Future-Net
MPLS, VPLS AND BEYOND

30 April – 3 May 2007, New York City

A ZERO-PACKET-LOSS, NETWORK-LAYER IP PROTOCOL, WITH MPLS INTERNETWORKING

Curtis A. Siller, Jr. Ph.D.
Chief Technology Officer
Rivulet Communications, Inc.

PRESENTATION OUTLINE

A Zero-Packet-Loss, Network-Layer IP Protocol, with MPLS Internetworking

- Problem Statement:
  - How to Provide End-to-End QoS, with MPLS Internetworking?
- QoS in the Context of this Presentation
- Typical Applications… and Their Requirements
- Quality of Service (QoS) Approaches
  - *Over provisioning*
  - *Prioritization*
  - *Time-based Resource Reservation* (TbRR)
    - Focus on *Autonomous Flow Scheduling* (AFS)
- AFS Protocol and Network Architecture
- Internetworking AFS and MPLS
- Case Studies
- Summary

WHAT IS QUALITY OF SERVICE?

- In the Context of this Presentation:
  - A Layer-3 (Network) attribute of IP switched/routed Networks
  - Does not include network *reliability*
  - An integral element of:
    - Quality of Application (QoA) and
    - Quality of Experience (QoE)
    - QoE = QoA + QoS
- Most Frequently Cited Characteristics and Goals:
  - Minimize *delay* (exclusive of propagation)
  - Minimize *delay variation* (aka, *jitter*)
  - Minimize *packet loss*
  - Minimize *packet loss variation*

TYPICAL APPLICATIONS AND THEIR REQUIREMENTS

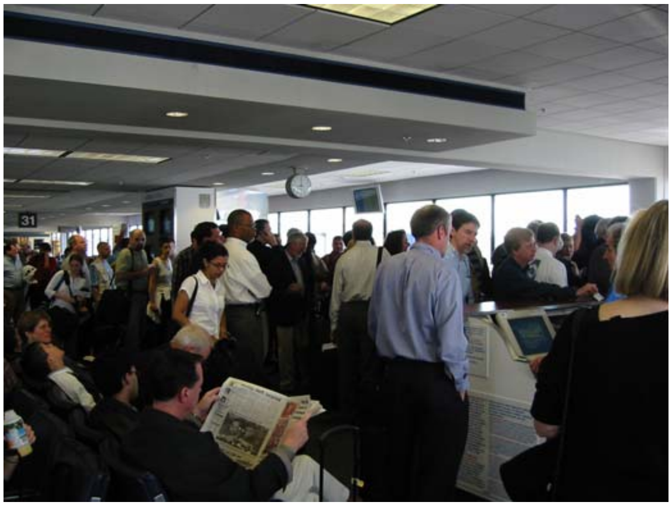
- High-Performance QoS is Most Often Associated with *Real-time* (Delay-Sensitive, Interactive) Applications
  - Voice
    - < 100 → 150 ms one-way delay (ITU-T standard)
  - Video Audio and Lip-Synch (Conferencing, VoD, IPTV)
- Associated *Real-time*, Application-Dependent Packet Loss Requirements
  - Voice: 0.5 → 1.0%
    - Packet loss concealment (PLC) is common
  - Video Conferencing: < 0.1% (ITU-T draft)
  - IPTV: < 0.00001%
- *Non-Real-time* Applications
  - Storage-Area Networks
    - Rapid, predictable back-ups related to packet loss
  - Private Line Emulation over IP (PLEoIP)
    - One-way delay ≈ 100 ms; PLR < 0.0001%

ACHIEVING QUALITY OF SERVICE

- Over Provisioning
  - Over provisioning estimates range as high as 12X
    - Corresponds to 8% utilization
  - Expensive – doesn't scale – no service guarantees
- Nearly All Other Methods Involve*:
  - Queue Management with Prioritization
    - DiffServ is a prominent example
    - Remains a subject of on-going research, with numerous approaches continually being proposed
- Time-based Resource Reservation (TbRR)
  - Well-suited to critical, especially time- and packet-loss-sensitive flows
  - Not limited to real-time applications
    - E.g., Storage area networks

