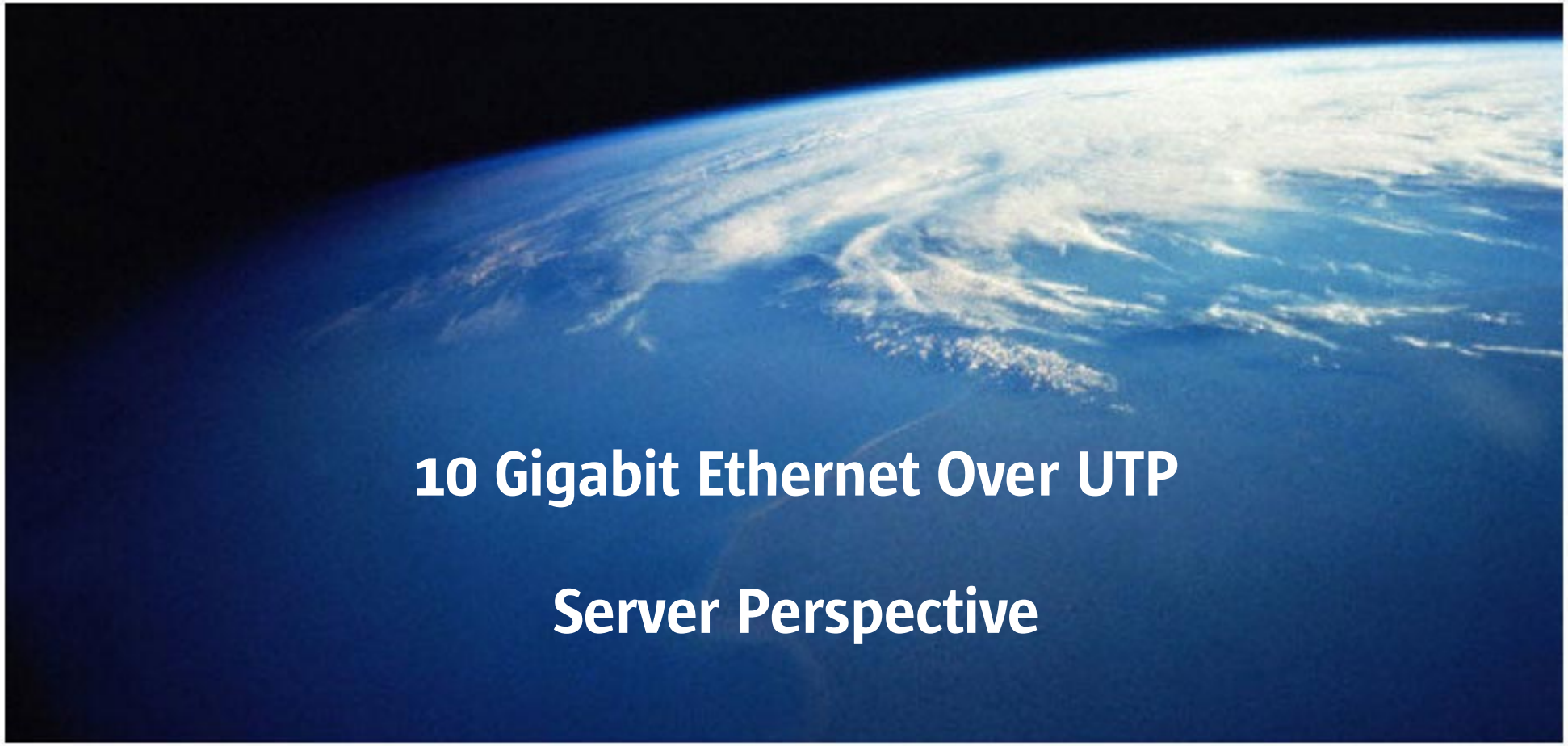


10 Gigabit Ethernet Over UTP

Server Perspective

Shimon Muller

Sun Microsystems, Inc.

10GBASE-T Study Group

January 9, 2003

Vancouver, BC

Sun microsystems

We make the net work.

