Building and Measuring a High Performance Network Architecture

for

SC 2000 Wrap Up Meeting

January 10, 2001

Bill Kramer

Contact: kramer@lbl.gov

510 486-7577

Agenda

- Reports
  - Written
- Review the technical activity and results
- Observations and Recommendations for future SC
  - SCinet 2001
  - Strategic
- Overall Summary

SCinet 2000 Goals

- Provide a stable, high quality network infrastructure for SC 2000 activities
  - support the basic infrastructure of the conference, the attendee's needs and the extensive education program.
- Provide reliable, high performance network for exhibitors
  - conference provides a showcase for the research and development activities of both the industry and research exhibitors.
- Provide experimental opportunities and demonstrations for the latest High Performance Networking technology
  - the opportunity to co-locate so much technology, expertise and so many varied applications seldom comes and a role of SCinet is to demonstrate the use of this technology in novel ways.
- Support and facilitate applications that make use of high performance networks
  it is one thing to build a large network, it is another to make it usable and still another to actually use it. SCinet works to accomplish all three objectives.

Four Separate Networks

- COMMODITY
- WIRELESS
- PRODUCTION
- EXPERIMENTAL

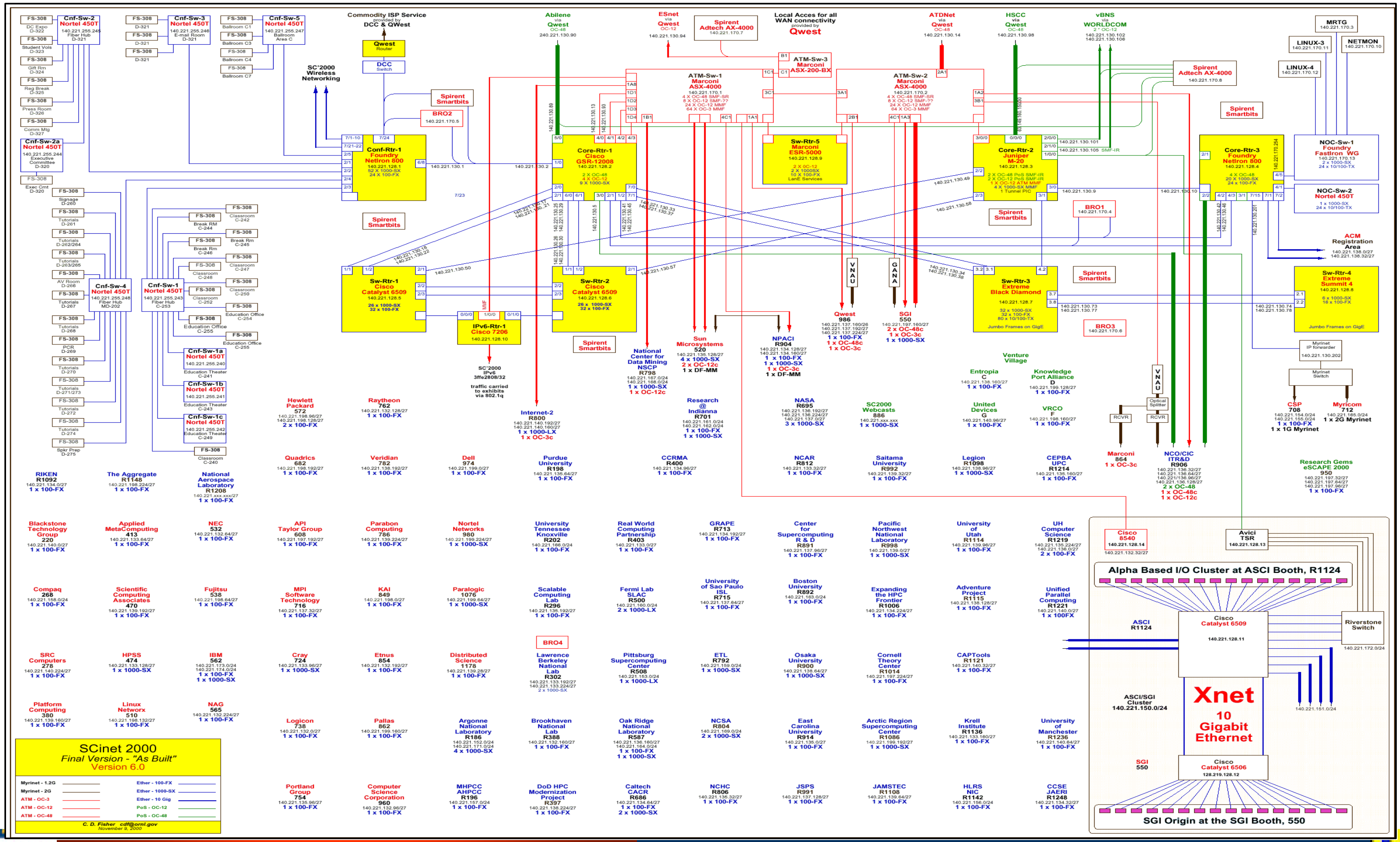
- SCinet is a full fledged ISP

Internal Connectivity

Type of Connection	Number of Connection	
	SC 99	SC 2000
OC-48c ATM	2	6
OC-48 PoS	1	5
OC-12c ATM	5	13
OC-12 PoS	0	2
OC-3c ATM	11	7
1,000 Mbps-LX	0	5
1,000 Mbps-SX	29	67
100 Mbps-FX	46	79
10 Mbps-FL	22	0

