



Expanding Cybersecurity and Infrastructure Beyond the Border

Deb Agarwal
DAAgarwal@lbl.gov
Lawrence Berkeley Laboratory



Outline



- Distributed Science is a Reality
- Distributed science software environment
- Infrastructure required
- Cybersecurity environment
- Issues that need to be addressed
- Research and operations can have dramatic impact when they work together
- Return on Investment-based decision making
- Conclusion



Cybersecurity and Infrastructure to Support Distributed Science



● Preserve

- Access to national user facilities
- Participation in international collaborations
- Ability to host scientific databases and repositories
- Innovation and prototyping capabilities

● Protect

- High performance computers
- Protect experiment systems
- Protect desktop and laptop systems
- Ability to do science

● ***Need to figure out how to preserve and support open science while protecting the resources from cyber incidents***



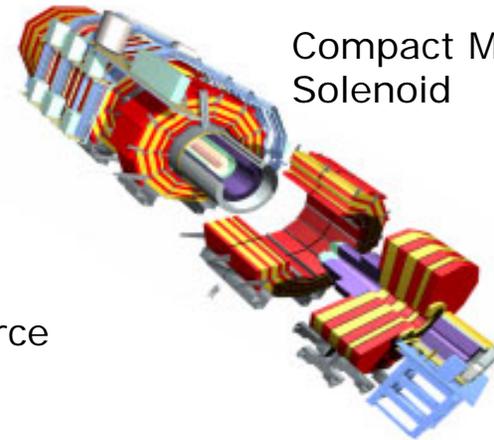
Experiments



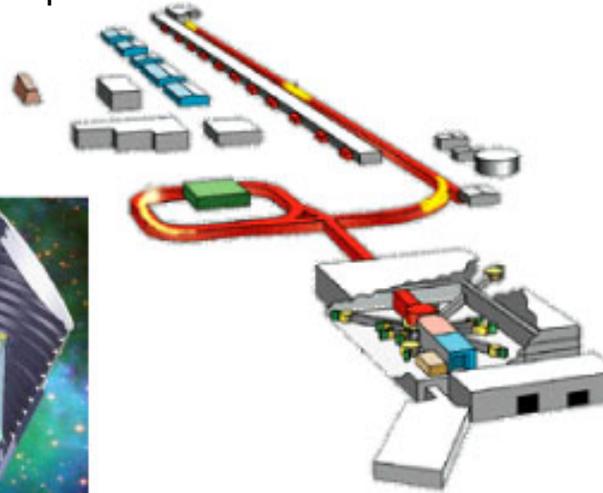
