Network Management & Monitoring

Network Performance Definitions

Metrics

Network performance metrics

- Channel capacity, nominal & effective
- Channel utilization
- Delay and *jitter*
- Packet loss and errors

Metrics

System performance metrics

- Availability
- Memory, CPU Utilization, *load*, *I/O wait*, etc.

Service performance metrics

- Wait time / Delay
- Availability
- Service-specific metrics
- How can I justify maintaining the service?
- Who is using it? How often?
- Economic value? Other value?

Common Network Performance Measurements

Relative to traffic:

- Bits per second
- Packets per second
- *Unicast* vs. *non-unicast* packets
- Errors
- Dropped packets
- Flows per second
- Delay (RTT)
- Jitter (delay variation)

Nominal Channel Capacity

The maximum number of bits that can be transmitted on a unit of time (eg: bits per second)

Depends on:

- Bandwidth of the physical medium
  - Cable
  - Electromagnetic waves
- Processing capacity for each transmission element
- Efficiency of algorithms in use to access medium
- Channel encoding and compression

Effective Channel Capacity

Always a fraction of the nominal channel capacity

Dependent on:

- Additional overhead of protocols in each layer
- Device limitations on both ends
  - Flow control algorithm efficiency, etc.
    - For example: TCP

Channel Utilization

What <u>fraction of the nominal channel capacity</u> is actually in use

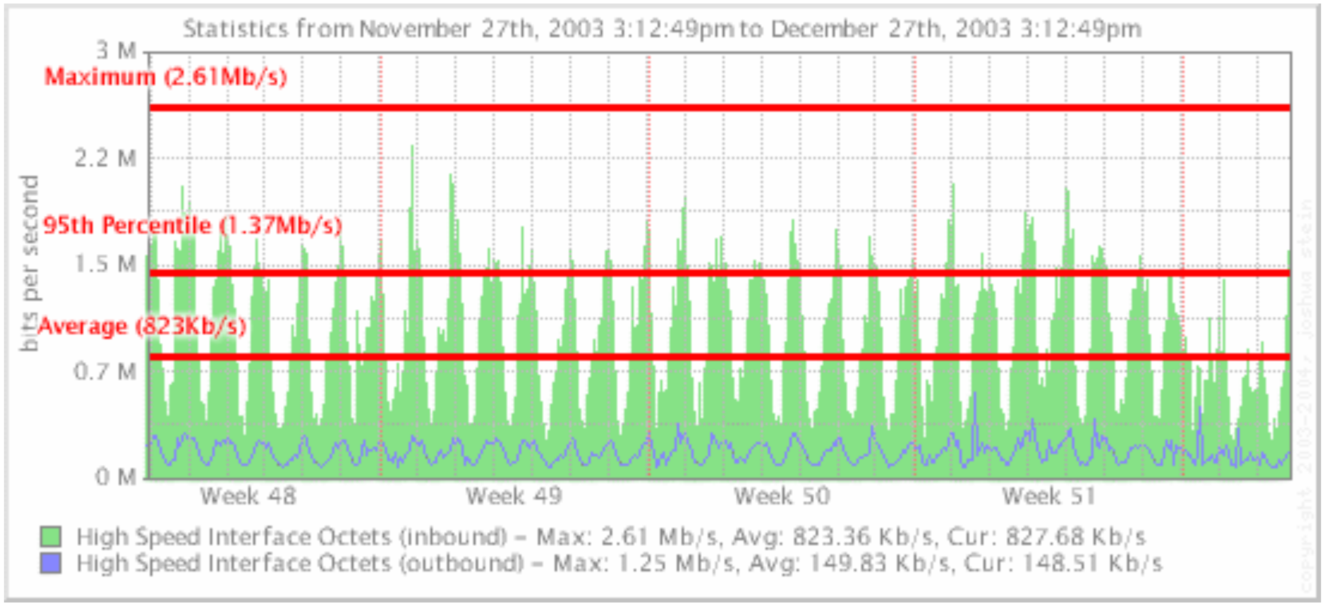
Important!

- Future planning
  - What utilization growth rate am I seeing?
  - For when should I plan on buying additional capacity?
  - Where should I invest for my updates?
- Problem resolution
  - Where are my bottlenecks, etc.