*FEC is being considered as an adjunct for IPTV. (See SMPTE standards)

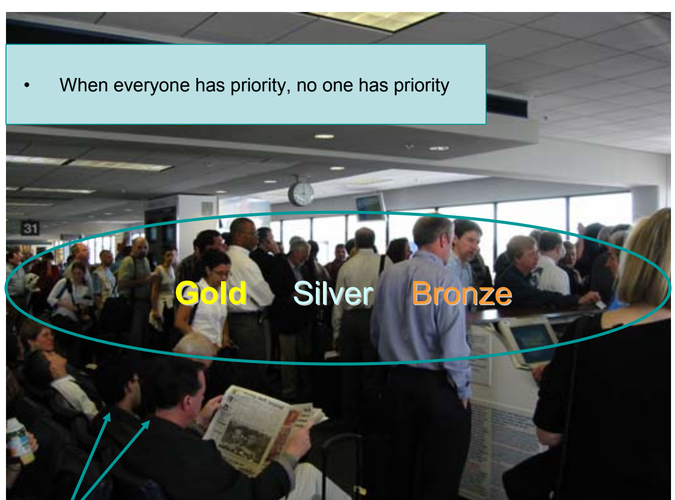
DOES PRIORITIZATION WORK?



WHAT WE KNOW ABOUT PRIORITIZATION

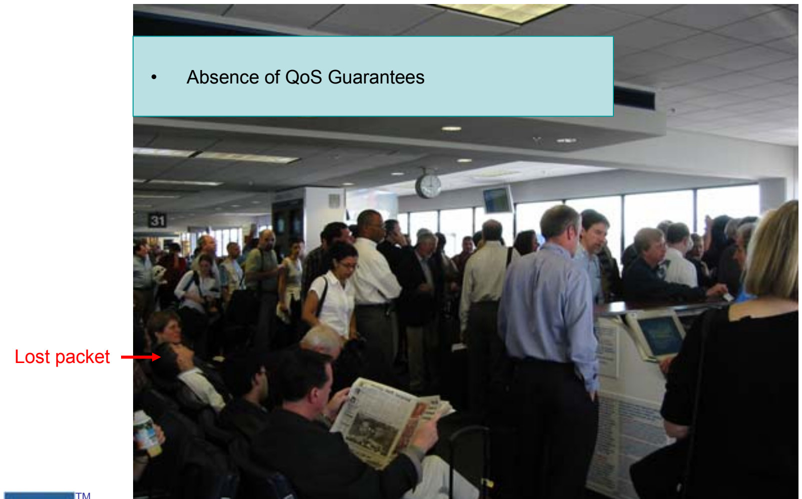
- "When Everyone has Priority – No One has Priority"
  - I.e., Doesn't scale to high levels of utilization
    - Degrades to "best effort"
- Highest Priority Starves Lower-Priority Traffic
- Extremely Difficult to Provide Multiple, Ordered, Service Classes (Gold, Silver, and Bronze) Simultaneously Constrained Across:
  - Packet loss, jitter, and queuing delay
- Prioritization Does Not Provide QoS Guarantees
  - Probabilistic
- Organizational Barriers Impede Multiple Applications on a Single Network
  - Concern that heterogeneous applications will contend among each other
  - Paraphrasing George Orwell: "All applications are equal, but some applications are more equal than others."

DOES PRIORITIZATION WORK?



Starved packets

DOES PRIORITIZATION WORK?



WHAT IS DiffServ?

