Optimizing Network Performance on Mac OS X

Speaker
Peter Sichel
Chief Engineer
Sustainable Softworks
psichel@sustworks.com

MacWorld Boston 2005
Session P232
14-Jul-2005 9:00 AM - 10:15 AM

Introduction

Speaker Background

Macintosh developer specializing in kernel level TCP/IP networking for the past 9 years.

- IPNetMonitorX - Macintosh internet tools
- IPNetTunerX - network optimization tuner
- IPNetRouterX - native IP connectivity suite
- 14 years prior experience as software architect at Digital Equipment Corporation. Emphasis on user interface design and communications.

Overview

What is this session about:

- Improving network throughput and availability
- What happens when you push TCP/IP to the extreme?

TCP tuning and optimization

- Why is it necessary to tune TCP/IP in the first place?
- Network stack background
- Optimization 101 - measure, adjust, repeat
- Common tuning problems and solutions

Beyond tuning

- When is tuning not enough?
- Packet Shaping and QoS (Quality-of-Service)
- Automatic fail over
- Dead gateway detection

Why is it necessary to tune TCP/IP?

- TCP/IP is not controlled by any manufacturer and is designed to work on almost any kind of underlying network with widely varying characteristics.

- To meet this challenge, the protocol designers made TCP adaptive. TCP is self monitoring and optimizes its own behavior to match the network environment. Adaptation takes time and the default settings cannot optimize for every possible kind of network.

- There are many independent implementations, the protocols continue to evolve in response to practical experience, some implementations have peculiar compatibility constraints.

Network Stack Background

Based on an Open Source BSD reference implementation

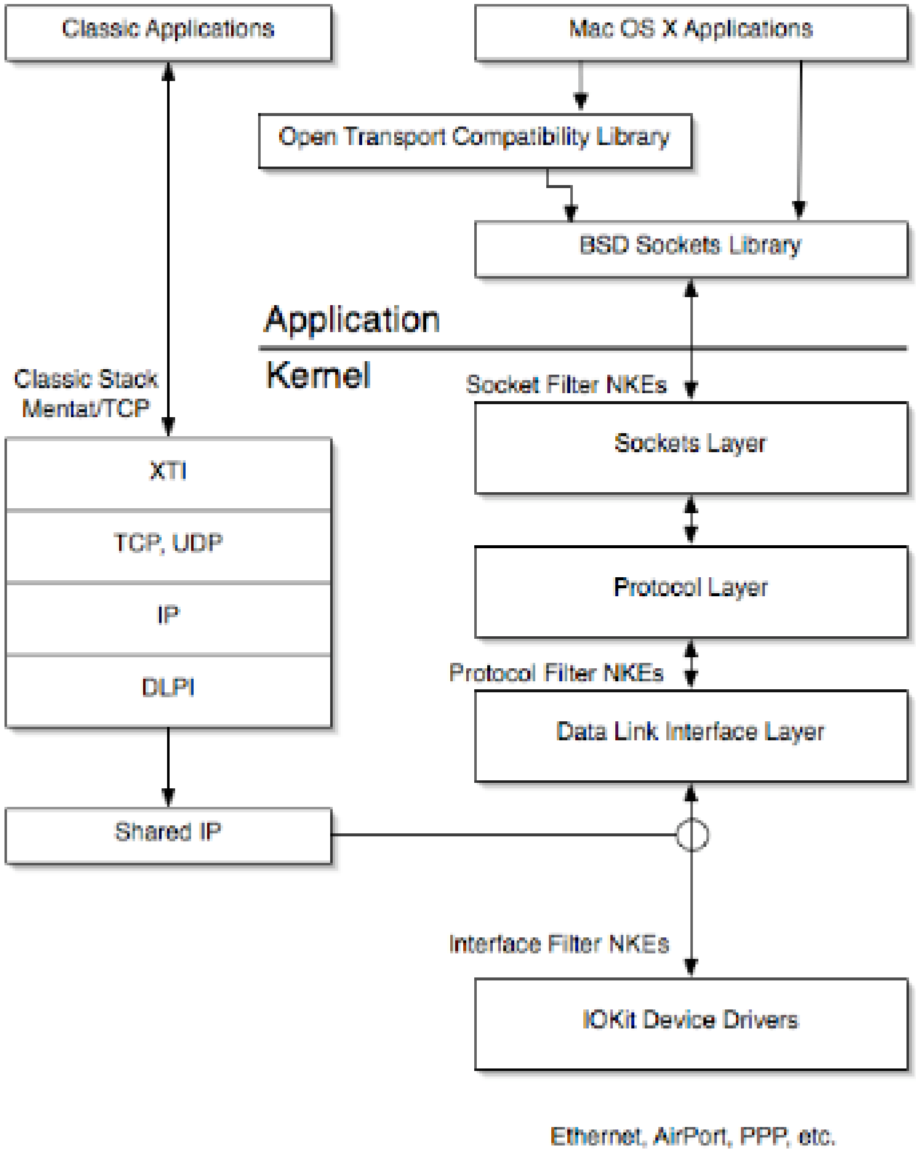
Apple enhanced for greater flexibility and user friendliness