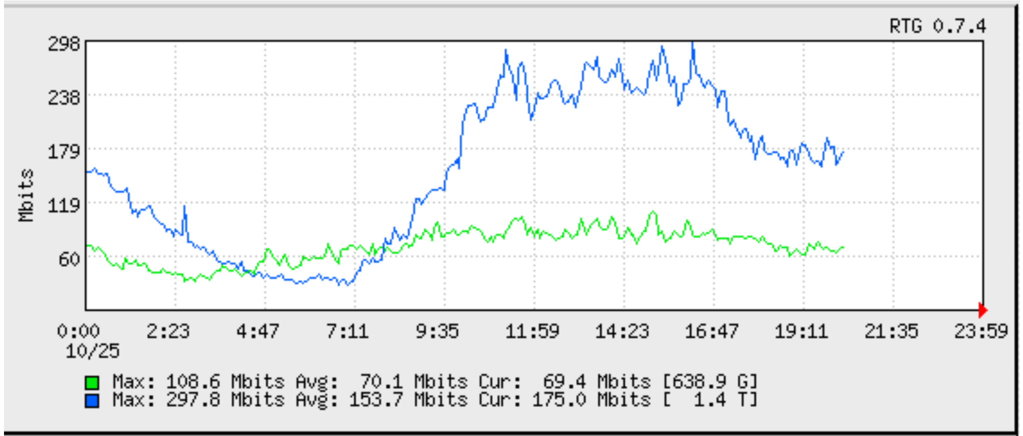
95th Percentile

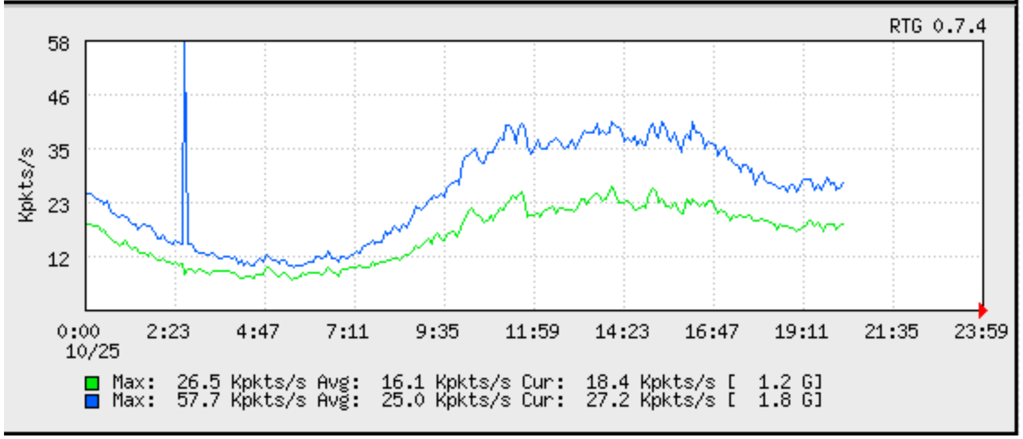
- The smallest value that is larger than 95% of the values in a given sample
- This means that 95% of the time the channel utilization is equal to or *less* than this value
    - Or rather, the peaks are discarded from consideration
- Why is this important in networks?
    - Gives you an idea of the standard, sustained channel utilization.
    - ISPs use this measure to bill customers with “larger” connections.

95th Percentile



Bits per second vs Packets p.s.





