Effect of Receive Buffer Size

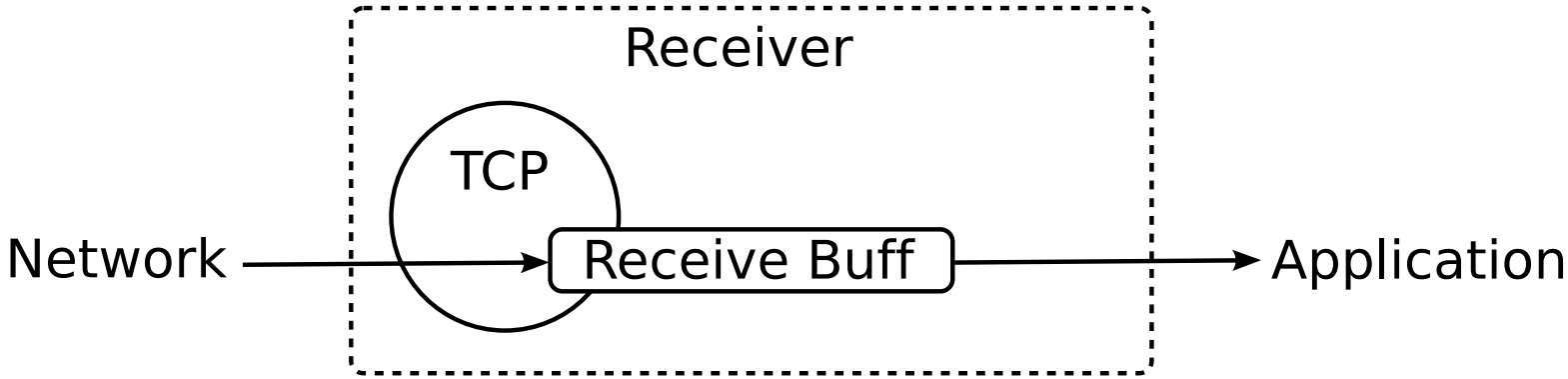
An OS-based Perspective

Jerome White and David X. Wei

California Institute of Technology

The Receive Buffer

- Between TCP and application



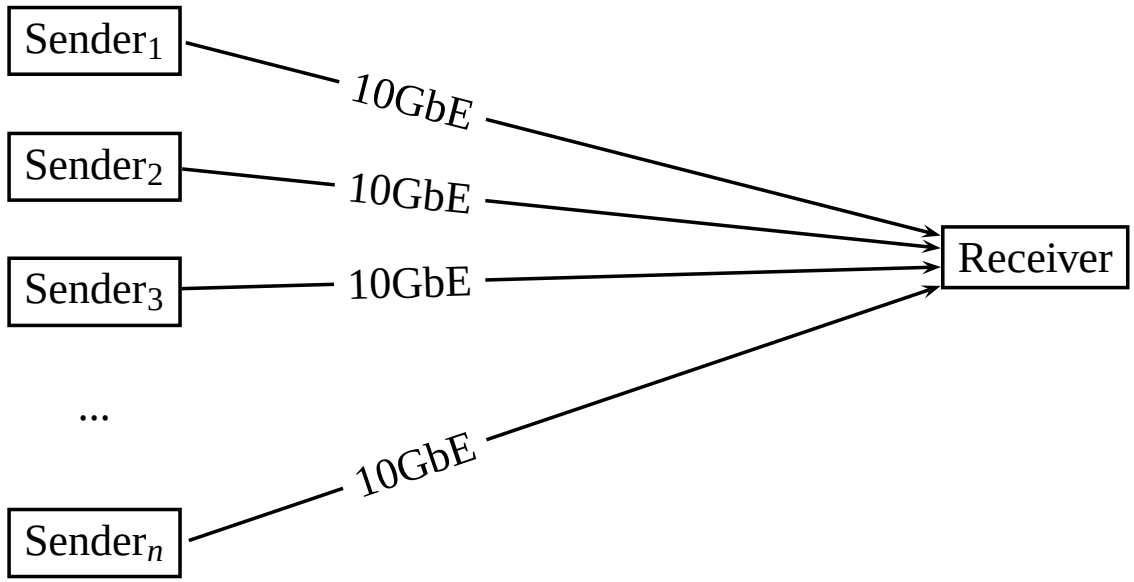
Motivation

- What is the best size?

$$Rx\ size = \begin{cases} < \text{BDP} & \text{bad: } \textit{network} \\ \geq \text{BDP} & \text{good} \end{cases}$$