- Adding wireless, Commodity and other – 130 Gbps of total bandwidth

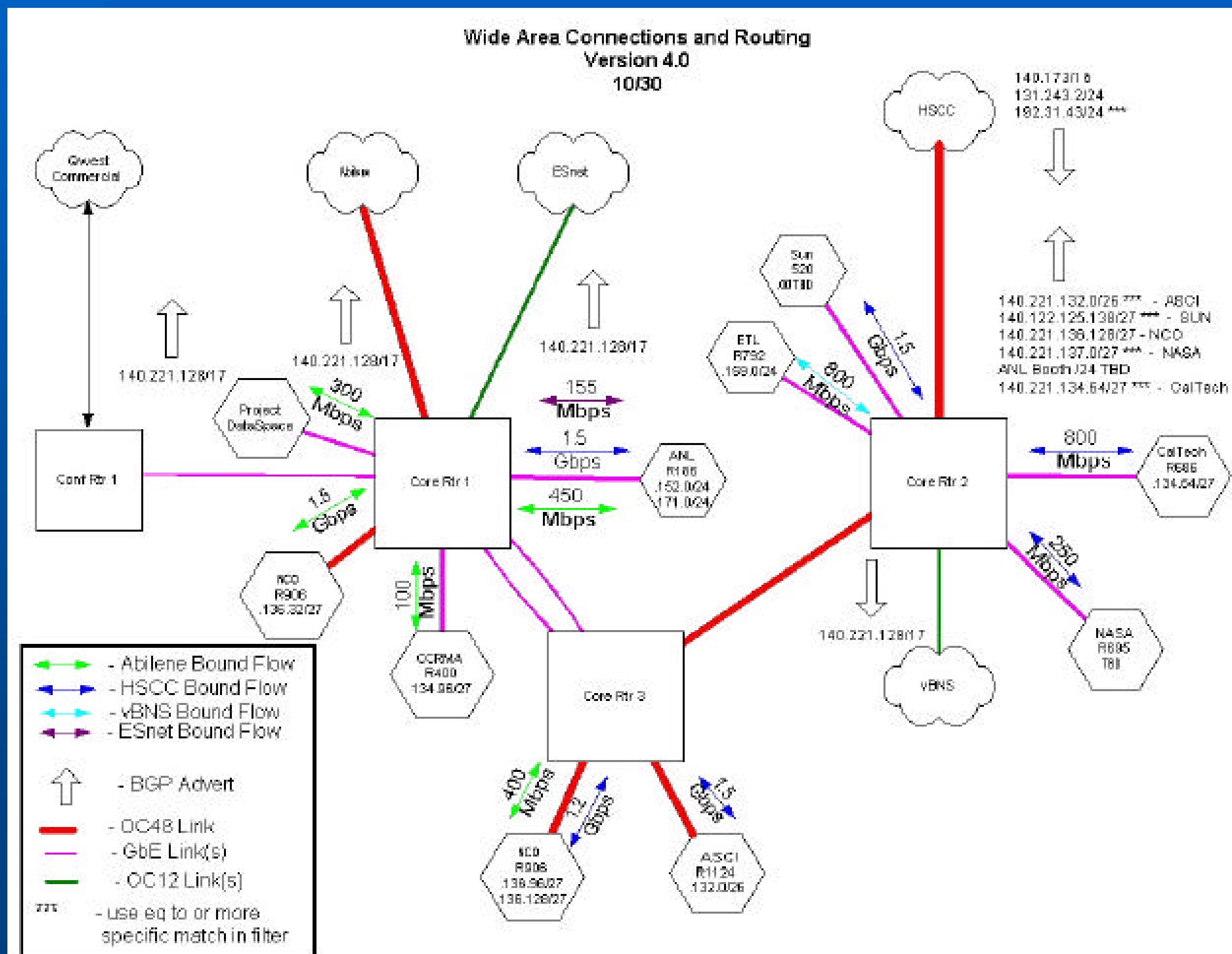
SCinet Design



External Connectivity

Network	Type	Maximum Speed
Abilene	OC-48 ATM	2.5 Gbps
ATDnet	OC-48 ATM	2.5 Gbps
HSCC	OC-48	1.5 Gbps
Esnet	OC 12 ATM	655 Mbps
vBNS	OC 12 ATM	655 Mbps
vBNS	OC 12 POS	655 Mbps
Commodity	ATM	12 Mbps
Total		8.477 Gbps