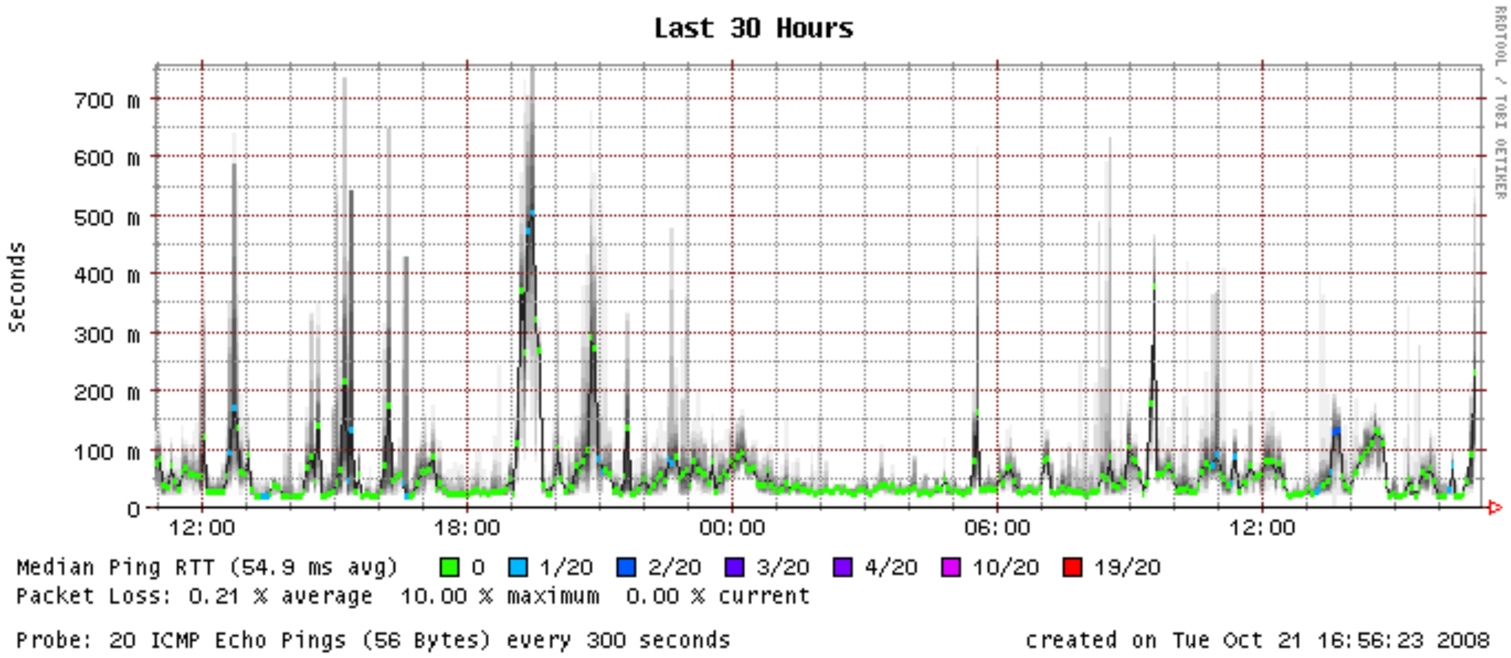
End-to-end Delay

The time required to transmit a packet along its <u>entire path</u>

- *Created by an application, handed over to the OS, passed to a network card (NIC), encoded, transmitted over a physical medium (copper, fibre, air), received by an intermediate device (switch, router), analyzed, retransmitted over another medium, etc.*

- The most common measurement uses *ping* for total round-trip-time (RTT).

Historical Measurement of Delay



```
Median Ping RTT (54.9 ms avg)   🟩 0  🟦 1/20  🟦 2/20  🟪 3/20  🟪 4/20  🟪 10/20  🟥 19/20
Packet Loss: 0.21 % average  10.00 % maximum  0.00 % current

Probe: 20 ICMP Echo Pings (56 Bytes) every 300 seconds                    created on Tue Oct 21 16:56:23 2008
```

Types of Delay

Causes of end-to-end delay:

- Processing delays
- Buffer delays
- Transmission delays
- Propagation delays

Processing Delay

Required time to analyze a packet header and decide where to send the packet (e.g. a routing decision)

Inside a router this depends on the number of entries in the routing table, the implementation of data structures, hardware in use, etc.

This can include error verification, such as IPv4, IPv6 header checksum calculations.

Queuing Delay

- The time a packet is enqueued until it is transmitted
- The number of packets waiting in the queue will depend on traffic intensity and of the type of traffic (bursty or sustained)
- Router queue algorithms try to adapt delays to specific preferences, or impose equal delay on all traffic.

Transmission Delay

The time required to push all the bits in a packet on the transmission medium in use

For N=Number of bits, S=Size of packet, d=delay

d = S/N

For example, to transmit 1024 bits using Fast Ethernet (100Mbps):

d = 1024/1x10e8 = 10.24 micro seconds

Propagation Delay

- Once a bit is 'pushed' on to the transmission medium, the time required for the bit to propagate to the end of its <u>physical trajectory</u>
- The velocity of propagation of the circuit depends mainly on the actual distance of the physical circuit
- In the majority of cases this is close to the speed of light.

