Flow Data at 10 GigE and Beyond

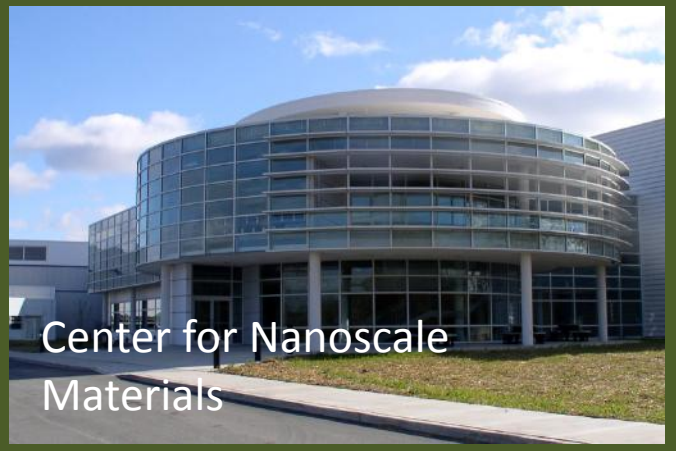
What can (or should) we do ?

Scott Pinkerton
pinkerton@anl.gov
Argonne National Laboratory
www.anl.gov

About me ….

- Involved in network design, network operation & network security for the last 10 years
- Flow data practitioner
- Campus perspective
- Our flow data uses typically include:
  - Real-time anomaly detection
  - Forensic analysis

User facilities



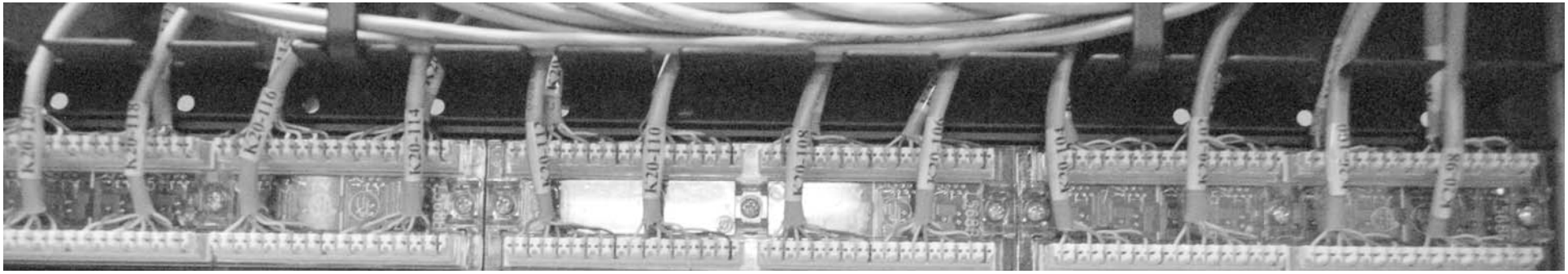
Using Flow Data in a campus environment

- In ~2000 started collecting Netflow data from all of the core campus network devices using the OSU Flowtools package
- By 2004, we were collecting Netflow data down in the distribution and access layers of the campus network
- Today, still consider flow data to be a critical part of our anomaly detection systems. Goals are to:
  - Protect the Laboratory computers from the Internet
  - Protect the Internet from the Laboratory computers
  - Have visibility into “lateral movement” of compromised hosts
- Campus environments can be large ….

Texas A&M Campus Network

- Wired Network
  - 10 Gbps backbone
  - 50,000 computers
  - 90,000 wired ports
- Gateway to regional and national networks

- Wireless Network
  - 11 million square ft. of wireless access
  - 340+ buildings with wireless access across 5200 acres



U of M Twin Cities Campus Network

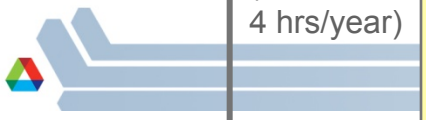
- 23 Cisco 6509s
- 4,323 Cisco 3750s
- 1,133 Switch Stacks
- 74,414 Switchports
- Redundant 10-Gigabit Backbone
- Topology: 18 layer-2 switched domains interconnected by a layer-3 MPLS-VPN backbone

A "big science" Perspective - driving speeds & feeds

- Data networks continue to evolve in support of the scientific mission
- Key drivers include:
  - Large Hadron Collider (LHC), CERN
    - CERN to US Tier1 data rates: 10 Gbps by 2007, 30-40 Gbps by 2010/11
  - Leadership Computing Facilities (LCF), ANL and ORNL
  - Relativistic Heavy Ion Collider (RHIC), BNL
  - Large-scale Fusion (ITER), France
  - Climate Science
    - Significant data set growth is likely in the next 5 years, with corresponding increase in network bandwidth requirement for data movement (current data volume is ~200TB, 1.5PB/year expected rate by 2010)