Atlas
Detector



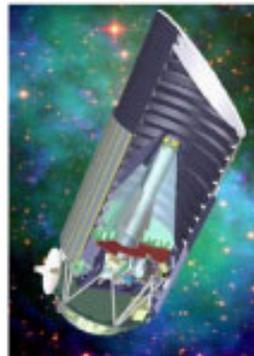
Compact Muon
Solenoid



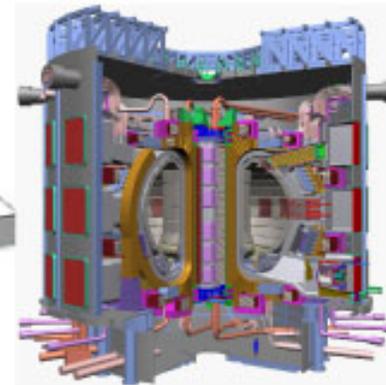
Spallation Neutron Source



Ultrahigh Voltage
Electron Microscope



Supernova/
Acceleration
Probe



ITER Tokamak



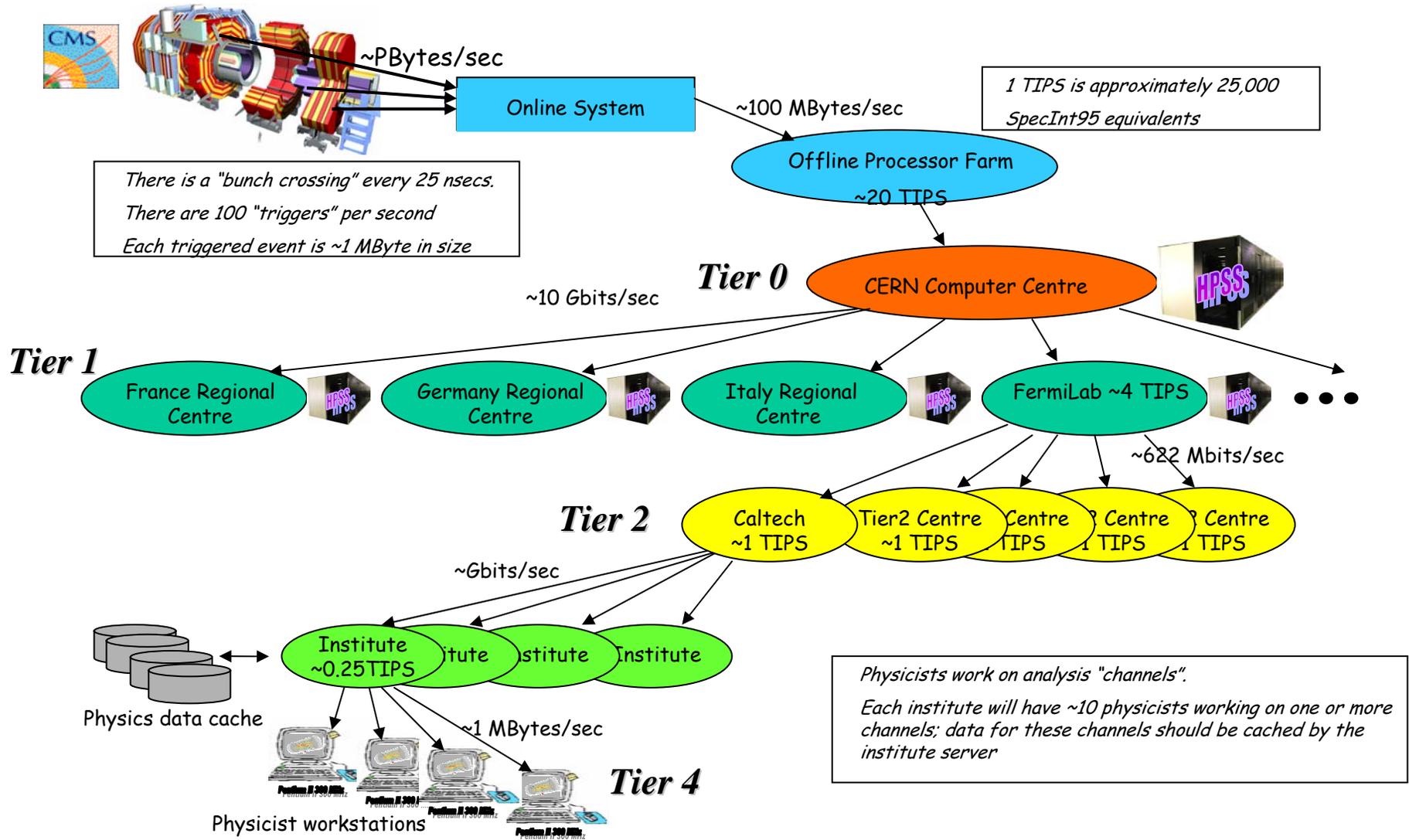
Science Requirements for Networks - 2003



Science Areas	2003 End2End Throughput	5 years End2End Throughput	5-10 Years End2End Throughput	Remarks
High Energy Physics	0.5 Gb/s	100 Gb/s	1000 Gb/s	high bulk throughput
Climate (Data & Computation)	0.5 Gb/s	160-200 Gb/s	N x 1000 Gb/s	high bulk throughput
SNS NanoScience	Not yet started	1 Gb/s	1000 Gb/s + QoS for control channel	remote control and time critical throughput
Fusion Energy	0.066 Gb/s (500 MB/s burst)	0.198 Gb/s (500MB/ 20 sec. burst)	N x 1000 Gb/s	time critical throughput
Astrophysics	0.013 Gb/s (1 TBy/week)	N*N multicast	1000 Gb/s	computational steering and collaborations
Genomics Data & Computation	0.091 Gb/s (1 TBy/day)	100s of users	1000 Gb/s + QoS for control channel	high throughput and steering



