May 3, 2006 / 10:15 - 11:15

Infrastructure and Services Main Track

Skinny or Fat Networks: Improving Remote User and Branch Office Performance

Eric Siegel, Senior Analyst, Burton Group
esiegel@burtongroup.com

INTEROP®
MAKES YOU
SMART

Agenda

- *The Problem: Performance over the WAN*
- Problem Analysis
- Solutions
- Vendors
- Panel Discussion with:
  - Mark Day, Chief Scientist, Riverbed Technology
    - » www.riverbed.com/
  - Efi Gatmor, CTO, Expand Networks
    - » www.expand.com/
  - Raj Kanaya, VP Application Networking, Citrix Systems
    - » www.citrix.com/
  - Tom Yohe, VP Engineering, Stampede Technologies
    - » www.stampede.com/

- *Keep a list of questions to ask our panel at the end of the session!*

The Problem: Performance over the WAN

- The Problem Appears
  - An application that's working well over a LAN
  - Is moved to a WAN, and:
    - Response time is much longer than expected
    - Application timeouts may appear
  - But the application cannot be altered
    - It's a packaged application, or
    - Application redesign is too expensive
  - At great expense, backbone bandwidth is massively increased
    - And the problem remains!
  - Now what do we do?
    - *Fortunately, performance of these applications can often be improved without making any changes to the application or to the WAN itself!*

Agenda

- The Problem: Performance over the WAN
- *Problem Analysis*
  - *Transmission Latency*
  - *Bandwidth Restriction and Fat Files*
  - *Error Rate*
  - *Ping-Pong Protocols*
- Solutions
- Vendors
- Panel Discussion

Problem Analysis

- Transmission latency is caused by physics
  - *Increasing the line speed decreases serialization latency, but it does not decrease transmission latency*

    0 ms. to serialize 1500 bytes onto 100 Mb/s Fast Ethernet

    8 ms. to serialize 1500 bytes onto a 1.5 Mb/s link

  - Typical one-way transmission latencies:

