NETWORK CAPACITY PLANNING

Why is it so different?

Andy Bolton
CEO, Co-founder
CAPACITAS
Agenda

- Introduction – why networks are important
- Questions on Network Capacity Planning
- Network technology primer:
  - LANs
  - WANs
  - Protocols
  - So what is necessary to capacity & performance?
- Network Modelling
- Capacity Planning
- Performance Management
- Typical network issues
- Answers on Network Capacity Planning
- Summary
Introduction
Why networks are important

- Over 1.4 billion telecoms subscribers in 30 OECD countries in 2001 (OECD)
- 5,299 fixed-line communications operators in OECD in 2002 (OECD)
- 65% of UK business have access to the Internet (OFTEL)
- Over 150m web hosts and 600m Internet users (ITU)
- Over 2.6m Broadband subscribers in the UK (OFTEL)
- Annual global telecommunications services revenue is now greater than $1,000,000,000,000 (ITU) whereas the total annual revenue of Microsoft, IBM, Oracle, Dell and HP is only $215Bn
- Virtually no modern IT application or system functions without some form of communications network
Questions on Network Capacity Planning

• What is so different about network capacity & performance and why does it scare system capacity & performance experts?

• How does network and system capacity planning differ?

• How does network and application performance analysis differ?
Network Technology Primer 1

Local Area Networks (LAN):

- **Bus based architectures e.g. Ethernet**
  - Everyone talks to everyone else – blocking architecture
  - Performance issues above 40% utilisation
  - Industry was moving away from bus architecture until recently

- **Ring bases architectures e.g. Token Ring**
  - Only one person can use the “token” – blocking architecture
  - Losing prominence due to ubiquity of Ethernet
  - Ring can be driven at higher utilisation than Ethernet

- **Switched architectures e.g. Ethernet (again!)**
  - A “segment” only talks to the destination “segment” – non-blocking architecture
  - Has become standard for most LANs
Network Technology Primer 2

Wide Area Networks (WAN):
- Access e.g. Metro “Leased Tails”
  - Method of connecting to the Customer Premises Equipment (CPE)
  - Usually expensive as involves civil engineering (digging roads)
  - Potentially being replaced by xDSL at lower speeds
- Aggregation Point e.g. PoP
  - Building or private area where multiple circuits are aggregated
  - Can be a switching facility – Internet Telehouse or Telephone Central Office
- Backbone e.g. Long-Distance Fibre
  - Carries large amount of traffic – typically 144 fibres x 160 wavelengths x 10Gbit (each 10GBit = ~100,000 simultaneous phone calls)
  - Submarine systems more complex due to power problems, although this is changing rapidly as signal reach increases
Network Technology Primer 3

Network Protocols (OSI 7-layer model)

- **Application**
  - Telnet, FTP, SMTP, HTTP
- **Transport**
  - TCP, UDP, SPX
- **Network**
  - IP, IPX
- **Data-link**
  - Ethernet, ATM, PPP, HDLC, FDDI
- **Physical**
  - SDH/SONET, GSM

<table>
<thead>
<tr>
<th>Application</th>
<th>Presentation</th>
<th>Session</th>
<th>Transport</th>
<th>Network</th>
<th>Data-Link</th>
<th>Physical</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© Capacitas 2003
So what understanding of network technology is needed?

- Thankfully only a basic knowledge of underlying technology needed to do a reasonable job of capacity analysis!
- However this isn’t the full story...
- Systems capacity and performance can be predicted with no underlying knowledge of the technology other than high-level view of CPU, Memory and Disk – True or False?
- In most cases for simple planning issues this can often be true...
- ...however an accurate analysis of system performance and capacity planning cannot be undertaken without a deeper knowledge of underlying technology i.e. process scheduling, threading, caching, swapping and paging, RAID etc.
Network Modelling
- A simple network
Network Modelling
- A typical complex network
Network Modelling
- model the logical & ignore the physical
Network Modelling
Routing - the capacity planner’s enemy

- Routing is non-linear
- Routing decision is on a packet-by-packet basis
- In packet-switched networks the routing decision usually made by an algorithm (e.g. Dijkstra)
- In circuit-switched networks the routing decision can be made automatically by a provisioning system, decided by pre-determined rules, or worse-case an ad-hoc decision by a field-engineer
Network Modelling
Routing - the capacity planner’s enemy
Network Modelling
- tools

• Don’t reinvent the wheel – lots of modelling tools and material available

• Acquiring a tool should be based on Reuse, Buy, Build strategy

• Typical carrier-grade tools include:
  – NPAT/IPAT (WANDL)
  – SPGuru (OPNET)
  – VPICServiceMaker (VPI Systems)
Network Performance