Distributed Science Infrastructure in High Energy Physics



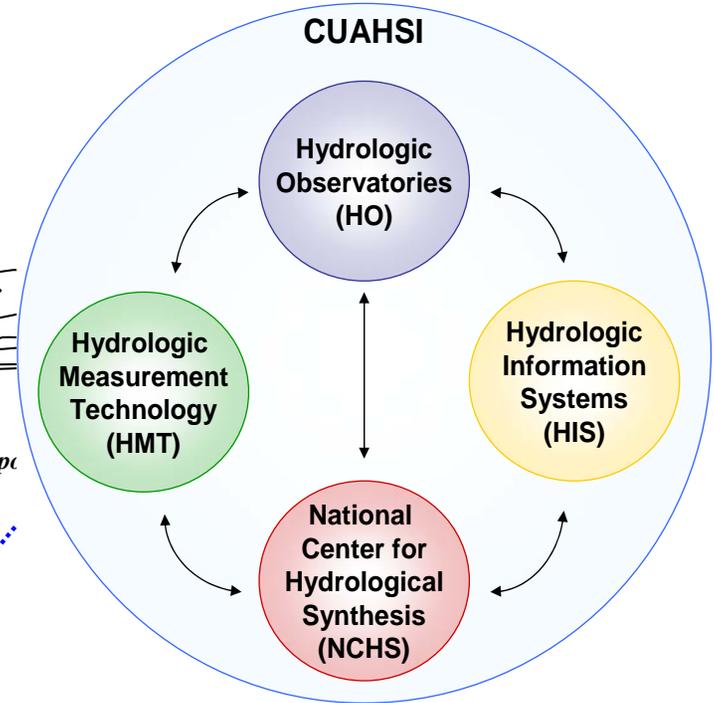
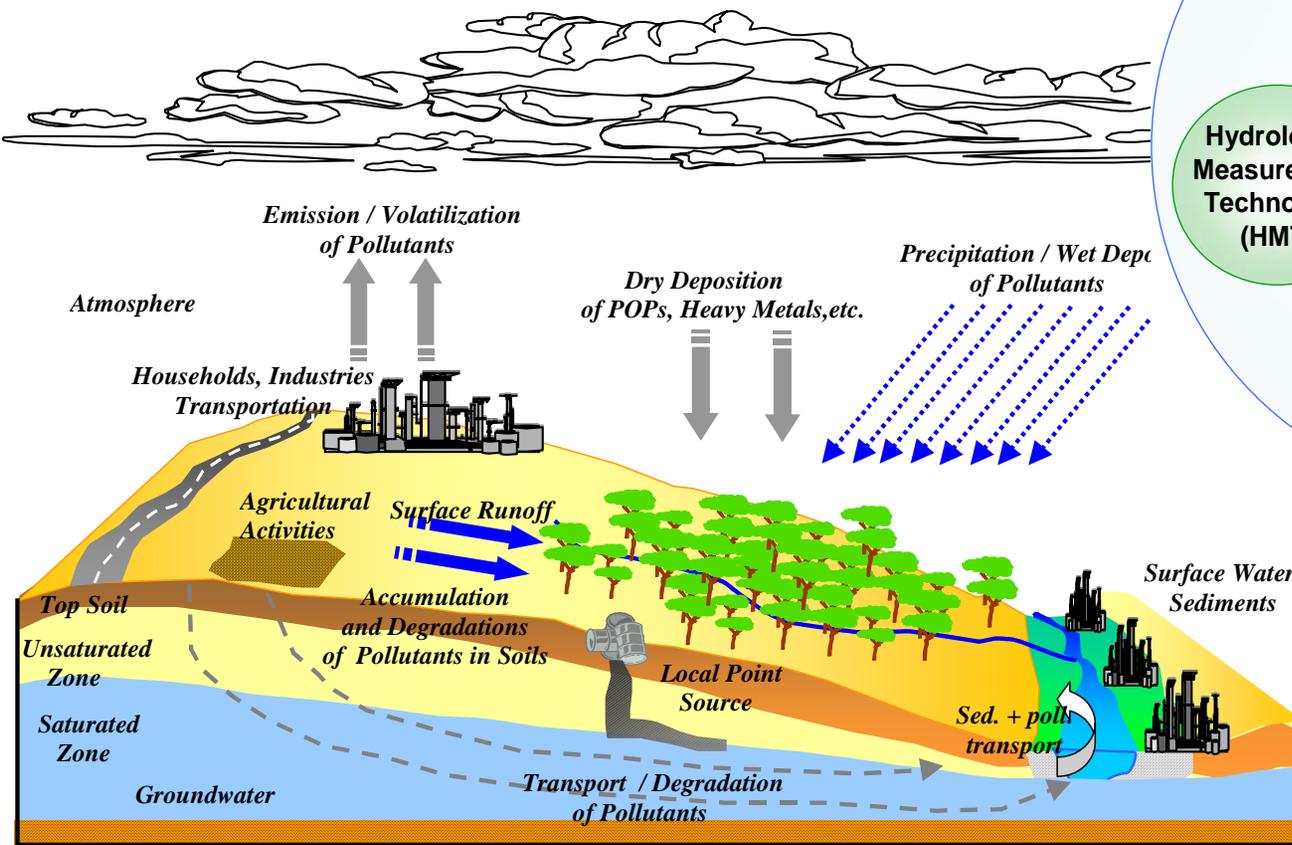
from Harvey Newman, CalTech