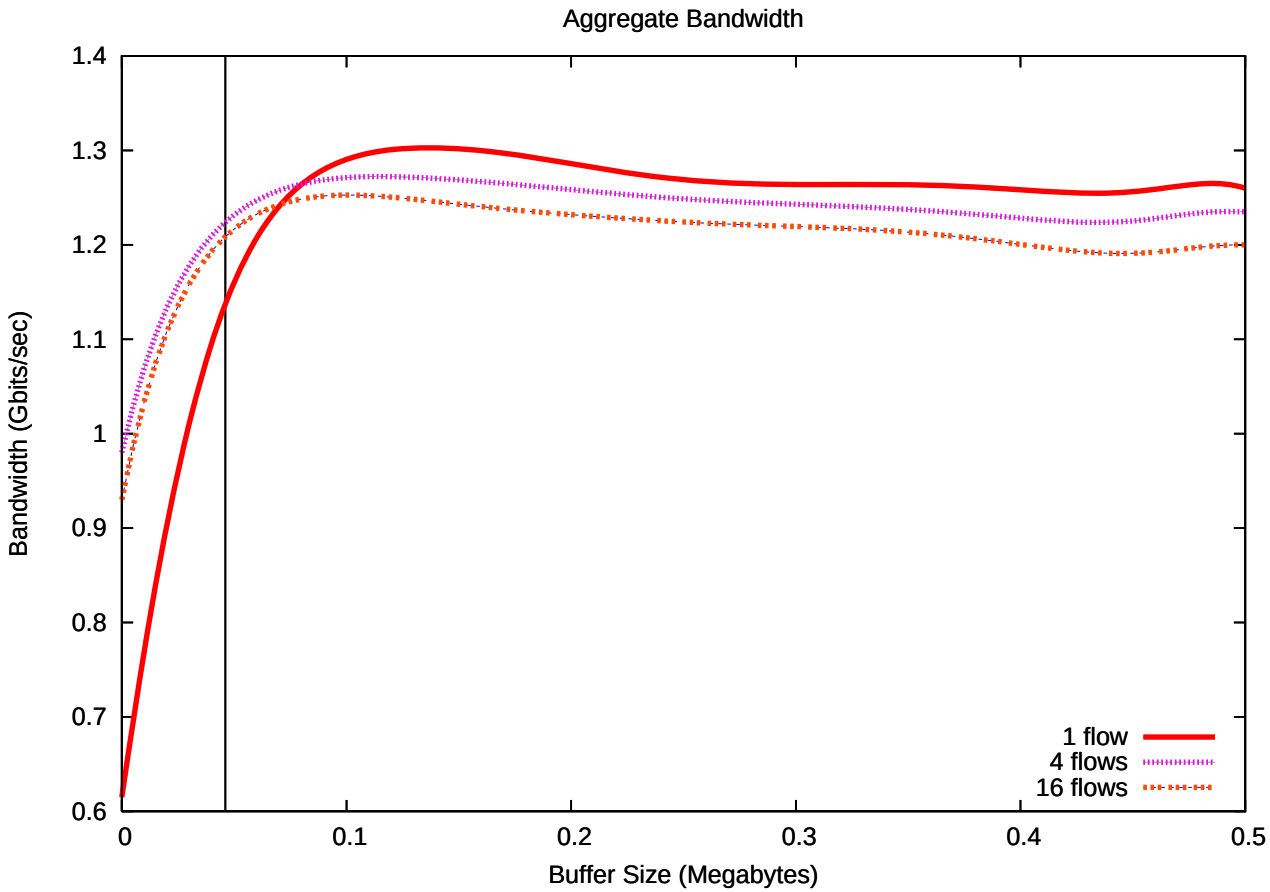
Motivation

- What is the best size?
- CPU becomes the bottleneck



Motivation

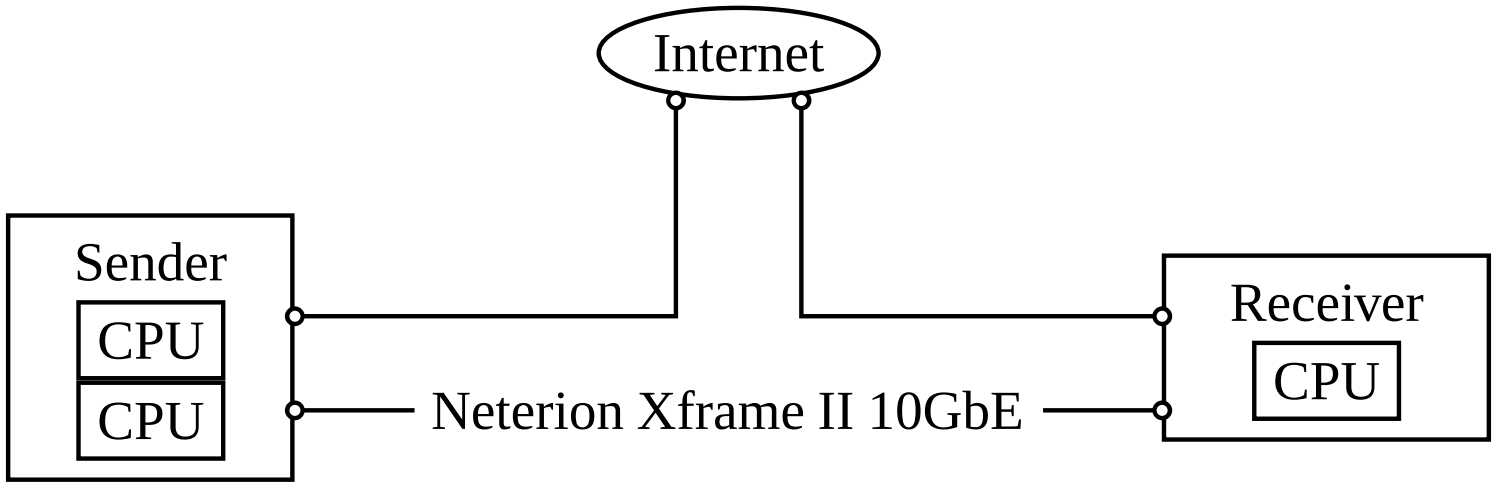
- What is the best size?
- CPU becomes the bottleneck



Motivation

- What is the best size?
- CPU becomes the bottleneck
- CPU-based metrics needed!