    0 ms. on a LAN

    30 ms. New York to San Francisco

    50 ms. New York to London

    125 ms. New York to Melbourne, Australia

    260 ms. per hop through a geosynchronous satellite

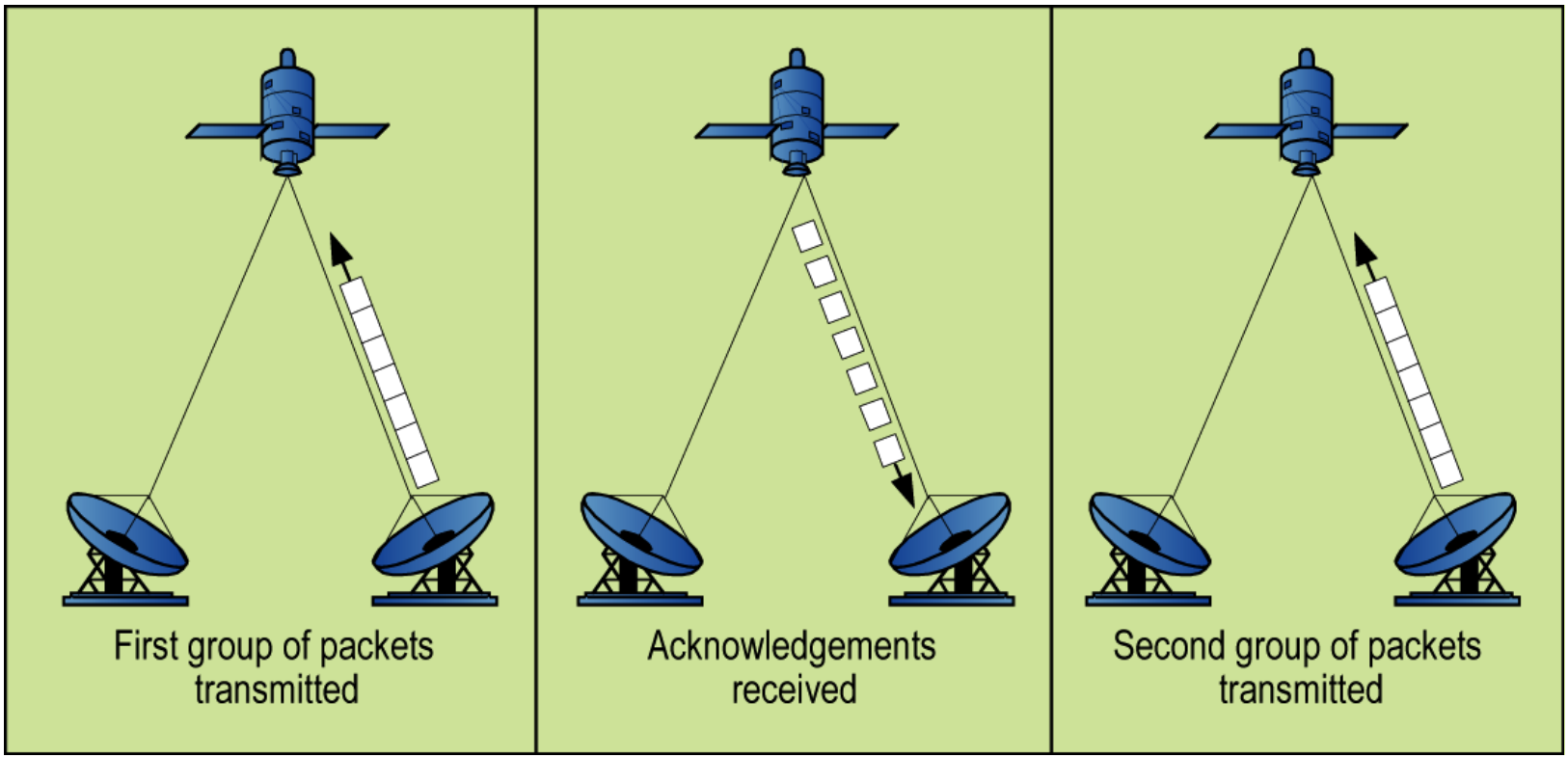
    - *There may be multiple hops for VSAT or other situations*

Problem Analysis

- Bandwidth restriction is typical of WANs
  - Applications that transmit massive amounts of data have problems with bandwidth-restricted WANs
    - NOTE: much of that data may be unnecessary ("fat files")
      - Repetitive transmission of Web images, formatting information, etc.
      - Identical file transmissions to multiple users at the same location
  - Effective bandwidth depends on:
    - The ability of the protocol to fill the transmission channel
      - Protocols pause if acknowledgements are not received in time
      - NOTE: The application layer may have its own flow control
        - » Windows file access protocol ("CIFS") transmits a maximum of 64KB before pausing for acknowledgement
    - The error rate (error discovery and recovery)

Problem Analysis

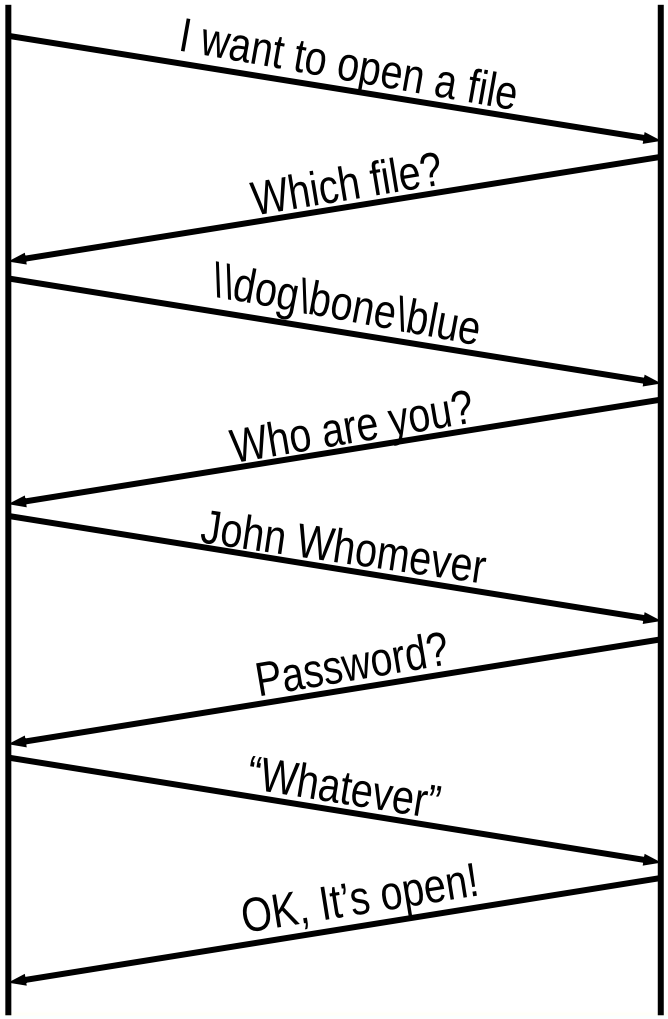
- Flow control issues on a high-latency, high-bandwidth link