Hydrology Synthesis – CUAHSI/NSF



HydroView



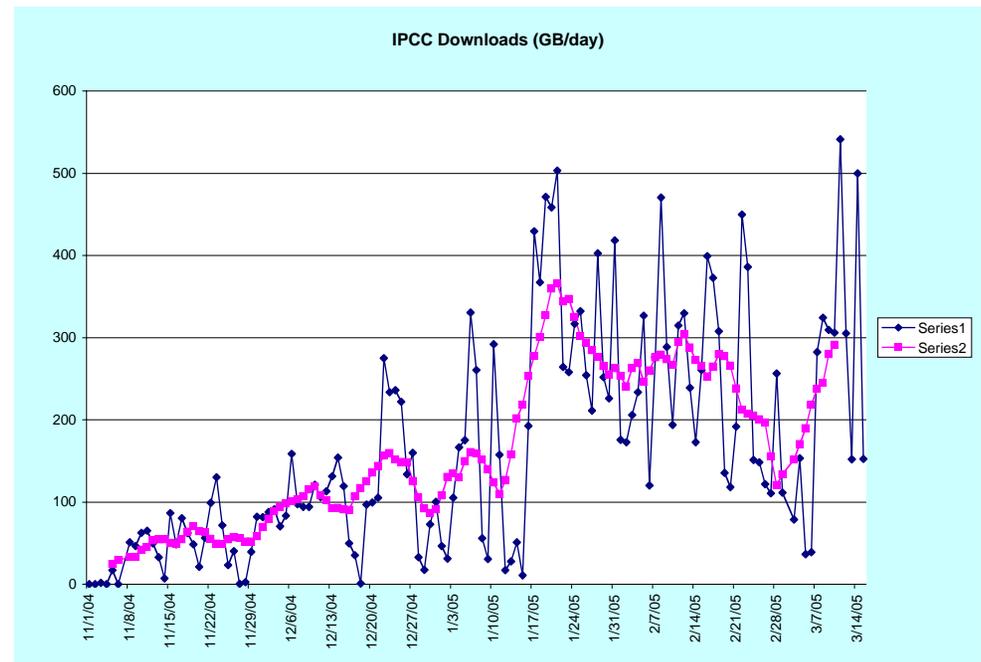


Delivering Climate Data



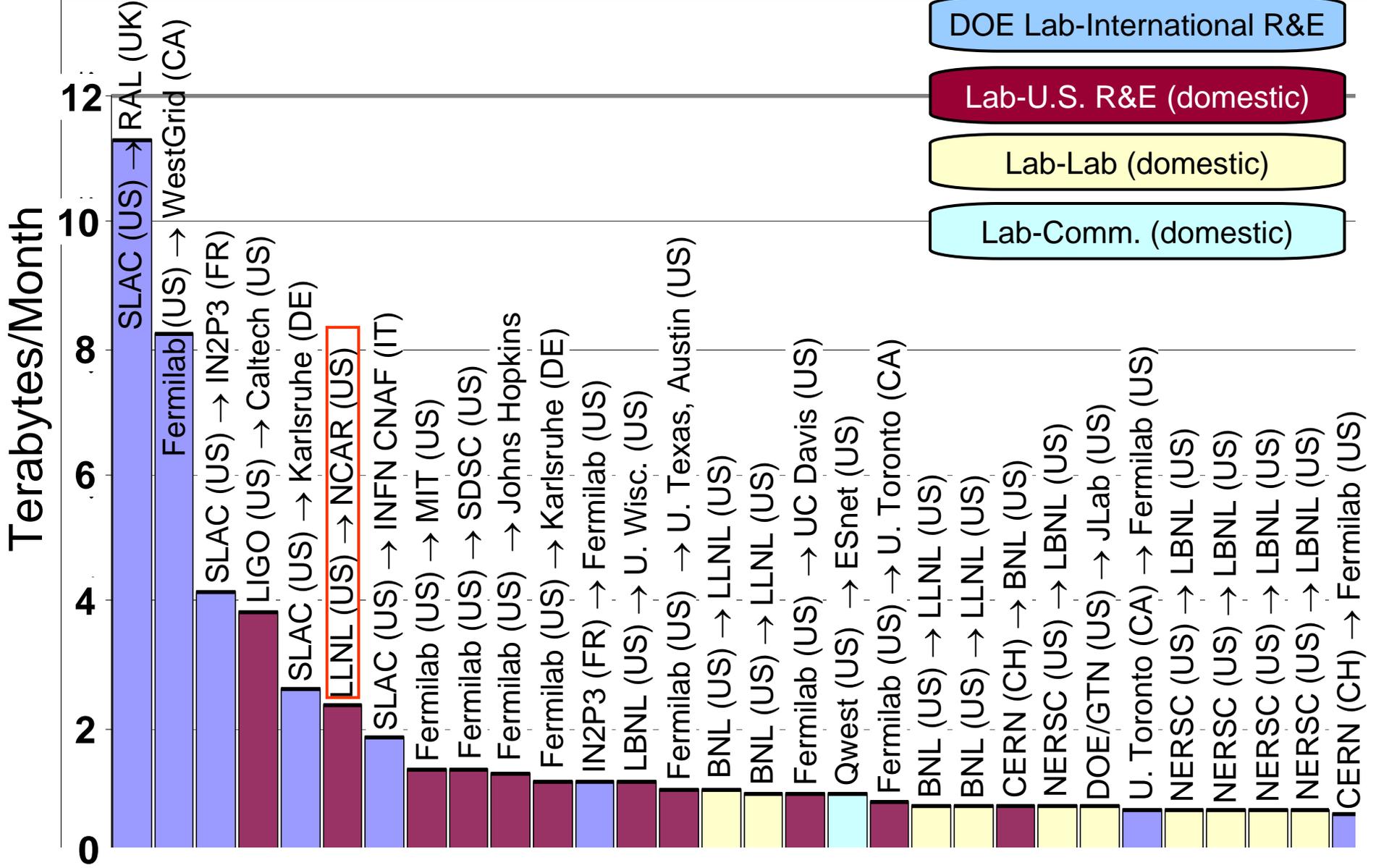
- Earth System Grid (ESG) provides production service (secure portal) to distribute data to the greater climate community.
 - Over 18 terabytes (~40k files) published since December 2004
 - About 300 projects registered to receive data
 - Over 22 terabytes of data downloaded (~125K files) with 300 gigabytes daily.
- Analysis results of IPCC data, distributed via ESG, were presented by 130 scientists at a recent workshop (March 2005).

