

NETWORLD INTEROP

Las Vegas 2004

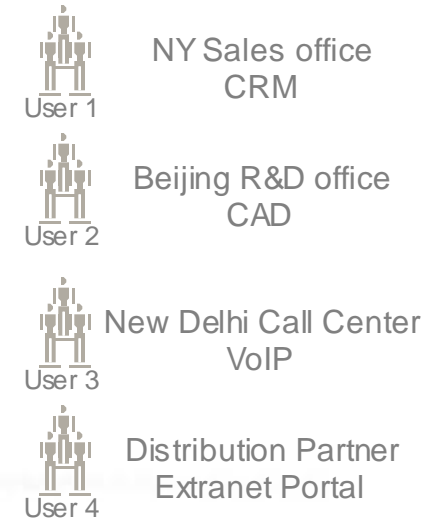
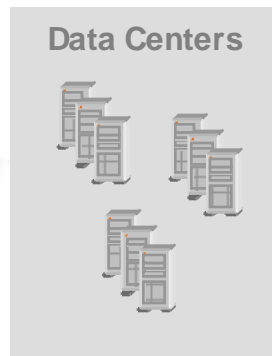
Performance and Data Center Conference

Performance Grace Under Applications Stress

Bad Things Happen: The Case for Adaptive Infrastructure

Tim Lee-Thorp, VP Marketing
13 May 04

Here is what I want...



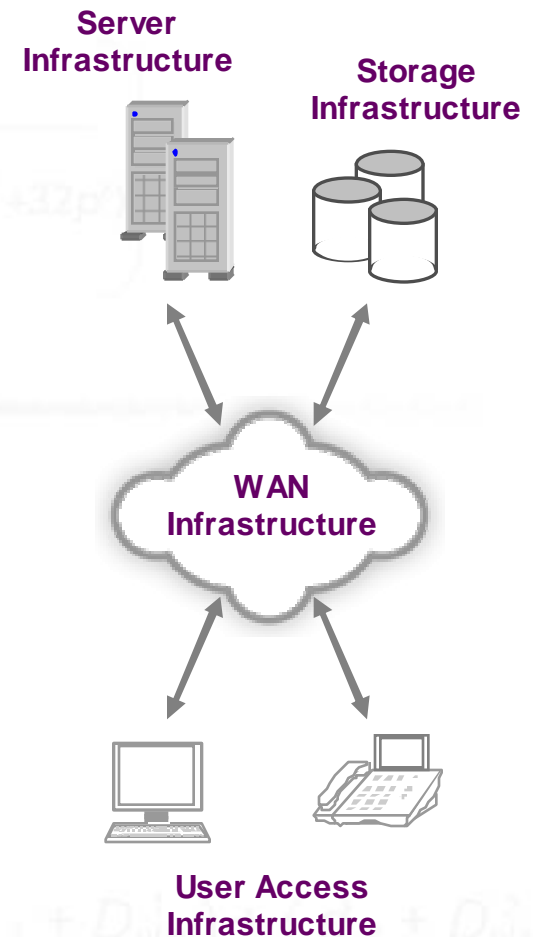
CEO: “Make sure my workers can get their jobs done as effectively as possible, while staying within expense budgets”

CIO: “Make sure that workers can use their apps and data wherever they are, while staying within expense budgets”

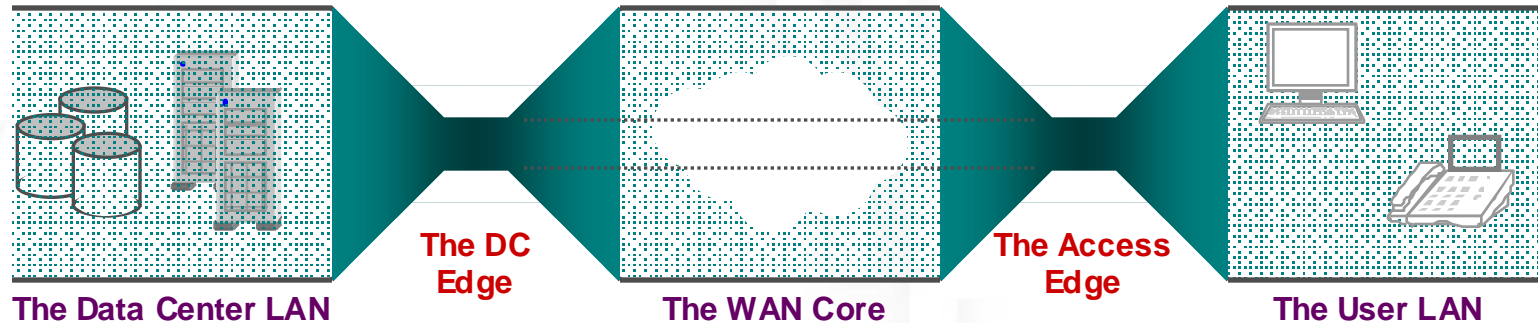
Ops: “Deploy the most redundant, most resilient architecture we can afford, and assume that bad things will still happen”