Methodology: Hardware



- Sender-receiver direct 10GbE connection
- 0.031ms RTT (ping)
- 2.4Ghz AMD Opteron/4Gb memory

Methodology: Software

- Traffic generated with iperf 2.0.2
- Buffer size controlled via `/proc` variables
- Modified Linux 2.6 kernel
  - Snapshot taken at start and end of a timeslice

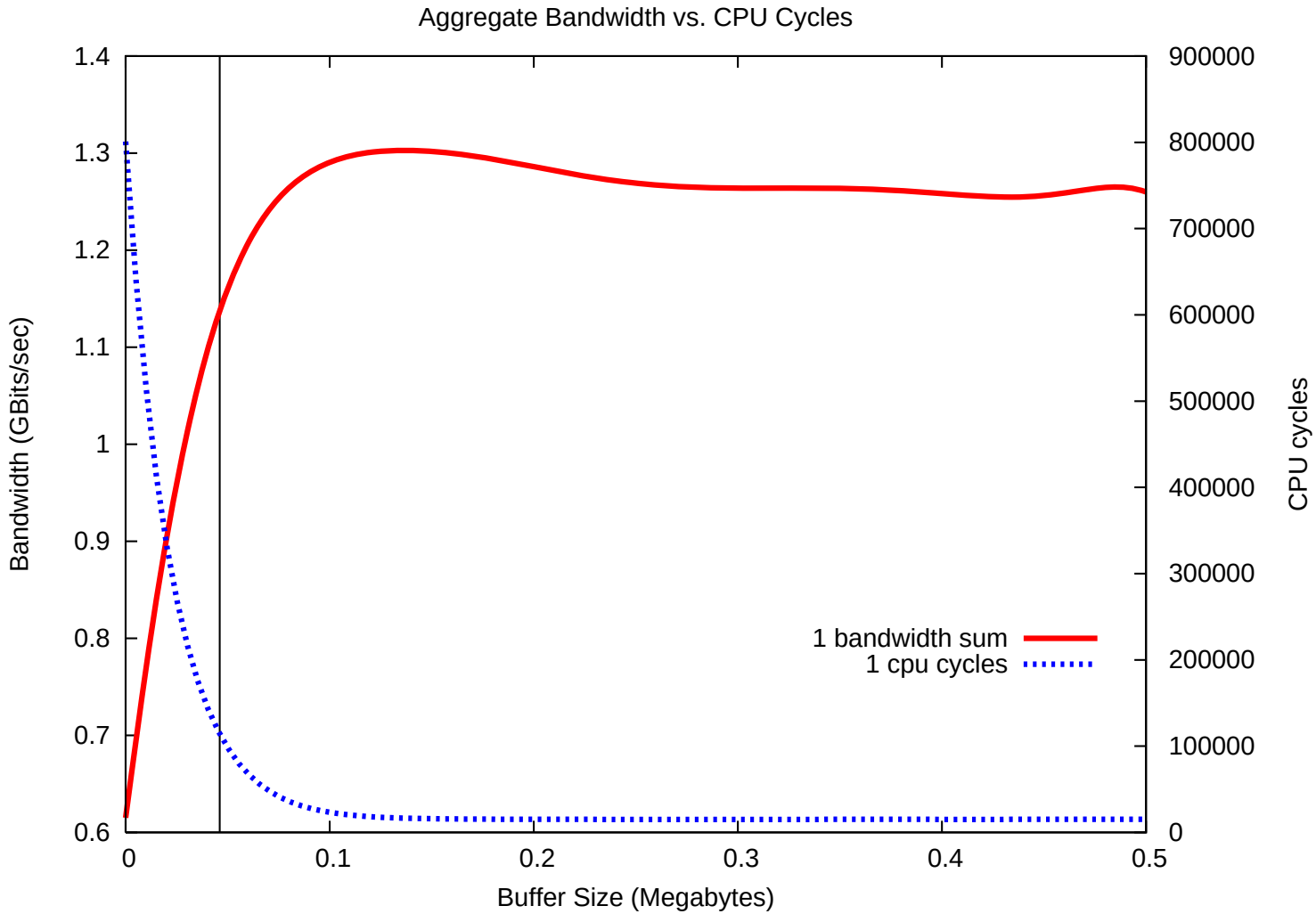
Observation: Overview

- Correlate bandwidth trends to OS trends
- Observe as receive buffer size increases

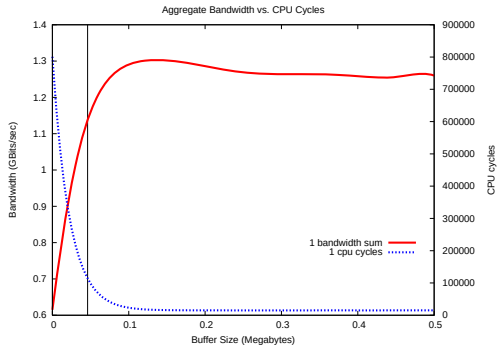
Cycle Count

- CPU timeslices obtained

Cycle Count


Aggregate Bandwidth vs. CPU Cycles
Bandwidth (GBits/sec)
CPU cycles
Buffer Size (Megabytes)
1 bandwidth sum
1 cpu cycles
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
900000
800000
700000
600000
500000
400000
300000
200000
100000
0
0
0.1
0.2
0.3
0.4
0.5