- Dynamic configuration
- Power Management
- Plug ins
- Multithreading

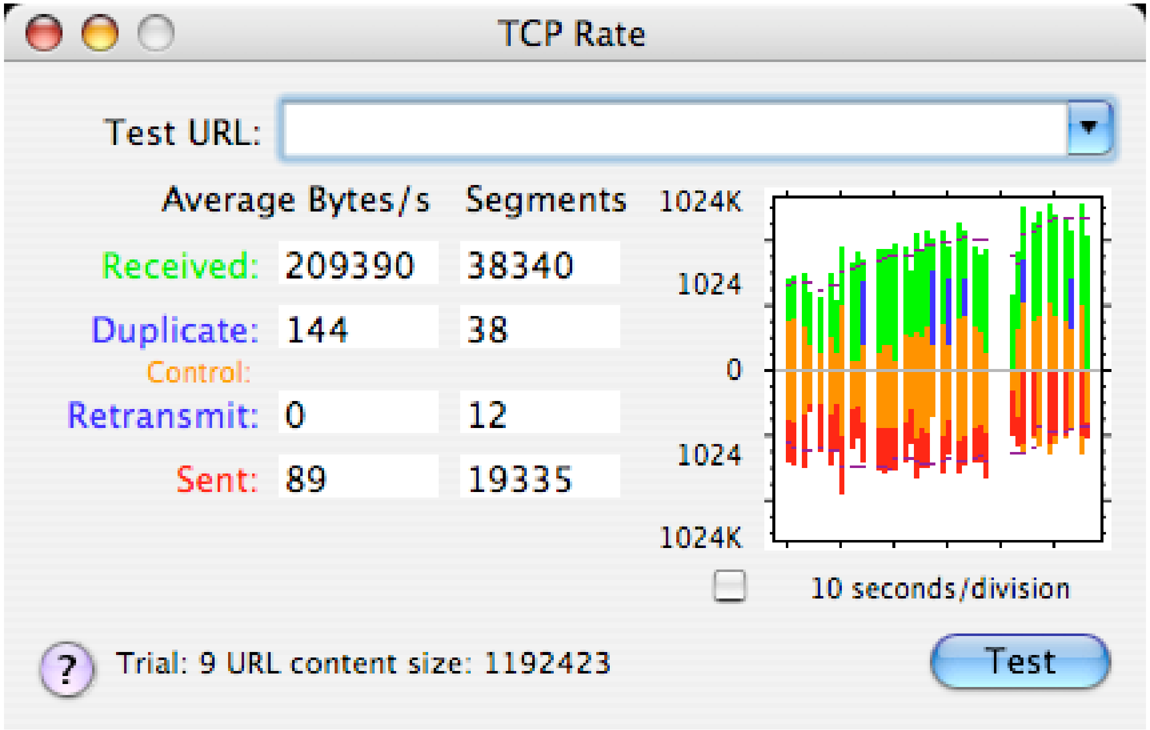
Apple’s goal is to make TCP/IP networking simple and transparent.

Tools like IPNetTunerX allow you to access the details when things don’t work as well as they should.