The four major IT infrastructure areas

- ...that support users and their apps
 - Servers
 - Storage
 - WAN
 - User access
- This session focuses on the WAN and how it both contributes to and impacts applications availability

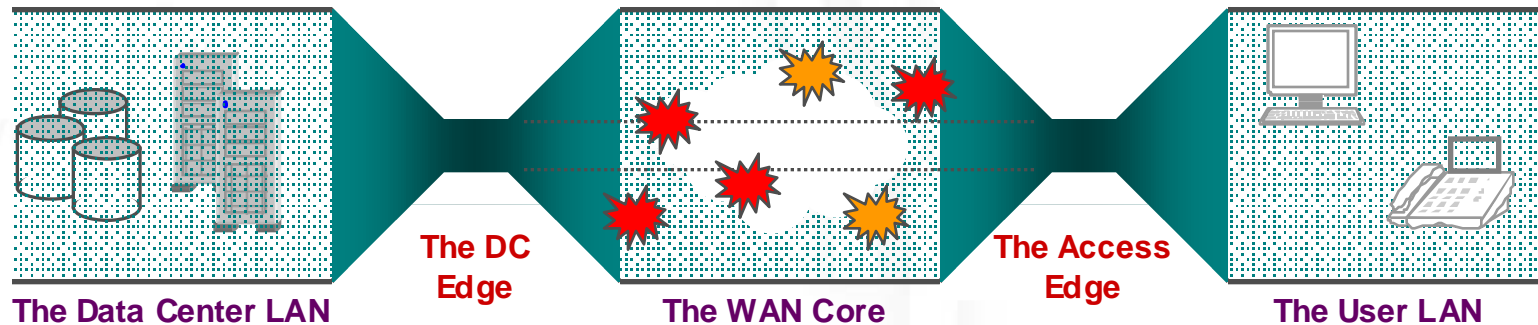


Architect to minimize the “Bow Tie” effect

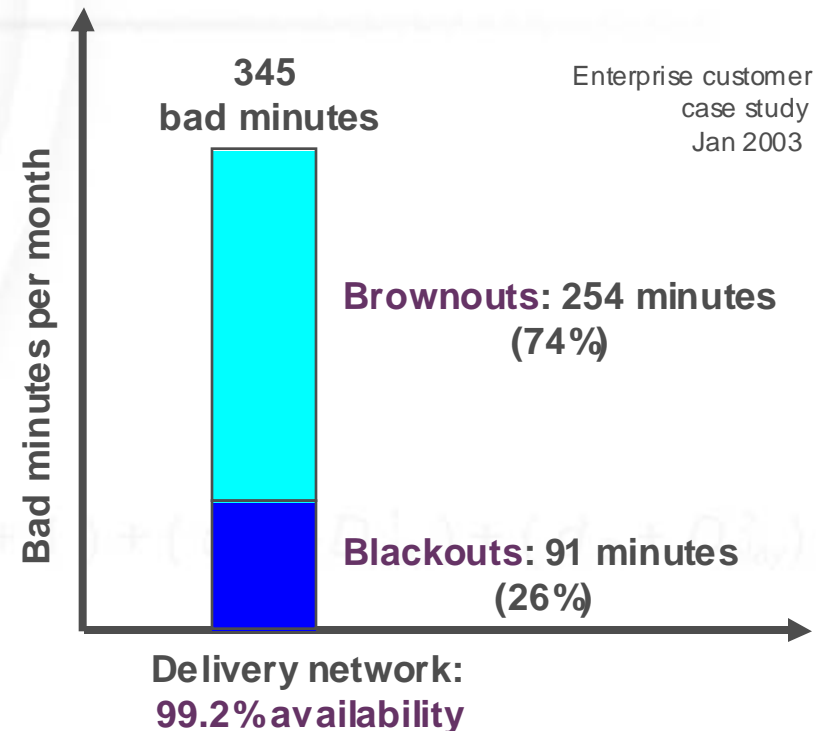


- Use top-of-the-line Cisco / Juniper edge gear
 - With redundancy
- When in doubt, over provision
 - OC12s everywhere w/ bullet-proof SLAs
 - **With redundancy**
- Stuff more stuff into a link
 - Compression
- Meter stuff into the link
 - Traffic shaping, QoS, filters
- Instrument everything for “management”