Enabling Access to Climate Data from the Intergovernmental Panel on Climate Change



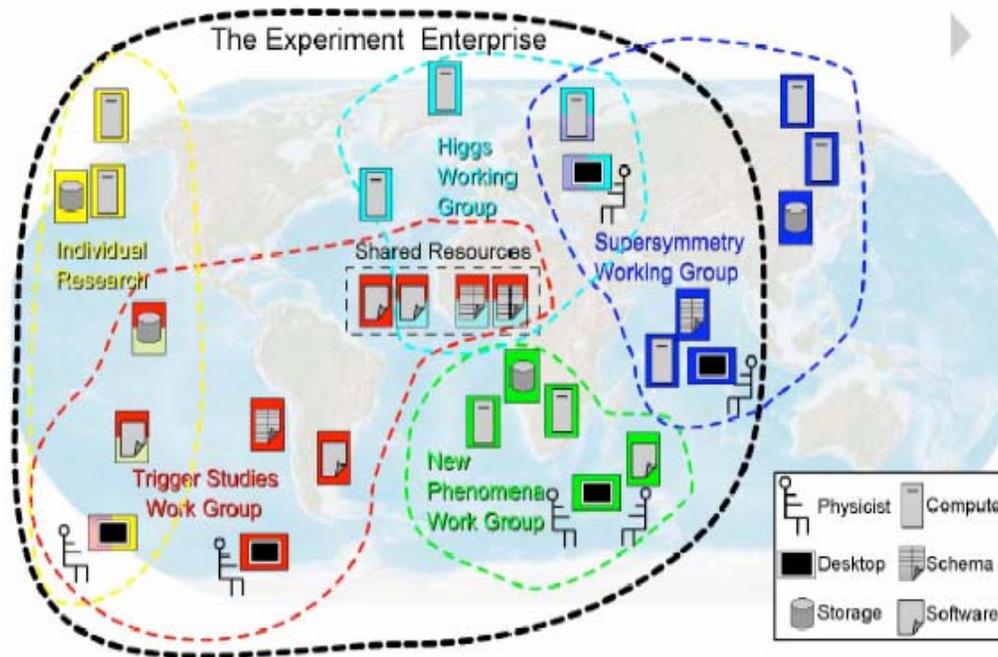
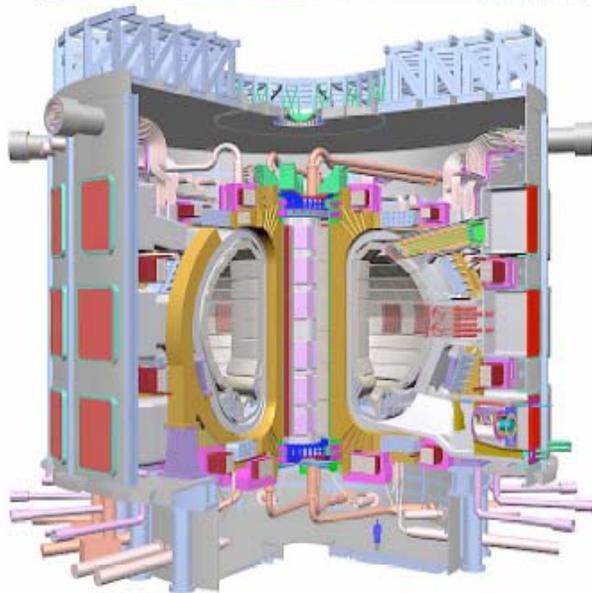
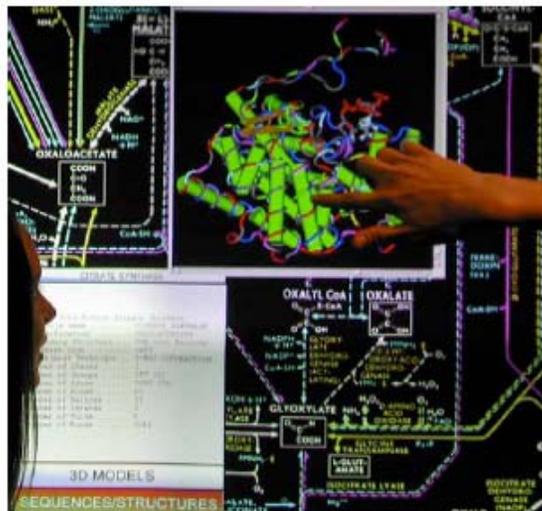