S.Muller

Introduction

- Why do we need 10 Gigabit Ethernet in our servers today and (yes!) in the desktops tomorrow
- Why do we need a 10 Gigabit Ethernet solution over Cu

S.Muller

Sun microsystems

10-Gigabit Ethernet: Market Perspective

S.Muller

10-Gigabit Ethernet – Market Reality

- 10-Gigabit Ethernet deployment is on to a very slow start
- Products:
  - Mostly uplink blades for existing Gigabit Ethernet switches
  - Few NICs
  - Not wire-speed
  - Expensive
- Networks:
  - A few deployments
  - Mostly long-haul point-to-point interconnects between data-centers
  - " Very High End" customers

10-Gigabit Ethernet — Issues

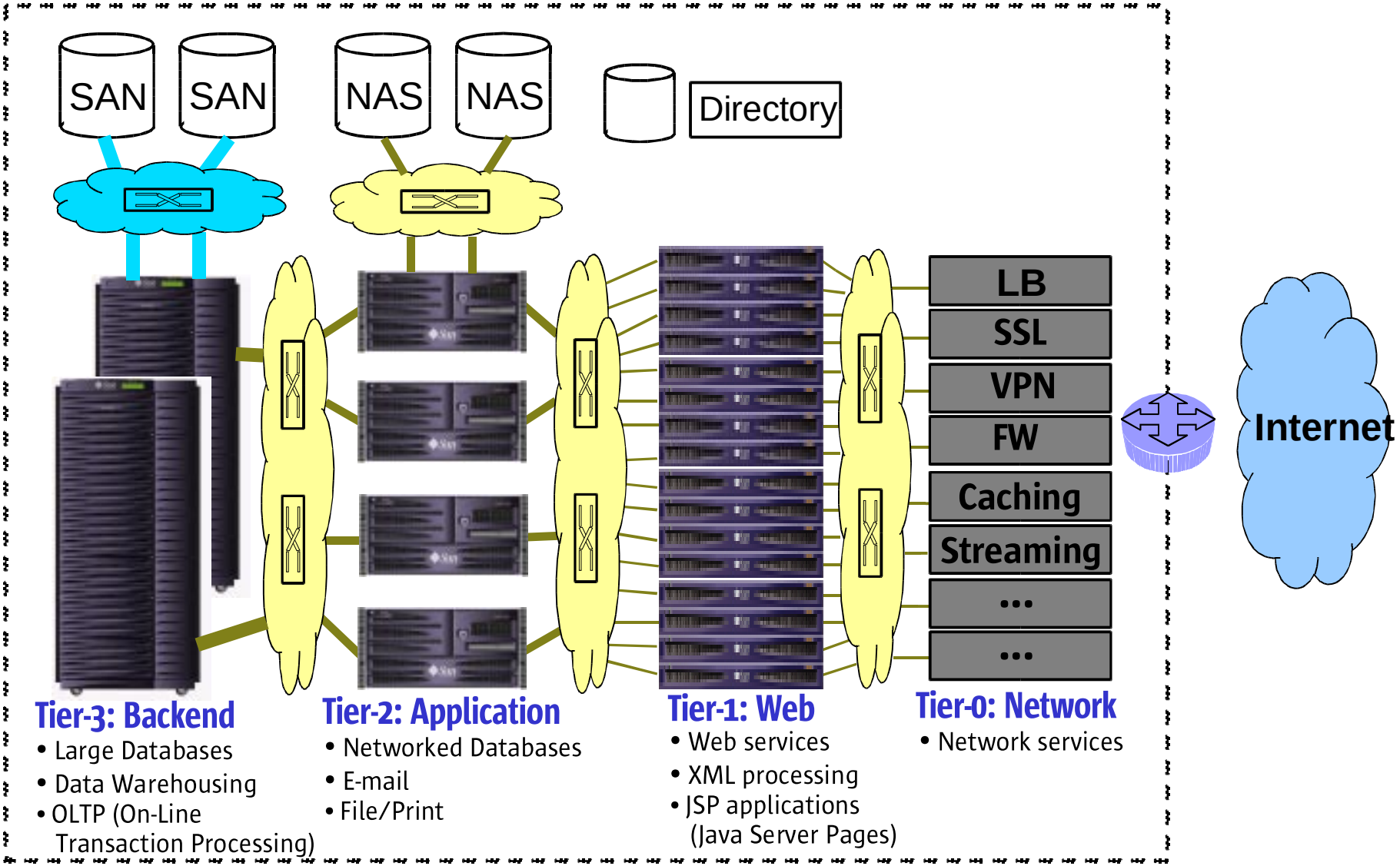
- Timing
  - 10 Gb/s rates put a strain on existing switches and servers
  - New generation of systems is required to take advantage of the faster network
    - Switches with backplanes and switch fabrics that can support multiple 10-GE ports
    - Servers with faster I/O subsystems (IB, 3GIO, etc.)
  - It's the economy, stupid...
- Need
  - Do we really need 10 Gb/s networking any time soon?
    - What is the killer application?
    - Where does it make sense?
- Performance
  - If we deploy it, will we really get 10 Gb/s?
    - Mostly an issue for computer vendors
    - More than just an I/O issue
    - The network is faster than the computer and is likely to stay that way
    - From a server perspective we need to address this

