IPCOMM2006
September 25-27 • Gaylord Opryland • Nashville, TN

Leap the Wizard Gap: Keys to Network Optimization

Loki Jörgenson
Chief Scientist
Apparent Networks

Tuesday, Sept 26 | 10:30am – 11:30am

What Will Be Covered

The "Wizard Gap":

The distance between typical network performance and what you should be able to attain.

If you have heard even a few IP telephony implementation horror stories, you know by now that a typical network isn't going to cut it. Identify the most common sources of degradation, learn how to couple user QoE to network performance, and discover a better way to troubleshoot and maintain the kind of reliable network you need to handle IP.

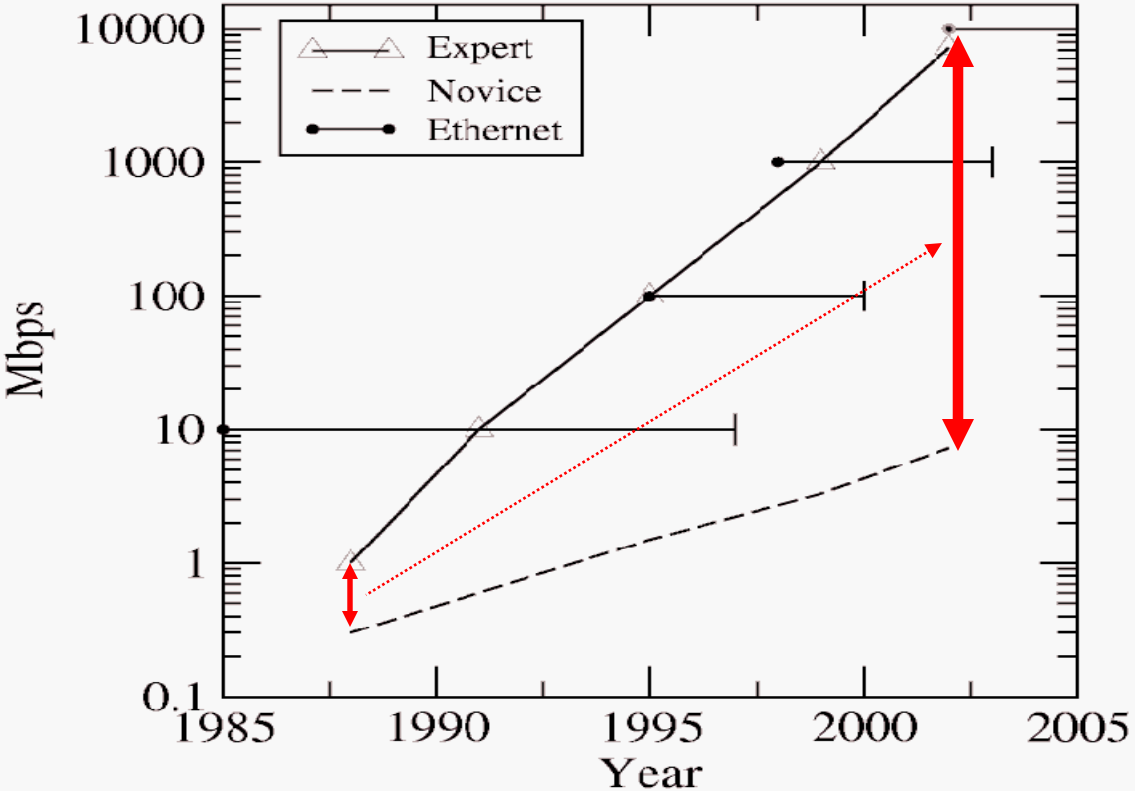
What Will Be Covered

- What is the "Wizard Gap?”
- Identifying the sources of degradation
- 3D definition of network performance
- Opportunities for optimization

Outline

- **Defining "Wizard Gap"**
  - **TCP buffer tuning**
  - **NICs/drivers**
  - **Duplex conflicts**
- **Three dimensions of visibility**
- **Real-time Autognostics**
- **Strategies**
  - **Develop end-to-end, continuous monitoring of app performance infrastructure**
  - **Diagnostics → Autognostics**
  - **Device and configuration management**

Wizard Gap



Reprinted with permission (Matt Mathis, PSC)
http://www.psc.edu/~mathis/

Wizard Gap

Working definition:

Ratio of effective network performance attained by an average user to that attainable by a network *wizard*….

TCP Steady State Model

Simple TCP throughput model

$$Throughput = \frac{MSS}{RTT} * \frac{0.7}{\sqrt{loss}}$$

MSS – Maximum Segment Size

RTT – Round Trip Time

Loss – rate of packet loss

Mathis, Semke, Mahdavi, Ott, "The Macroscopic Behavior of the TCP. Congestion Avoidance Algorithm" **CCR July** 1996

Wizard Gap

Primary performance inhibitors :

- TCP buffer tuning
- NICs/driver
- Duplex conflict

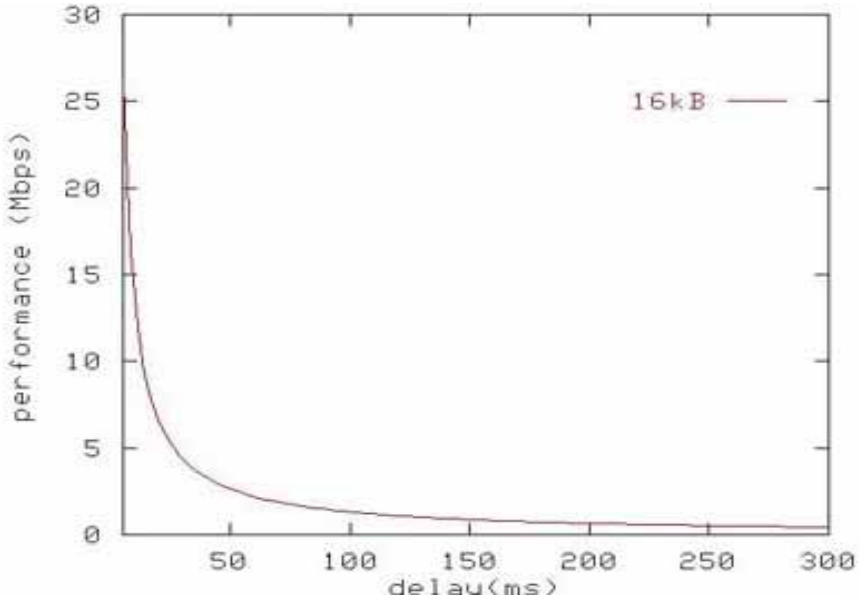
TCP Buffer Tuning

- Congestion window (cwnd) defined by
  - Slow start
  - Congestion avoidance behaviors
- Typically TCP suffers under high bandwidth and/or large latency
- BW-delay product
  - End-to-end transfer rate
  - End-to-end latency (RTT/2)

TCP Buffer Tuning

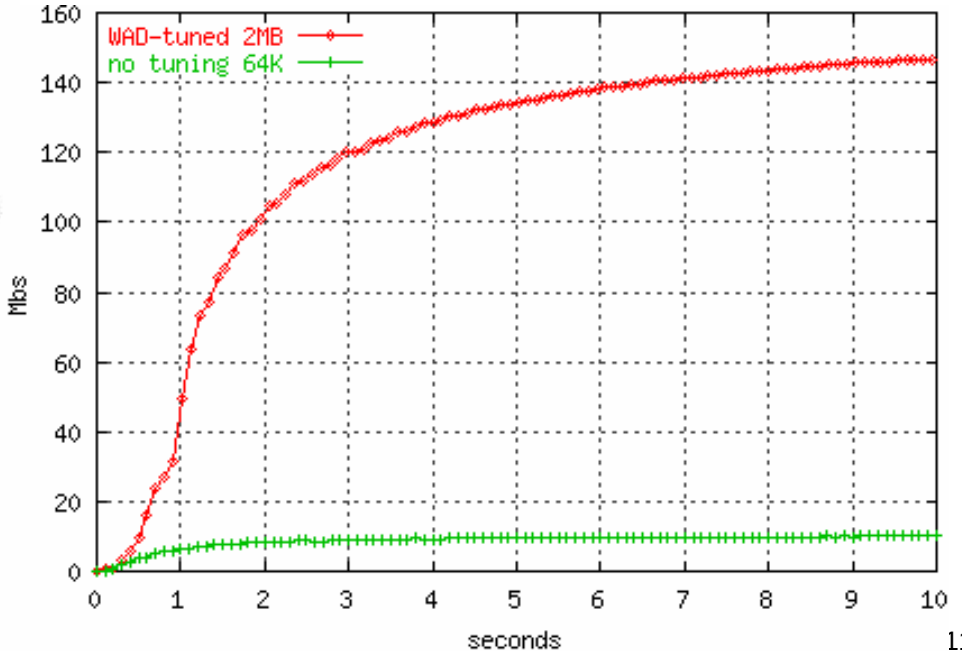
- **Socket buffer sizes act as limiter**
  - **Require ~2 x BW x delay**
- **Ex. 100 Mbps on 100 msec**
  → ~2 Mbytes
- **Typical sizes ~ 10 kbytes**
  → 400 kbps!!