- DiffServ = Differentiated Services
- A Means to *Classify/Mark* IP Packets for Per-Hop Behavior
  - An extension of prior Type of Service (ToS) bits
  - Alone, classifying/marking IP packets does not provide QoS
- DiffServ Code Points (DSCPs) are a 6-bit Field in the IP Packet Header
  - DSCPs infer PHB
  - Implementation of PHB and queuing left to vendors
- DSCPs are Summarized in the Next Slide…

DSCPs

DSCP 101 110 Corresponds to **Expedited Forwarding**, EF. EF intended for low-loss, low-delay, low-jitter services (e.g., "real time")

The following DSCPs are Allocated for **Assured Forwarding**, AF. AF-PHB used to guarantee bandwidth to an AF class

Drop Precedence	**Class 1**	**Class 2**	**Class 3**	**Class 4**
Low Drop	AF11 = 001010	AF21 = 010 010	AF31 = 011 010	AF41 = 100 110
Medium Drop	AF12 = 001 100	AF22 = 010 100	AF32 = 011 100	AF42 = 100 100
High Drop	AF13 = 001 110	AF23 = 010 110	AF33 = 011 110	AF43 = 100 110

Ascending (Better) Drop Precedence

Ascending (Higher) Class Priority

DSCP 000 000 is **Default** (e.g., Best Effort)

GENERIC TIME-BASED RESOURCE RESERVATION TECHNIQUES

- Reserve IP "Resources" for Most-Critical Traffic on the Basis of *Time*
  - Link throughput is reserved before flow initiation, eliminating packet contention
  - No need for complex packet buffer management
  - Requires synchronization (widely available GPS, BITS, CDMA, PCR)
- Implications:
  - Significantly reduces delay
    - Queued packets are a major source of end-to-end delay
  - Minimal to no jitter
  - Without contention, there is no Layer-3 packet loss
    - Loss concealment techniques are unneeded
  - QoS guarantees, rather than statistical
- Traditional Prioritization Available for Less Important Applications
- A Variety of TbRR Techniques ("Time-driven Priority", "Layer-1 Switching", "Sequencing", "Time-Synchronized Multi-Layer Switching")
  - Few comply with communications standards
  - Most require specialized switch/routers and "state machines"

THE "HOW" OF AUTONOMOUS FLOW SCHEDULING

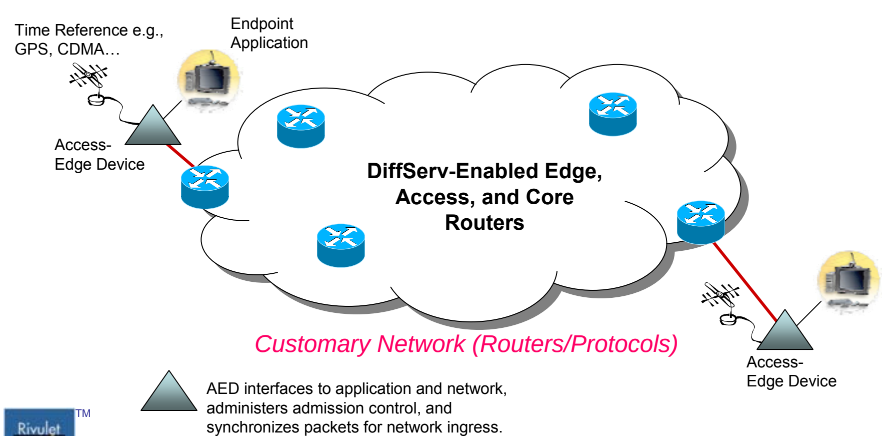
- Uses DiffServ
  - Commercial, standards-compliant, existing routers and customary routing protocols
- Hierarchical Approach to Ordering Traffic
  1. Critical Traffic (Highest Priority Traffic Uses *Expedited Forwarding, EF, with DSCP 101110*)

  No resource contention among real-time flows, because…

  2. Discovery *Reserves* "Time Slots" (2nd Priority; *Assured Forwarding AF*)
  3. Other traffic receives lesser priority (Lesser *AF classes, and Default – "Best Effort"*)
     – Use with traditional prioritization techniques

AFS NETWORK ARCHITECTURE

– Exceptionally Simple
– Synchronized Endpoints
– Existing DiffServ-Enabled Routers