Cycle Count: Too Small



- Too small - lots of context switching

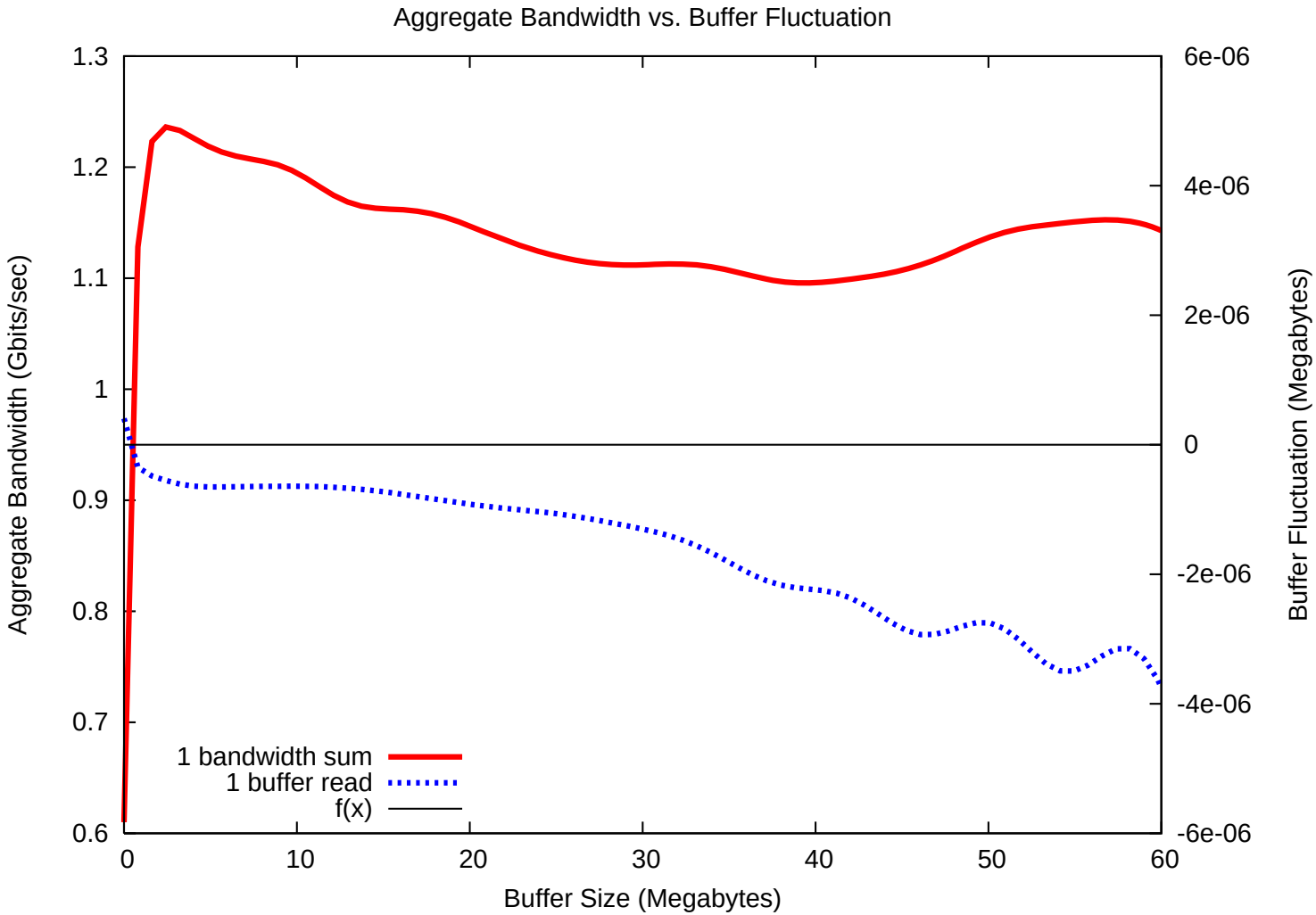
Buffer Fluctuation

- Average buffer change during process activation:

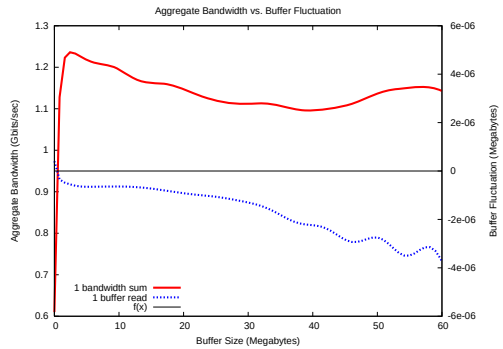
$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)$$

- Is data added or removed?