TCP Buffer Tuning



PSC 1997

ORNL 2003



TCP Buffer Tuning

- **Linux 2.6.17 has Tx/Rx auto-tuning and *default* 4 Mbyte maximum window size**
  - **100 Mb/s on a 300 ms path**
  - **1 Gb/s on a 30 ms**
  - **assuming extremely loss-less network**
- **MS touts 64 Kbyte window for Vista**

TCP Tuning Article

http://www.onlamp.com/pub/a/onlamp/2005/11/17/tcp_tuning.html

NICs/Drivers

- Rated at 10/100/1000 Mbps
- Not able to put packets back-to-back at line speed
- Peak (2-way) bandwidth limited

NICs/Drivers

- **Inefficiencies in down-level drivers**
- **Undetectable without measurement**
- **Is it the network, OS, or NIC?**

NICs/Drivers(cont.)

Standard	Standard Link Speed	Framing Overhead	Theoretical (Calculated) Bandwidth	Optimal (Realistic) Bandwidth
DS0 or ISDN	64Kbps	3.9%	0.123Mbps	(same)
ISDN dual channel	128Kbps	3.9%	0.246Mbps	(same)
T1 (HDLC+ATM)	1.544Mbps	11.6%	2.73Mbps	2.65-2.75Mbps
T1 (HDLC)	1.544Mbps	3.5%	2.98Mbps	2.80-2.98Mbps
E1	2.0Mbps	3.5%	3.86Mbps	3.72-3.90Mbps
T3	45Mbps	3.5%	86.85Mbps	85.0-86.9Mbps
10M Ethernet HDX	10Mbps	2.5%	9.75Mbps	9.6-9.8Mbps
10M Ethernet FDX	10Mbps	2.5%	19.5Mbps	19.4-19.6Mbps
100M Ethernet HDX	100Mbps	2.5%	97.5Mbps	97-98Mbps
100M Ethernet FDX	100Mbps	2.5%	195Mbps	180-195Mbps
GigE Ethernet	1Gbps	2.5%	1.95Gbps	0.4-1.8Gbps

100 Mbps NICs

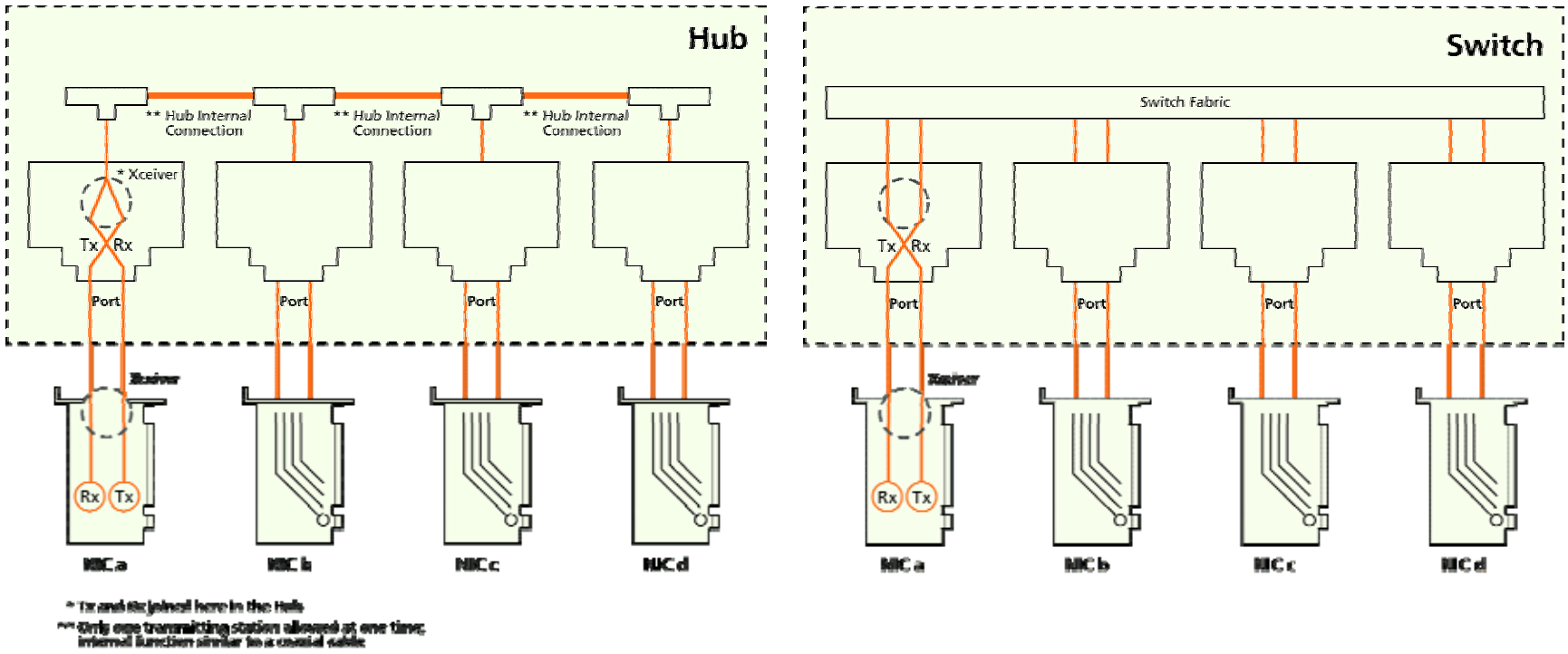
NIC	driver version	OS	2-way BW
3Com 10/100 FE575C	MS 5.5.0.0	Win2K	182
Intel PRO/100 VE	Intel 5.41.27.0	Win2K	182
Intel Pro/100 + MiniPCI, 10/100 NIC	Intel 6.1.3.0	Win2K	185
3Com 3905B or C 10/100 32 bit PCI	MS 5.0.2170.1	Win2K	185
Intel 82559 10/100	Intel 5.41.27.0	Win2K	181
Intel 82559 10/100	MS 4.1.67.0	Win2K	96
3Com 3CN3ACx5566 10/100	3Com 1.10.14.0	Win2K	178
3Com 3C930 10/100	3Com 4.8.0.0	Win2K	176
Linksys PCMPC200, PCMCIA	MS 5.5.0.0	Win2K	151
Xircom CBE2-100, PCMCIA, 32 bit	MS 2.58.2.2	Win2K	117
Xircom CardBus Ethernet 10/100	Xircom 3.12	Win2K	110
3Com 3CCSH572BT PCMCIA, 16 bit	MS 2.0.3.4000	Win2K	12
3Com/USR Robotics 3CCFE574BT	MS 2.0.3.4000	Win2K	12

Duplex Conflict

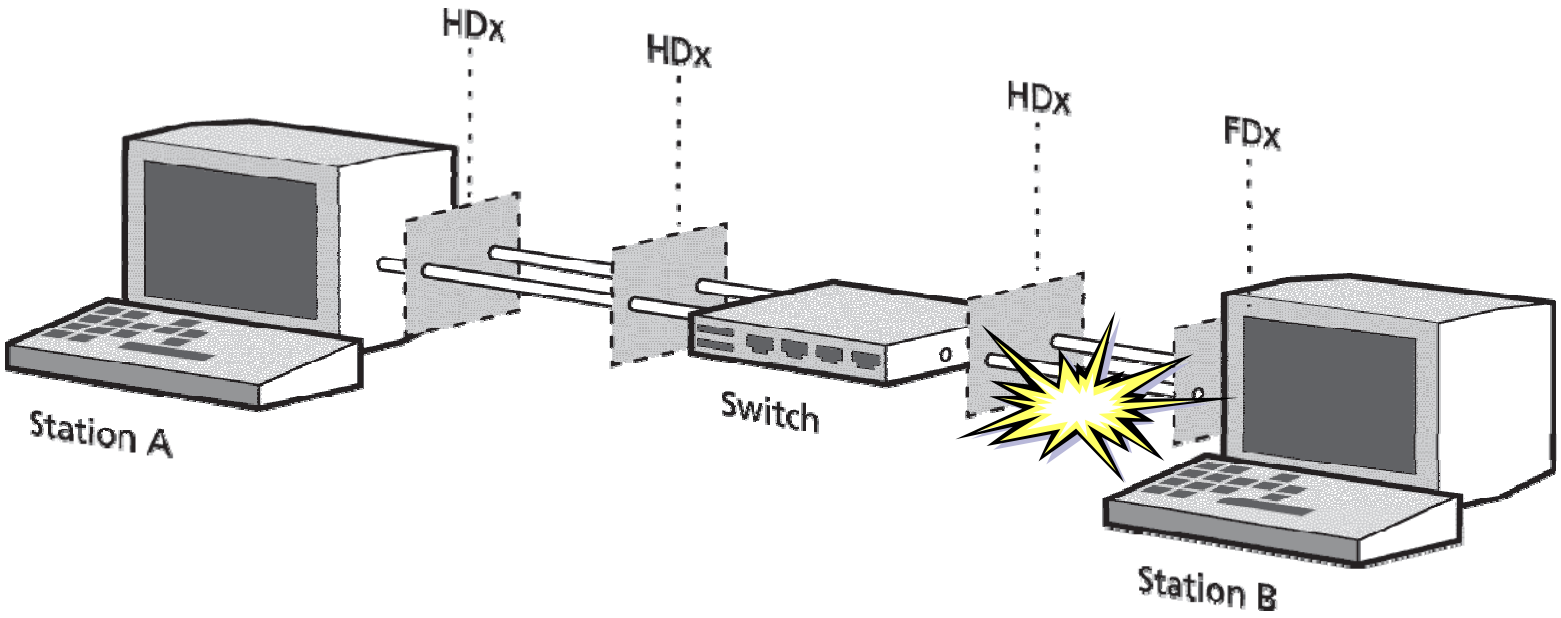
- Two connected Ethernet interfaces use different duplex modes
  - Half duplex (CSMA/CD) for shared media
  - Full duplex for separate Tx/Rx media
- Auto-negotiation fails or manual used
- Connection succeeds, Ping works
- Collisions during data transfers

  → excess of 60% packet loss!!

