

SharkFest'17 US

Understanding Throughput & TCP Windows

A Walk-Through of the Factors that can limit TCP Throughput Performance

Kary Rogers

Director, Staff Engineering | Riverbed Technology

Agenda

- TCP ownership
- Level set terms
- Receiver side factors
- Sender side factors
- Other factors
- Checklist



Why Do We Care About TCP?

**TCP:
The Red
Headed Step
Child
of the Protocol
Stack**

© Alan Shapiro

What is Bandwidth?

bandwidth 

noun | band·width | \ˈbænd-ˌwɪð, -ˌwɪθ\

Definition of BANDWIDTH for English Language Learners

: a measurement of the ability of an electronic communications device or system (such as a computer network) to send and receive information

What is Throughput?

throughput 

noun | through·put | \ˈthrü-put\

Definition of THROUGHPUT for English Language Learners

: the amount of material, data, etc., that enters and goes through something (such as a machine or system)

What is the Difference?

- Bandwidth is the total, raw capacity
- Throughput is what you're actually getting
- We will focus on throughput of a single connection

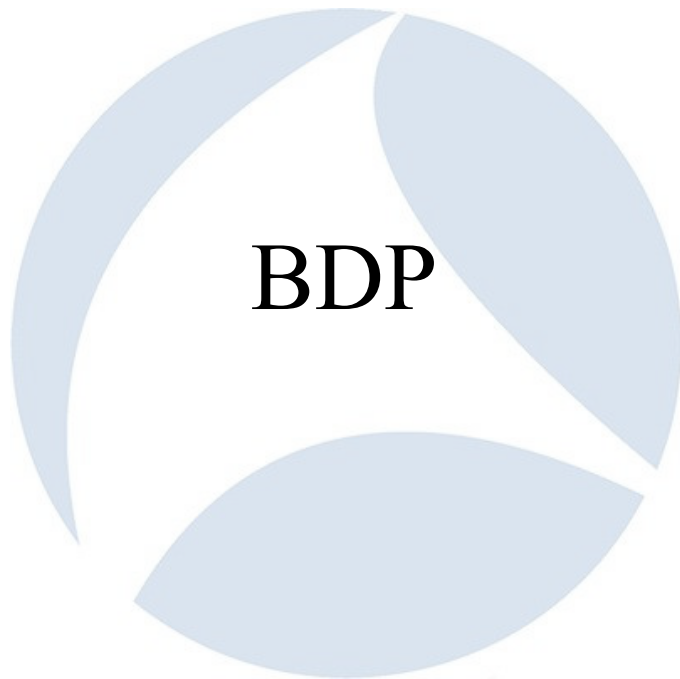
What is BDP?

The amount of data TCP should have within the transmission path at any one time in order to fully utilize the available capacity

i.e. golf balls in a PVC pipe



How Much Data Can We Send?



The Setup

..... Unsent data

256k

allthecats.zip

What are the Factors?

- Amount of data the receiver says it can receive
- Amount of data the sender thinks it can send

Receiver

- Receive window - RWIN, RWND, SO_RCVBUF
- Amount of free space in the receiver's socket buffer
- The application reads data from the buffer to free space
- Advertised in every packet in the TCP header

Header Length: 32 bytes

▶ Flags: 0x010 (ACK)

Window size value: 3045

[Calculated window size: 779520]

[Window size scaling factor: 256]

