Wednesday – February 27, 2008

A Pathway to a Connected California —
"The Value of Converged Networks"

Impact of Unified Communications

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

Issues

The unified-communications network is real-time, multi-service, and mission-critical

- The infrastructure carries real-time VoIP, videoconferencing
- And it carries complex transaction and web services flows

- *Some flows use large amounts of bandwidth*
- *Some flows have stricter performance requirements than others*
- *Problems must be detected and fixed very quickly*

- *Increasing demands on bandwidth will create capacity challenges!*

Implications of Unified Communications

Agenda

- Issues
- *Optimization Technologies*
  - *Quality of Service*
  - *WAN Performance Optimization ("Application Acceleration")*
  - *"Application Delivery"*
- Fast Problem Detection and Repair

Quality of Service

Quality of Service (QoS) classifies network traffic, then gives some of it special handling

- Useful at congestion points
- For example, give priority to VoIP delay-sensitive traffic
  - Less sensitive traffic is delayed, but protocol modifications and acceleration (discussed next!) may make that delay less apparent.
- Typical QoS design:
  - High priority for VoIP and video teleconferencing
  - Medium priority for streaming video and transactions
  - Low priority for bulk data transfer
  - "Policing" to discard traffic that shouldn't be on the network at all
- Performance requirements may be associated with user or source
  - E.g., when the governor speaks...

WAN Performance Optimization

5

WAN Performance Optimization ("Application Acceleration")

- In many cases, application performance can be greatly improved *without making any changes to the application or to the WAN*.
  - This can release bandwidth for use by applications (e.g., VoIP) that need priority by QoS because they can't be tricked by optimization
- Set expectations properly! There are rapid changes in technology and market; *optimization solution may have a short life (< 5 years)*.
  - Fortunately, WAN performance optimization technology can often pay for itself in under one year.

Types of optimization:

- Compression
- Protocol Acceleration (TCP, CIFS, NFS, WAFS, email, etc.)
- Application Remoting
- Streaming Media (audio, video)
- Individual Remote User Optimizations

WAN Performance Optimization:

Compression

6

New "Data Reduction" Compression Technology

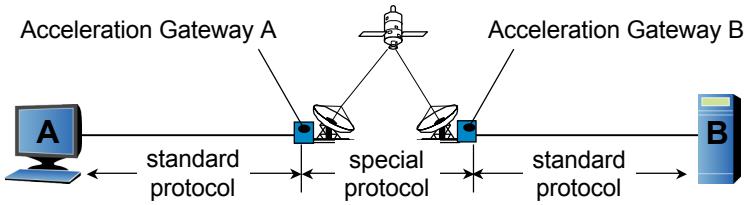
- Proprietary compression algorithms
- Detects and compresses duplicate segments that are megabytes long and separated by gigabytes of other data
- Duplicate strings are stored in dictionaries at both ends of path
- Usage Examples:
  - Short, similar transmissions to and from a remote call center
  - Files with internal repetition
  - Large files sent between two locations, changing only slightly each time
  - Backups or file synchronization to a central location
  - SAN over WAN

WAN Performance Optimization:

Protocol Acceleration

Protocol acceleration to improve flow and error control and to reduce ping-pong behavior

- Hide network delay and inefficient protocols by using gateways to convert to a protocol that is more suitable for the WAN
  - (Improve performance of TCP, HTTP, Microsoft CIFS, NFS, email, etc.)
- Wide Area File Services (WAFS) is a type of protocol acceleration
  - 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
    - May provide other branch functions (print, etc.)