Buffer Fluctuation



Buffer Fluctuation: Too Large

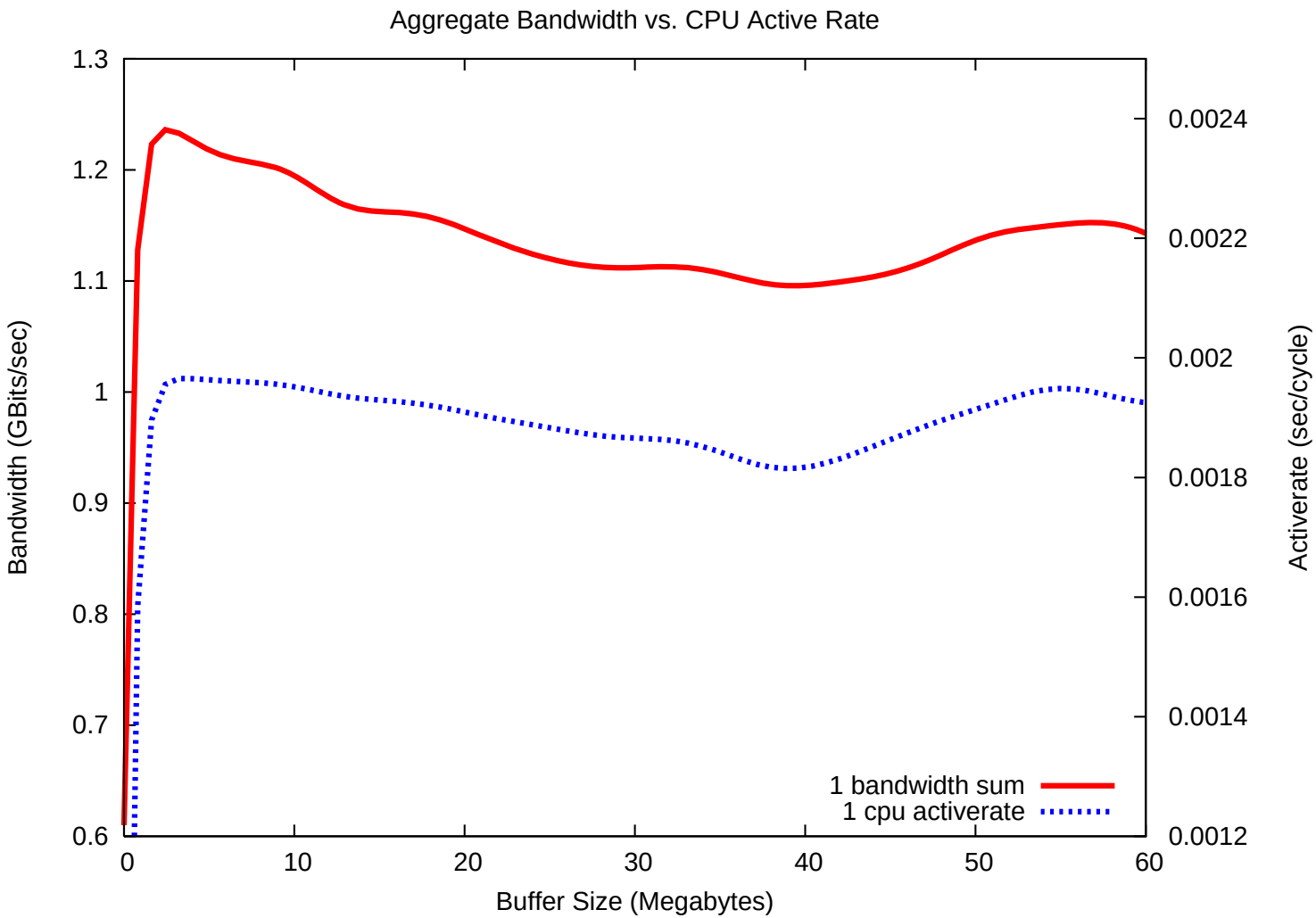


- Good determinant for when buffer sizes are too large
- A such thing as too much data?
  - Disconnect between advertised window and receive OS
  - Maintaining process-ability is important!

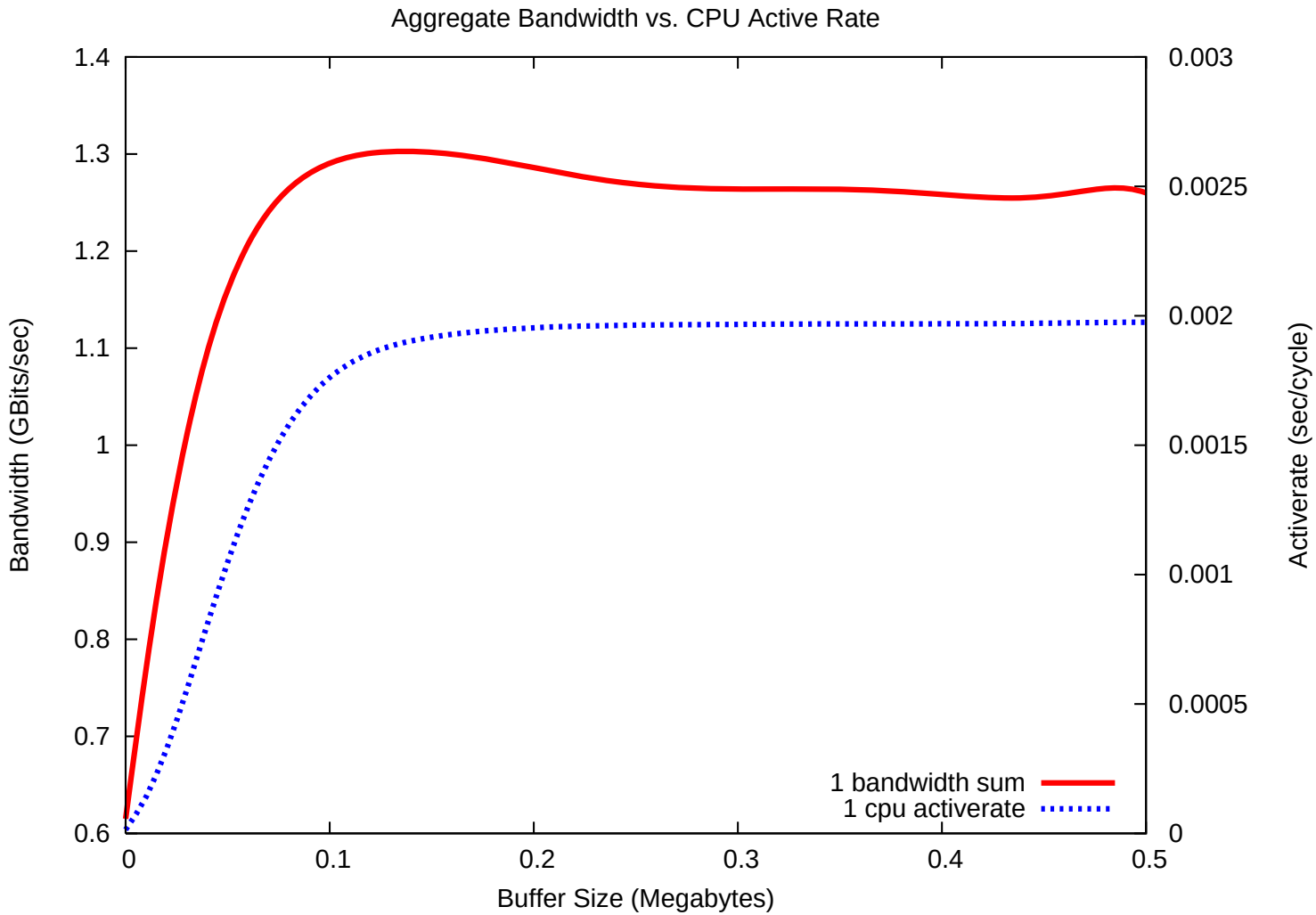
Active Rate

- Process Active Rate
  - Duration of active period
  - milliseconds per timeslice

