

COMP 519 – Future of Ethernet

Jeffrey Shafer

Future of Ethernet

- Ethernet standard (802.3) first published in 1983
- Much of the original standard has been discarded at 1Gbps and above:
 - No more shared bus or thick coax cable, only point-to-point links
 - No more Carrier Sense Multiple Access or collisions
 - No more Manchester encoding



“Today's Ethernet technology is extremely diverse and has very little in common with what appeared in '74. The good news is that they still call it Ethernet, and that's my word.”

Bob Metcalfe, 2003

Future of Ethernet

- Some parts remain
 - Ethernet frame format
 - Business model
 - Companies compete with proprietary designs
 - IEEE standards ensure interoperability
 - Standards evolve rapidly (but with backwards compatibility)



“What Ethernet is today is more than a packet format or media access algorithm--it is a business model”

“If they want to call 802.11 wireless Ethernet, I'm all for it, especially because it's reminiscent of the Aloha network from which 802.11 is derived”

Bob Metcalfe, 2003

Future of Ethernet – 10 Gbps

- Diversity of physical layer options (just like previous Ethernet!)
 - 6 fiber optic standards + 3 copper standards
- Marketplace will determine which survive
 - Possibly 10GBASE-T, which can use normal twisted-pair Ethernet cables
 - Currently expensive! (rare outside of datacenters)