Hub vs. Switch



Conflict on Switch



So, what is ‘auto-negotiation’**???**

Other Wizard Gaps

throughput → Data apps

jitter → Real-time apps

RTT → Transactional apps

Loss → all

Sources of Degradation

- Capacity bottlenecks
- Congestion-like behaviors
- Serialization and propagation

  ……
- Loss
- Reordering

Sources of Degradation

- Many sources
- End-host, mid-path, media, edge
- Layer 1, 2 and 3
- Configuration, malfunction, dysfunction, emergent

Cost of Performance

Annual Spend

- **on Networks = $1.3 Trillion**
- **on Network/Systems Mgmt = $9 Billion**

…. and yet

Cost of Performance

- **82%** of the time, IT discovers network problems by end users complaining about application performance
  (Network World)

- **38%** of 20,000 application support tests showed serious application impacting network issues
  (Apparent Networks)

- **78%** of network problems are “beyond our control”
  (TELUS)

Cost of Performance

- **50%** of network alerts are false positives
  (Netuitive)
- **85%** of existing networks are not ready for VOIP
  (Gartner 2004)
- **60%** of IT problems due to human error
  (Networking/CompTIA 2006)
- **18:1** cost of IT staff to capital expenditures
  (ITWorldCanada 2006)

Performance Degradation

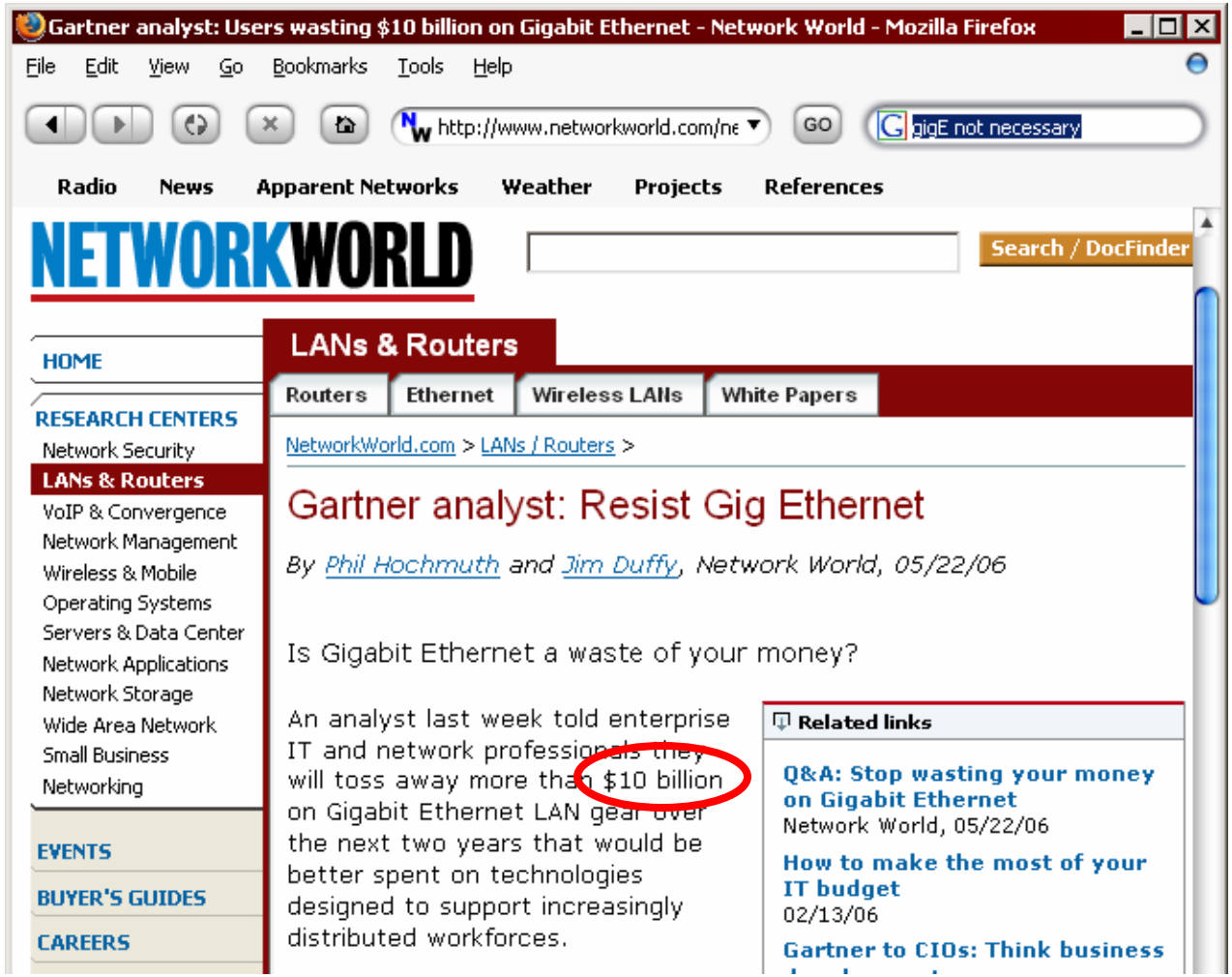
- Based on a survey of 20,000 customer tests with application performance issues, network identified 38% of cases:
  - 20% of networks have bad NIC card drivers
  - 29% of devices have packet loss, caused by:
    - **50% high utilization**
    - **20% duplex conflicts**
    - **11% rate limiting behaviors**
    - **8% media errors**
    - **8% firewall issues**

Throwing Bandwidth

When was it a mistake to throw anything at the problem?

→ When you don't know what the problem is……. bandwidth can easily be the right (or wrong) answer.

Desktop GigE Denied



Bandwidth vs. QoS

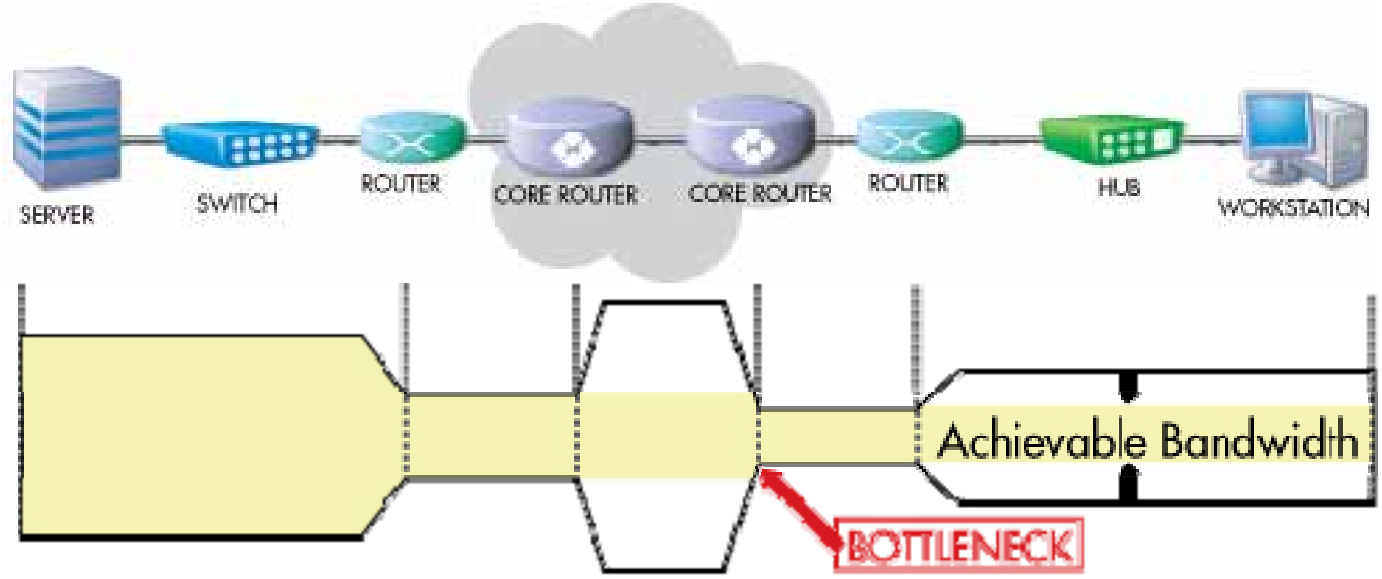
Outline

- **Defining "Wizard Gap"**
  - **TCP buffer tuning**
  - **NICs/drivers**
  - **Duplex conflicts**
- **Three dimensions of visibility**
- **Real-time Autognostics**
- **Strategies**
  - **Develop end-to-end, continuous monitoring of app performance infrastructure**
  - **Diagnostics → Autognostics**
  - **Device and configuration management**