(0.26 seconds one-way latency through geosynchronous satellite)
Effective throughput with Windows/XP defaults is 35 KBytes/sec on a 45 Mbps link!

Problem Analysis

- "Ping-Pong" protocol response time is affected by link latency



	LAN	WAN (50 ms one-way latency)
	0 ms.	50 ms.
	0 ms.	150 ms.
	0 ms.	250 ms.
	0 ms.	350 ms.
	0 ms.	400 ms.

Agenda

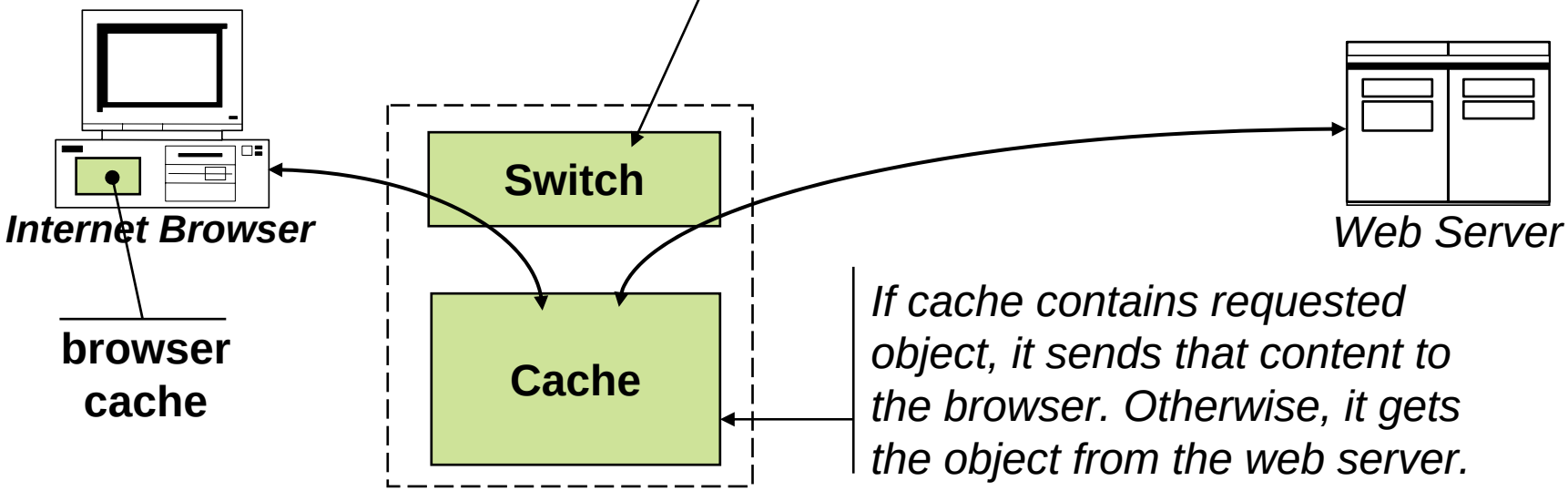
- The Problem: Performance over the WAN
- Problem Analysis
- *Solutions*
  - *Caching*
  - *Compression*
  - *Protocol Acceleration*
  - *Wide Area File Services*
  - *Client Optimization*
- Vendors
- Panel Discussion