15

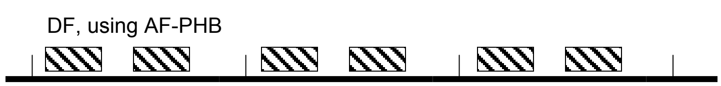
ESTABLISHING THE FIRST AUTONOMOUSLY SCHEDULED FLOW

(a) Common Time Interval (User Traffic Not Shown)

16

ESTABLISHING THE FIRST AUTONOMOUSLY SCHEDULED FLOW

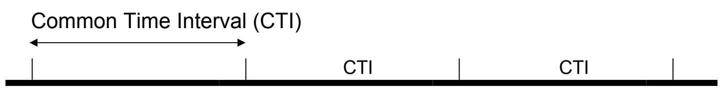
(a) Common Time Interval (User Traffic Not Shown)



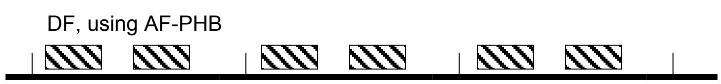
(b) Discovery Flows

17

ESTABLISHING THE FIRST AUTONOMOUSLY SCHEDULED FLOW



(a) Common Time Interval (User Traffic Not Shown)

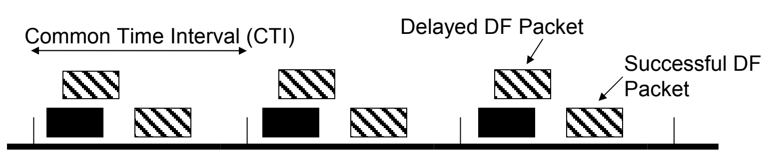


(b) Discovery Flows

(c) Establishment of First ASF; Less-Critical Traffic Not Shown

18

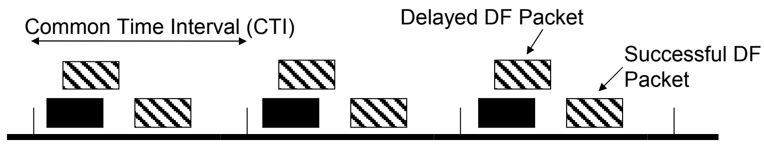
ESTABLISHING A SUBSEQUENT ASF



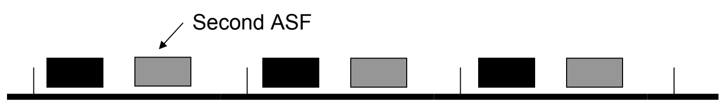
(a) Introducing Discovery Flows, in the Presence of an ASF

19

ESTABLISHING A SUBSEQUENT ASF



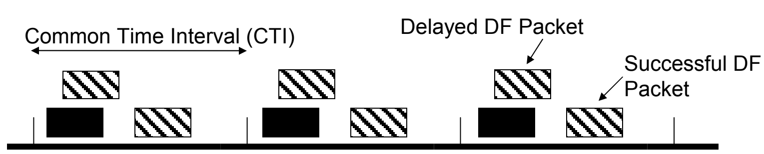
(a) Introducing Discovery Flows, in the Presence of an ASF