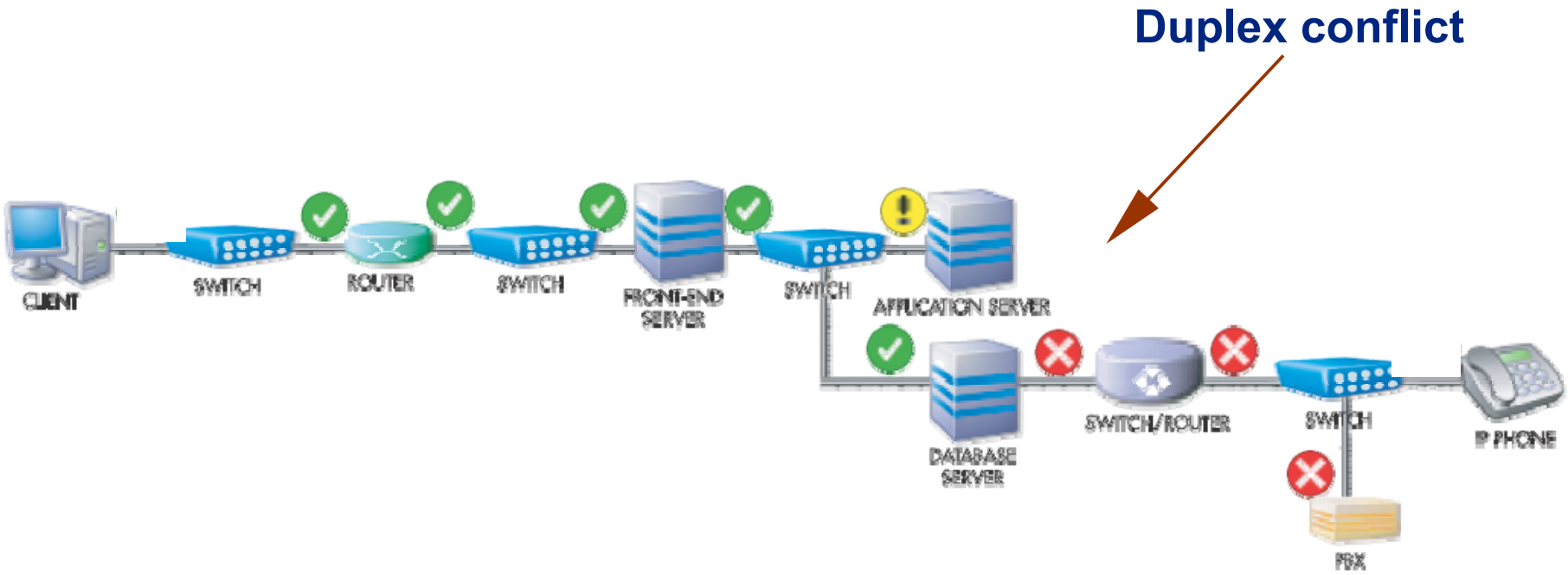
End-to-end Visibility

Network path characterization

- How big?
- How long?
- How much traffic

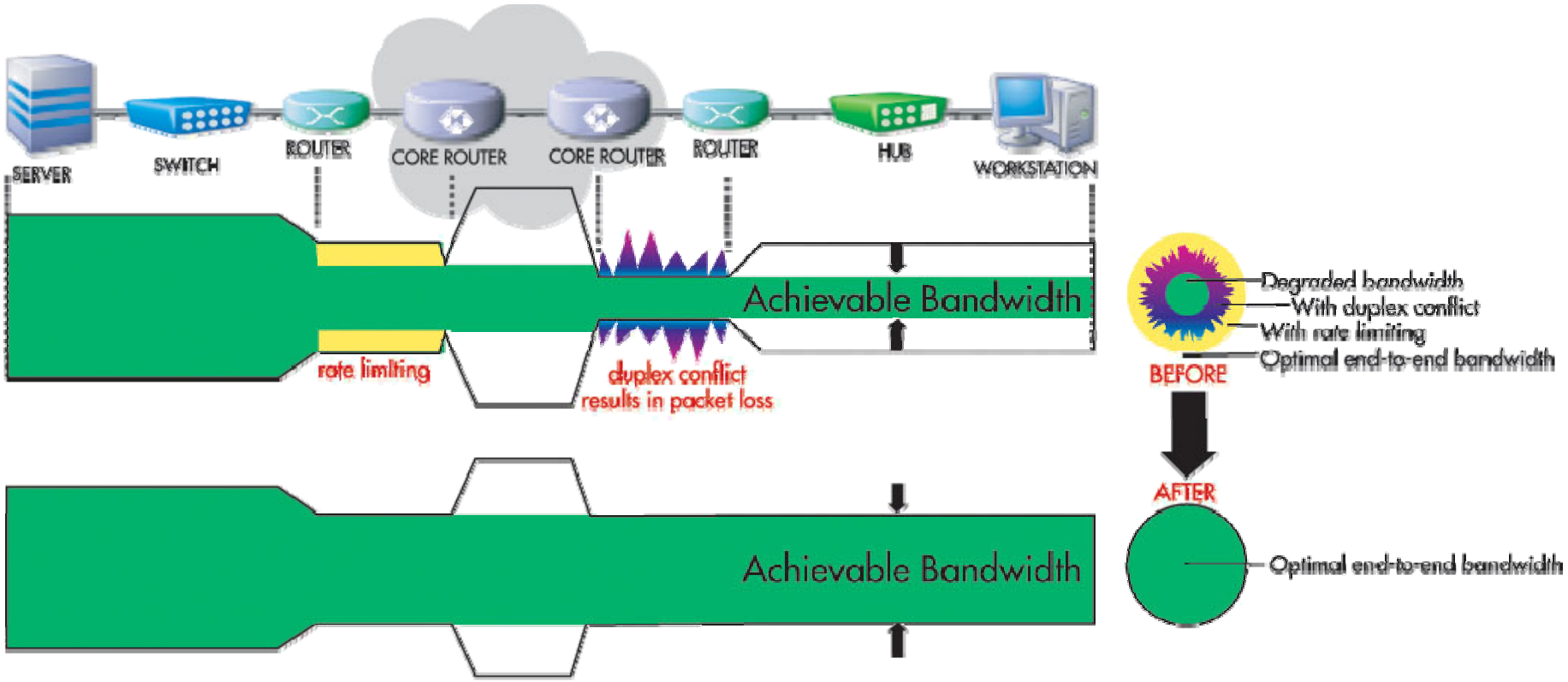


Pin-point Diagnostics



- **Where in the path is the degradation?**
- **What is causing it?**
- **What is needed to remediate it?**