SCinet WAN Flow Diagram



SCinet Timeline

Event	Date	Time from Start
Arrival	10/30/00 9:00 AM	
First Fiber Lifted	10/30/00 4:00 PM	7.00
First Newspaper article	11/1/00 8:00 AM	
First Light to DCC	11/1/00 7:01 PM	58.02
First Power to NOC Racks	11/1/00 7:48 PM	58.80
First Light Extended to Electronics (GSR)	11/1/00 8:43 PM	59.72
Foundry NetIron on-line	11/1/00 8:56 PM	59.93
Cisco GSR, Foundry NI on-line	11/2/00 12:17 PM	75.28
Fore ASX40000 #1 on-line	11/2/00 1:32 PM	76.53
Fore ASX 4000 #2 on-line	11/2/00 1:35 PM	76.58
Juniper M20 on-line	11/2/00 3:20 PM	78.33
Extreme Black Diamond on-line	11/2/00 4:10 PM	79.17
Abilene Circuit up to Dallas POP	11/2/00 4:10 PM	79.17
Extreme Summit 4 on-line	11/2/00 4:15 PM	79.25
Marconi ESR 5000 on-line	11/2/00 4:20 PM	79.33
Cisco 7507 on-line	11/2/00 4:22 PM	79.37
SPIRENT gear (collectively) on-line	11/2/00 5:09 PM	80.15
Abilene Circuit Completed - First OC 48 for SC2000	11/2/00 5:14 PM	80.23
Abilene Peering up	11/2/00 5:25 PM	80.42
Cisco 6509 #1 on line	11/2/00 5:44 PM	80.73
Cisco 6509 # 2 on-line	11/2/00 5:45 PM	80.75
vBNS OC12 Packet over Sonnet up	11/2/00 6:40 PM	81.67
Best Power UPS on-line; All power changes complete	11/3/00 1:40 PM	100.67
HSCC OC-48 POS up	11/3/00 2:30 PM	101.50
Wireless Aps installed in Education area	11/3/00 5:00 PM	104.00
Address data purified and in the DB	11/4/00 7:15 AM	118.25
GPS Working for network monitoring	11/4/00 12:00 PM	123.00
Second vBNS OC-12 POS up, ATM Juniper, Marconi Up	11/4/00 12:45 PM	123.75
Completed Helpdesk Software	11/4/00 12:59 PM	123.98
Began accepting drop requests	11/4/00 1:00 PM	124.00
First Video Conference DCC to NSF at 30 frames a second	11/4/00 2:00 PM	125.00
ESnet up	11/5/00 1:35 PM	148.58
Bro 3 tap	11/5/00 3:00 PM	150.00
First TV report on local ABC affiliate	11/5/00 5:30 PM	152.50
260 Wireless Clients	11/6/00 11:00 AM	170.00
Bandwidth challenge 1.56 Gbps 1 second sample peak	11/7/00 10:15 PM	205.25
SCinet Production Network Shut down	11/9/00 4:00 PM	247.00
SCinet completely torn down and shipped	11/10/00 6:30 PM	273.50
Total Time from set up to tear down		11 Days, 12 Hours, 30 Seconds

What does it look like

Measurement and Testing

- Spirent Systems "Smart Bits" and Adtech technology to monitored and measured aspects of SCinet.
- The Indiana University Network Administration Suite (IUNAS) Internet-2 “Weathermap” technology for monitor wide area flows.
- Further measurement by Netflow software package
- The "Bro" package from LBNL used to monitor network traffic for intrusion.
- The SCinet team created custom software to measure other aspects of the network such as the wireless usage.
  - Spirent, Adtech and Bro used optical splitters to tap into the actual network connections at various points in the network.

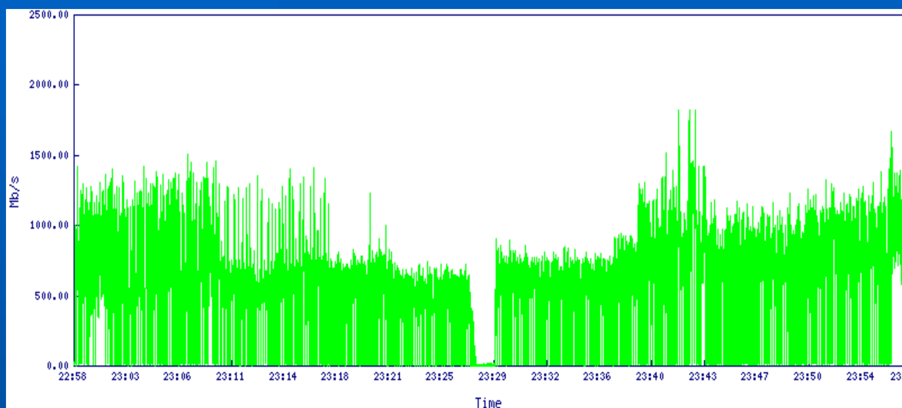
Network Bandwidth Challenge

Applications specifically challenged to "use all the network bandwidth"