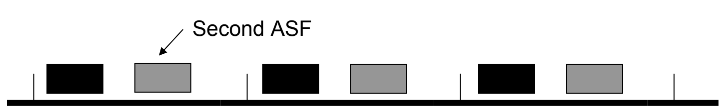
(b) Two ASFs; Non-ASF Traffic Not Shown

20

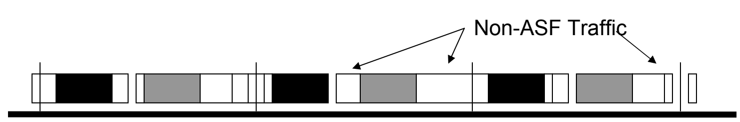
ESTABLISHING A SUBSEQUENT ASF



(a) Introducing Discovery Flows, in the Presence of an ASF



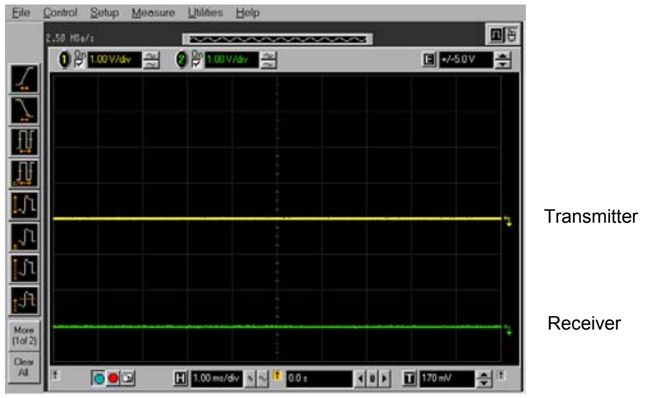
(b) Two ASFs; Non-ASF Traffic Not Shown



(c) Two ASFs, *with* Non-ASF Traffic (on the link and queued)

21

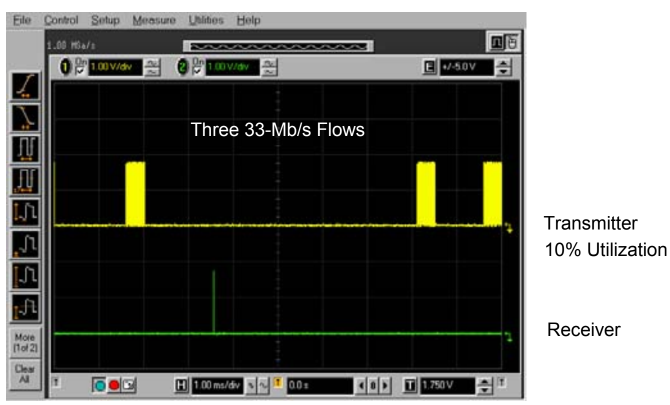
PICTORIAL CAPTURE OF LINK UTILIZATION



CTI: 10 milliseconds, 1 GigE link

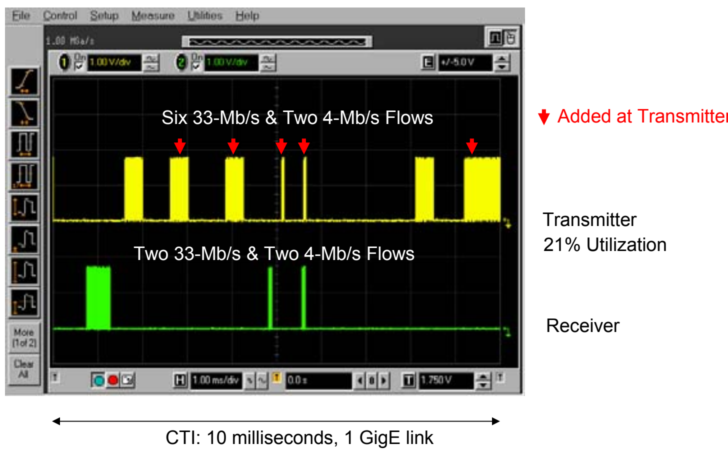
22

PICTORIAL CAPTURE OF LINK UTILIZATION



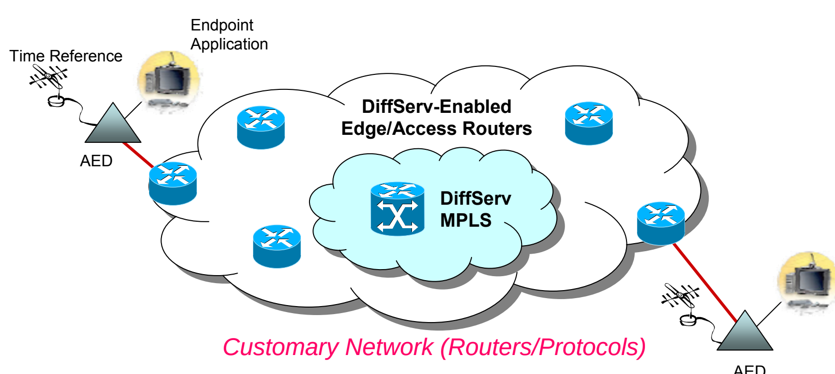
CTI: 10 milliseconds, 1 GigE link

PICTORIAL CAPTURE OF LINK UTILIZATION



ASF NETWORK ARCHITECTURE WITH MPLS

- **– Assume DiffServ Throughout**
  - **– Synchronized Endpoints**
    - **– ASF and MPLS Readily Meld**