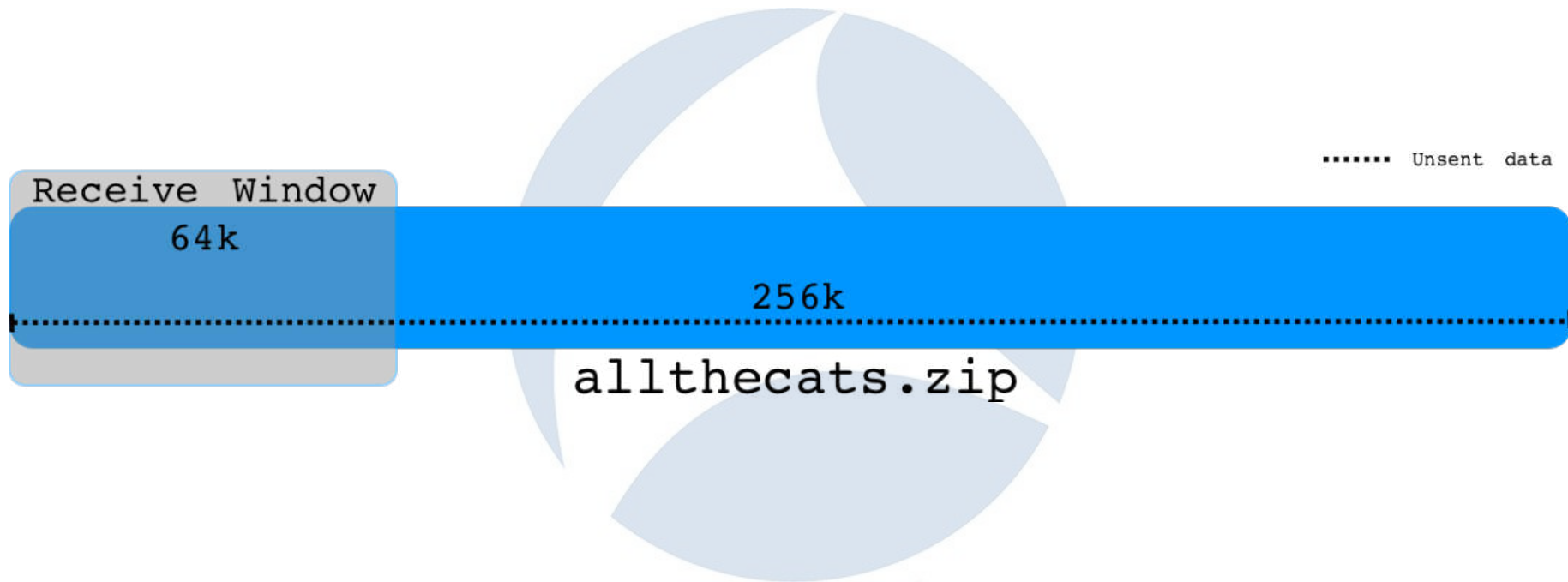
0000	00	50	56	b7	36	1e	00	24
0010	00	34	38	13	40	00	78	06
0020	9f	25	ec	1c	01	bd	96	09
0030	0b	e5	44	25	00	00	01	01
0040	c2	77						

Receiver

- 16 bit integer – 64k max
- Window scaling allows up to 1G – RFC 1323
- Must be supported by both hosts

```
▼ Options: (12 bytes), Maximum segment size 0000 f7 14 be 41 08 00 45 00
  ▶ Maximum segment size: 1460 bytes         0010 37 e6 0a e8 e8 be 0a 2d
  ▶ No-Operation (NOP)                       0020 4d 34 00 00 00 00 80 02
  ▶ Window scale: 8 (multiply by 256)       0030 05 b4 01 03 03 08 01 01
  ▶ No-Operation (NOP)                       0040
```

Receiver



How Much Data Can We Send?



$\text{MIN}(\text{RWIN}, \text{BDP})$

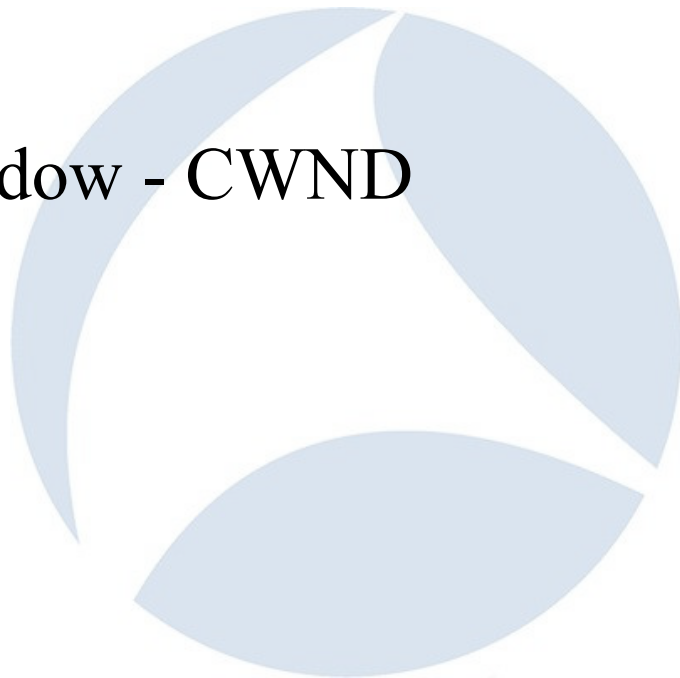
Receiver



To the pcap!