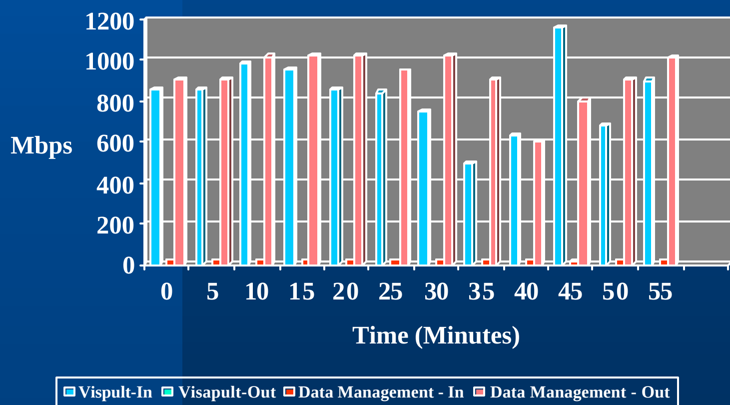
Examples:

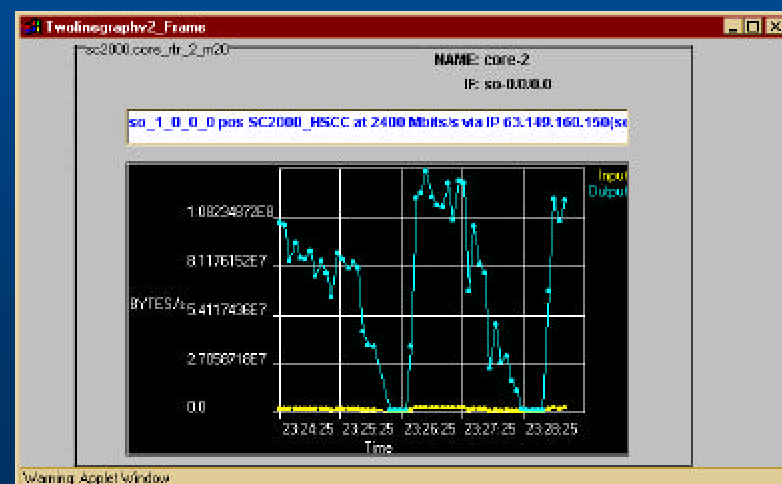
- **Visapult - Using High-Speed WANs and Network Data Caches to Enable Remote and Distributed Visualization**
- **QOS – Enabled Audio Teleportation**
- **A Data Management Infrastructure for Climate Modeling Research**
- **NRL HDTV Video Application**

