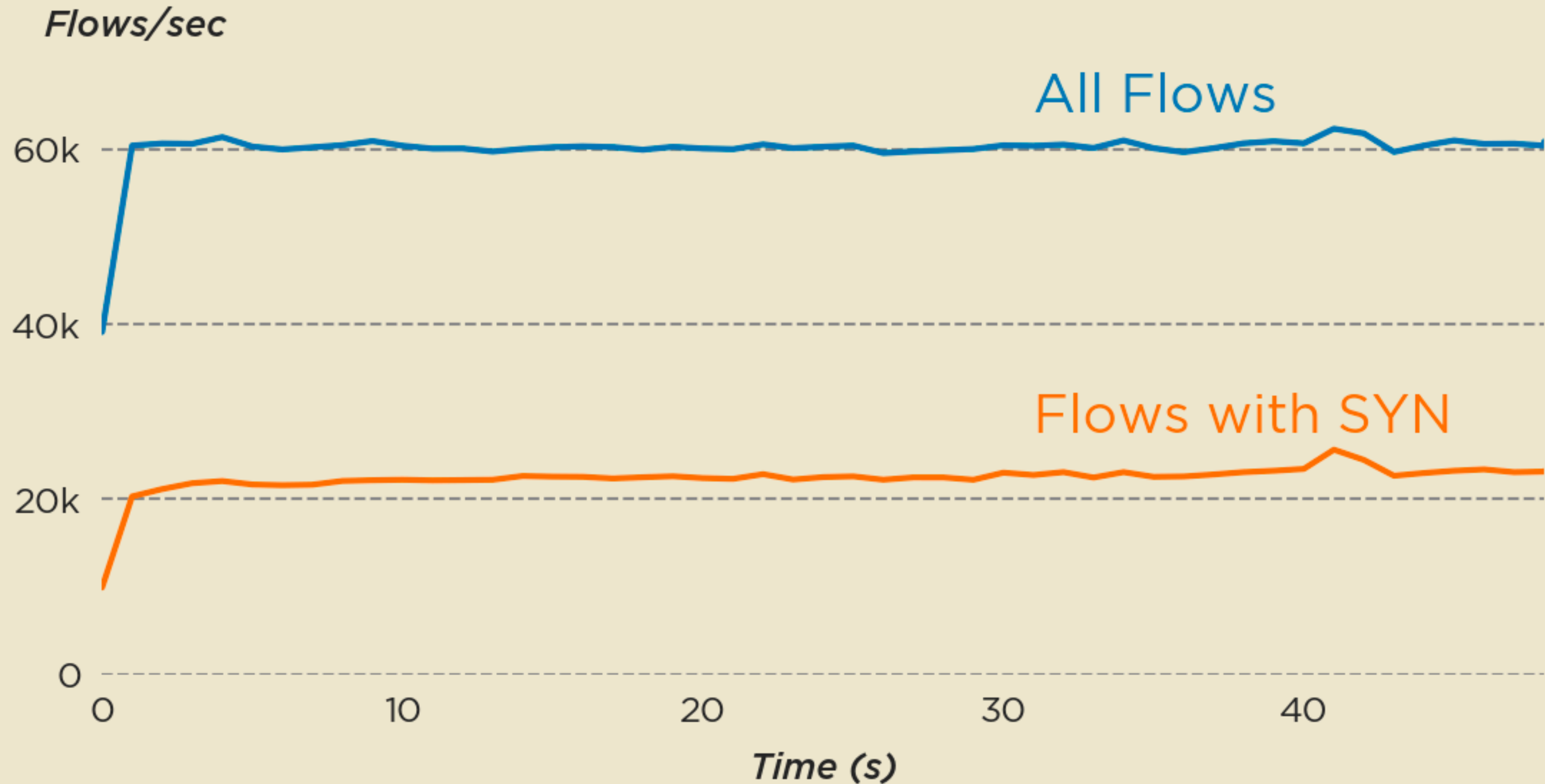
1. A flow starts with a SYN
2. Ends with a FIN/RST
3. Sends data in between
4. We get to observe 1-3

What is the number of flows?