Mac OS X BSD Stack



Optimization 101 - measure, adjust, repeat



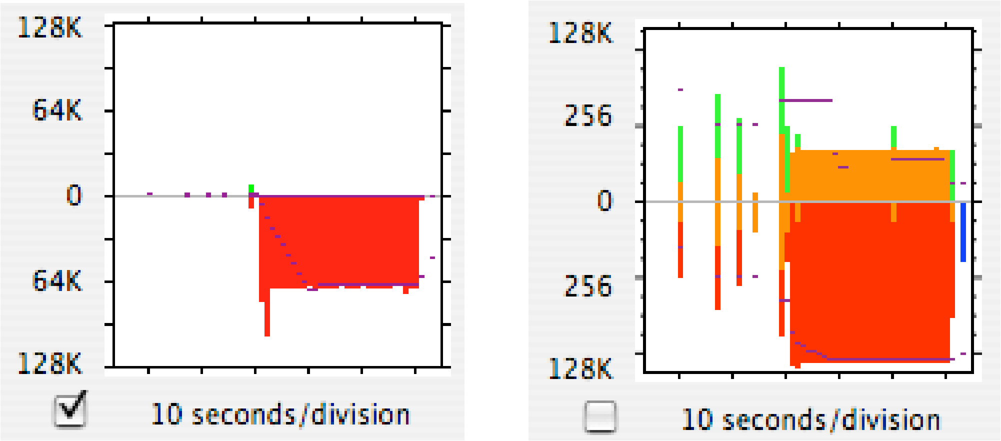
TCP Rate

Test URL:

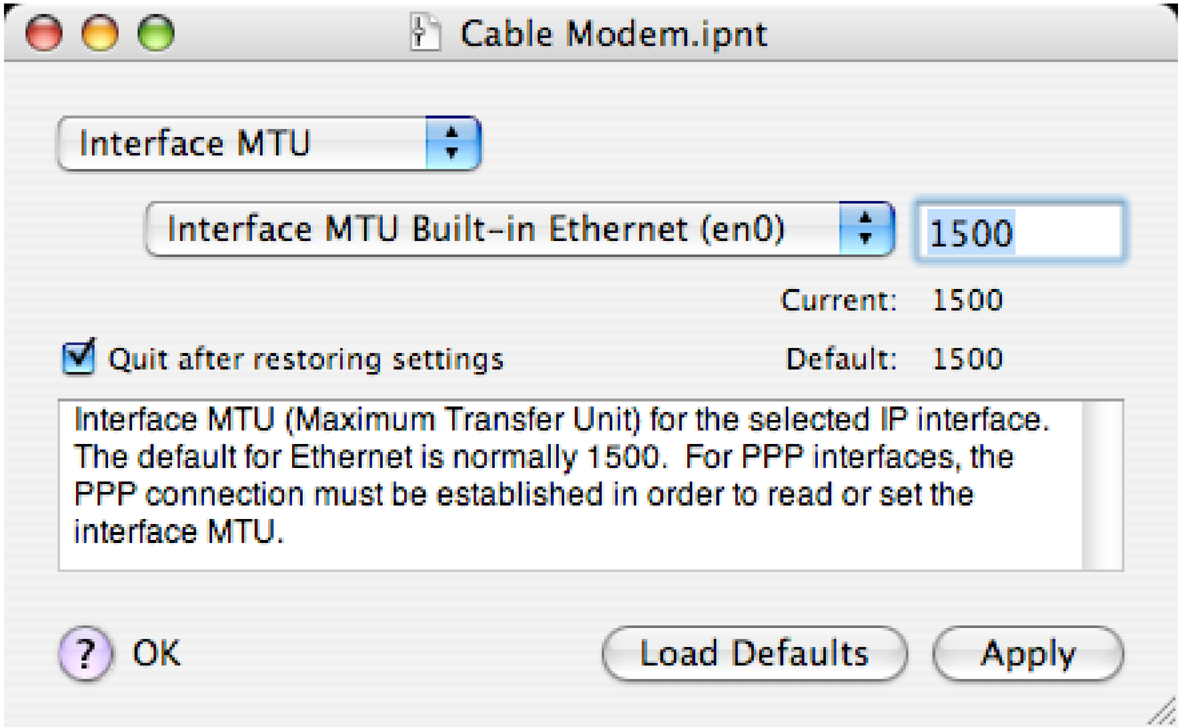
	Average Bytes/s	Segments
Received:	209390	38340
Duplicate:	144	38
Control:		
Retransmit:	0	12
Sent:	89	19335

Trial: 9 URL content size: 1192423

Problematic transfer with duplicate/retransmit data



Efficient transfer (Linear and Log scale)



Interface MTU

Interface MTU Built-in Ethernet (en0) 1500

Current: 1500

☑ Quit after restoring settings

Default: 1500

Interface MTU (Maximum Transfer Unit) for the selected IP interface. The default for Ethernet is normally 1500. For PPP interfaces, the PPP connection must be established in order to read or set the interface MTU.