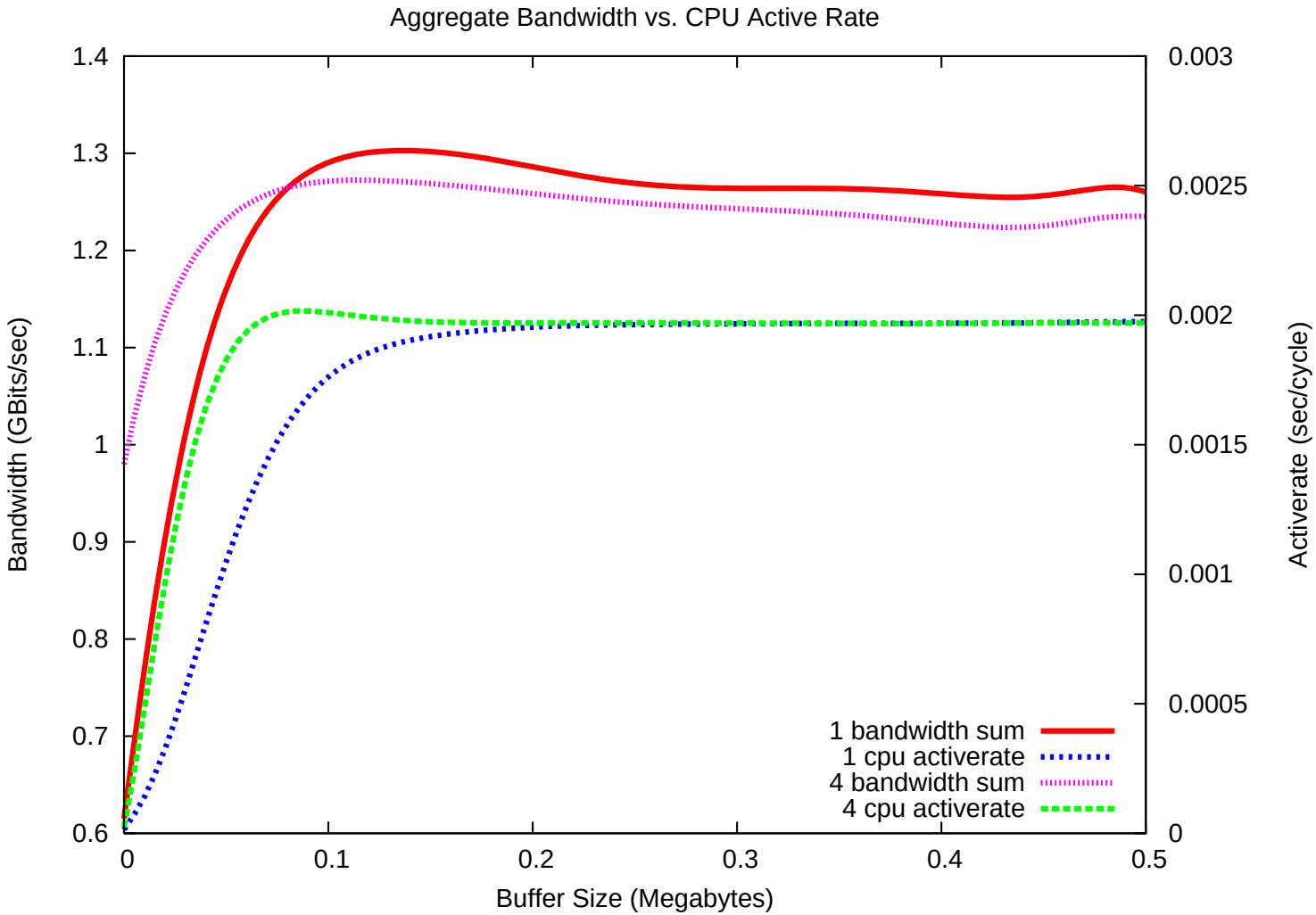
Active Rate


Aggregate Bandwidth vs. CPU Active Rate
Bandwidth (GBits/sec)
Activerate (sec/cycle)
Buffer Size (Megabytes)
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.0024
0.0022
0.002
0.0018
0.0016
0.0014
0.0012
0
10
20
30
40
50
60
1 bandwidth sum
1 cpu activerate