In practice:

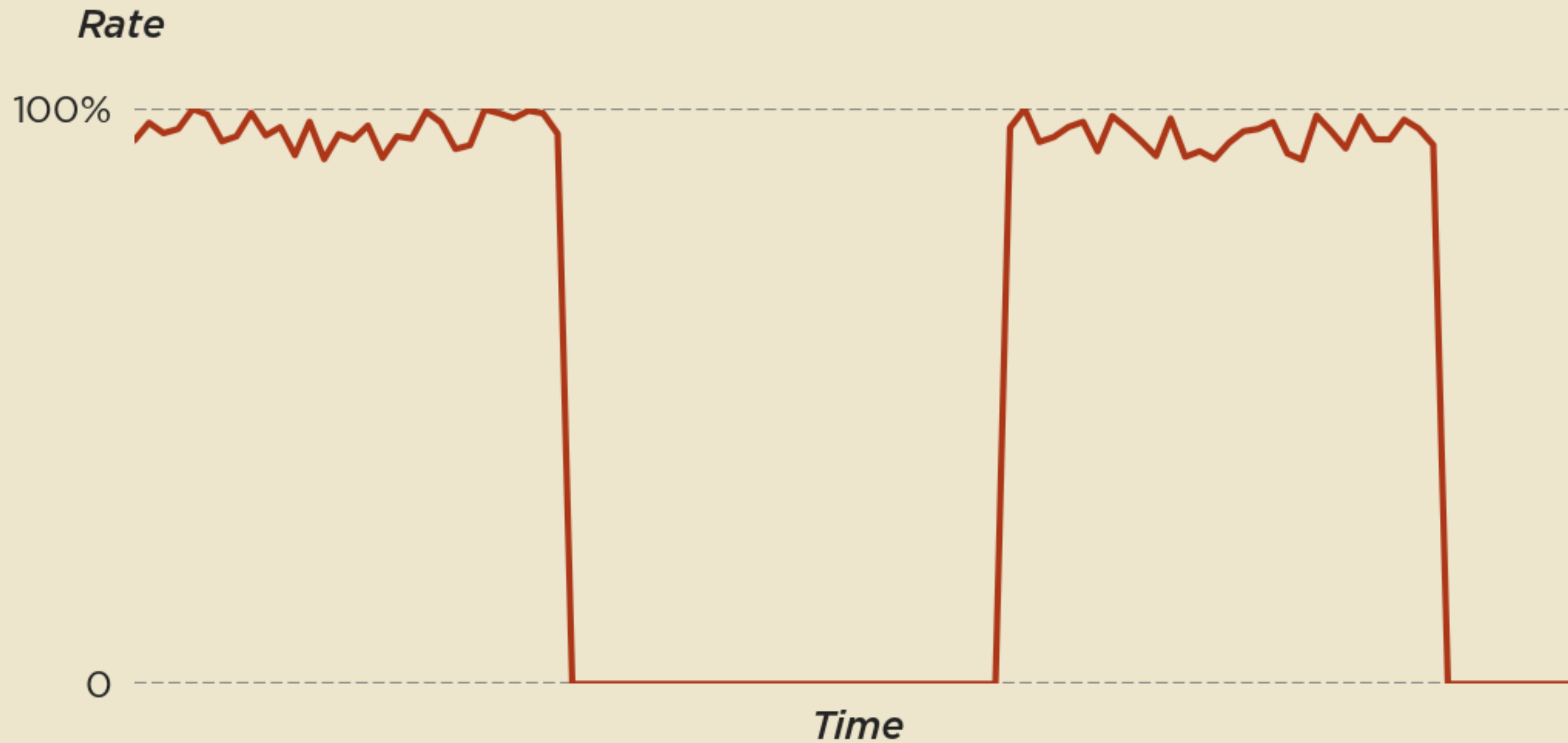
1. ~~A flow starts with a SYN~~
2. ~~Ends with a FIN/RST~~
3. ~~Sends data in between~~
4. ~~We get to observe 1-3~~