IPNetTunerX tuner settings document

Common tuning problems and solutions (1)

TCP window size

- The receiver tells the sender its receive buffer size.

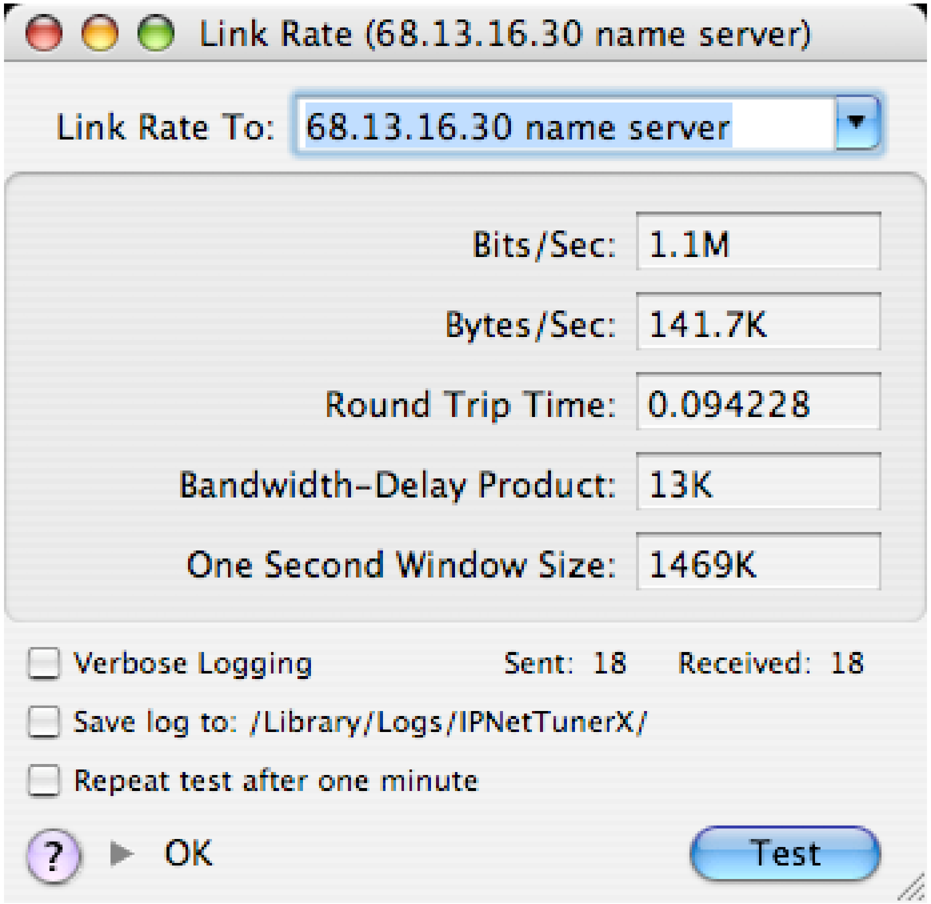
- The sender may transmit additional segments without waiting for an acknowledgment up to the receive buffer size.

- The capacity of the pipe is equal to the Bandwidth-Delay product.
  We want the window to be large enough to keep the pipe full.
  Speed-Limit = Window-Size / Round-Trip-Time

- net.inet.tcp.recvspace, net.inet.tcp.sendspace

- Increase window size to at least Bandwidth-Delay product.
  Do not exceed “1 second Window size”.

- Can use Link Rate tool to measure and calculate these values.



Link Rate (68.13.16.30 name server)