WAN Performance Optimization:

Application Remoting

"Application Remoting" carries only the screen image, keystrokes, and mouse movements over the WAN

- The application and its files remain together on the server
  - Avoids pulling huge files and applications across the WAN
  - Saves money by using fewer application licenses
  - Examples: Citrix ICA, Microsoft Terminal Server

WAN Performance Optimization:

Streaming Media

Streaming Media

- Special caches
- Content Distribution Networks
- Consider peer-to-peer technologies

WAN Performance Optimization:

Individual Remote User Optimizations

Individual mobile users outside offices need optimization

- For clients on the public Web:
  - Content Distribution Networks (CDNs) and access services
- For all clients:
  - Specialty "site-to-user" optimization solutions provide:
    - Advanced caching
    - Advanced bi-directional compression
    - Enhanced transmission between server and client
  - Consider use of Application Remoting, or conversion to Web or Java

Application Delivery

Paradigm shift from "application deployment" to "application delivery"

- Citrix and others plan to create integrated systems that will automatically evaluate the user's platform and link to the servers, then select the optimum combination of:
  - Application remoting
  - Downloading the application and its data to the user
  - WAN performance optimization
  - Client and server virtualization
  - VPN
  - Measurement and session-management tools

... to provide a high-performance, secure experience for the end user without any need for that user to log on differently

  - And it will make system management easier

Implications of Unified Communications

Agenda

- Issues
- Optimization Technologies
  - Quality of Service
  - WAN Performance Optimization ("Application Acceleration")
  - "Application Delivery"
- *Fast Problem Detection and Repair*

Fast Problem Detection and Repair

... "Problems must be detected and fixed very quickly" ...

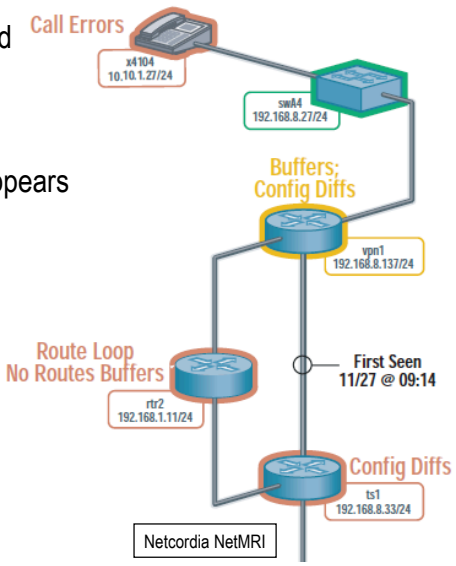
but how?

- Configuration Audit
  - Configuration Management Data Base
  - Audit tool ("Change Detector")
- Triage Metrics
  - For quick incident detection and "triage" (isolation to a subsystem) in a complex, rapidly-changing, partially-opaque infrastructure

Fast Problem Detection and Repair

Audit

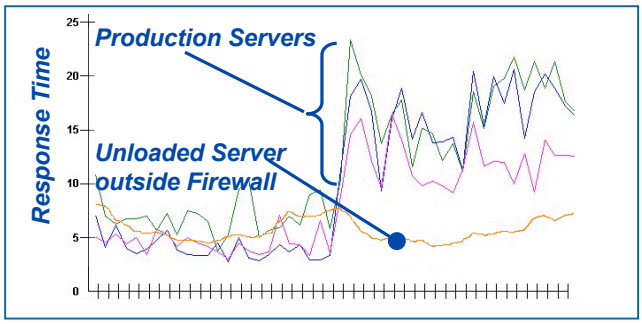
- Find topological, software and configuration changes before they cause problems
- Create record of changes for reference if an incident appears



Fast Problem Detection and Repair

An example of "triage"

- "Performance is poor. Should the Network Operations Center call the server people or the network people?"
  - Using active synthetic response-time measurement from branch office:



- Note: this is not diagnostic; we don't know why there is a problem. But we do know which team is probably responsible!

Fast Problem Detection and Repair

Triage design

- Three basic goals:
  - 1. Recognize well-understood incidents (and the quick fix for them!)
  - 2. Isolate subsystem problems to the responsible organization *credibly*
  - 3. Recognize complex, major-crisis problems *quickly*

- Design approach:
  - "Look at what you do during a meltdown, but without the stress!"
  - Create clear diagnostic demarcation points
    - Design the metrics for easy incident detection and triage
    - Examples: DNS, Internet access, authentication database, message queues to mainframes, branch office access
    - NOTE: If you can't create clear demarcation points, maybe you should redesign your system architecture!
  - Collect trendlines; build easy-to-use triage plans for the NOC