Application Performance

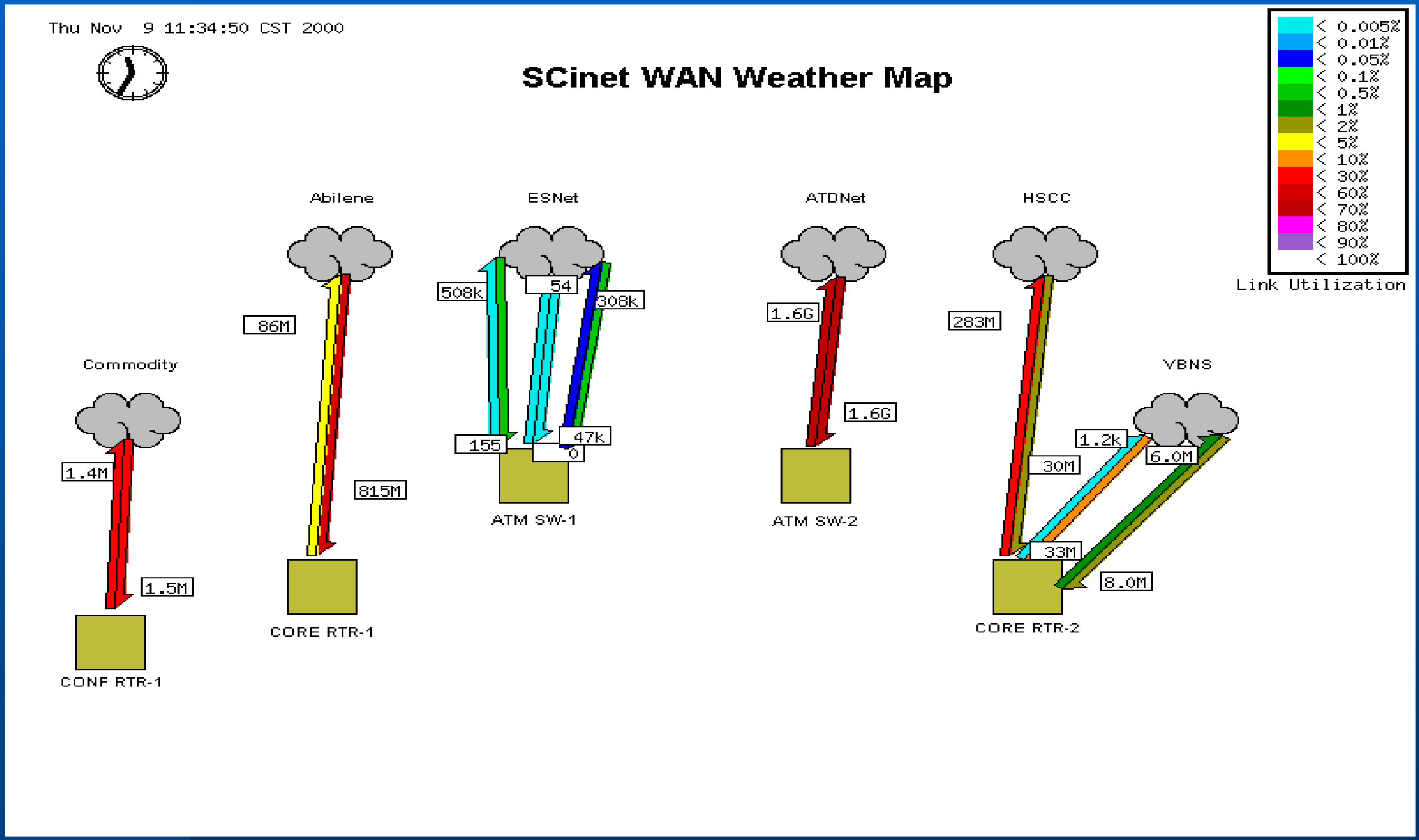


Bandwidth Demonstration of Vispault Application
Application reported 1.48 Gbps
5 second measurements reported 1.56 Gbps
.1 second measurement reported 1.72 Gbps
262 GB transferred in 1 hour (582 Mbps sustained)



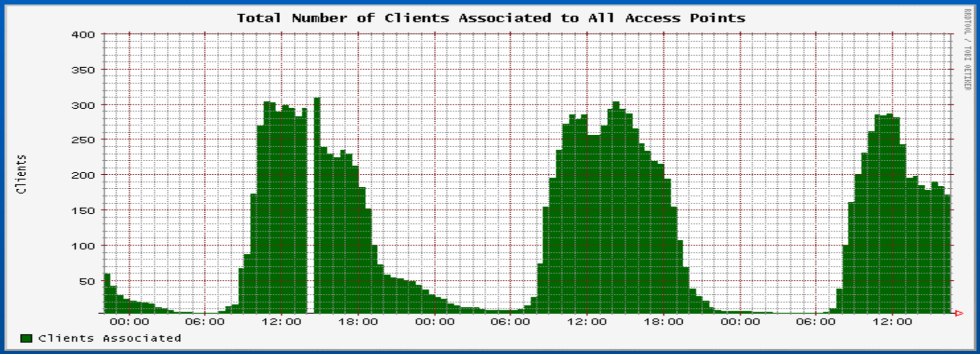


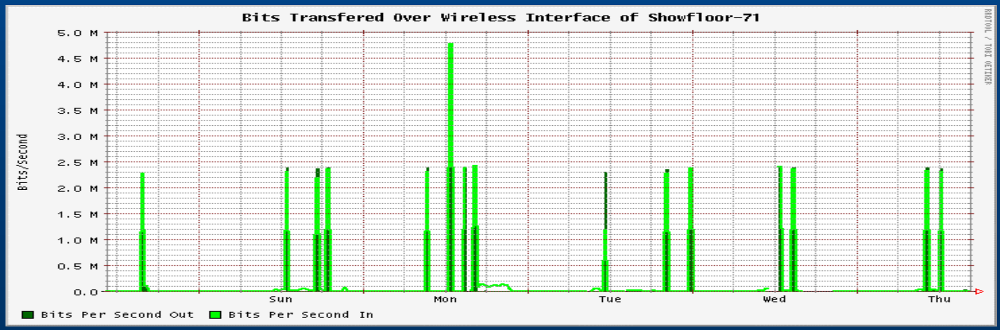
Overall WAN Usage



4.92 Gbps highwater mark

Wireless


Total Number of Clients Associated to All Access Points
Clients
400
350
300
250
200
150
100
50
00:00
06:00
12:00
18:00
00:00
06:00
12:00
18:00
00:00
06:00
12:00
Clients Associated


Bits Transfered Over Wireless Interface of Showfloor-71
Bits/Second
5.0 M
4.5 M
4.0 M
3.5 M
3.0 M
2.5 M
2.0 M
1.5 M
1.0 M
0.5 M
0.0
Sun
Mon
Tue
Wed
Thu
Bits Per Second Out
Bits Per Second In