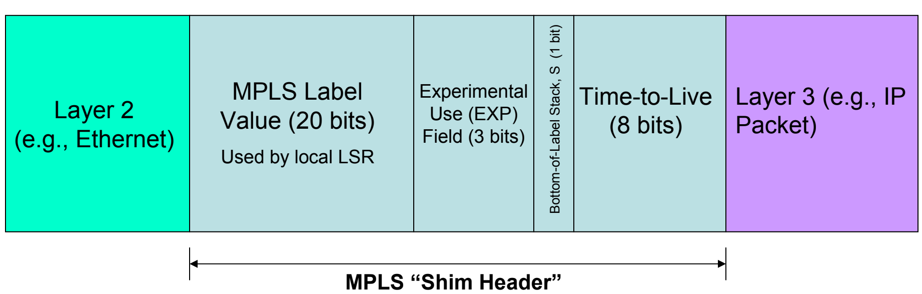


MPLS

- A Multiprotocol Transmission Technology for Label Switching
- A Logical Basis for VPNs
- Generally Deployed in the Backbone (“Core”) of Carrier Networks
  - The “average” packet traverses 10 – 12 nodes
    - Packets typically traverse only 3 – 5 nodes in the core
      - High speed routers/switches; “fat pipes”; little or no packet loss or jitter
    - The QoS problem largely arises in the edge/access
      - Commonly 5 or more nodes
      - Slower, smaller routers; “thin pipes”; lots of congestion; substantial packet loss, jitter, and queuing delay
- Potential Difficulty Scaling MPLS to Include Access/Edges Routers
  - Labels must be intelligently applied
    - The larger the network, the greater the problem
    - Potentially complex and expensive traffic engineering

MPLS *SHIM* HEADER

- Between Layer 2 (Data Link) and Layer 3 (Network Layer) IP Packet

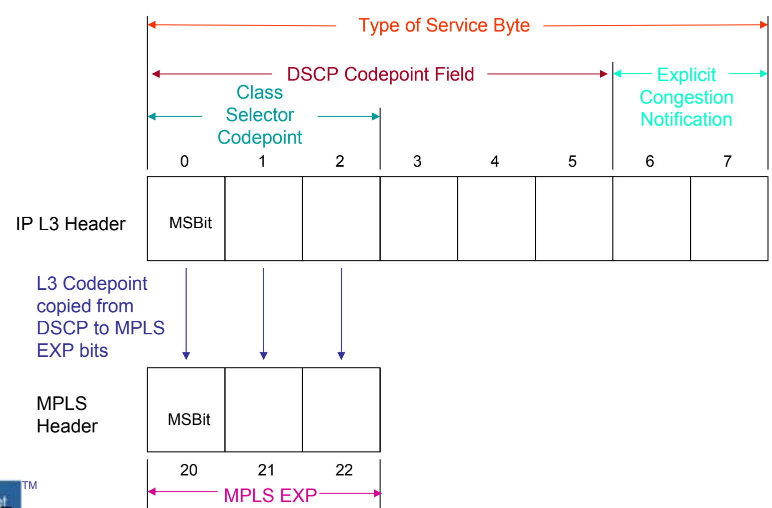


MPLS - DiffServ and ASF

- IETF RFC 3270, " Multi-protocol Label Switching (MPLS) Support of Differentiated Services"
  - Provides for a "Uniform Mode" of Tunnel Operation
  - As MPLS packets move through an MPLS domain, labels can be added, removed and manipulated
  - However, the 3-bit EXP field does not change.
  - The 3 most-significant EF, AF**x**y (Class), and Default bits in the IP packet header are directly mapped to EXP in the MPLS Shim Header
    - EF 101 – EXP 101
    - AF4 100 – EXP 100
    - AF3 011 – EXP 011
    - AF2 010 – EXP 010
    - AF1 001 – EXP 001
    - Def 000 – EXP 000