Science Network Requirements Aggregation Summary

Science Drivers / Science Areas / Facilities	End2End Reliability	Connectivity	2006 End2End Band width	2010 End2End Band width	Traffic Characteristics	Network Services
Advanced Light Source	-	• DOE sites • US Universities • Industry	1 TB/day 300 Mbps	5 TB/day 1.5 Gbps	• Bulk data • Remote control	• Guaranteed bandwidth • PKI / Grid
Bioinformatics	-	• DOE sites • US Universities	625 Mbps 12.5 Gbps in two years	250 Gbps	• Bulk data • Remote control • Point-to-multipoint	• Guaranteed bandwidth • High-speed multicast
Chemistry / Combustion	-	• DOE sites • US Universities • Industry	-	10s of Gigabits per second	• Bulk data	• Guaranteed bandwidth • PKI / Grid
Climate Science	-	• DOE sites • US Universities • International	-	5 PB per year 5 Gbps	• Bulk data • Remote control	• Guaranteed bandwidth • PKI / Grid
High Energy Physics (LHC)	99.95+% (Less than 4 hrs/year)	• US Tier1 (DOE) • US Tier2 (Universities) • International (Europe, Canada)	10 Gbps	60 to 80 Gbps (30-40 Gbps per US Tier1)	• Bulk data • Remote control	• Guaranteed bandwidth • Traffic isolation • PKI / Grid



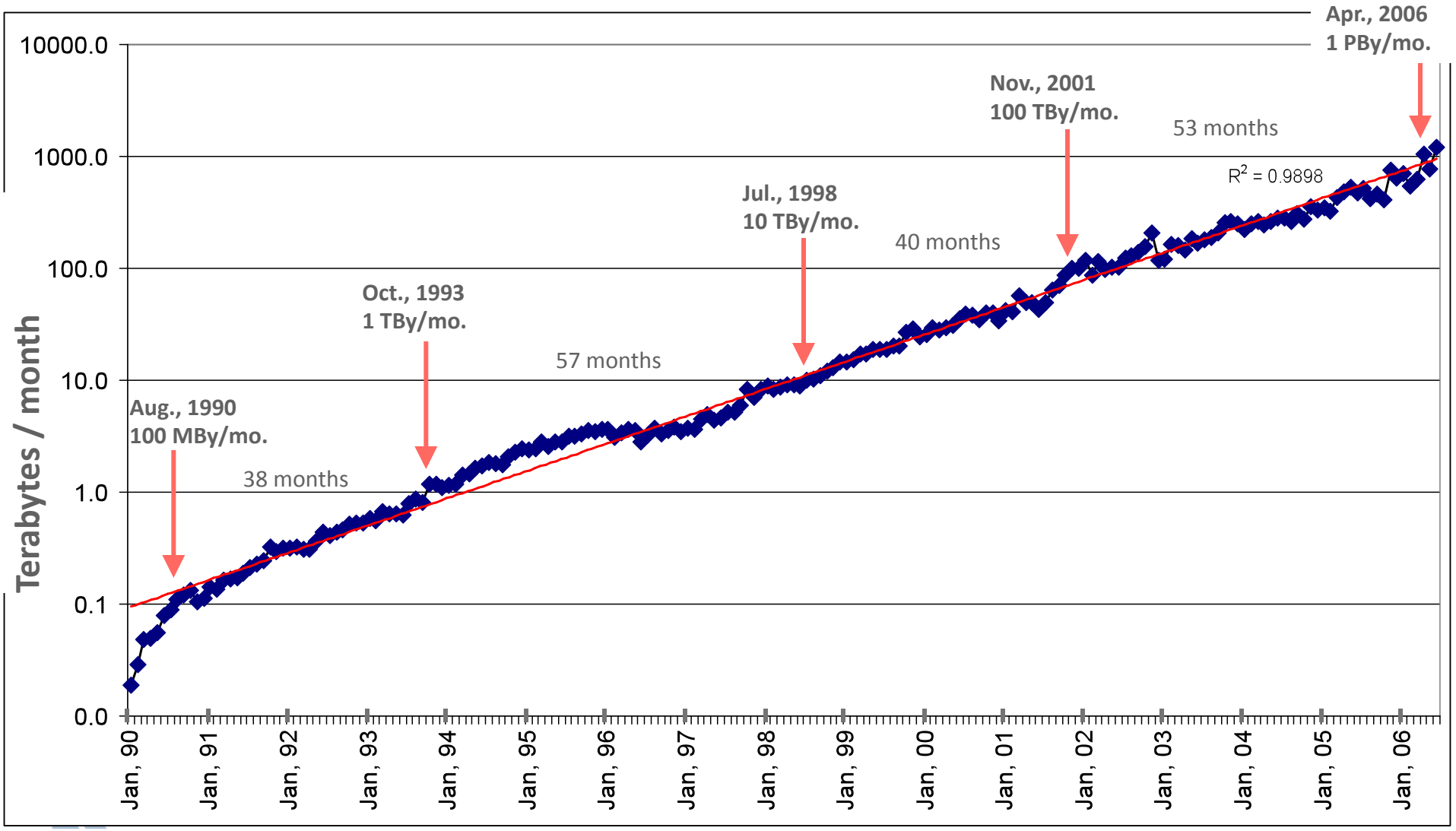
Science Network Requirements Aggregation Summary