10-Gigabit Ethernet – Issues (continued)

- Cost
  - Too expensive...
  - Too expensive...
  - Too expensive...
    - Does not yet follow the traditional Ethernet model of 10x performance for 3x cost
  - The physical layer is a major contributor to the high cost
    - Today only 10GBASE-LR and 10GBASE-ER are shipping
    - The low-cost alternative for the data-center is not really “low cost”
  - Very little installed cabling infrastructure to support interesting distances
    - Adds to deployment cost
  - Prices will drop over time, not clear how fast

10-GE Technology: Addressing the Issues

Service Point Architecture



S.Muller

Defining the Tiers

- Tier 3:
  - Big/Fat systems
  - Lots of fast processors (32 and above)
  - Lots of power
  - Scaling: Mostly vertical, very little or none horizontal
    - Single OS image executing on SMP cache-coherent architecture
  - Cost sensitivity: Low
    - Performance is King!

- Tier 2:
  - Mid-range systems
  - Moderate number of processors (typically 8-24)
  - Power density: Moderate
  - Scaling: Combination of vertical and horizontal
  - Cost sensitivity: Moderate
    - Price/Performance is the major consideration

S.Muller

Defining the Tiers (continued)

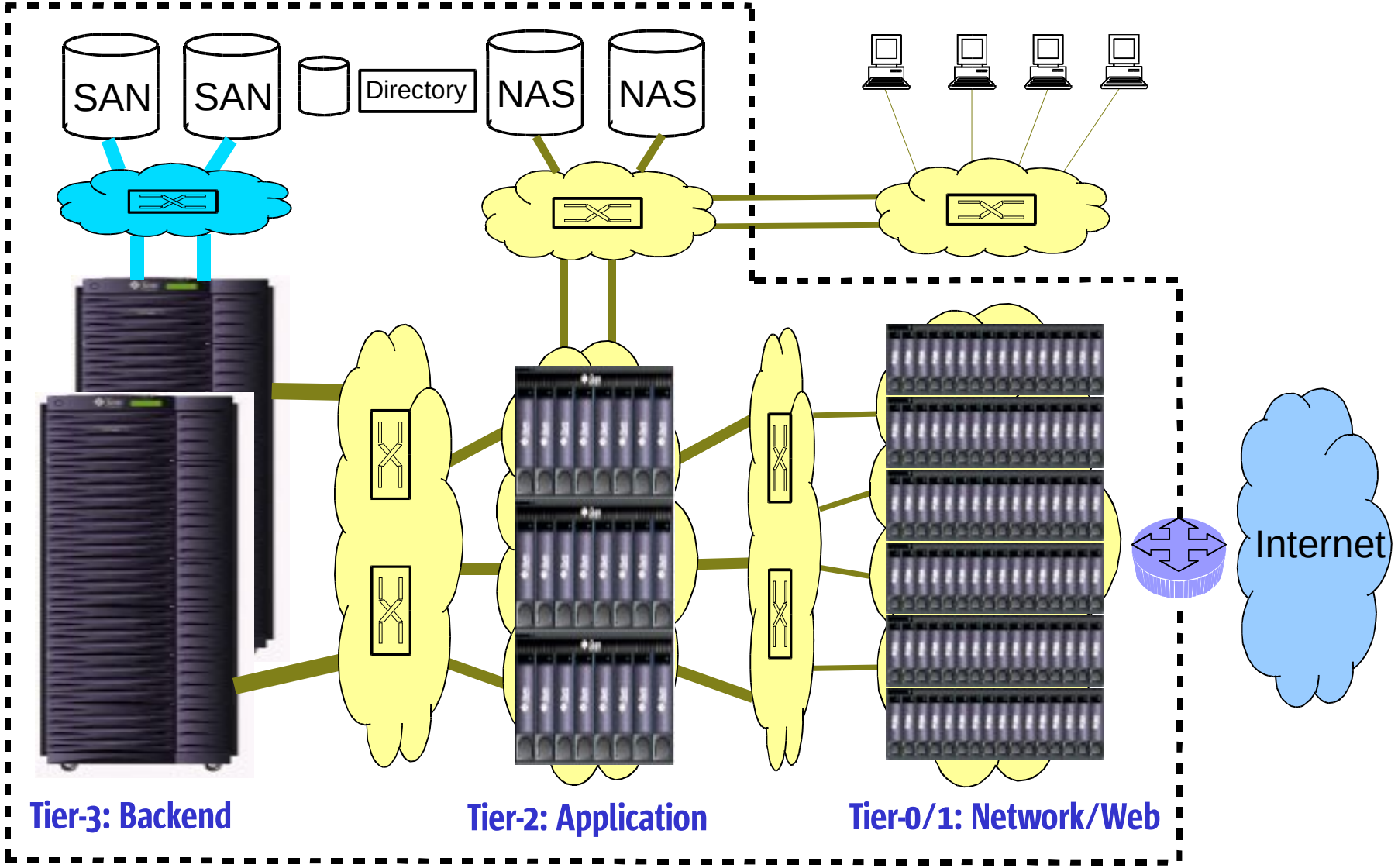
- Tier 1:
  - Low-end systems
  - Few processors (1-4)
  - Power density: Critical
    - Measured in Watts per square foot
  - Scaling: Horizontal
    - Multiple OS images executing on multiple servers in a networked architecture
  - Cost sensitivity: Critical
  - Cabling is a major problem
- Tier 0:
  - Network services
    - Multiple evolving functions between the Web tier and the network cloud