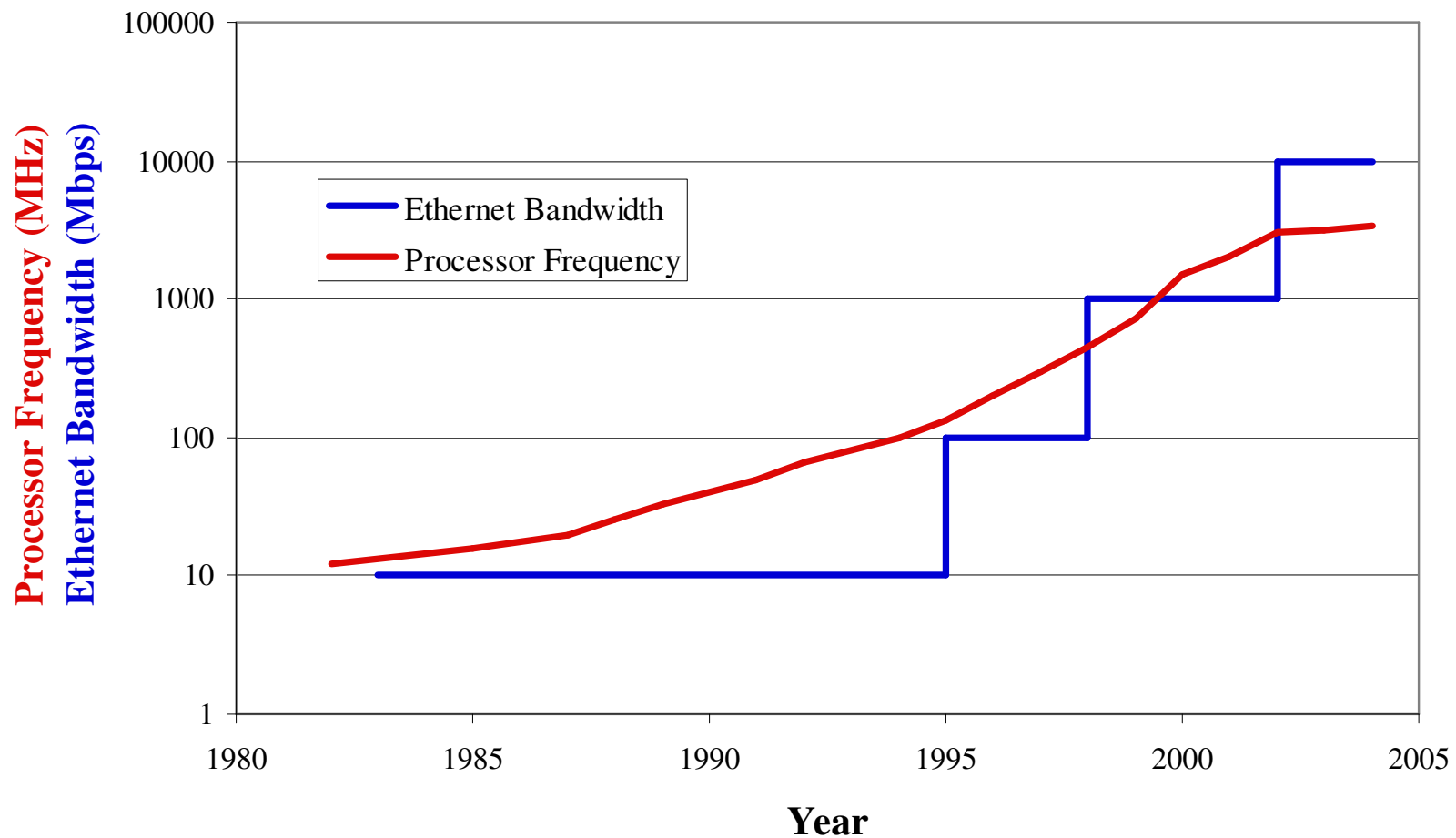
Future of Ethernet – 40/100 Gbps

- Data centers are already aggregating multiple 10Gbps lines together – Always demand for more bandwidth!
- Design objectives set in 2006 – research ongoing
 - 100 Gbit/s (at the client interface)
 - Range of 100 m on OM3 multi-mode optical fiber
 - Range of 10 km on single-mode optical fiber (SMF)
 - Full-duplex operation only
 - Preserve 802.3 / Ethernet frame format at the MAC level
 - Preserve current minimum and maximum frame size
 - Low bit error rate: below 10^{-12}
- Current status
 - Design demonstrated by joining together 10 10Gbps links in parallel
 - Standard to be finished by 2010

Network Performance – Achievable?

- In 2010, will I be able to buy a 100Gbps NIC, plug it into my computer, and expect to get 100Gbps of throughput?
 - Not even close!
 - Challenging to produce/consume that much data
 - Challenging to produce/consume headers for that many frames
 - 81,300 frames/s at 1Gbps
 - 813,000 frames/s at 10Gbps
 - 8,130,000 frames/s at 100Gbps

Network vs Processor Performance



Network performance is outpacing processor (core) performance

Network Performance – Achievable?

- Better NIC designs are needed
 - Transmit data path:
 - TCP Segmentation Offload (TSO)**
 - Send the NIC a large buffer (64kB)
 - Have NIC segment data into multiple packets
 - Receive data path:
 - Large Receive Offload (LRO)**
 - More efficient for network stack to process a large buffer of data (from a single stream) than many small buffers
 - Data must be aggregated either on the NIC or in software
 - Surprisingly, even software method can improve performance by reducing overhead of higher layers of the network stack

Network Performance – Achievable?

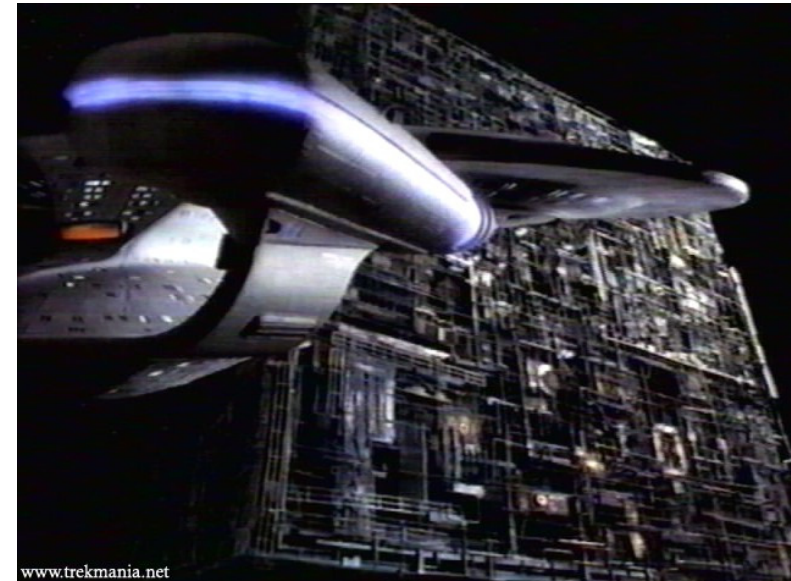
- Better OS architectures are needed
 - Multicore is a given
 - Efficiently parallelized network stacks are required
 - How many cores should the network subsystem scale to?
 - How do we divide the work?
 - Per connection?
 - Per message?

Future of Ethernet – Jumbo Frames?