Not all flows have SYN packets



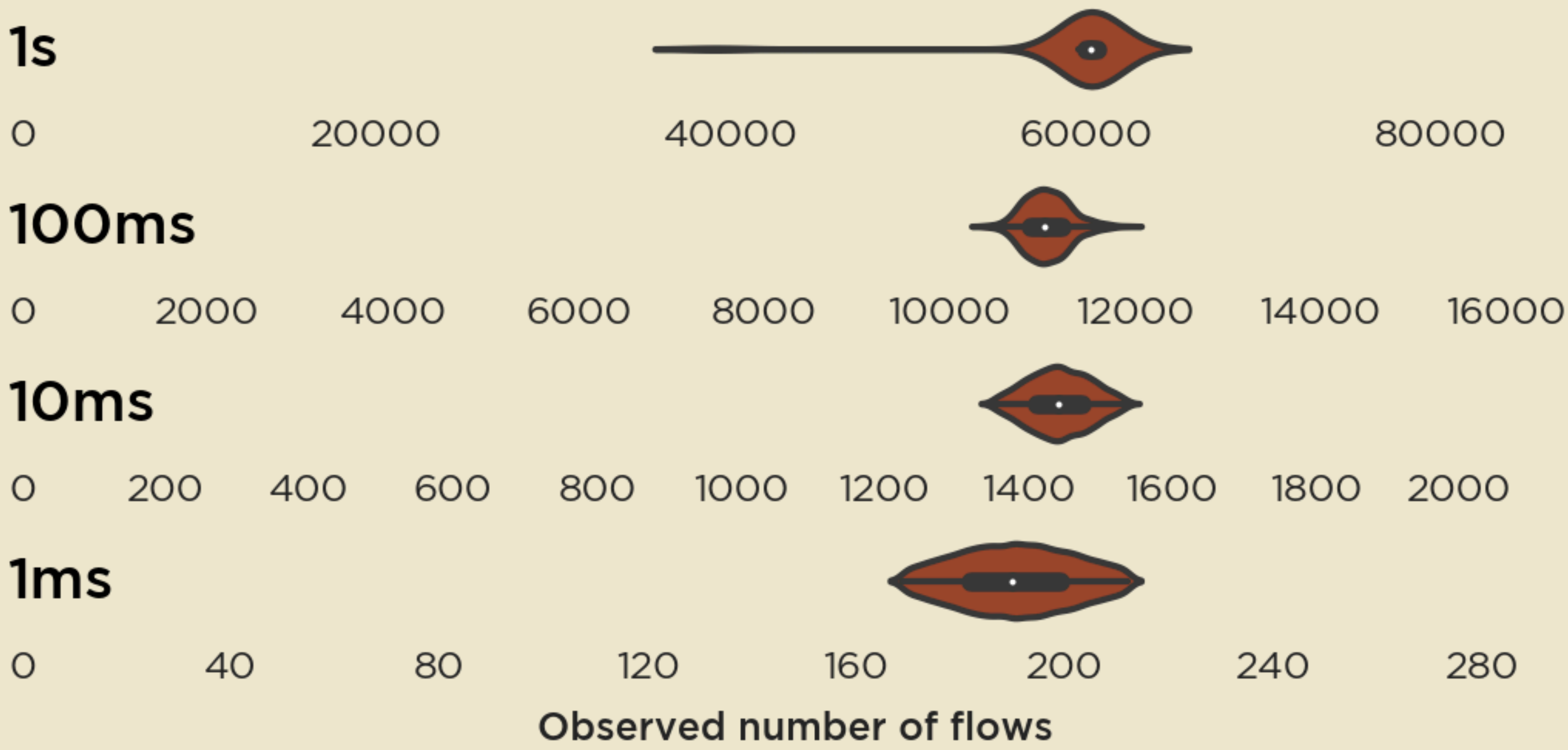
Suggestion: count the number of unique
five tuples

Not all flows send data in between



Suggestion: count the number of unique
five tuples *in a measurement interval*

Number of flows depends on interval



Suggestion: count the number of unique five tuples in *a few* measurement intervals