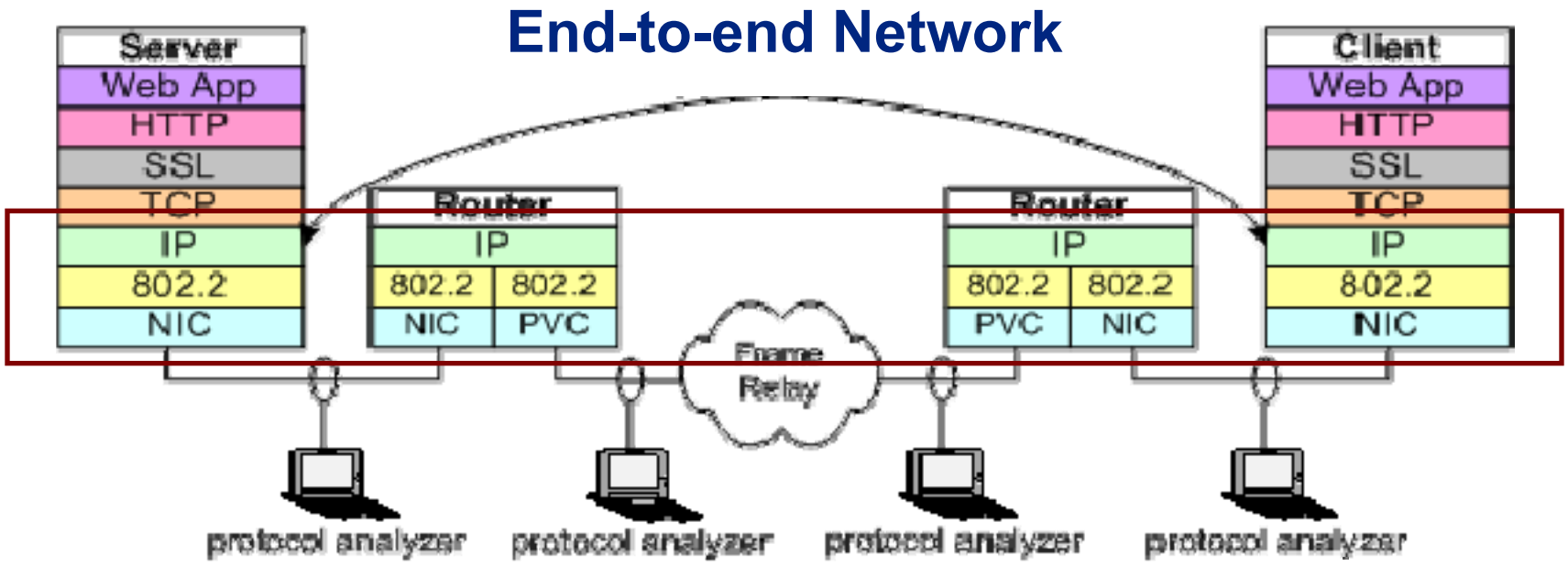
Iterating to Performance



Network Visibility

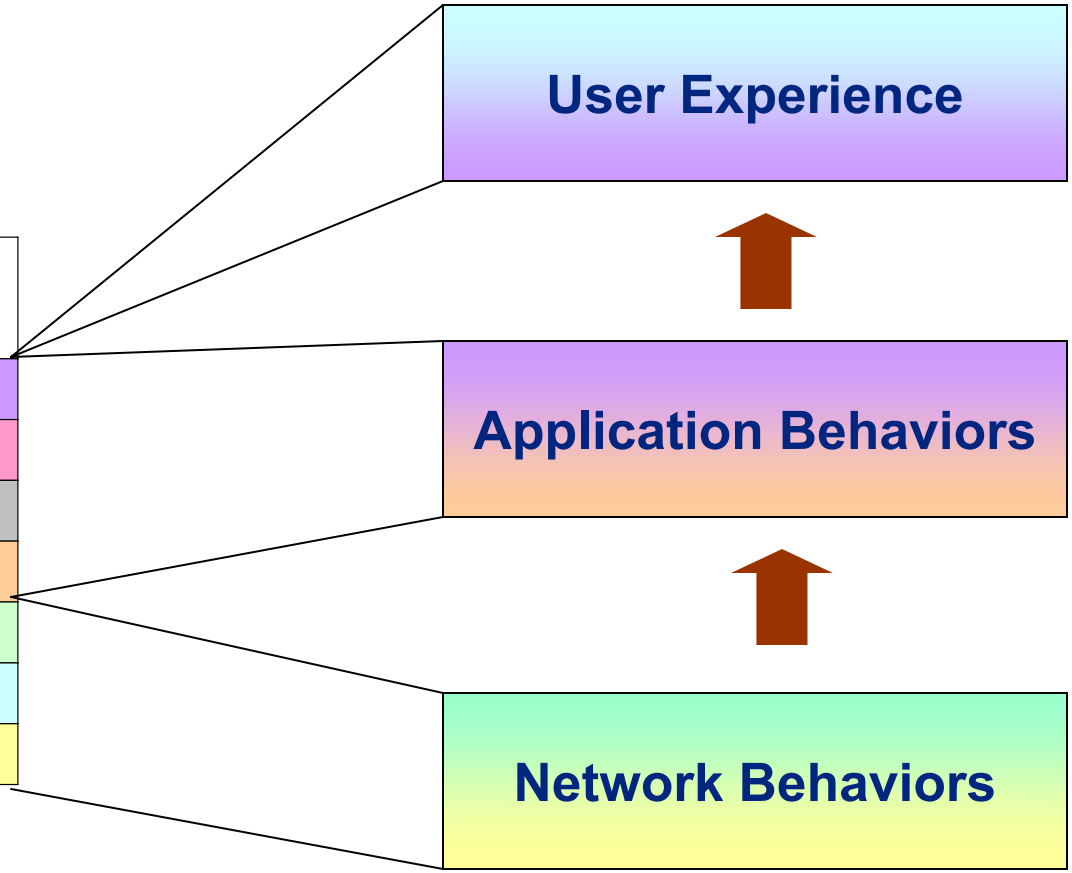
- Need to be able to “see” performance
- End-to-end
- Application-specific
- Dynamically
- Under field conditions

End-to-End Layers

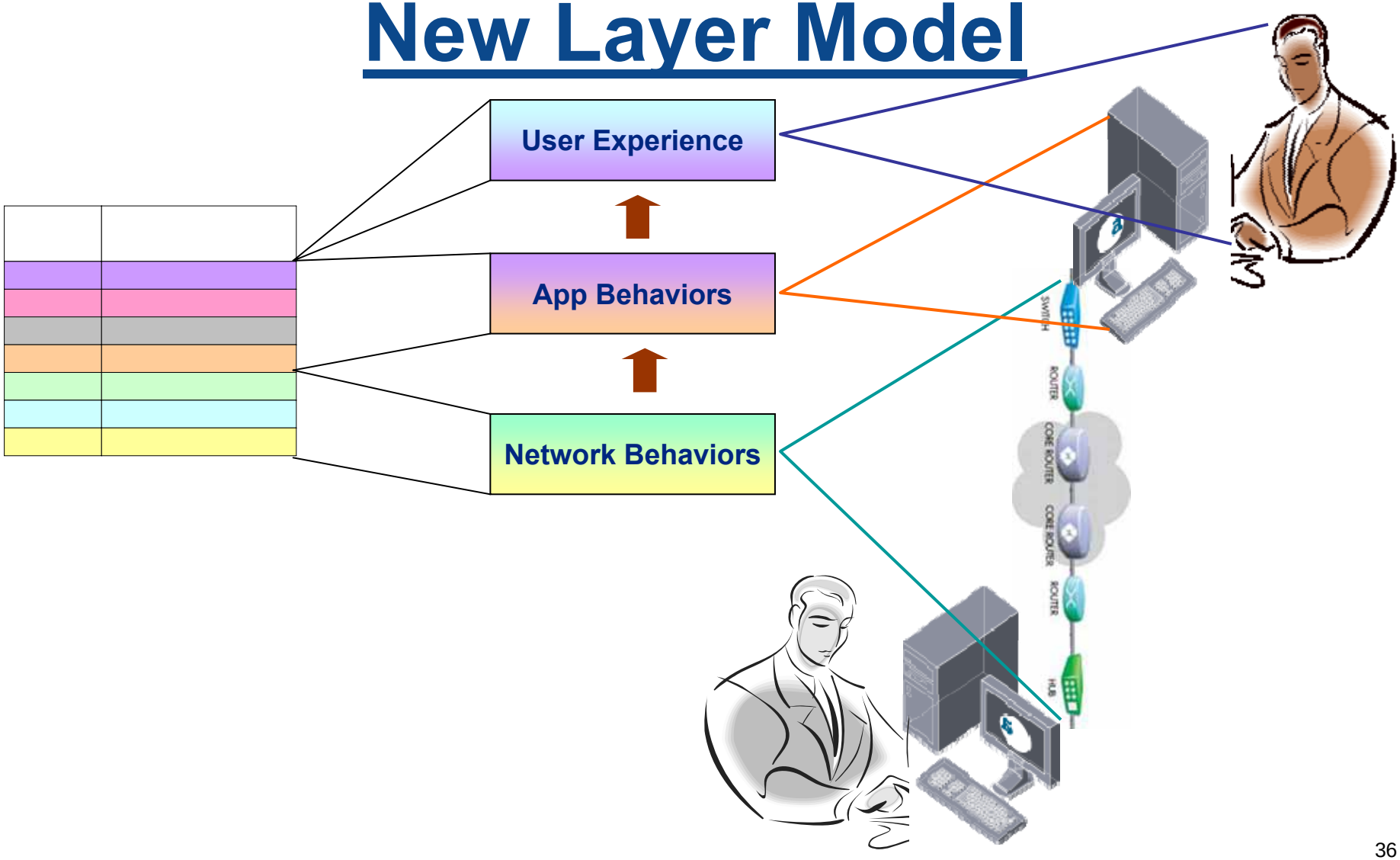


New Layer Model

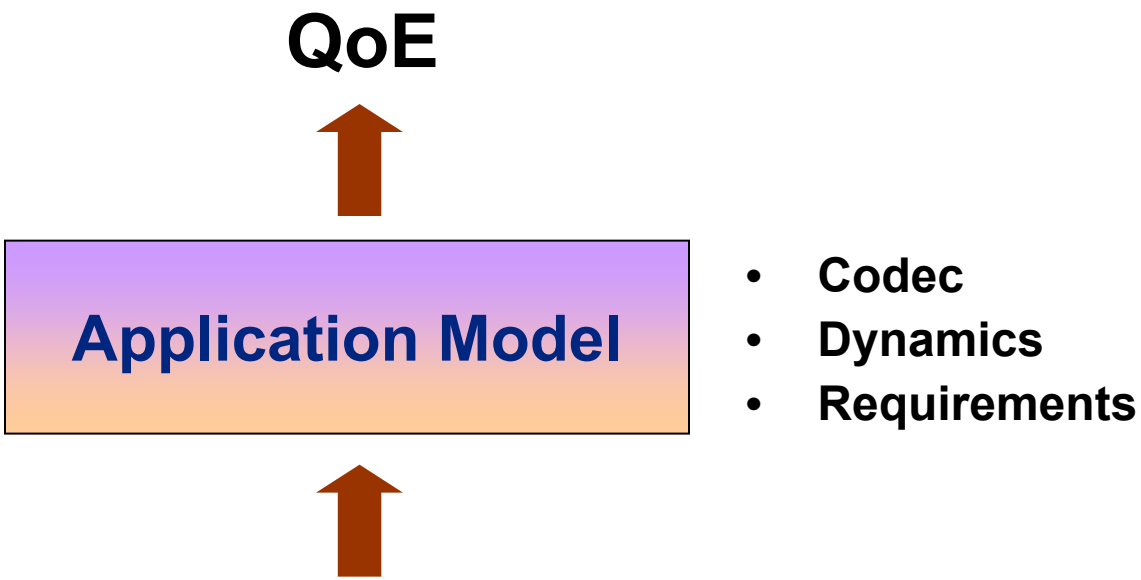
OSI Layer	Description
7	Application
6	Presentation
5	Session
4	Transport
3	Network
2	Data Link
1	Physical