Summary of SCinet Activity

Category	
Total Miles of Fiber	82.2 miles
Time from first lift to Outside Connectivity	59.7 hours
Miles of Fiber per hour	1.4 fiber miles per hour
Time to first OC-48 Connection (Abilene)	80.2 hours
Total Theoretical Peak External Bandwidth	8.477 Gbps – self limited, 9.477 Gbps actual interface
Estimated Theoretical Peak Show Floor Bandwidth	More Than 130 Gbps
Wireless Coverage Area	Entire Show Area - 207,338 square feet
Total effort of volunteers	Total 75 people = 11.27 people years 1.5 FTE years spent on site before anyone else even shows up
Value of volunteer efforts @ $200,000 / FTE year	$2,225,000
Estimated Value of donated equipment and services for the tested	Greater than $25,000,000

Observations

- **Two network disruptions**
  - **Tuesday Morning – the Universal Power Supply on the DNS server and commodity router (located in the bowels of DCC)**
    - **Impact was on the opening session and web casting**
  - **Routing problems Friday morning – from within the Qwest backbone due to equipment failures in Chicago**
    - **Was not due to network teardown**
  - **Both outages were at times staff had to be called in from hotels**

Observations

- **Xnet – 10 Gigabit Ethernet**
  - Point-to-point network between booths, using Cisco's pre-production 10 Gigabit Ethernet blades for their 6500 series switching routers.
  - CISCO selected the parallel interface to showcase at SC2000.
    - Less than full serialization by intercepting 4 parallel streams and running them out directly as parallel data streams on optical ribbon cable.
    - 20-CPU storage cluster (hooked to the switch through 20 separate Gig-E interfaces) feeding data through a pair of the 10-Gigabit Ethernet cards to a 20-processor compute cluster which was processing the data and rendering images
  - The issues were purely fiber related, or at best ,fiber interface related.
    - Error rate came down to a fully usable level, and the system actually delivered the promised level of performance
    - The 10 GE blades performed well, and did not require swapping or replacement.
  - The demonstration implementation used "striping" at the application layer and none of the drivers or stacks have been optimized.
  - Eight 1 Gbps streams between a pair of 10 GE ports was consistently demonstrated.
  - Due to fiber limitations it was not possible to utilize all four 10 GE boards.
    - The maximum that would have been transferred was 8 Gbps due to the number of GE feeder ports.

Observations

- Commodity
  - Provided excellent levels of service
  - Integrated with Wireless
- Wireless –
  - It is feasible to provide wireless connectivity to the entire conference
  - Interference with rogue devices is very problematic
    - Other Access Points
    - Other Wireless Devices
  - Resolving interference problems is non-trivial
  - Load balancing
    - Some level feasible with power adjustments
    - Not sufficient resolution to provide balanced service

Observations

- Production Network
  - Intentional network complexity demonstrated interesting aspects of performance and interoperability
  - BGP on all the core routers worked well
  - Unexpected issue of having HSCC and Commodity traffic both use quest backbone made things complex
  - Layer 3 switch performance under real load was interesting to observe (and discrete conclusions about vendor products determined)
  - IPv6 support needs improvement – most vendors still do not do it correctly or at all
  - Auto configuration of GE does not work in general

Observations

- Monitoring and Measurement –
  - Bro works very well and can accommodate Gigabit speeds
  - Spirent/Adtech provided excellent monitoring at interfaces
    - Network complexity required vendor development staff to make a number of modifications/improvements
    - Wide area measurement coordination feasible
  - SNMP polling is useful in a number of ways
  - Detailed monitoring data is significant and complex to evaluate

Observations

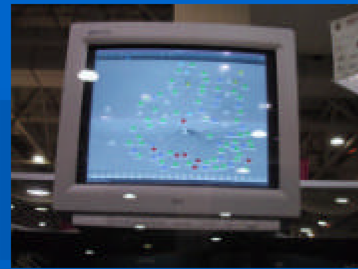
- Two major single applications ran near the limit of the individual external networks –
  - Data Man used 70% of the bandwidth they were allowed and
  - Visapult 100% used of the bandwidth they were allowed.
  - both had the limit of 1.5 Gbps since they were running over HSCC.
- High resolution of our measures saw more higher peak usage
  - .1 second samples indicated peaks that were 16% higher than the 5 second samples.

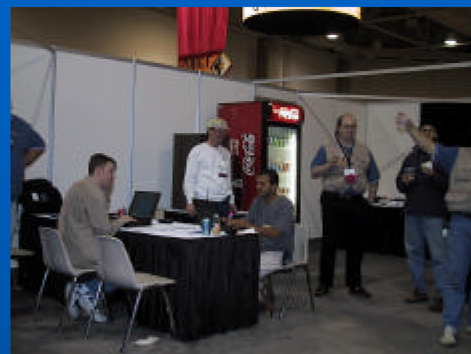
Technical Summary

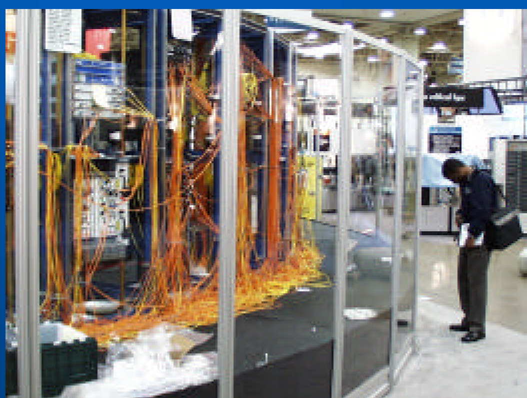
- It is feasible to build complex networks in short time-frames
  - WAN connectivity is a major factor for success
  - Planning and coordination are mandatory for success
  - Expertise is essential
- Details of Monitoring and Measurement are feasible at the 10’s of Gbp/s range
  - Application specific
  - System wide