Link Rate To: 68.13.16.30 name server

Bits/Sec: 1.1M

Bytes/Sec: 141.7K

Round Trip Time: 0.094228

Bandwidth-Delay Product: 13K

One Second Window Size: 1469K

□ Verbose Logging Sent: 18 Received: 18

□ Save log to: /Library/Logs/IPNetTunerX/

□ Repeat test after one minute

Measuring RTT and Bandwidth-Delay product

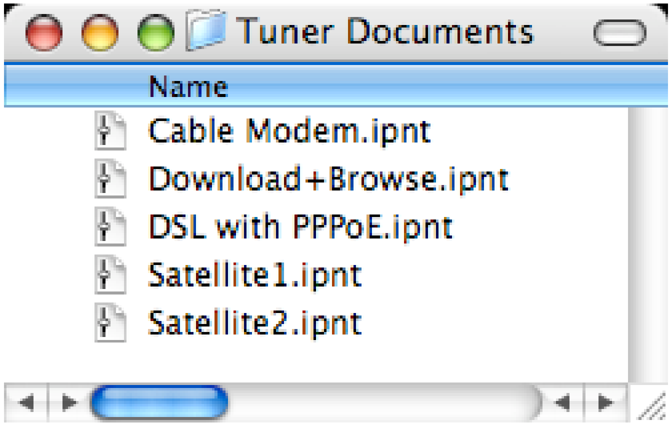
Common tuning problems and solutions (2)

TCP delayed ack

- net.inet.tcp.delayed_ack
- Turn off for links with high latency

Path MTU discovery

- net.inet.tcp.path_mtu_discovery
- Turn off if not supported



Tuner Documents

Name

Cable Modem.ipnt
Download+Browse.ipnt
DSL with PPPoE.ipnt
Satellite1.ipnt
Satellite2.ipnt

Common tuning problems and solutions (3)

Match the Network MTU

- The MTU or Maximum Transmission Unit defines the largest segment or packet that can transmitted across a given network (without being split into smaller segments).

- For best performance, we normally want to use the largest packet size possible to reduce per packet overhead.

- BSD stacks use an automatic path MTU discovery technique, but if your network connection requires a specific MTU, it may be necessary to set this explicitly for best performance or compatibility.

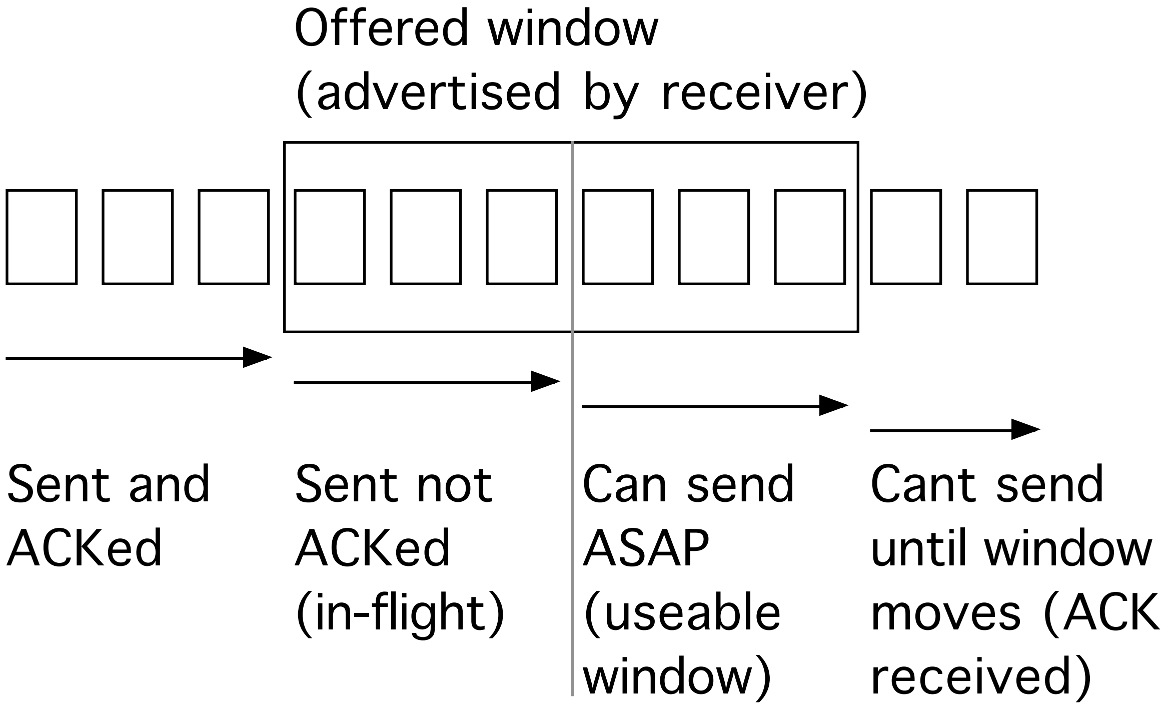
- May need to reduce manually for PPPoE and some VPNs.

- Use 1492 for PPPoE, try 1450 if some web pages fail to load.