Solutions: Caching

- Caching works on identical, complete files
  - Forward caches improve response time and decrease bandwidth
  - The request to the server will come from the cache, and the cache can use more efficient transmission parameters (window size, packet size, protocol, etc.) than the typical end user

Interception: *Port 80 traffic goes through cache regardless of destination*

Proxy: *Browser explicitly sends Port 80 traffic directly to cache*



Traffic is diverted by external or internal switch or router

Solutions: Caching

- Cache must not serve stale (obsolete) content!
  - Web server controls object expiration by using HTTP headers

```
HTTP/1.1 200 OK
Date: Sat, 06 May 2000 18:06:07 GMT
Expires: Mon, 30 Oct 2000 14:19:41 GMT
```

  - Dynamic and secure objects are not usually cached
    - Web caches won't store cgi, .asp, .jsp, GET query-string, POST
- Many public Web access ISPs provide free forward caching
- Content Distribution Networks offer forward caching as a service

Solutions: Compression

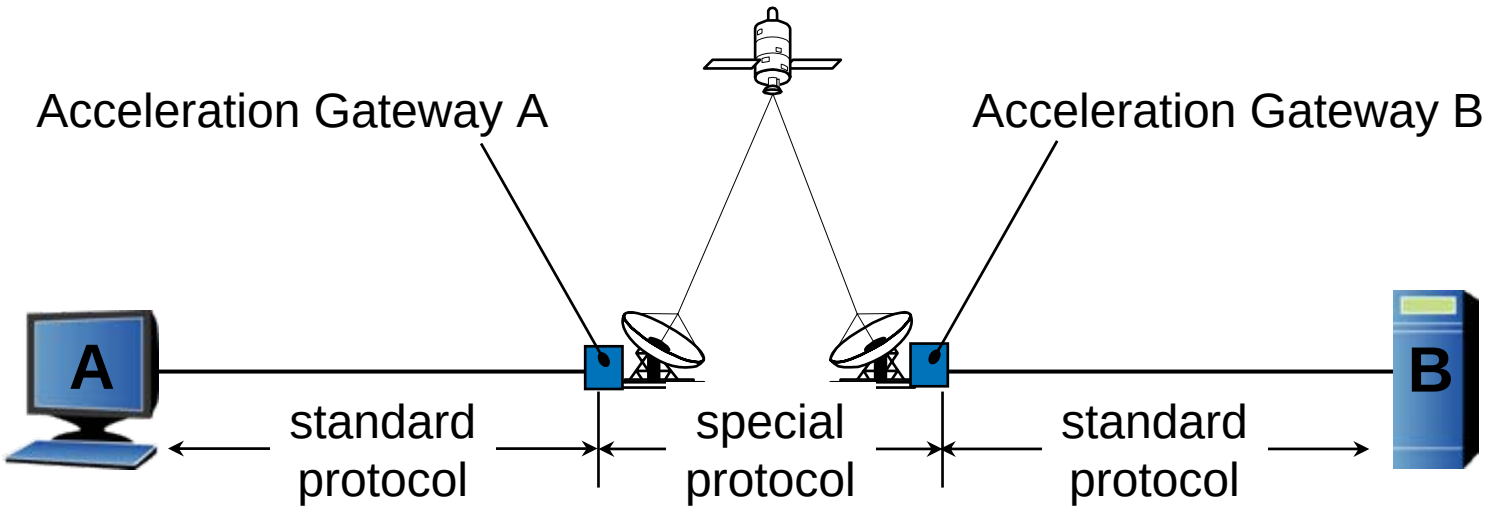
- Compression decreases the bandwidth needed
  - It will usually also improve transfer time
  - Compression efficiency is sensitive to:
    - The algorithm (possibly tuned to the particular protocol or application)
    - The length and frequency of any repeating patterns in the data
    - The length of the data window that can be examined
    - The amount of storage for shared compression directories
  - Encrypted data appears to be "random"; *it cannot be compressed*
  - Standard compression techniques are commonly available:
    - gzip/deflate compression (included in all modern browsers)
    - Media compression (PNG, JPEG, GIF, MPEG ...)
  - *And Now*... Advanced Compression! There have been *major* advances in compression technologies over the past three years!