- Ethernet frame size - 1500 bytes
- Ethernet Jumbo frame size - 9000 bytes
 - 32-bit CRC loses effectiveness above 12000 bytes
 - Pros
 - Can send more data with the same amount of overhead (in creating and routing based on the header)
 - Improves efficiency of host computers, NICs, and switches
 - Supported by most gigabit NICs and switches
 - Cons
 - More data must be re-sent if frame is corrupted
 - Not an IEEE standard
 - Not supported across commercial Ethernet
 - 9000 byte length is not universally accepted
 - Jumbo frames must be segmented to send across internet, losing the efficiency advantage
 - Used primarily in cluster computers / datacenters, where all equipment supports jumbo frames
 - Competing networks (e.g. Myricom) also support jumbo frames

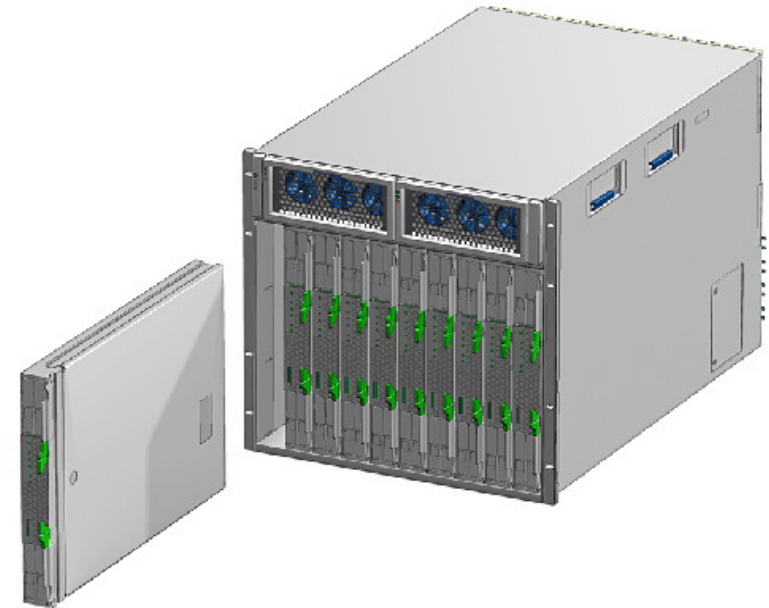
Future of Ethernet

- Battle plan
 - Attack on all fronts!
- Targets
 - Backplane Technology – Blade servers
 - Short distance, < 1 meter
 - SAN - Storage Area Networks
 - Short distance, inside datacenter
 - MAN - Metropolitan Area Network
 - Long distance, tens of km
- Marketplace will determine if these new products succeed



Future of Ethernet

- Expand into backplane market (short distance < 1 meter)
 - Applications
 - Blade server chassis – link blade to network, storage, and management I/O
 - Routers and switches with removable line cards
 - Goals
 - Replace vendor-specific implementations with IEEE Ethernet standard (including frame format and frame size)
 - IEEE 802.3ap standard with 10GBase-KX4 / 10GBase-KR physical layers
 - Reasons to migrate
 - Multi-vendor interoperability – Mix and match blades
 - Cost savings due to efficiencies of scale and leveraging existing Ethernet switching technology
 - High performance – 10Gbps



Future of Ethernet

■ Expand into SAN market

- Storage Area Networks (short links inside datacenter)
- Goals
 - Replace Fibre Channel as link between storage array and storage servers
- Reasons to migrate
 - Cost – Fibre Channel is expensive! (used only by large corporations and supercomputer centers)
 - Standardization – Why have a different network standard that is only used to link storage and servers?
 - Ease of deployment – Fibre Channel is very specialized and requires training, but most institutions already have staff with Ethernet experience
- Transition plan
 - iSCSI – SCSI over TCP/IP (and thus, over Ethernet)
 - Deployed by small/medium-sized businesses (i.e. institutions that did not already have Fibre Channel SANs)

Future of Ethernet

- Expand into MAN market
 - Metropolitan Area Network (Long distance, tens of km)
 - Goals
 - Replace SONET (Synchronous Optical Networking) as telecom backbone
Allows institutions to run Ethernet end-to-end, even across multiple physical sites
 - Reasons to migrate
 - SONET frames are voice-oriented, not data-oriented
 - Large frame sizes that increase with network speed
 - 87 bytes for OC-1
 - 6,704 bytes for OC-192
 - Unused frame capacity is wasted
 - Inefficient when carrying packet data (Ethernet / IP frames)
 - High cost of SONET equipment
 - Save \$\$ on network administration by having single technology
 - Transition plan
 - Ethernet frames can be encapsulated and sent over SONET with a lightweight header
 - Allows telecoms to migrate first before replacing existing equipment

Scaling Ethernet

- Can I have a single switched Ethernet network spanning the entire world?
 - Commercial switches only have ~16,000 entry forwarding table
 - How do the switches find the destination computer?
 - Broadcast to every computer in the world?
 - Ethernet scalability has limits
- Routing / Higher-Layer Protocols Needed
 - Partition network into discrete LANs
 - Link to other LANs may also be Ethernet, but link is not accessible via a switch, but instead a router