- **AFS's Hierarchical Use of EF, AF and Default (Best Effort) is Preserved Across MPLS**

DSCP - to - MPLS EXP MAPPING



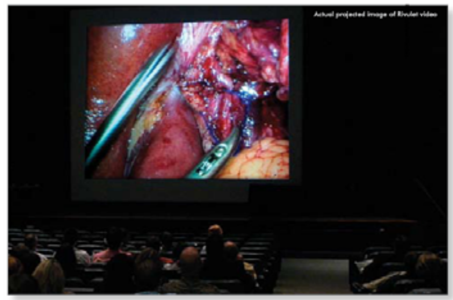
RANK ORDERING QoS for DIFFERENT IP/MPLS TECHNOLOGY ALTERNATIVES

Environment	Approach	Rank 0 (Worst) – 3 (Best)	Issue(s)
IP Environment	IP – Best Effort	0	The "original" internet without any QoS mechanisms
IP Environment	IP – DiffServ	1	Standards-based prioritization; potential for packet contention (jitter, delay, and loss) remains
IP Environment	Autonomous Flow Scheduling (AFS)	3	Unique use of DiffServ: no packet loss, low delay, little or no jitter
MPLS Environment	MPLS	0	Adding labels are basis for VPNs, but no explicit improvement in QoS
MPLS Environment	MPLS – DiffServ	1	QoS generally equivalent to IP – DiffServ; MPLS supports backbone switching
MPLS Environment	MPLS – DiffServ – TE	2	Potentially superior to MPLS-IP-DiffServ; requires traffic engineering
MPLS Environment	MPLS – AFS	3	Equivalent to AFS; fully interoperable with MPLS-DiffServ or MPLS-DiffServ-TE

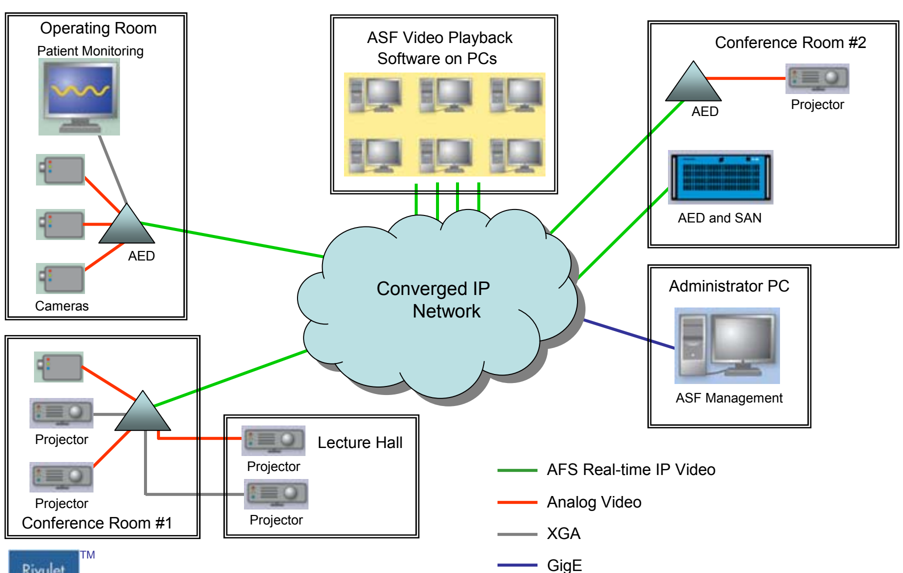
CASE STUDIES – MAJOR MEDICAL CENTERS IN NEW YORK CITY

TWO CASE STUDIES – MAJOR MEDICAL CENTERS IN NEW YORK CITY

- Case A
  - Multi-facility environment
  - Converged IP network supporting: data, voice, medical information, videoconferencing, and video, with upgrade to high definition (HD)