Sender

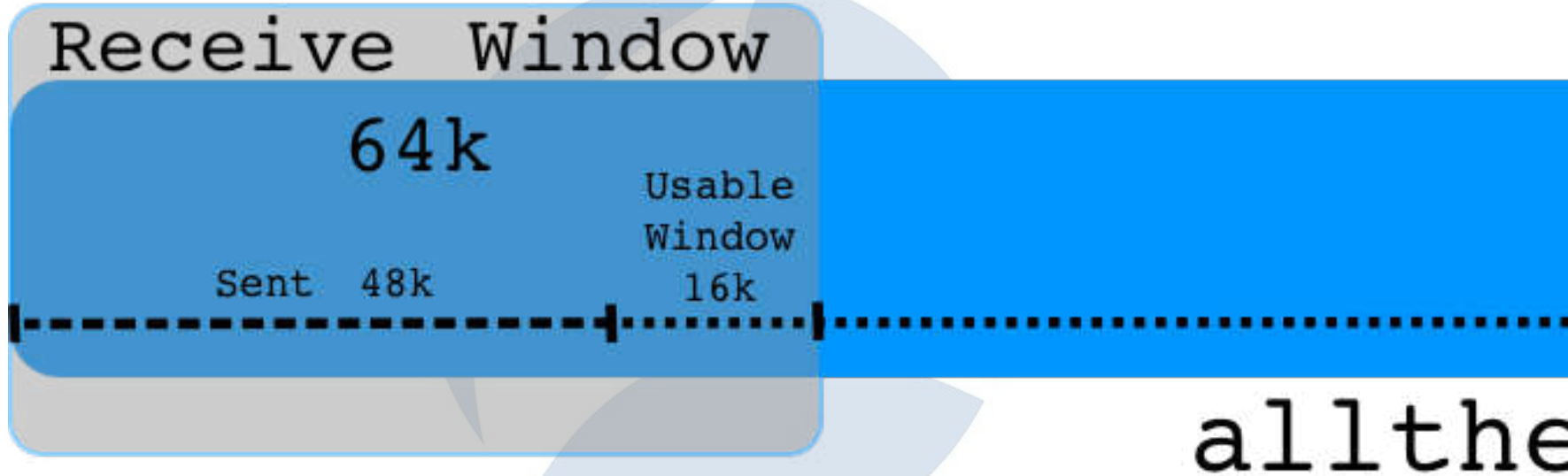
- Bytes in flight
- Congestion window - **CWND**
- Send buffer



Sender – Bytes in Flight

- Sender side of RWIN
- Amount of data that has been sent on the wire but not yet acknowledged i.e. in flight

Sender – Bytes in flight



Sender – Bytes in Flight

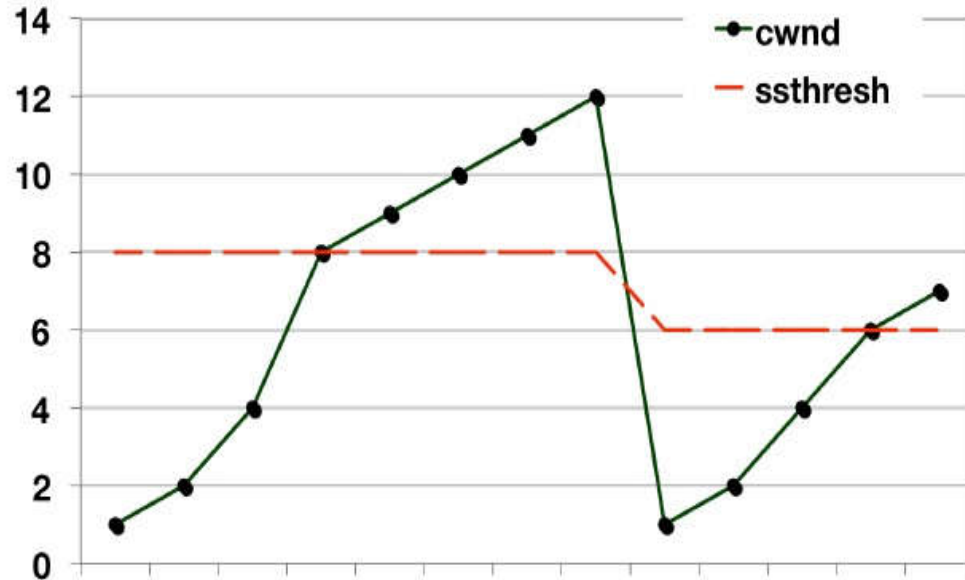


To the pcap!