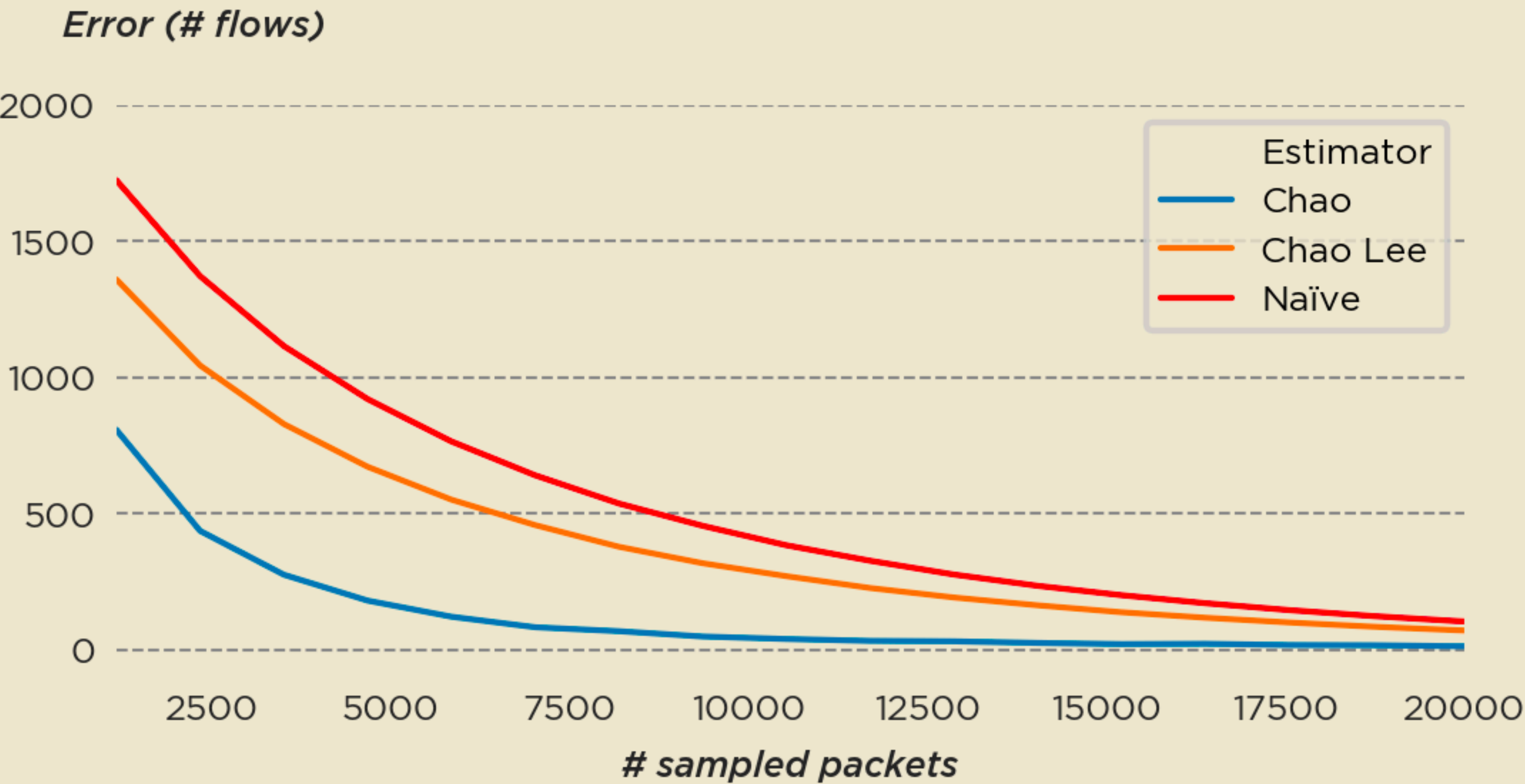
We don't get to see all packets



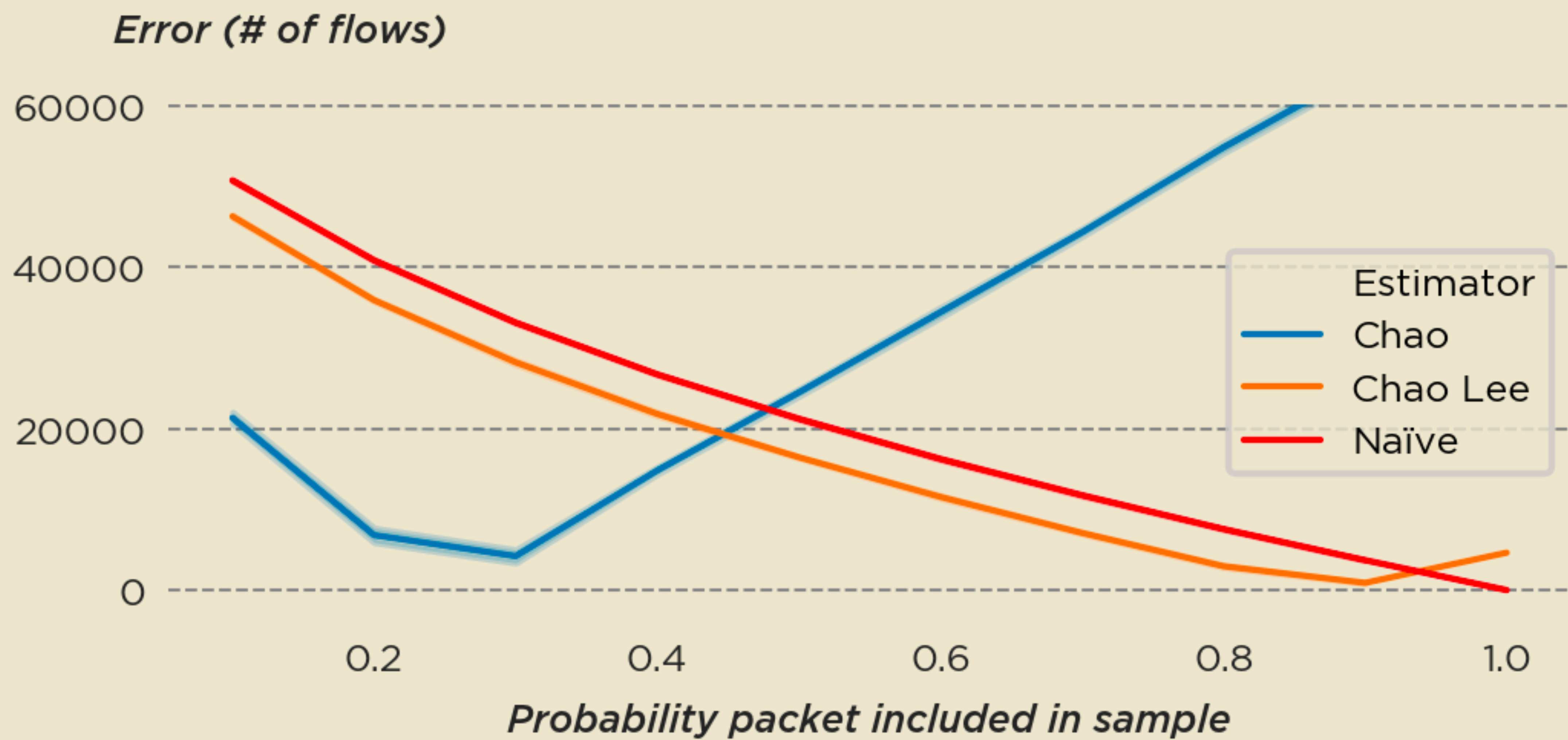
Estimating the unseen



Biological techniques give a good estimate of the number of flows (in certain settings)



**But biological estimators give a bad estimate
in the network setting**



Suggestion: don't sample
(or solve a cool statistics problem)

Suggestions

Count the number of unique five tuples over a few RTTs

If infeasible, count SYNs and FIN/RSTs over a very long period and sanity check.

Thanks!