Still, bad stuff happens in fabric you don't own



- Problems / obstructions / constipation occurs
- Sending packets onto a link that has a pot-hole down the road will result in a bump



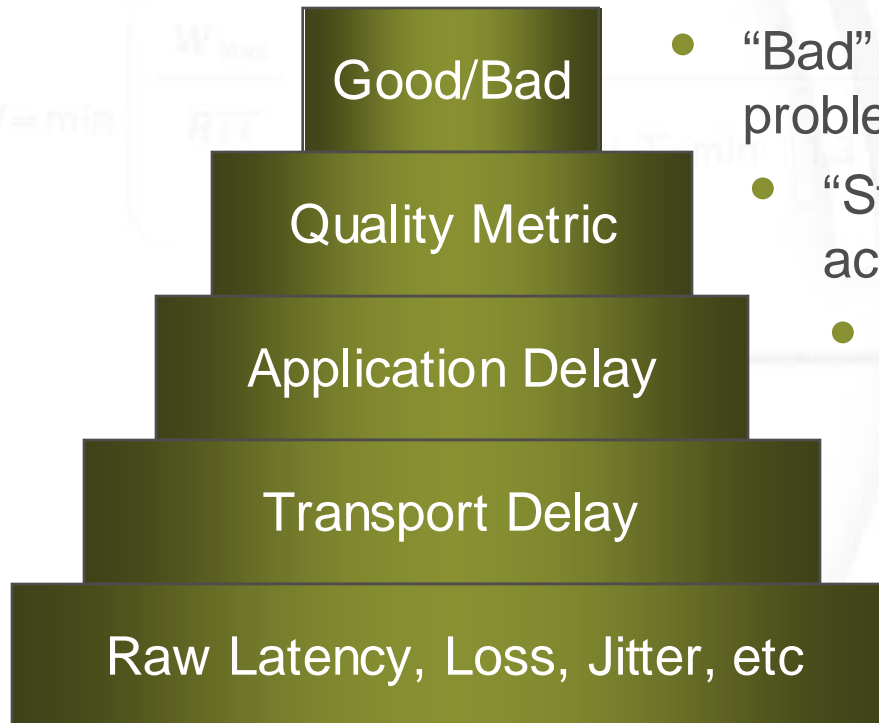
So, how do you plan to avoid these problems

- Like the servers and the disk drives, *virtualize*
 - Multiple servers front-ended with load balancers, job schedulers, FEPs, app accelerators, etc.
 - Racks of disk drives abstracted by RAID controllers and load balancers
 - The key is redundancy with intelligent “oversight”
- In the WAN, of course you need multiple paths
 - A combination of private and public links, architected as appropriate
- Define policy for availability requirements
 - Manage the performance – cost continuum
 - Monitor / assess / adapt / report

Oversight: static vs adaptive policy enforcement

- Static policy is a fixed set of specific rules that are enforced in a “local” environment
 - The “gatekeeper” model
 - No facility for a feedback loop
- Adaptive policy uses overall objectives that abstract away from low level mechanics
 - What is the “best” we can do right now
 - Monitor and evaluate the current and the alternative choices
 - Use “the best”
 - Re-adjust when things change
 - Report on trends; alert on policy violations

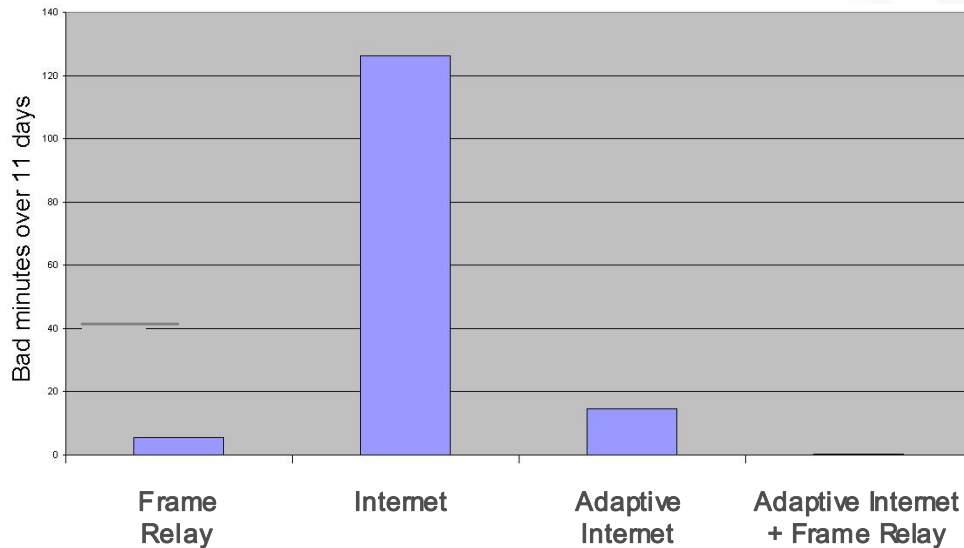
App Models: The key to adaptive networking