    - Follows the appliance model
  - The appliances are typically connected in series
    - The weakest link in the chain is the performance bottleneck
    - Cost sensitivity: Driven by Tier 1
      - Amortized across multiple servers in Tier 1

S.Muller

Computing Trends

- Massive Horizontal Scaling in Tiers 1 and 2
  - Takes the horizontal scaling paradigm to the next level of performance
    - Better compute density
    - Better power density
    - Better price/performance
- Strong Desire to Collapse Tiers 0 and 1
  - To some extent also Tier 2
  - More efficient use of Tier 0 functions
    - Better performance
    - Better price/performance
    - Ease of use: Fewer boxes to manage

Tomorrow's Datacenter/Server Farm



S.Muller

Implications for Networking Requirements

- Intra-Tier Communication
  - A significant portion of network traffic is contained within a single enclosure
    - Inter-Blade links
    - " Network in a box"
  - Network services provisioning in the compute box will create additional pressure on the communication links
    - " Packet pipeline"
    - Network traffic will be required to traverse the links multiple times
- Inter-Tier Communication
  - Blade computing increases b/w pressure on the external communication links
    - Shifts the network traffic aggregation points from external devices to the compute box
    - Increases traffic aggregation at least by an order of magnitude in the lower tiers
    - In the higher tiers network requirements are driven by much more powerful systems
- Desktop
  - Will use anything they can get "for free"