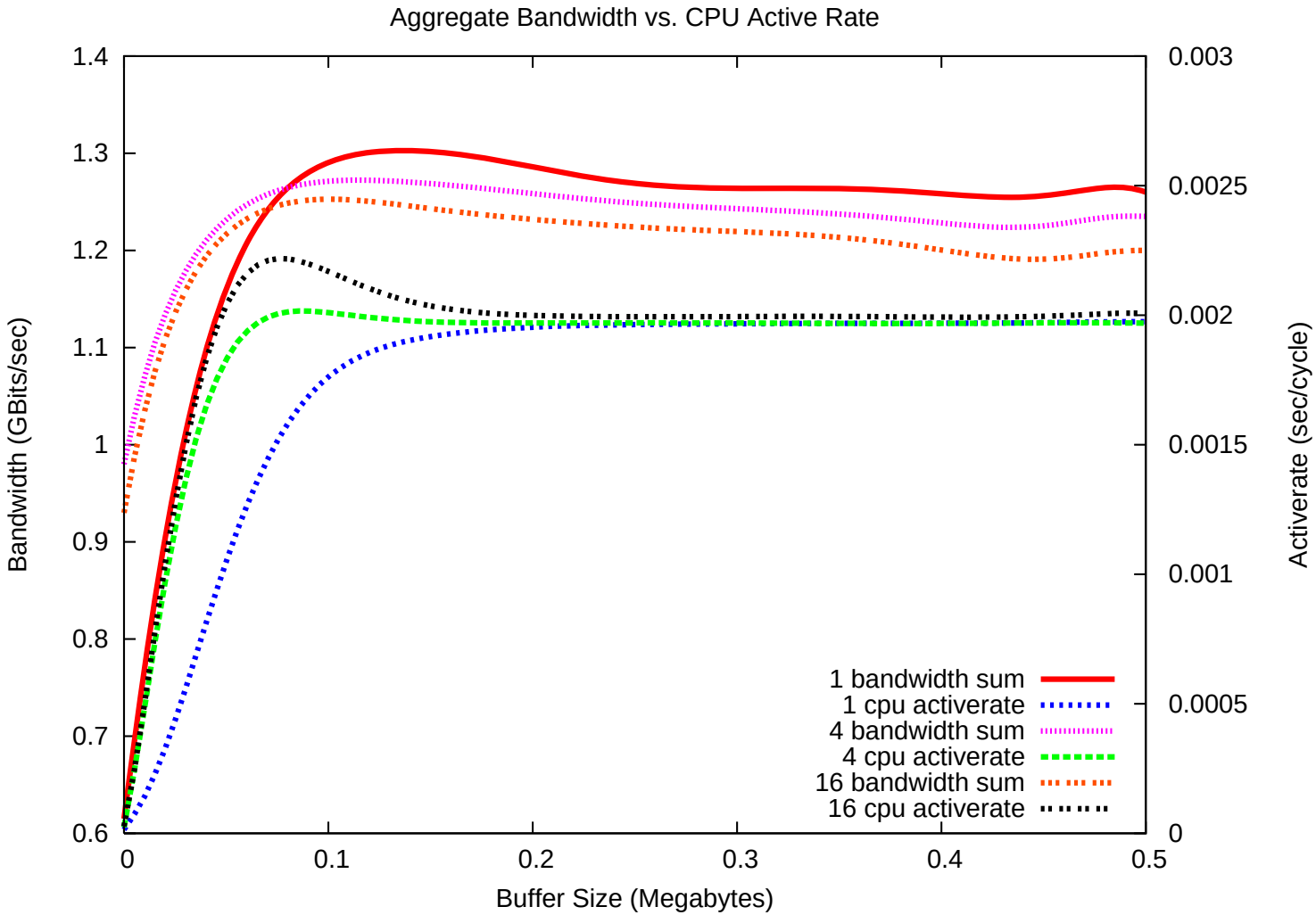
Active Rate


Aggregate Bandwidth vs. CPU Active Rate
Bandwidth (GBits/sec)
Activerate (sec/cycle)
Buffer Size (Megabytes)
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.003
0.0025
0.002
0.0015
0.001
0.0005
0
0
0.1
0.2
0.3
0.4
0.5
1 bandwidth sum
1 cpu activerate

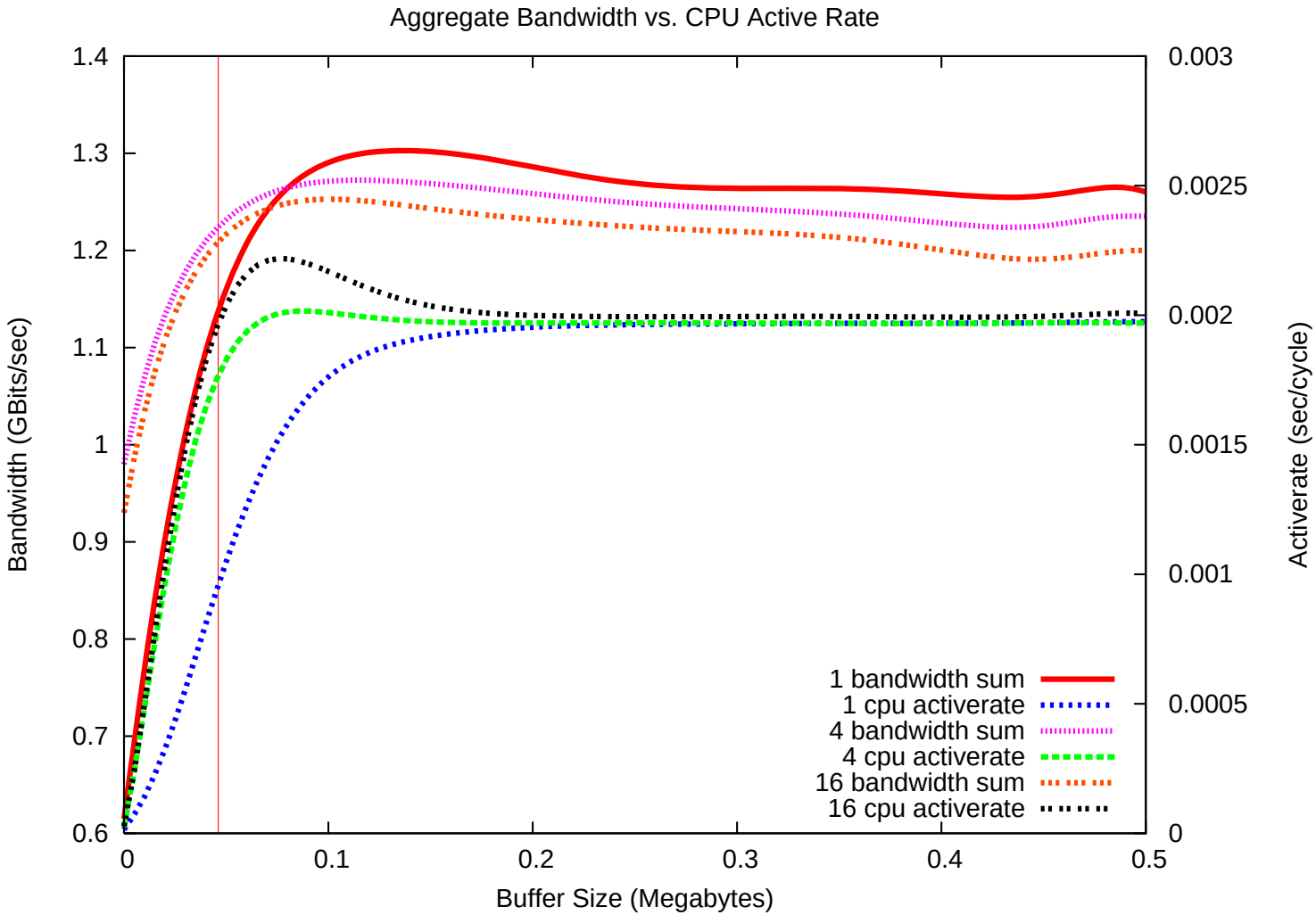
Active Rate


Aggregate Bandwidth vs. CPU Active Rate
Bandwidth (GBits/sec)
Activerate (sec/cycle)
Buffer Size (Megabytes)
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.003
0.0025
0.002
0.0015
0.001
0.0005
0
0
0.1
0.2
0.3
0.4
0.5
1 bandwidth sum
1 cpu activerate
4 bandwidth sum
4 cpu activerate