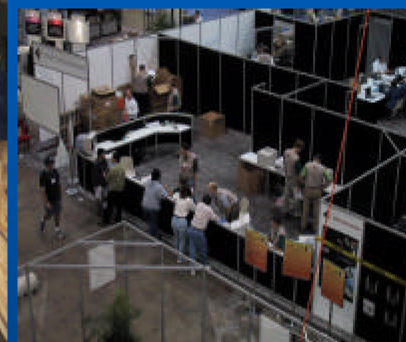
Technical Summary

- The insight gained at unique events like SC is extremely valuable
  - Provide a test bed for other network researchers
  - Provides a gathering place for many vendors to try out inter operation of advanced equipment
- Real Scientific Applications demonstrated very substantial bandwidth utilization over wide area – at least when operating from a relatively controlled environment.
- It is a lot of work to provide high levels of service in a flexible, technology aggressive environment
  - Can support 1,000 of users, 100’s of projects and 10’s of intensive demonstrations - but it is not easy



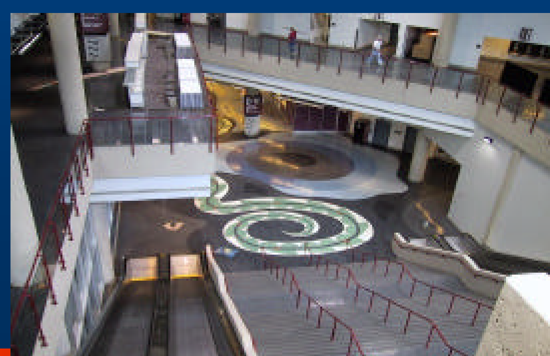







10/31/2000


10/31/2000



Observations for SC 2001

- There should be better coordination with exhibit setup
- Wireless was highly successful
  - Early deployment was very helpful to exhibitors, attendees and SC committees
  - Could be ready for full service
    - SC should fund equipment just like they do for commodity hardware
    - SCinet should coordinate all RF devices
- NOC Tours well received
  - More than 100 people signed up and attended
- NOC and Help should be staffed after Sunday party and when sessions start, not just when exhibit is open
- NOC and Help Desk should be staffed during gala
- NOC panels were highly appreciated by sponsors
- NOC power should be simplified
  - Create individual rack power distribution and pigtails

Observations for SC 2001

- Router configurations should be pre-assigned
- Wireless PDAs for drop teams and trouble teams
- Have a full job position for WAN coordination
  - Last Mile and peering
- SCinet should have redundant DHCP and DNS servers as part of its permanent inventory (e.g. not vendor loaned)
  - SC should have a permanent mail server for staff
- The separation of network function into the 4 areas worked very well
  - The basic production networks were highly effective
- The intentional complexity of the network was valuable and interesting

SCinet Badging

- **Currently there are three indicators of SCinet access, almost all of which overlap to a significant degree.**
  - **SCinet Staff sticker label – given out before the conference registration starts, it provides allows access to the NOC and allows complete access to all areas of the conference. It also allows transporting equipment in and out of areas.**
  - **SCinet registration – provided to 35 people. This is not a registration to the technical program and does not provide a proceedings not attendee gift.**
  - **SCinet ribbon – allows access to the NOC and allows complete access to all areas of the conference. It also allows transporting equipment in and out of areas.**
- **Large number of the people really do not need complete SCinet ribbon access but that is the only level available.**

SCinet Badging Improvements

- **An improved access plan would include four tiers.**
  - **SCinet Key Personnel – these are the workers year round who also spend a number of weeks at the conference site. The number should be flexible but budgeted to 40. These workers should get the registration goodies. These people would also have complete access to the entire conference, either with" blinkies" or some other indicator**
  - **SCinet Ribbon – a ribbon acknowledging significant contributions to SCinet, but a not a key personnel. The ribbons are honorary**
  - **SCinet Support engineers – these people are field engineer and vendor personnel that just need access to the NOC for installing, maintaining and deinstalling the equipment. Many of these people are involved for only one or two days, or are on call for problems.**
  - **NOC Access – many exhibitors need some access to the NOC to do network experiments, help with set up and work with SCinet staff. They do not need the complete access allowed by the traditional SCinet ribbon.**

SCinet Badging Improvements

- Purchase one of the laminated ID card systems.
  - Generate picture IDs for all SCINET staff, increasing security over the sticker system.
  - The badge backgrounds can be modified year to year, so the investment can be depreciated.
  - Consider magnetic strips on the IDs that can be used with door locks to control access to the NOC.