Real-time ASF Video in a Converged IP Network



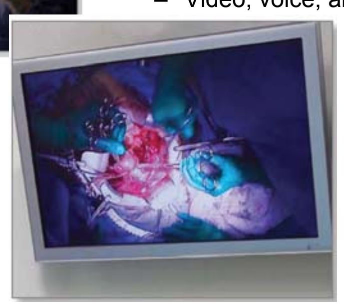
CASE A: END USER ASSESSMENT

"…..joins the very small group of elite institutions moving toward video on a fully converged network. The technical and operational quality on display today was apparent to every conference attendee, and underscores the importance of your work to the Medical center. You have ushered in a new era of digital media facilities on campus."

Assistant Dean of Advanced Applications, xxxxxx School of Medicine

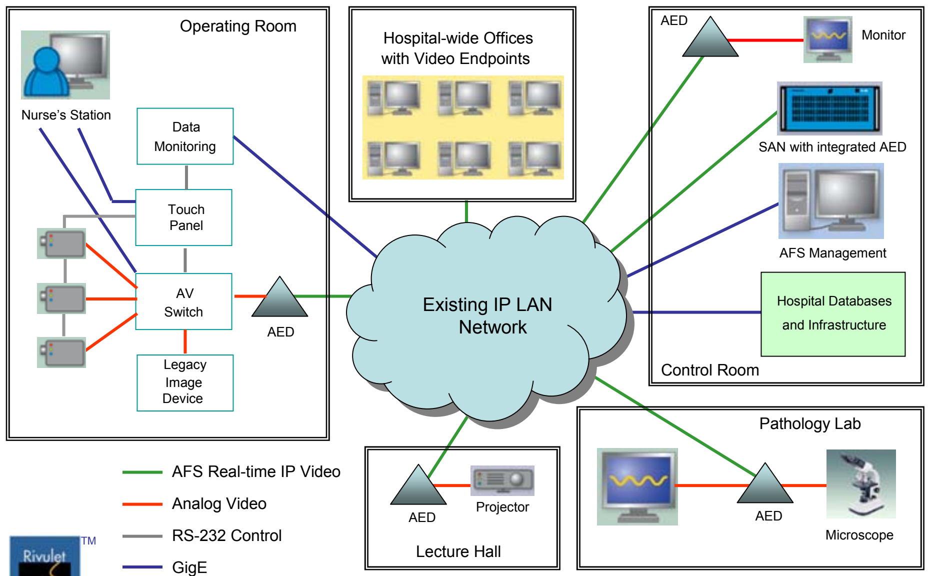
TWO CASE STUDIES – MAJOR MEDICAL CENTERS IN NEW YORK CITY

- Case B
  - Upgrade to existing IP LAN
    - Unsatisfactory video performance (packet loss) at <10% utilization
      - **With AFS, tested to ≈ 90%**
  - 21 Operating Rooms
  - Video, voice, and archival storage



35

Hospital Campus-Wide Video Distribution



Rivulet™

Convergence that works. Really.™ www.rivulet.com

36

CASE B: END USER ASSESSMENT

"After months of operation, hundreds of hours of video flows, not a single lost packet."*

"…The best news is, since we went live, I haven't heard from the IT guys."

Engineering Specialist, xxxxxxxx Center

* Service turned up May 1, 2006; thousands of surgeries; no packet loss

Rivulet™

Convergence that works. Really.™ www.rivulet.com

AUTONOMOUS FLOW SCHEDULING

- QoS for Critical Applications (e.g., Audio, Video and Other Real- and Non-Real-time Applications)
  - No packet loss
  - Minimal jitter
  - Low delay
- High Network Utilization
- Scalable
  - Uses existing Internet routing protocols
- Switching/Routing Technology
  - Works with existing, standards-compliant DiffServ-enabled routers
  - Interoperable with MPLS
- Synchronization
  - Readily available AED timing
    - GPS, CDMA, BITS, Rb Oscillators

Thank You!

& Q&A