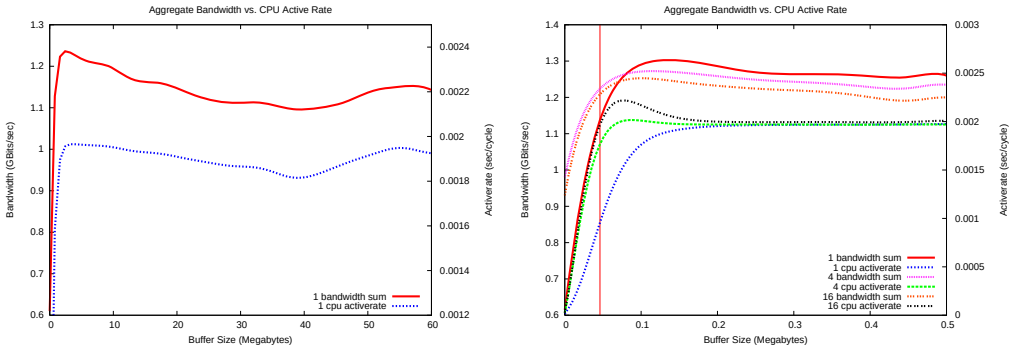
Active Rate


Aggregate Bandwidth vs. CPU Active Rate
Bandwidth (GBits/sec)
Activerate (sec/cycle)
Buffer Size (Megabytes)
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.003
0.0025
0.002
0.0015
0.001
0.0005
0
0
0.1
0.2
0.3
0.4
0.5
1 bandwidth sum
1 cpu activerate
4 bandwidth sum
4 cpu activerate
16 bandwidth sum
16 cpu activerate

Active Rate


Aggregate Bandwidth vs. CPU Active Rate
Bandwidth (GBits/sec)
Activerate (sec/cycle)
Buffer Size (Megabytes)
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.003
0.0025
0.002
0.0015
0.001
0.0005
0
0
0.1
0.2
0.3
0.4
0.5
1 bandwidth sum
1 cpu activerate
4 bandwidth sum
4 cpu activerate
16 bandwidth sum
16 cpu activerate

Active Rate: Just Right



- High correlation between Active Rate and bandwidth
- Good determinant for when buffer sizes are too small

Future Work

- Round out study
  - Include send buffer
  - Non-active time dynamics
  - Correlation or causality?
- Turn observations into solutions!

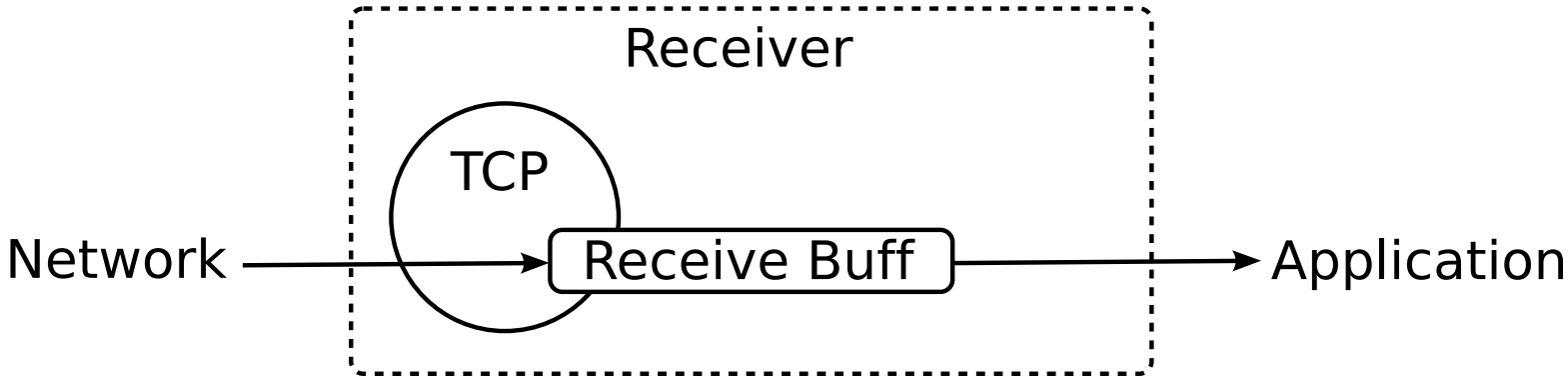
Conclusion

- CPU rate a good lower bound
- Buffer utilization a good upper bound

$$Rx\ size = \begin{cases} < \text{BDP} & \text{bad: network} \\ \leq \text{BDP} & \text{bad: context-switchs} \\ > \text{BDP} & \text{good: maximize active rate} \\ \gg \text{BDP} & \text{bad: Too much data} \end{cases}$$

Conclusion

- CPU rate a good lower bound
- Buffer utilization a good upper bound
- Paying attention to the OS is a good thing



Conclusion

- CPU rate a good lower bound
- Buffer utilization a good upper bound
- Paying attention to the OS is a good thing