Common tuning problems and solutions (4)

Selective ACK

- Retransmit missing segments only
- Windows option not currently supported on Mac OS X



Offered window
(advertised by receiver)

Sent and
ACKed

Sent not
ACKed
(in-flight)

Can send
ASAP
(useable
window)

Cant send
until window
moves (ACK
received)

When is tuning not enough?

When data links are saturated.

When you want to allocate bandwidth based on business priorities.

When you want redundancy, automatic fail over, or load balancing to increase available bandwidth.

Packet Shaping and QoS (1)

Each TCP connection tries to use as much bandwidth as it can get, and backs off when packets are lost or delayed.

When a data link becomes heavily saturated by multiple competing applications:

- ACKs can be delayed limiting throughput
- Efficiency drops as data is retransmitted
- Connections back off to avoid congestion

The purpose of Packet Shaping and QoS is to adapt network behavior to increase TCP/IP efficiency over the “slow link” or “last mile”.

Packet Shaping and QoS (2)

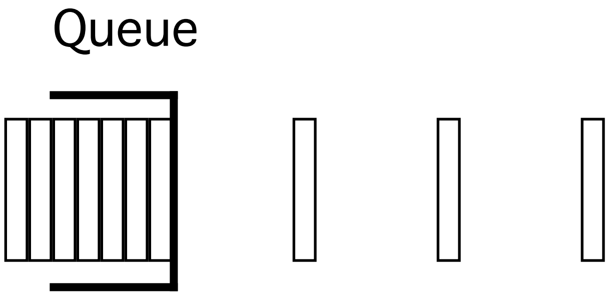
Three basic techniques:

- Queuing (buffering and waiting) to limit send rate. Since TCP is adaptive, it will self adjust to the impaired data link.
  `Examples: throttled, LinkSys WRT-54, Xincom`

- Expedite higher priority traffic by sending out of order or reserving dedicated bandwidth. May support TOS field in IP header.
  `Example: throttled`

- Use TCP's built-in flow control mechanism to modulate the send rate for maximum efficiency.
  `Examples: Packeteer, IPNetRouterX`

Packet Shaping Techniques (1)

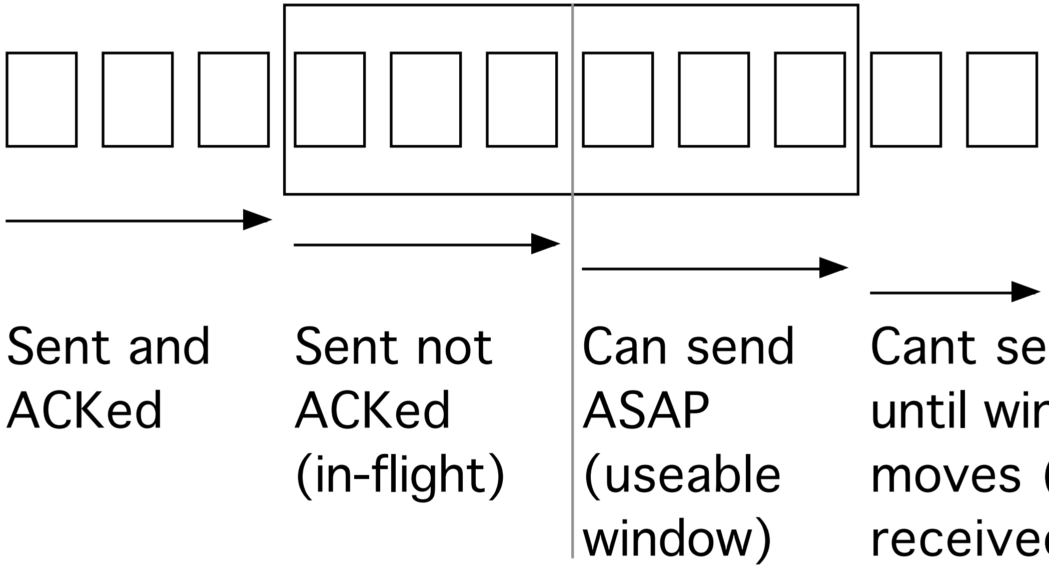


Queuing (buffering and waiting)

- Easiest to implement, can be applied to any protocol
- Can only rate limit send traffic
- Limited number of queues
- “Shaping” amounts to re-ordering and dropping packets
- Can be effective, but sacrifices some efficiency