Sender – CWND

- Sender side flow control based on network conditions and capacity
- A state variable – not in the TCP header
- Referred to in multiples of MSS
- Starts small – 2, 4, 10
- Increases exponentially during slow start
- Increases linearly after slow start ($cwnd > ssthresh$)
- Increases until RWIN or packet loss

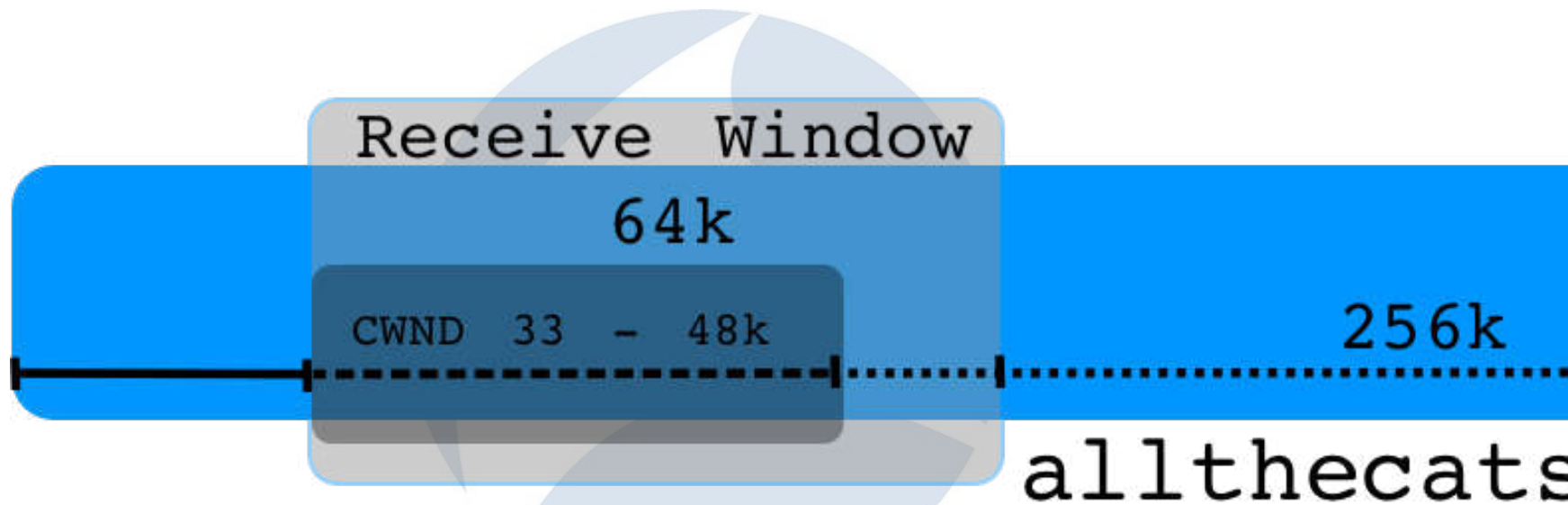
Sender – CWND



Sender – CWND

- Sender is bound by CWND even if RWIN is large
- Growth of CWND in recovery depends on algorithm
 - Windows client uses New Reno
 - Windows server uses DCTCP or CTCP
 - Linux uses CUBIC
- CWND visibility
 - TCP Analyzer tool on Windows
 - ss tool on Linux

Sender – CWND



How Much Data Can We Send?