Science Drivers / Science Areas / Facilities	End2End Reliability	Connectivity	2006 End2End Band width	2010 End2End Band width	Traffic Characteristics	Network Services
Magnetic Fusion Energy	99.999% (Impossible without full redundancy)	• DOE sites • US Universities • Industry	200+ Mbps	1 Gbps	• Bulk data • Remote control	• Guaranteed bandwidth • Guaranteed QoS • Deadline scheduling
NERSC	-	• DOE sites • US Universities • Industry • International	10 Gbps	20 to 40 Gbps	• Bulk data • Remote control	• Guaranteed bandwidth • Guaranteed QoS • Deadline Scheduling • PKI / Grid
NLCF	-	• DOE sites • US Universities • Industry • International	Backbone Band width parity	Backbone band width parity	• Bulk data	
Nuclear Physics (RHIC)	-	• DOE sites • US Universities • International	12 Gbps	70 Gbps	• Bulk data	• Guaranteed bandwidth • PKI / Grid
Spallation Neutron Source	High (24x7 operation)	• DOE sites	640 Mbps	2 Gbps	• Bulk data	

Table 1: ALCF requirements summary

Feature	Key Science Drivers		Anticipated Network Requirements	
Time Frame	**Science Instruments and Facilities**	**Process of Science**	**Local Area Network Bandwidth and Services**	**Wide Area Network Bandwidth and Services**
Near-term (0-2 years)	• ALCF production resources (intrepid)	• Large file transfers. Other labs and computing centers are common targets, but it can be any institution based on INCITE users needs. • Some real-time video, computational steering, real-time control apps possible.	• Node to node is handled by proprietary vendor interconnect. 425 MB/s per link. • Node to storage is approx. 1,000 ports of 10 gigabit. • Other local needs are primarily admin-related and are trivial.	• 10s of TB/day • 10-30 Gbps
2-5 years	• Next major machine upgrade	• Large file transfers. Other labs and computing centers are common targets, but it can be any institution based on INCITE users needs. • Real-time video, computational steering, real-time control apps more common, but still relatively small in comparison to file transfers.	• Node to node is handled by proprietary vendor interconnect. 1-5 GB/s per link. • Node to storage is likely InfiniBand-based and on the order of 3K-5K ports • Other local needs are primarily admin-related and are trivial.	• 100s of TB/day • 100-300 Gbps
5+ years	• Push towards exascale computing	• Massive data sets are common. File transfers still dominate, but WAN file systems, distributed databases use grows. • Machines are sufficiently powerful that computational steering, real time simulations are used regularly • Use of collaboration tools continues to grow	• Node to node is probably still handled by proprietary vendor interconnect, but could be standards based, such as InfiniBand. • Node to storage is likely InfiniBand or other standards-based interconnect. • Other local needs are primarily admin-related and are trivial.	• Petabytes per day • Terabit networks