Packet Shaping Techniques (2)

Offered window
(advertised by receiver)



Sent and
ACKed

Sent not
ACKed
(in-flight)

Can send
ASAP
(useable
window)

Cant send
until window
moves (ACK
received)

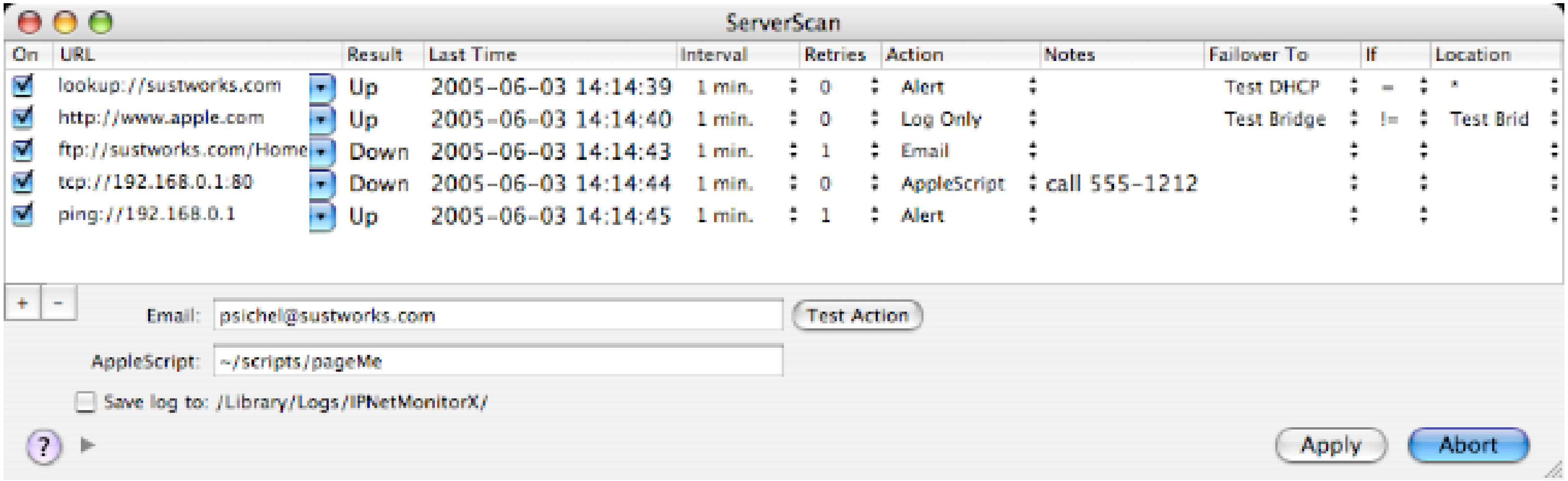
TCP Rate Limiting

- Adjust receive window on the fly
- Withhold and reinsert ACKs and modify window size.
- Can rate limit send and receive traffic
- Avoids congestion for best use of available bandwidth.

Automatic fail over

Network Location “Automatic”

- Dynamic configuration for immediately attached data links
- Can be augmented with Server Scan and Location Helper tool