New Layer Model



App-to-Net Coupling



Application Ecology

- **Paraphrasing ITU categories**
  - **Real-time**
    - Jitter sensitive
  - **Synchronous/transactional**
    - Response time (RTT) sensitive
  - **Data**
    - Bandwidth sensitive
  - **Best-effort**
    - Not sensitive

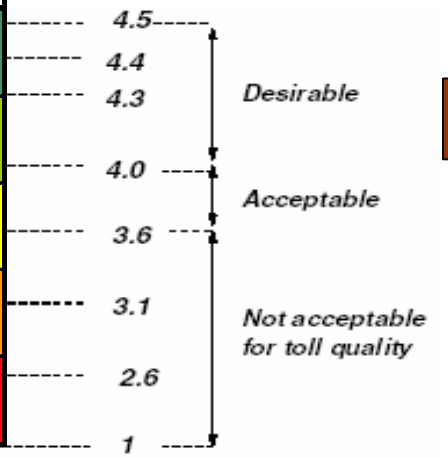
E-Model Mapping: R → MOS

E-model generated "R-value" (0-100)

- maps to well-known MOS score

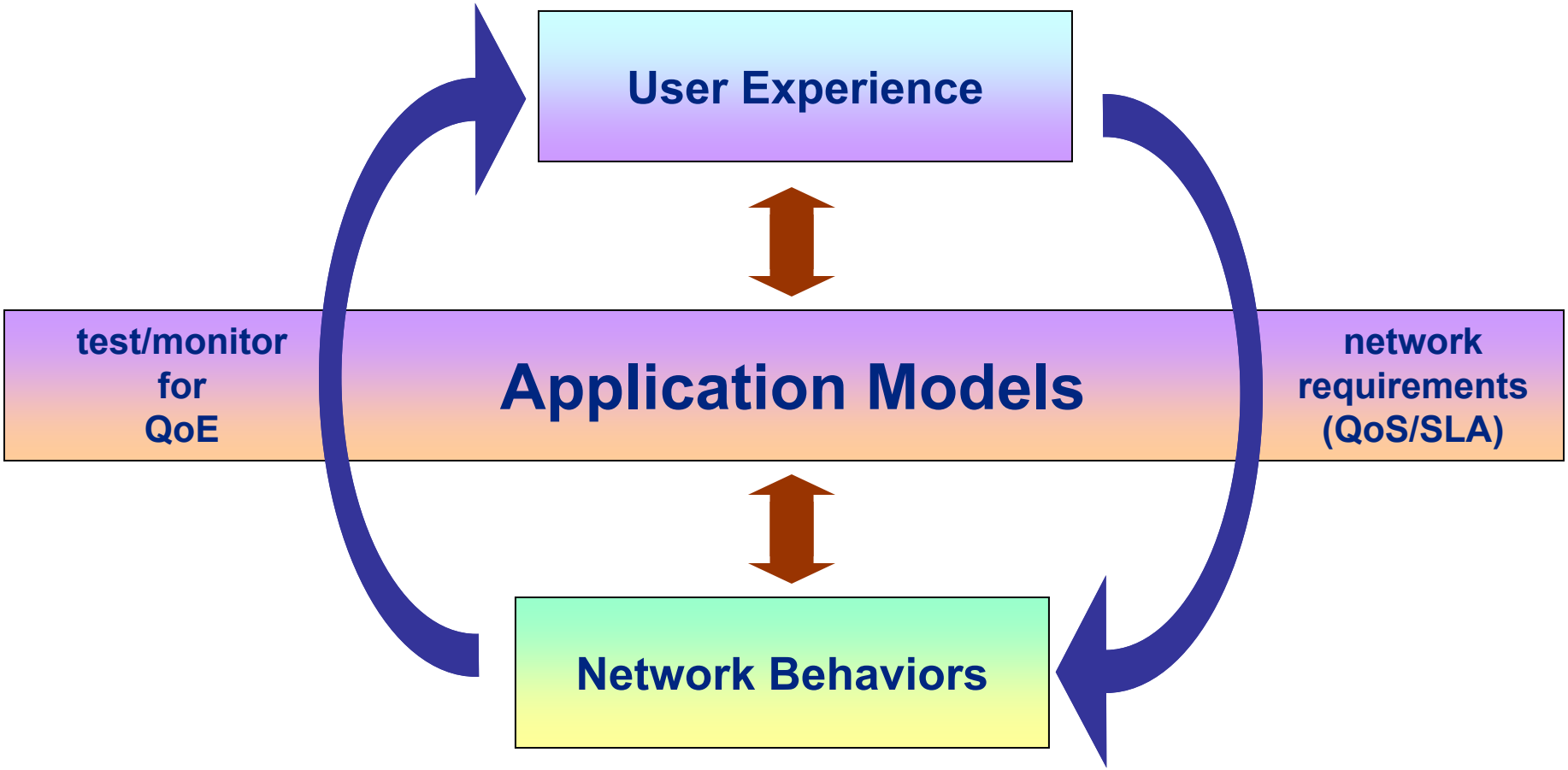
E-Model Analysis

R -value range	speech transmission quality category
90 - 100	best
80 - 90	high
70 - 80	medium
60 - 70	low
0 - 60 *	(very) poor

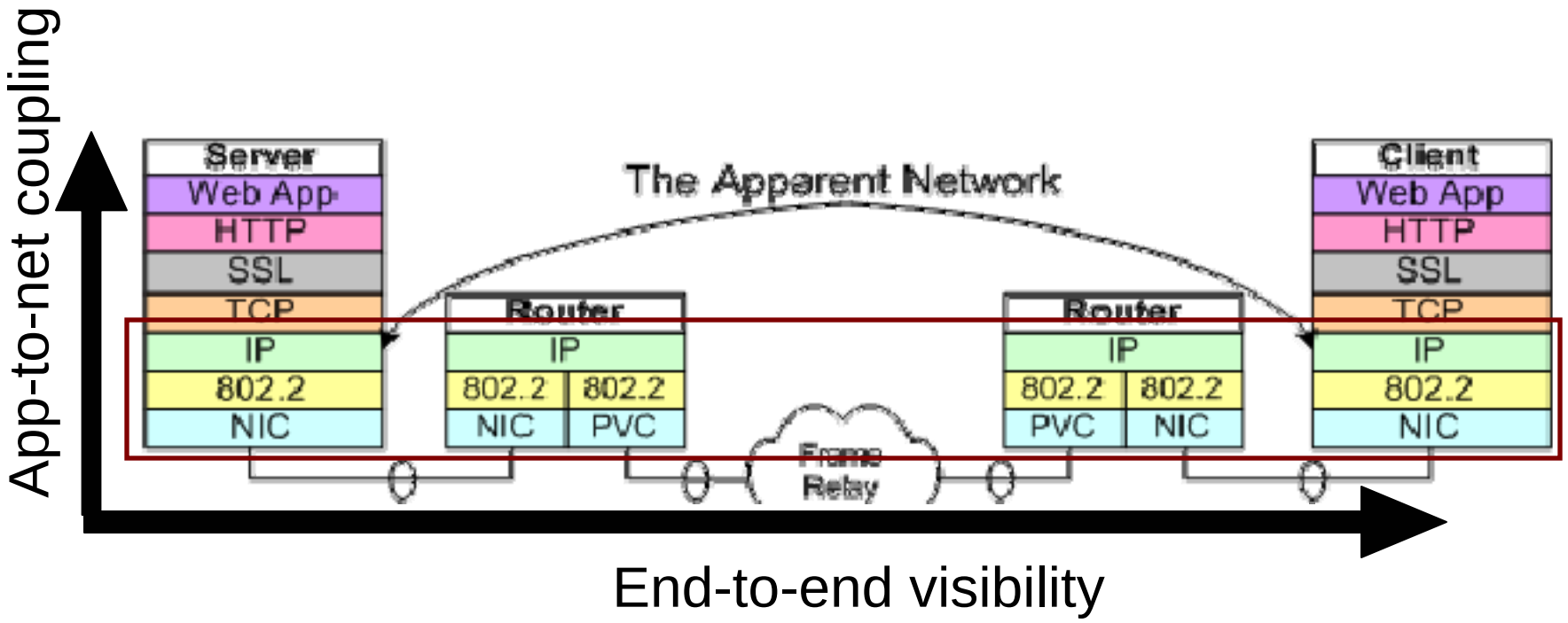


MOS (QoE)

Coupling the Layers



Network visibility



Outline

- **Defining "Wizard Gap"**
  - **TCP buffer tuning**
  - **NICs/drivers**
  - **Duplex conflicts**
- **Three dimensions of visibility**
- **Real-time Autognostics**
- **Strategies**
  - **Develop end-to-end, continuous monitoring of app performance infrastructure**
  - **Diagnostics → Autognostics**
  - **Device and configuration management**