ESnet Traffic has Increased by 10X Every 47 Months, on Average, Since 1990



Log Plot of ESnet Monthly Accepted Traffic, January, 1990 – June, 2006

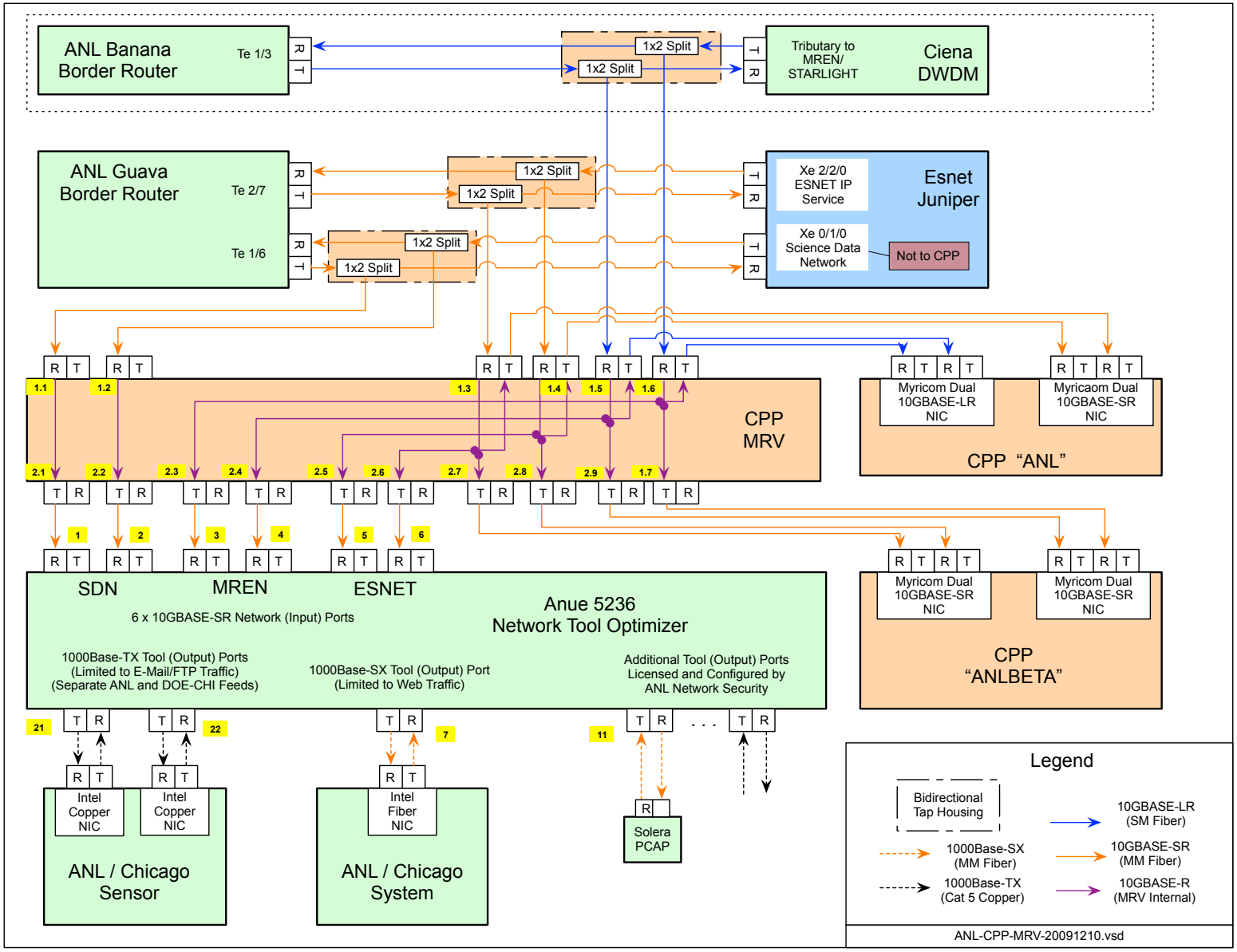
Key Take Aways

- Building networks for the future – takes a lot of planning
- Or, maybe more importantly it takes a lot of predicting (future requirements)
- Without the planning (and the predicting) how can the vendors gear up to provide the necessary capabilities ?
- Are we doing a good job communicating future requirements for flow data ?

Future of non-sampled Flow data seems bleak (IMHO)

- Speeds and feeds increasing to keep pace with scientific demand
- Many/most vendors are struggling to provide non-sampled flow data directly from the switches or routers just @ 10 Gbps (much less at 40 or 100 Gbps)
- Can optical taps really scale up to provide the needed number of monitor points ?
  - For me, I think the answer is no

Leveraging taps to create monitor points



What Can We Do - Process Perspective ?

- Identify our needs/requirements
- Write it down
- Communicate it to the vendors

What are our needs/requirements/drivers ?

- Strong support for “existence” analysis
- Scalable
  - From a campus perspective (monitoring at the border & internally)
  - From a “big science” perspective (speeds/feeds & large file txfer)
  - Equals – built in to the switches & routers (IMHO)
- Non-sampled data
  - Sampled data has its place (traffic engineering & other)
  - Painful to perform forensic analysis with sampled data

What if ideas ?

- Leveraging “cores” internal to the switches & routers for custom applications
  - Bloom filter ?
  - What info/data types would we want available to an internal app ?
- Adapting to the Very Very Large data txfers
  - Do we need a new scale for active timeout ? Was 30 seconds, now 30 minutes or 3 hours ?
- Notifications at start of a new flow – first packet ?

As a community – what can we do ?

- Should we try and develop future requirements ?
- Do we have enough energy/motivation to do it ?
- Can we agree on requirements ?
- Can we influence the networking equipment purchase decisions ?
- Your thoughts ??