  For d = distance, s = propagation velocity

  PD = d/s

Transmission vs. Propagation

Can be confusing at first

Consider this example:

Two 100 Mbps circuits

- 1 km of optic fiber
- Via satellite with a distance of 30 km between the base and the satellite

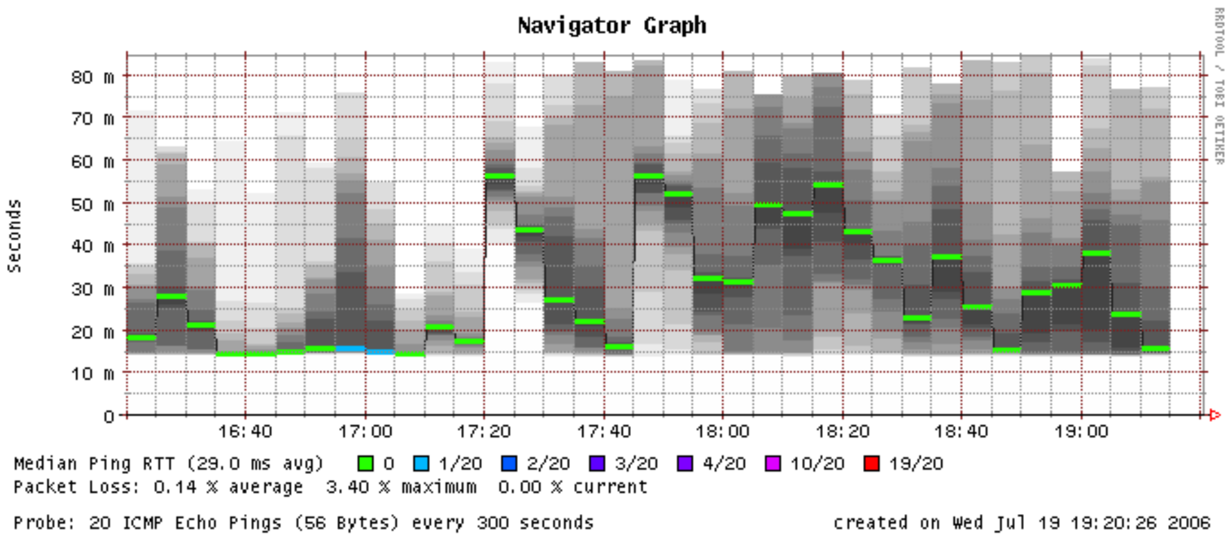
For two packets of the same size which will have the larger transmission delay? Propagation delay?

Packet Loss

Occurs due to the fact that buffers are not infinite in size

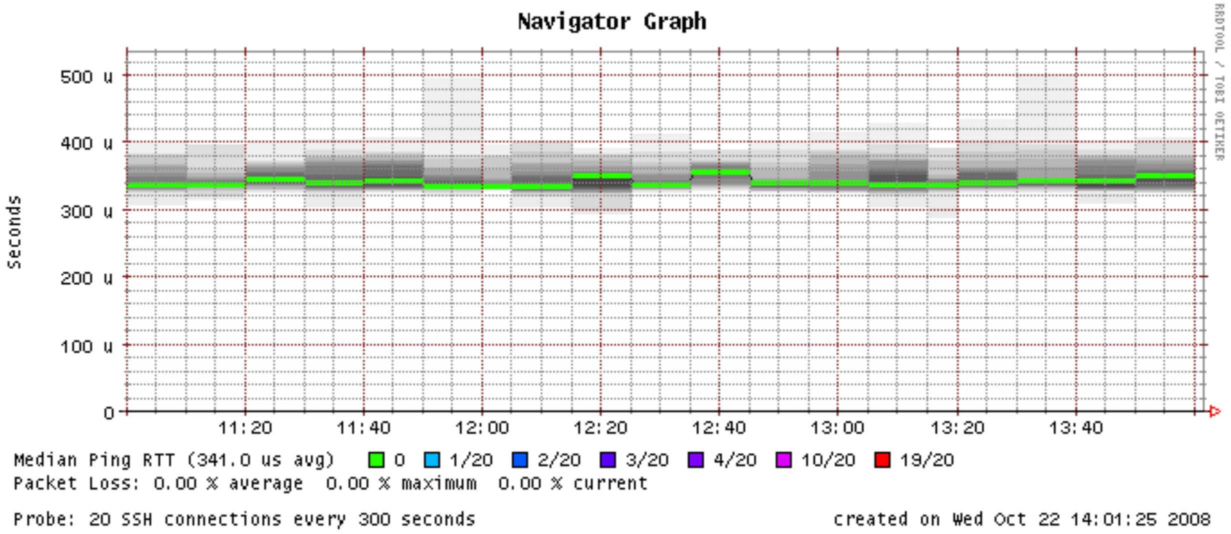
- When a packet arrives to a buffer that is full the packet is discarded.
- Packet loss, if it must be corrected, is resolved at higher levels in the network stack (transport or application layers)
- Loss correction using retransmission of packets can cause yet more congestion if some type of (flow) control is not used (to inform the source that it's pointless to keep sending more packets at the present time)