Solutions: Compression

- Advanced Compression:
  - Requires installation of software or an appliance at both ends
  - "Data reduction" compression builds compression dictionaries at both ends of the path, then substitutes pointers for data strings
    - Can remember sequences that are megabytes long and are separated by gigabytes
    - Some similarities to caching, BUT:
      - Works on any data sequence, on any application
        - » Can find identical sequences across different applications
        - » Does not require identical, complete, identically-named files
      - No stale data problem
      - Proprietary; requires software or appliance at both ends
  - Compression engines can now run at extremely high speeds (e.g., 45+ mbps) with extremely low latency (e.g., 2 ms.)

Solutions: Protocol Acceleration

- Protocol acceleration to improve flow and error control and to reduce ping-pong behavior
  - Hide network latency and chatty ping-pong protocols by using gateways to convert to a more efficient protocol
    - A simpler form is sometimes called “spoofing”
      - NOT the security community’s “spoofing”; they mean “impersonation”
    - Requires external box or plug-in at receiver



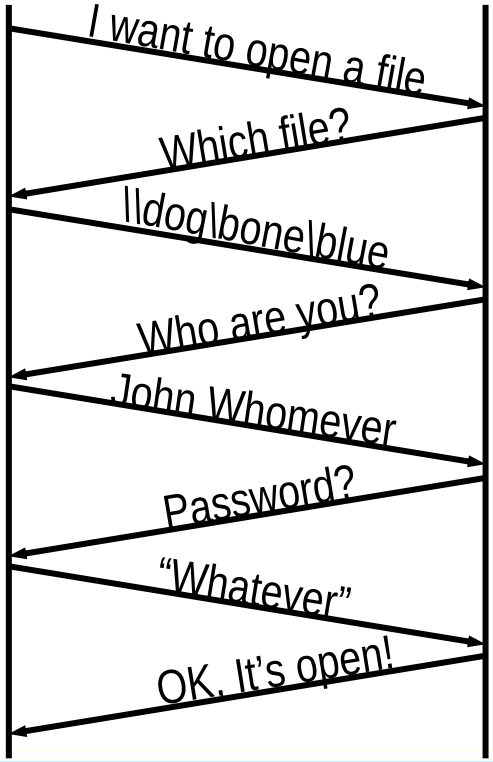
Solutions: Protocol Acceleration

- Protocol acceleration to improve transport flow and error control
  - Improve flow control by tuning the protocol; for example:
    - Larger windows
    - Avoid TCP's Congestion Avoidance algorithm
  - Improve error detection and recovery
    - Forward Error Correction (FEC) codes
    - Specialized protocols (e.g., SCPS-TP) to handle error bursts