Autonomic Computing

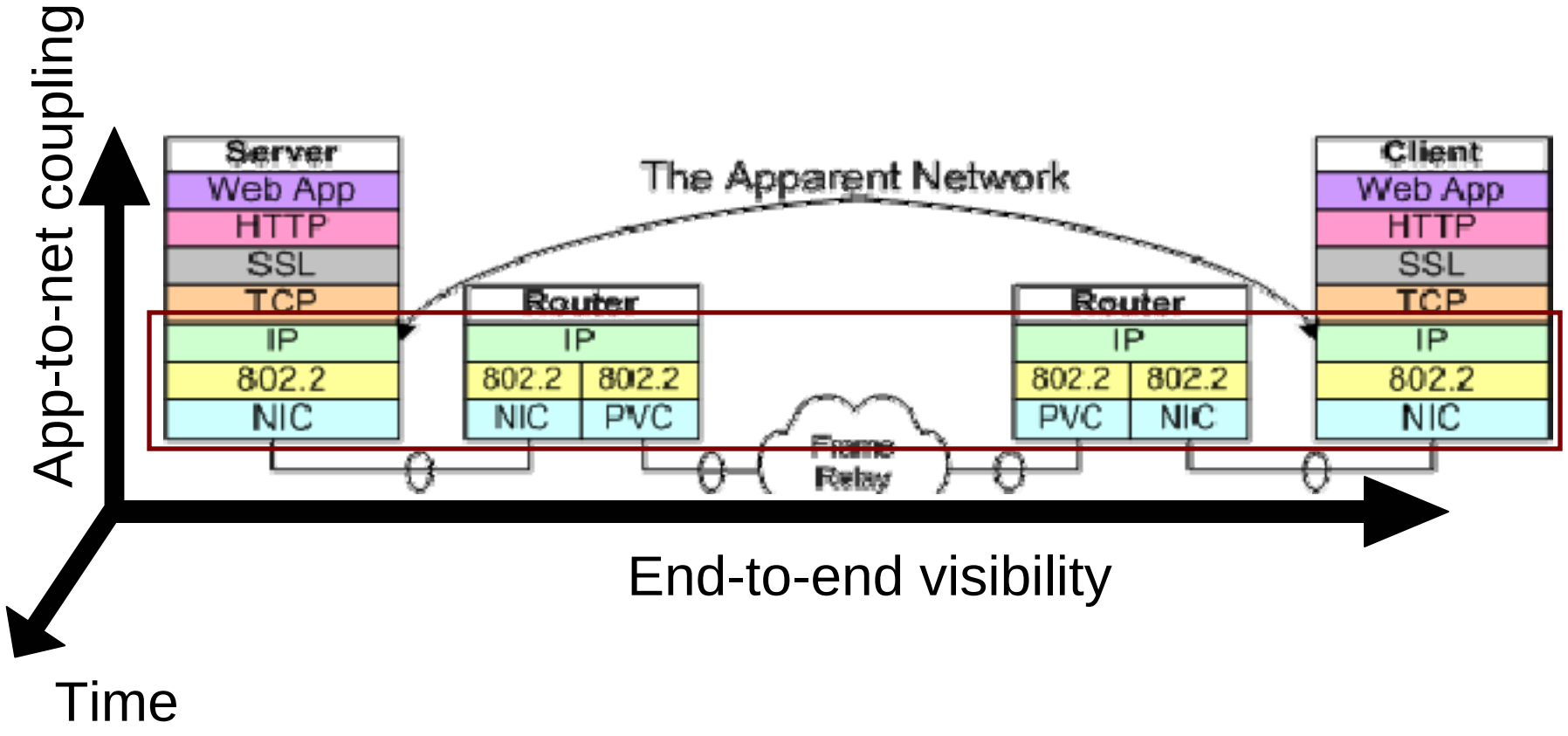
- IBM, HP, Microsoft, Juniper, Cisco
- 18:1 cost ratio for IT staff to capital
- IT skills pool diminishing

Autonomic Networks

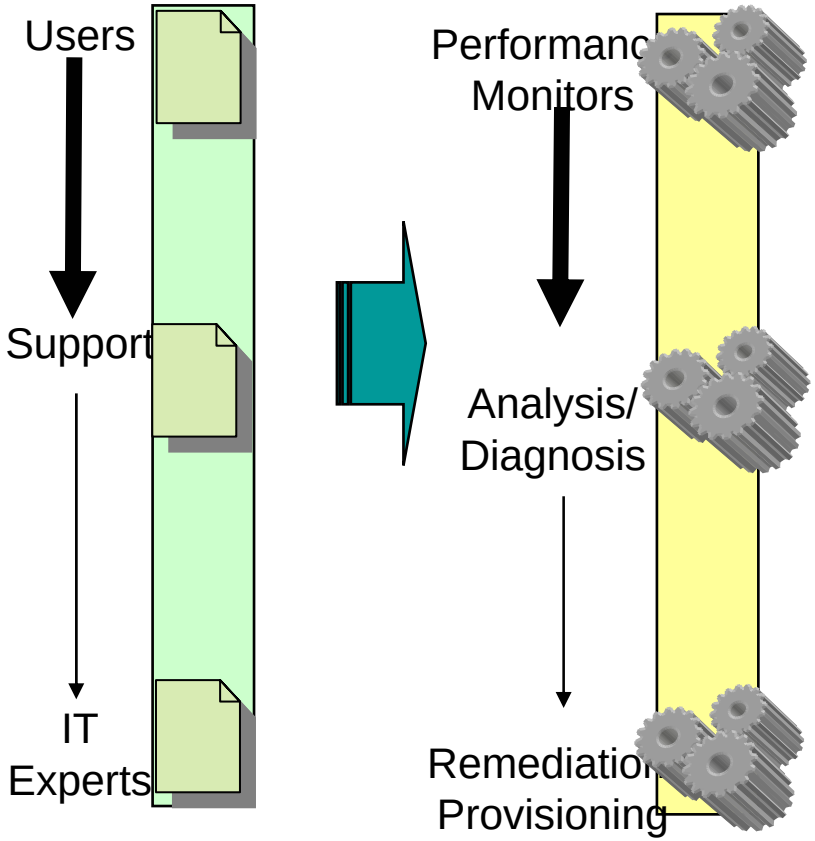
What will it take?

- Solutions must be scalable
- Humans removed from critical steps
- Must be robust and reliable
- Must be real-time
- Non-intrusive and pervasive
- Adaptive and application specific
- Continuously aware

Network visibility ….. in 3D!!