Jitter



Median Ping RTT (29.0 ms avg) 🟩 0 🟦 1/20 🟦 2/20 🟪 3/20 🟪 4/20 🟪 10/20 🟥 19/20
Packet Loss: 0.14 % average 3.40 % maximum 0.00 % current

Probe: 20 ICMP Echo Pings (56 Bytes) every 300 seconds created on Wed Jul 19 19:20:26 2006



Median Ping RTT (341.0 us avg) 🟩 0 🟦 1/20 🟦 2/20 🟪 3/20 🟪 4/20 🟪 10/20 🟥 19/20
Packet Loss: 0.00 % average 0.00 % maximum 0.00 % current

Probe: 20 SSH connections every 300 seconds created on Wed Oct 22 14:01:25 2008

Flow Control and Congestion

- Limits the transmission amount (rate) because the receiver cannot process packets at the same rate that packets are arriving.
- Limit the amount sent (transmission rate) because of loss or delays in the circuit.

Controls in TCP

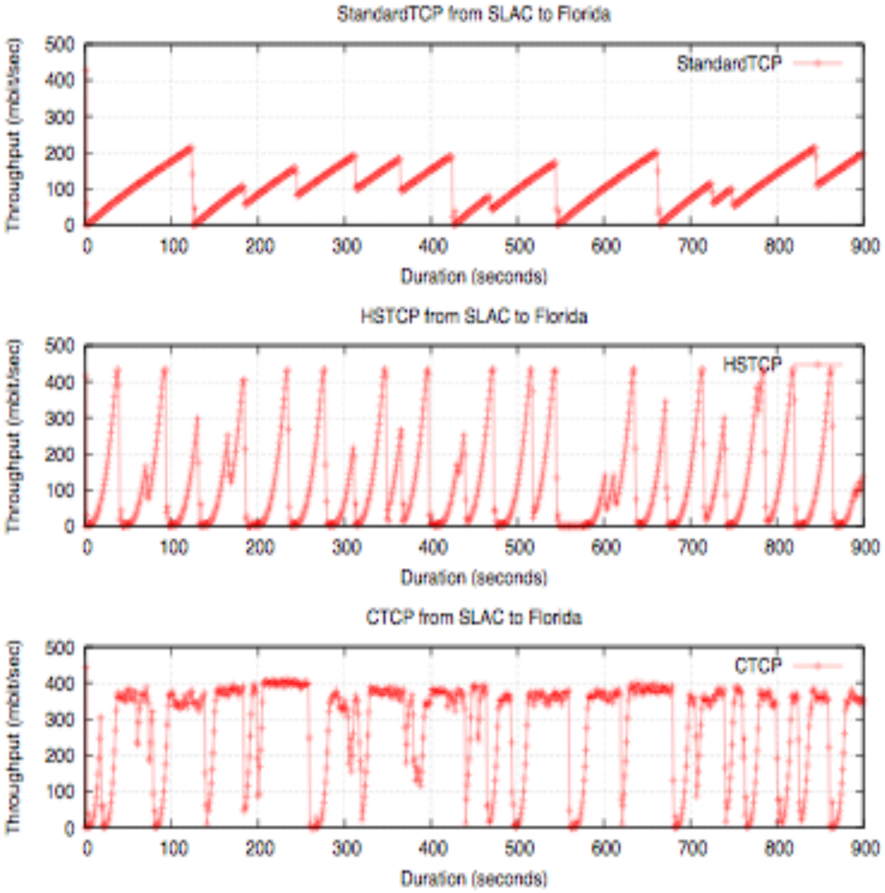
IP (Internet Protocol) implements service that is not connection oriented.

- There is no mechanism in IP to deal with packet loss.

TCP (Transmission *Control* Protocol) implements flow and congestion control.

- Only on the ends as the intermediate nodes at the network level do not talk TCP

Different TCP Congestion Control Algorithms



Questions?

?

Congestion vs. Flow in TCP

Flow: controlled by window size (RcvWindow), which is sent by the receiving end.

Congestion: controlled by the value of the congestion window (Congwin)

- Maintained independently by the sender
- This varies based on the detection of packets lost
  - Timeout or receiving three ACKs repeated
- **Behaviors:**
  - Additive Increments / Multiplicative Decrements (AIMD)
  - Slow Start
  - React to *timeout* events