Solutions: Protocol Acceleration

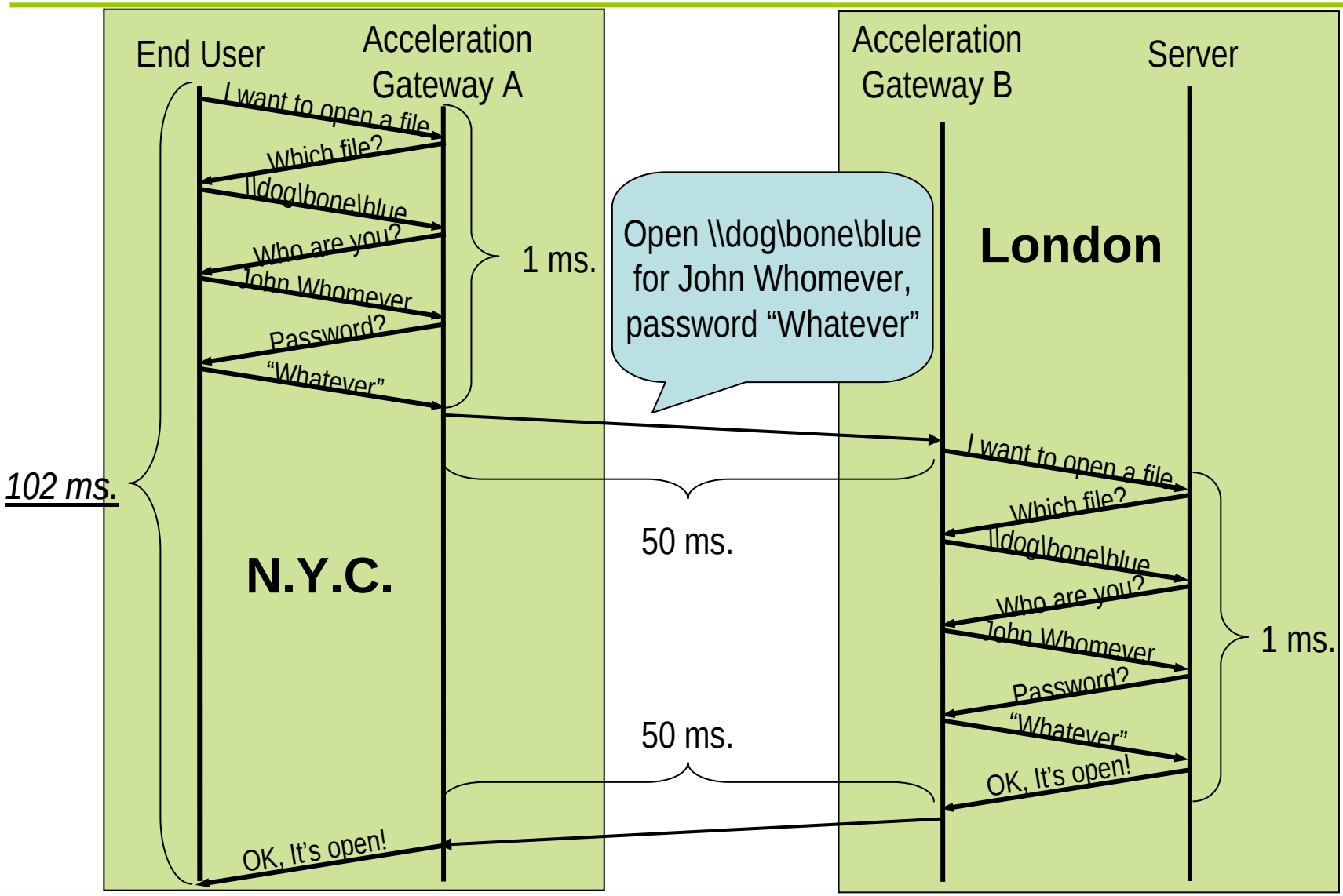
- Protocol acceleration to reduce ping-pong behavior
  - Useful for Microsoft's CIFS (Common Internet File System), MAPI (Messaging Application Program Interface), Unix's NFS (Network File System) and many other high-level protocols (database, etc.)
  - May also perform read-ahead and other complex optimizations

Do you remember this?
It took 400 ms.



Let's see how acceleration helps...

Solutions: Protocol Acceleration



Solutions: Wide Area File Services

- Wide Area File Services (WAFS) optimizes access to files over a large geographic area
  - Remote appliance replaces remote file server
    - Contains true file server functionality and a huge disk-based file cache
      - Can function even if the central site is unavailable
      - Can provide local file storage (e.g., for temporary files)
    - Communicates with WAFS coordination appliance at central site using enhanced protocol and compression techniques
    - May provide other branch functions (print, DNS, DHCP, etc.)
  - Primarily used to centralize backup and security controls
    - Can also manage shared files (locking is controlled at the central site)
    - Pure WAFS solutions did not handle client-server, MAPI, etc.
  - Overlap between WAFS vendors and protocol acceleration vendors
    - WAFS sold to server groups; protocol acceleration sold to network groups