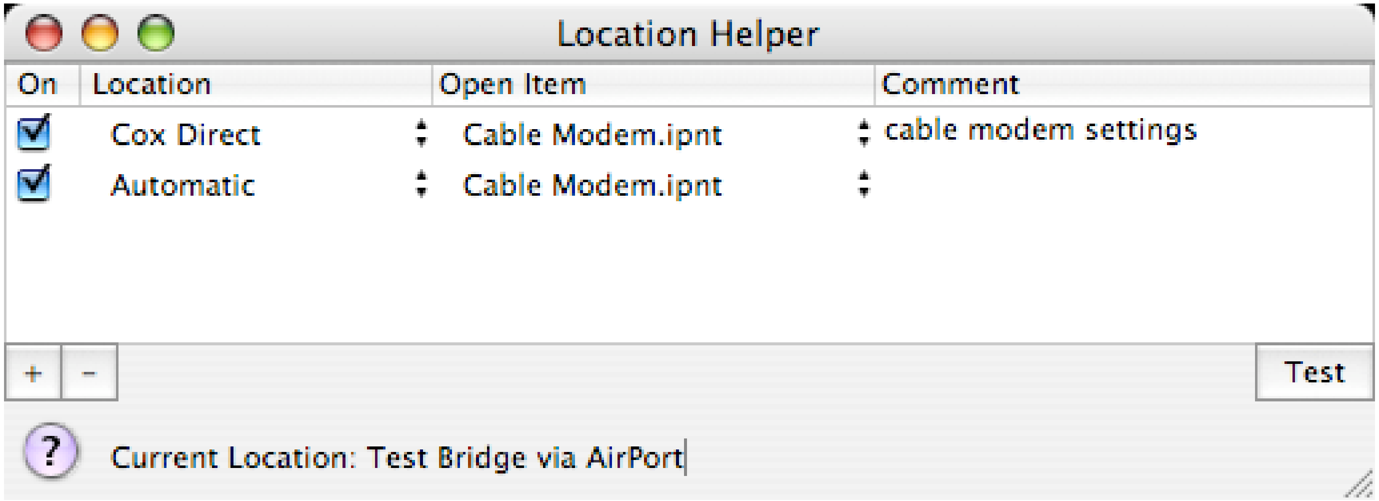
ServerScan

On	URL	Result	Last Time	Interval	Retries	Action	Notes	Failover To	If	Location
☑	lookup://sustworks.com	Up	2005-06-03 14:14:39	1 min.	0	Alert		Test DHCP	=	*
☑	http://www.apple.com	Up	2005-06-03 14:14:40	1 min.	0	Log Only		Test Bridge	!=	Test Brid
☑	ftp://sustworks.com/Home	Down	2005-06-03 14:14:43	1 min.	1	Email				
☑	tcp://192.168.0.1:80	Down	2005-06-03 14:14:44	1 min.	0	AppleScript	call 555-1212			
☑	ping://192.168.0.1	Up	2005-06-03 14:14:45	1 min.	1	Alert				

Email: psichel@sustworks.com

AppleScript: ~/scripts/pageMe

□ Save log to: /Library/Logs/IPNetMonitorX/



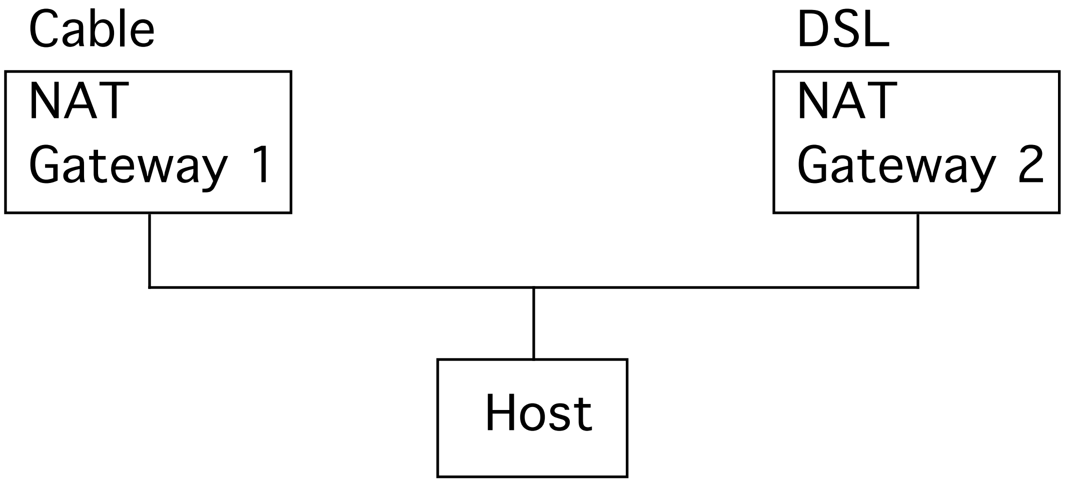
Location Helper

On	Location	Open Item	Comment
☑	Cox Direct	Cable Modem.ipnt	cable modem settings
☑	Automatic	Cable Modem.ipnt	

Current Location: Test Bridge via AirPort

Dead Gateway Detection

- Similar to Windows behavior
- When a connection attempt is retransmitted for the 3rd time, select an alternate gateway by creating a new host route on the fly.
- If more than half of active connections have failed over to alternate gateway, make it the default gateway by selecting a new location.
- Transparent fail over before connection attempt times out.



Questions & Answers

More Information:
http://www.sustworks.com
psichel@sustworks.com

Updated slides available here:
http://www.macworldexpo.com/BOS05conference