Making the Case for Automation



– Users **acting as "performance monitors"**

who are calling

– Support staff **acting as "expert systems"**

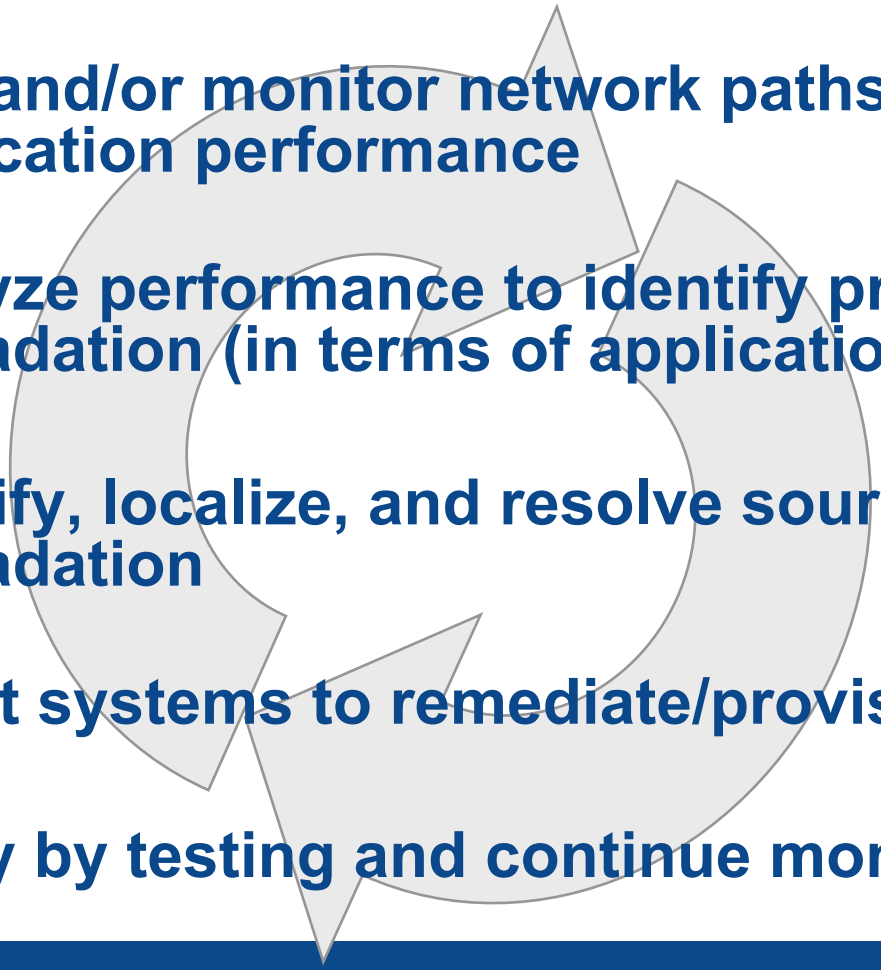
who are alerting

– IT experts **acting as "control systems"**

who are remediating/ provisioning

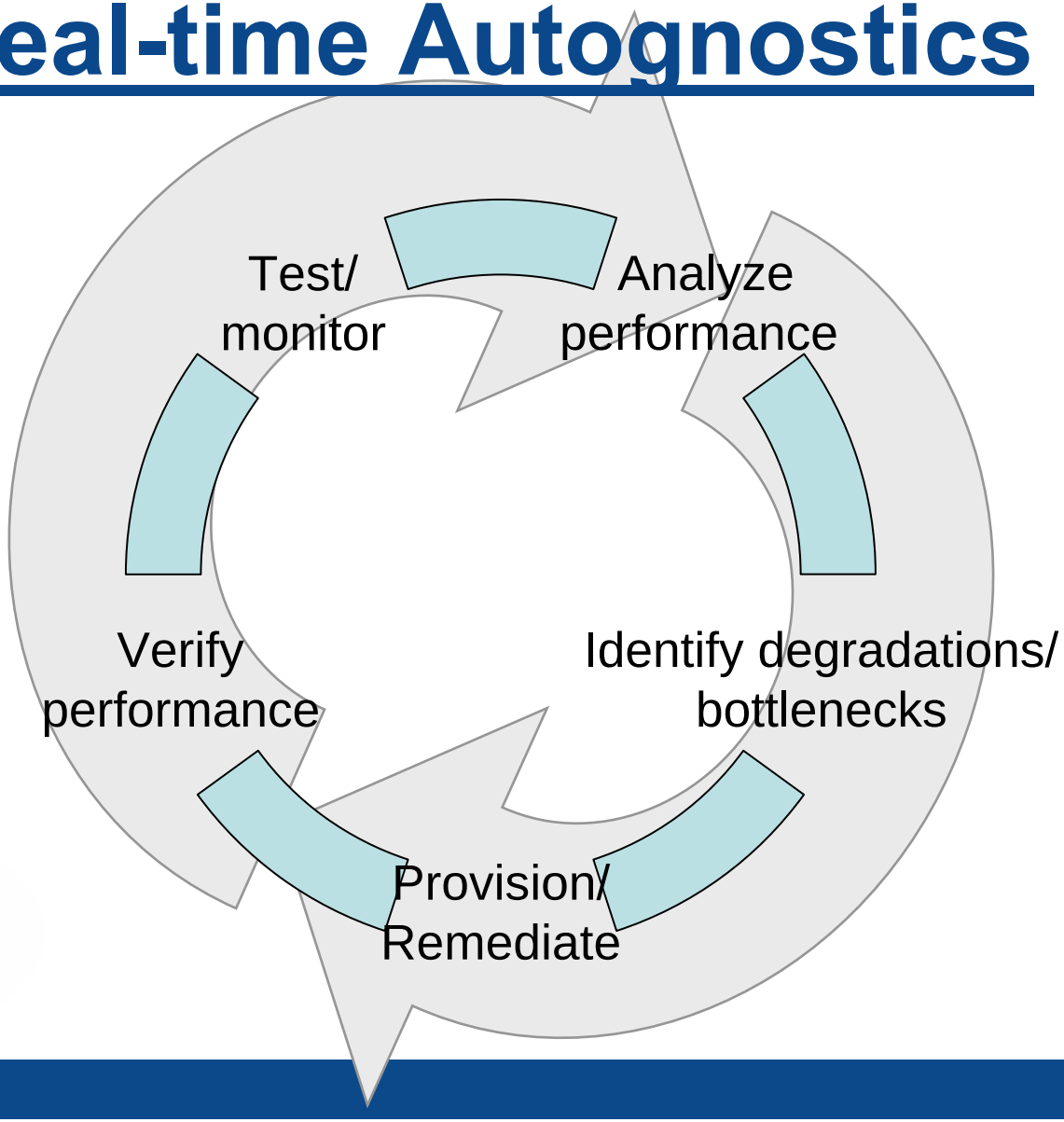
Autognostic Performance Cycle

Necessary steps:



- ✓ Test and/or monitor network paths for application performance
- ✓ Analyze performance to identify presence of degradation (in terms of application)
- ✓ Identify, localize, and resolve sources of degradation
- ? Direct systems to remediate/provision
- ✓ Verify by testing and continue monitoring

Real-time Autognostics



Outline

- **Defining "Wizard Gap"**
  - **TCP buffer tuning**
  - **NICs/drivers**
  - **Duplex conflicts**
- **Three dimensions of visibility**
- **Real-time Autognostics**
- **Strategies**
  - **Develop end-to-end, continuous monitoring of app performance infrastructure**
  - **Diagnostics → Autognostics**
  - **Device and configuration management**

Leaping the Wizard Gap

- **Continuous monitoring of application performance**
- **Diagnostics → Autognostics**
- **Remediation/provisioning**

Continuous Monitoring

- **Challenges**
  - Instrumenting the network
  - Monitoring for application performance
  - Networks you don’t own
  - Dynamic monitoring and scalability
  - Intelligent detection and alerts

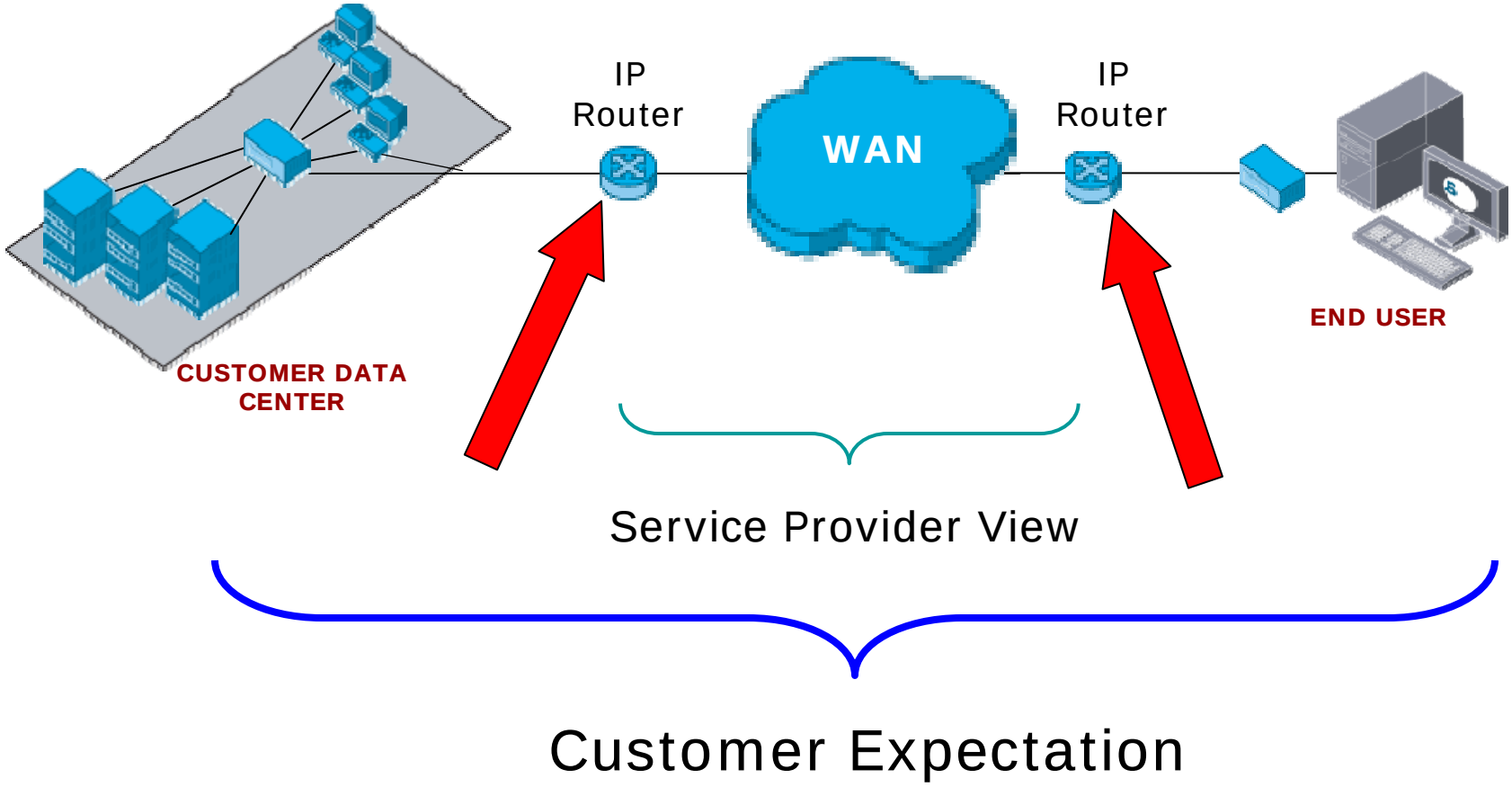
Diagnostics/Autognostics

- **Challenges**
  - Broad range of degradations
  - Found at different layers
  - Specific to application types
  - Not all related to device configuration
  - Need to be specific and targeted
  - Re-occurring (entropy and sticky fingers)

Remediation/Provisioning

- Challenges
  - Not possible to automate
- Device/configuration management
- Standardized interfaces
  - Netconf v1.0
  - SNMP v3.x
- Verification and validation
- Dynamic and configurable SLAs

The Ownership Dilemma



…. one step short of a leap

Key Points to Take Home

- Wizard gap is large and growing
- Different gaps for different apps
- Seeing is essential to optimizing
- Application modeling critical to performance monitoring
- Requirements for automation
- Missing remediation/provisioning

IPCOMM2006
September 25-27 • Gaylord Opryland • Nashville, TN

QUESTIONS?

Contact:
Loki Jörgenson
ljorgenson@apparentNetworks.com
(604) 433-2333

www.apparentNetworks.com

Apparent Networks