S.Muller

Server Networking Performance

- Challenge: What would it take to fill a 10Gb/s pipe on a server
  - Easiest: Horizontally scaled systems
    - Lots of aggregation
    - Relatively low-speed connections
    - Load balancers (L4/L7) will do the job
  - Hardest: Vertically scaled single SMP systems
    - Orders of magnitude fewer connections
    - Single high-speed connection performance is more important for relevant applications
    - System memory latency gets in the way of efficient handling of network traffic
    - Must find ways to achieve significant throughput speed ups for this type of systems

Server Networking Performance (continued)

- Where we are today
  - Bulk data throughput for a single TCP connection: 1 Mbps/MHz
    - Has been demonstrated by several vendors

- Issues:
  - Per packet processing overheads
    - Limiting factor for throughput in both bulk and transactional traffic patterns
  - Per byte processing overheads
    - Today – mostly copies
    - Contribute to excessive CPU and memory bus utilization
  - Scalability
    - Not linear across multiple CPUs in a single OS image
  - Performance penalty paid for protocol stack layering and software modularity

Server Networking Performance (continued)

- Future
  - Soon (2003):
    - 2 Mbps/MHz throughput will be achieved with relatively little effort
      - Most of the known issues can be effectively addressed today
  - Longer Term (2006):
    - 4 Mbps/MHz throughput can be achieved by the time 10-GE becomes ubiquitous
      - Hardware protocol acceleration/termination engines, RDMA, etc.
        - Not a good solution for general purpose networking
        - A viable option for specialized implementations (iSCSI, NAS, NFS, etc.)
    - Others... :-)
    - Brute force
      - Processors are getting faster
        - 4x speed up by 2006

What About the Desktop

- Question: Why would we need 10Gb/s on the desktop any time soon?
  - This is the wrong question to ask
  - The right question should be not “WHY?” but rather “WHY NOT?”
- The Answer:
  - Today's server is tomorrow's desktop
  - Network application performance is not only about sustained throughput
    - Latency is just as important
    - Massive overprovisioning of dirt cheap bandwidth is a viable alternative to QOS
- Ignoring 10 Gigabit Ethernet connectivity to the desktop may prove to be a mistake in the long term

10-GE Deployment Issues Revisited

- Timing
  - May be somewhat ahead of its time, but not by too much
    - New systems are being designed today that will be able to support 10 Gb/s rates
- Need
  - No new killer application, but the need is there
    - Not much different than Gigabit Ethernet four years ago
    - Blade computing is a new driver for technology adoption
- Performance
  - Not an issue for Tiers 0/1
  - Solvable problem for the upper tiers
    - Not much different than Gigabit Ethernet four years ago

10-GE Deployment Issues Revisited (cont.)

- Cost
  - This is the main roadblock for 10-Gigabit Ethernet deployment in datacenter
  - Paradox: 10 Gb/s connectivity is most urgent and the throughput is easiest to achieve in the tiers that are most cost sensitive
  - Current cost of 10-Gigabit Ethernet physical layers is at least an order of magnitude higher than can be justified for most datacenter deployments
  - The cost decline for optical link solutions has historically been much slower than that for Cu
  - *We need a low cost copper interconnect for 10-Gigabit Ethernet*

Why 10GBase-T?

- The most cost effective solution in the foreseeable future
  - Low initial cost
  - Will follow Moore's Law
    - Remember 1000Base-T?
- Provides a single uniform solution for 10-Gigabit Ethernet connectivity for all short-reach applications from the backplane to the desktop
  - Will drive the economies of scale
  - Will solve the datacenter deployment dilemma
    - Which cable to use where with which laser...
- Cabling Type/Reach
  - 100m reach is important not just for the desktop, but also to cover all datacenter deployments
  - Preservation of the installed cable plant
    - Highly desirable for the datacenter and server farm
    - A must for the desktop

That's All, Folks!

Shimon Muller

Sun Microsystems, Inc.

10GBASE-T Study Group

January 9, 2003

Vancouver, BC

Sun microsystems

We make the net work.