Observations for SC 2001

- Bandwidth Challenge was highly successful
  - Both as motivation, PR, vendor support and network demonstrations
- Improvement Suggestions
  - Contest Categories – establish different categories
    - high bandwidth challenges (throughput and peak)
    - applications that compete for more efficient bandwidth and
    - innovative use of bandwidth.
  - Planning – integrate challenge with other aspects of SCinet early
    - A/V displays and monitoring, bandwidth scheduling and time-sharing, contestant performance baseline measures.
  - Contestant Performance Baseline –
    - insure that performance baselines have been established for SCinet interconnects to high bandwidth contestants.
    - Throughput measures from the contestant's booth through the SCinet switches and routers should be verified and documented.
  - Scheduling access to time-shared bandwidth –
    - at SC2000, this almost became a full-time job.

Other Conference Components

- The rest of the conference remains much more focused toward computation than networking.
    - Only one paper session out of 23 devoted to networking (or 3 out of 63 papers)
    - No networking tutorials
    - No networking State-of-the-Field
    - Only one half of one of the nine Masterworks presentations had a networking theme
    - One of the nine panel sessions was a networking topic.
- Disturbing to network vendors and providers
    - For them to stay involved they need their customers coming to SC for the technical program as well and the computer vendors
- The steering committee should set an explicit goal of having at least 40% of all the conference activities devoted to network themes – mostly to networking over the wide area.
    - The establishment of an award program at the level of the Gordon Bell prizes is a step in the right direction.
    - Should the goal be set for SC 2001 or progressively?
- If the conference is not close to balance by SC 2004, it would be a missed opportunity.

Future Site Selection

- Conference facilities now recognize providing internet services is a money making activity
  - akin to SC coming and saying we need special food so we will cook our own.
  - By the middle of this decade, conference sites may be getting about that much revenue from ($750 per day per network IP address??) networking.
  - SC must continue to control and provide its own networking services, since clearly the growth in demand and capability far exceed anything a normal conference site can provide (despite their perceptions).
- This results in conflict that should be resolved up front
- The recommendations for the steering committee are that
  - The selection committee should include a networking expert
    - a volunteer or a paid consultant.
    - create the criteria used in evaluating sites relative to their ability to provide the networking needs of the conference in the out years and to participate in that evaluation and subsequent contract discussions.

Steering Committee Recommendations

- The selection committee should include a networking expert
  - a volunteer or a paid consultant.
  - responsibilities would be to create the criteria used in evaluating sites relative to their ability to provide the networking needs of the conference in the out years and to participate in that evaluation and subsequent contract discussions.
- Be skeptical that conference sites that have a good networking infrastructure in the conference site for standard conferences are good for SC.
  - may be the case that this is more of a hindrance than a benefit, particularly if the site is not provisioned for quick and cheap expansion.
- Consider repeating sites more often because there is good networking infrastructure in place.
  - A 10-year return cycle is of little benefit since the technology installed will clearly be out of date.
  - By returning to a site often, SC would know the infrastructure and be able to accomplish increment improvement.
  - Also there is more likelihood of consistency in the local support staff for networking .
  - Having a more frequent return arrangement might make it easier to negotiate better overall contracts

Site Contract Recommendations

- The base contract with the conference site must provide the ability for SC to do its own networking (run wires and fibers, bring in connections, control the area of the show, etc.).
  – This should negotiated done early and up front.
  – Just the same as there needs to be long term conference management expertise involved with negotiating these contracts, there should be long term network expertise involved so that the conference.
- Without these strategic steps in site selection and arrangements, it is likely the SC conference series will face an ever increase cost of doing networking on the scale needed to make the conference succeed.
  – Many of the documents and information from SC2000 could be used as basis for requirements

SCinet Team

- Networking experts were "stunned" at the complexity and capability of SCinet.
- For networking to continue to be a key SC component, the amount of network equipment, bandwidth and infrastructure loaned, given or bought does not matter. <u>The only critical resource that is irreplaceable are the volunteers of SCinet.</u>
- That critical resource must be carefully nurtured and supported

Overall Summary

- SC-2000 was a rousing success from the networking perspective.
  - Income more than $100,000 for the first time
- SCinet 2000 demonstrated conclusively
  - Very high performance, complex networks can be created and run effectively at the conference
  - Valuable applications that use very substantial amounts of the bandwidth provided.
    - Some experiments run at the show could not have been done otherwise.
  - This level of networking and usage is essential to make high performance networking an equal partner with high performance computing at SC.
- SC 2000 achieved other steps as well
  - Doubled the number of IAC members from networking companies
  - Set the foundation for what could be a very productive long-term relationship with Qwest.
  - Engaged new partners who made substantial contributions like Internet2, SBC DataCom and ATDnet.
  - Increased in participation among industry exhibitors from networking companies. In this sense SC-2000 was a rousing success from the networking perspective.

Off into the Sunset (so to speak)