- Need to know the performance SLA requirements first
- For basic answers break network down into key queuing areas:
  - Local Area Network (e.g. Switched Ethernet)
  - LAN connection to router (e.g. from Ethernet switch to router)
  - Router (e.g. Cisco)
  - Connectivity to next router (usually negligible delay)
  - Remote Router
  - Remote LAN
- Model each as a separate queue
- Use benchmarks for each component
Network Performance

End-to-end network time

Fastest routing time

Slowest routing time
In most cases networks can be abstracted to a simple M/M/N queue – however complex networks with many routers (or “hops”) may need closer analysis for more accurate answers.
Network Performance

• Competing demands on a network can cause severe contention and hence service delays
• Many modern networks have methods to prioritise and manage these competing demands:
  – Class of Service
  – Quality of Service
  – DiffServ
  – Weighted Fair Share Queuing
  – Multicast
  – Resource reSerVation Protocol (RSVP)
  – MultiProtocol Label Switching (MPLS)
Network Delays

A typical file download process:

1. Spawning a network connection thread
2. Translating a name to IP address using DNS
3. Routing the packet across the network
   1. Storing the packet
   2. Reading the packet header
   3. Determine destination route from a large routing table
4. Awaiting ack (acknowledgment)
5. Establishing a connection with a higher protocol layer
6. Sending data request (e.g. HTTP, FTP, POP3)
7. Receiving data
8. Closing the remote connection
9. “Crashing” the local established connection stack processes and threads

∴ Transmission time can be a small component of the overall end-to-end time.
Capacity Planning

• Usual systems capacity planning methods:
  – No forward planning – only react to threshold exceptions
  – Trending a particular utilisation metric over time to determine run-out dates
  – Demand-driven capacity planning using a suitable model to translate unit demand into resource usage

• All of the above methods are also applicable to network capacity planning
Capacity Planning

Process steps to network capacity planning:-
1. Understand the SLA performance requirements
2. Understand the network infrastructure to an appropriate level
3. Create a performance model to determine how the end-to-end response time varies with component loading
4. Use benchmark or testing data to populate performance model
5. Determine target thresholds for measurable components based on performance model
6. Implement an exception alerting system based on calculated thresholds
7. Create trended capacity models to calculate when individual thresholds will be exceeded if sufficient statistical data is available
8. Determine source of accurate demand forecast data (usually marketing)
9. Create a demand driven capacity model incorporating the performance model, predetermined thresholds and the demand data
10. Verify the model monthly against actual demand data and recalibrate
11. Base forecasts and budgets on the demand driven model output
12. Arrange upgrades based on the trended capacity model output
13. Trigger upgrades on the threshold alerting system
Typical Network Issues

- Not able to get the router utilisation info as router owned by outsourcer
- Network performance is small part of end-to-end response time
- Network appears to large and complex to measure, or is invisible as owned by outsourcer
- Service Level Agreement only specifies a “network ping” to measure network performance
- Increased need for network cost reduction has caused “convergence” – the use of one network for many different services such as application, web, messaging, voice and video conferencing
- Use of Virtual Private Networks causes:
  - Sharing of networks with external traffic and hence contention
  - Major encryption technologies (e.g. Triple-DES) required for both ends of link causing significant performance issues
- Little understanding of traffic flows or application workload profiles
Answers on Network Capacity Planning

• What is so different about network capacity & performance and why does it scare system capacity & performance experts?
  - It isn’t that different – the same concepts are used but with different technology and acronyms
  - Therefore it shouldn’t scare any system capacity planner or performance analyst

• How do network and system capacity planning differ?
  - Technology and some areas of potential uncertainty such as routing, but even these are well defined by algorithms

• How do network and application performance analysis differ?
  - Terminology and Acronyms
  - Um....
  - ......that’s it!
Summary

• Capacitas believes that Network and IT capacity planning are entwined in the modern distributed systems paradigm and this relationship will integrate further

• System Capacity Planners have the skills and tools at their fingertips already to undertake network capacity planning – they just need a basic understanding of the technology

• Application Performance Analysts have the skills and tools at their fingertips already to undertake network performance analysis – they just need a basic understanding of the technology

• Don’t be afraid – get on with it!
Bibliography

Books:
- Telecommunications Network Modelling, Planning and Design, 2003, Edited by Sharon Evans, IEE (BT Exact Series)
- Capacity Planning for Web Services, 2003, Daniel Menascé and Virgilio Almeida, Prentice Hall

Whitepapers:
- Automated Threshold Alerting using Statistical Filtering, 2002, Daniel Quilton, Capacitas (www.capacitas.co.uk)
- The Challenge of Planning Multiple Services over a Single Network Infrastructure, 2003, Dr. Manzoor Mohammed, Capacitas
- How to Outsource your Network without losing control, 2002, Andy Bolton, Capacitas

© Capacitas 2003