Solutions: Client Optimization

- For Individual Clients:
  - For clients on the public Web:
    - Content Distribution Networks (CDNs) and access services
  - Web browser optimization
    - All browsers handle gzip/deflate decompression
    - Specialty solutions provide:
      - Advanced caching; decrease the time lost testing for object expiration
      - Advanced bi-directional compression, “delta encoding”
      - Enhanced transmission between server and browser
      - Protocol pre-fetching and acceleration; e.g., for HTTP and CIFS
      - ... and they do it with plug-ins (some can be loaded automatically)
  - TCP socket-level optimization
    - For client-server applications
  - Terminal Emulator (thin client)
    - Leave the application and its files on the host!
    - But you’ll want a low latency connection

Agenda

- The Problem: Performance over the WAN
- Problem Analysis
- Solutions
- *Vendors*
- Panel Discussion

Vendors

These are brief, incomplete summaries! Visit vendors on the show floor!

- Certeon
  - *S-Series*: advanced compression; transport, HTTP, XML, Oracle, WebDAV, MS SharePoint protocol acceleration; QoS; SSL decrypt/encrypt inside box
- Cisco
  - *Wide Area Application Engine*: includes WAFS (from Actona)
  - *Application Velocity System*: for client optimization (from FineGround)
- Citrix Systems
  - *Thin client terminal emulator*: "application virtualization" using ICA protocol
  - *AppCompress Extreme*: for client optimization (from NetScaler)
- Expand Networks
  - *Accelerator*: advanced compression, transport protocol optimization, QoS, forward caching, fully-integrated WAFS (from DiskSites), print, DNS, DHCP; IP and non-IP
- F5
  - *WANJet*: advanced compression, transport & CIFS protocol acceleration, QoS
  - *WebAccelerator*: for client optimization
- Juniper Networks
  - *WX*: advanced compression; transport, CIFS, HTTP protocol optimization; QoS; link sharing with "policy-based multipath" (from Peribit)

Vendors

- Orbital Data
  - *Orbital 6500*: advanced compression, transport & CIFS protocol acceleration
  - *Orbital Edge*: for client optimization, includes CIFS protocol acceleration
- Packeteer
  - *Xpress, SkyX*: advanced compression, transport protocol acceleration, QoS
- Riverbed Technology
  - *Steelhead*: advanced compression; transport, CIFS, MAPI, MS-SQL, HTTP protocol acceleration
- Silver Peak Systems
  - Advanced compression, transport and CIFS protocol acceleration, QoS
- Stampede Technologies
  - *TurboGold*: Lotus Notes and Lotus Domino acceleration
  - *WebRider*: for individual remote Web, WebDAV, TCP socket users
- Tacit Networks
  - WAFS, print, forward caching, DNS, DHCP, files attached to MAPI email
  - Advanced compression, transport protocol acceleration
- AND.... *CDN and remote access (may be able to optimize SSL traffic also)*:
  - Akamai's Content Delivery Services and Application Performance Services
  - Netli's NetliOffload and NetLightning

Agenda

- The Problem: Performance over the WAN
- Problem Analysis
- Solutions
- Vendors
- *Panel Discussion with:*
  - Mark Day, Chief Scientist, Riverbed Technology
    - » www.riverbed.com/
  - Efi Gatmor, CTO, Expand Networks
    - » www.expand.com/
  - Raj Kanaya, VP Application Networking, Citrix Systems
    - » www.citrix.com/
  - Tom Yohe, VP Engineering, Stampede Technologies
    - » www.stampede.com/

Panel Discussion

- A few questions to start:
  - "I'd like to centralize 20 Microsoft file servers servers to improve control and backup. Please compare protocol acceleration (e.g., Riverbed) to WAFS (e.g., Expand/DiskSites)"
  - "I have traveling employees using Web-enabled and client-server applications that have poor WAN performance. Please discuss terminal emulators (e.g., Citrix ICA) and client optimization products (e.g., Citrix's AppCompress Extreme and Stampede's WebRider)"
  - "Which situations are handled better by a terminal emulator, and which are handled better by WAFS or a protocol accelerator?"
  - "I've heard that WAN optimization solutions hide traffic inside tunnels. Will that interfere with QoS, multihoming, MPLS, access control lists, router redundancy, problem diagnosis, etc.?"
  - "What are each vendor's vulnerabilities? How do they handle them?"
  - "How do I cost-justify WAN optimization?"