- “Bad” means an application quality problem caused by the network
- “Star ratings” comparable across apps
- Delay for a typical app transaction
- Transport layer impact of low level scores
- Low level measures

- Individual network-level metrics do not determine absolute quality
- What is “good” for one application type may be “bad” for another

Impact of adaptive networking on VoIP

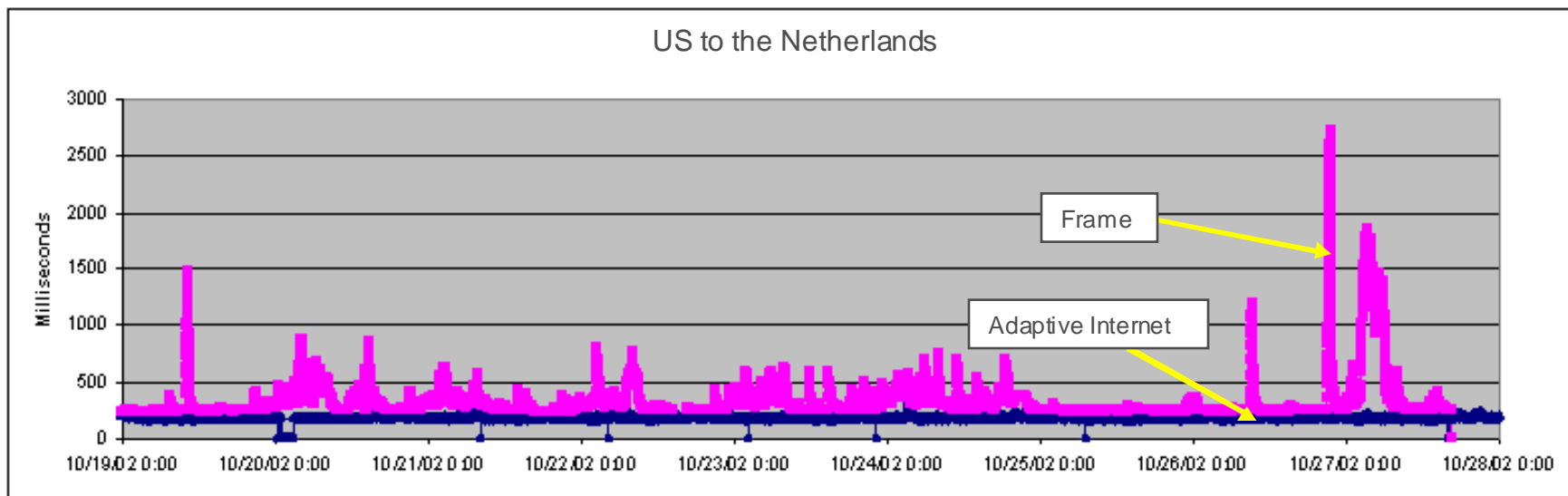


Configuration	Bad Minutes	Reliability (%)
Frame Relay	5.4	99.966
Internet	126.2	99.203
Adaptive Internet	14.7	99.907
Adaptive Internet + FR	0.4	99.997

- Sidestep brownouts on “in-flight” calls
- Deliver a 10-fold to 100-fold increase in availability
- Reduce bad minutes by 99%
- Enable VoIP over VPNs without network upgrades

Creating “Business Class Internet”

- International Frame Relay links are expensive
 - Conventional Internet VPN was not a viable replacement
- Adaptive networking delivered a reliable, “predictable” alternative
- Double the bandwidth at half the cost with equal or better performance



Next step: Integrate all the WAN control points

- Bring the link optimizers under an overall umbrella of adaptive WAN management
 - As traffic is rerouted to different links based on high-level availability / performance / cost considerations, adjust the local policies on-the-fly as appropriate
- Integrate link policy and enforcement w/ overall WAN policy and enforcement

The Grail: edge-to-edge

- In a perfect world, there would be a processor-to-glass “user experience” budget
 - How long should a transaction take
 - What is it worth to the business
- Dynamically renegotiate the budget across domains
 - Servers / storage / WAN / access
 - As load varies
 - As problems occur
- The global IT policy store
 - What is “better” worth right now
 - Global tradeoff of availability / performance / cost



Thank you

$$RTT = 2(T_f + 1) + (d_1 + D_{play}^1) + (d_2 + D_{play}^2)$$