$\text{MIN} (\text{CWND}, \text{RWIN}, \text{BDP})$

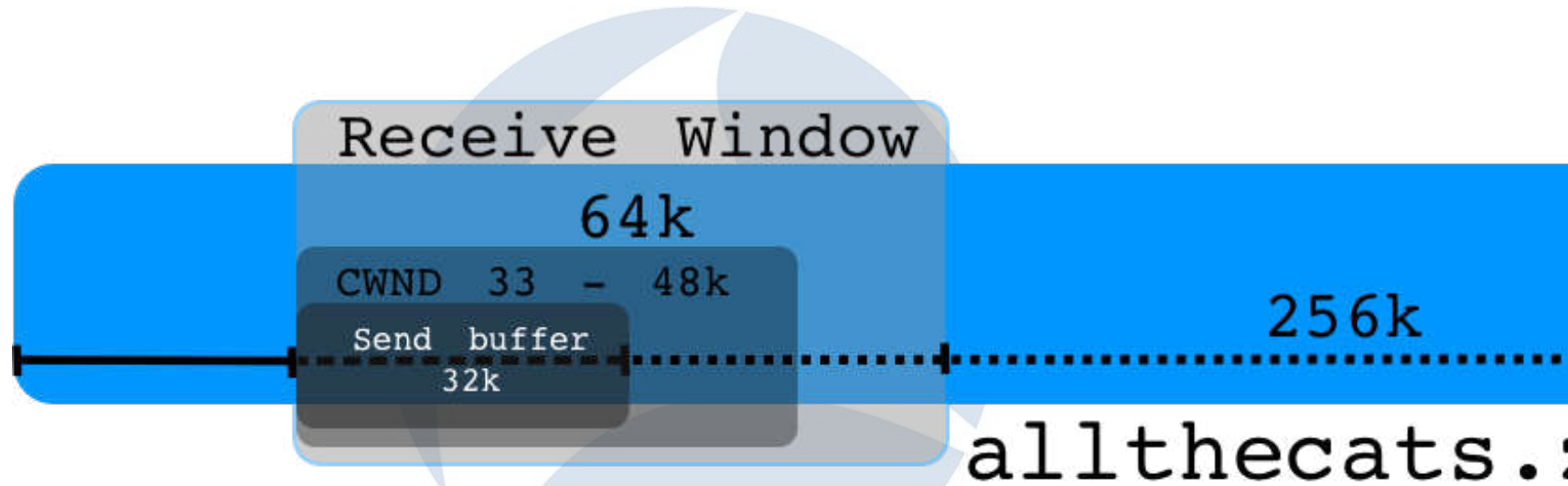


To the pcap!

Sender – Send Buffer

- Size of socket buffer application is writing to
- OS default or set by application
- Optimal size depends on BDP
- Send buffer limit determines how much data is kept outstanding in TCP for one blocking or non-blocking send request
- More info on [MSDN](#) in the Remarks sections

Sender – Send Buffer



How Much Data Can We Send?



$\text{MIN} (\text{SNDBUF}, \text{CWND}, \text{RWIN}, \text{BDP})$

Sender – Send Buffer



To the pcap!

Throughput Checklist

- What is the BDP? Bandwidth x Latency
- What is the advertised window of the receiver?
- Are bytes in flight (cwnd) increasing and reaching BDP or RWIN?
- No?
 - Does the sender stop and wait for ACKs after sending the same amount of data over and over? - send buffer
 - Is there packet loss (retransmissions, dup ACKs) preventing cwnd from growing? - congestion
- Yes? – Great!

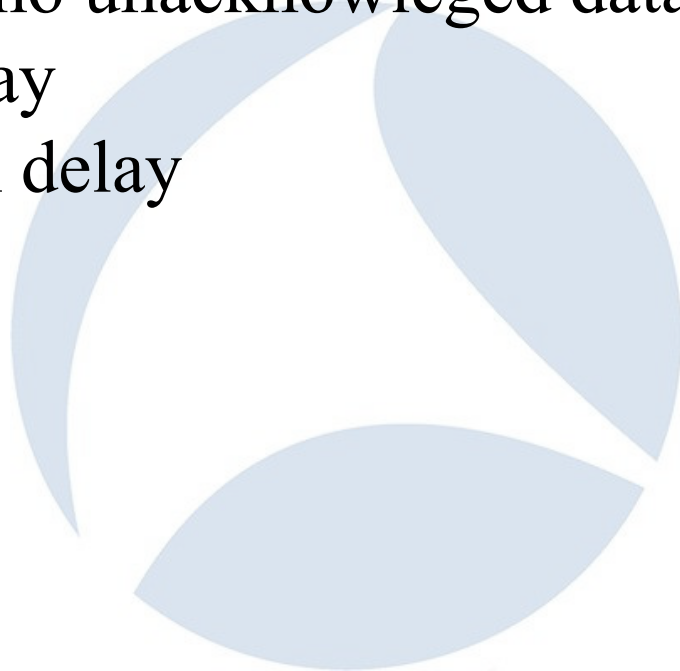
Throughput Checklist

- What are we missing?
- Delayed ACKs
- Nagle
- [Hansang's video](#)



Throughput Checklist

- What if there's no unacknowledged data but a delay?
- Application delay
- User interaction delay



Contact

- kary@packetbomb.com
- [@packetbomb](#)
- Packetbomb.com