Source and Destination of the Top 30 ESnet Flows, Feb. 2005





Science Has Become a Team Sport



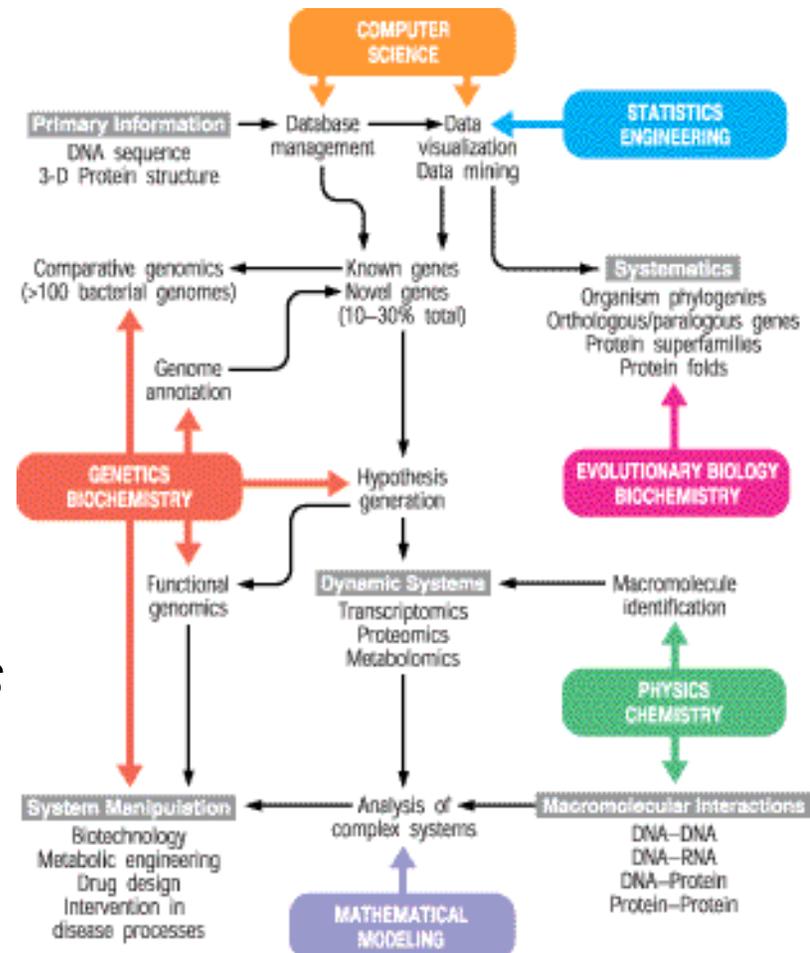
from Dave Schissel, GA



Teams Sharing Data and Expertise



Systems Biology: “studying biological systems by systematically perturbing them (biologically, genetically or chemically); monitoring the gene, protein, and informational pathway responses; integrating these data; and ultimately formulating mathematical models that describe the structure of the system and its responses to individual perturbations” (Ideker et al., 2001 Annu. Rev. Genom. Hum. Genet. 2:343)



Konopka, 2004 *ASM News* 70:163



Robust Science Support Framework



Web Services, Portals, Collaboration Tools, Problem Solving Environments

Authentication and Authorization

Resource Discovery

Secure Communication

Event Services And Monitoring

Data Transfer

Scheduling

Data Curation

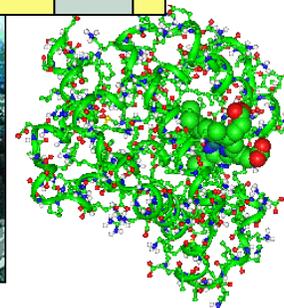
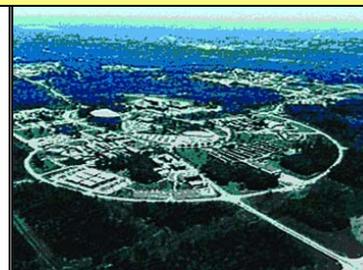
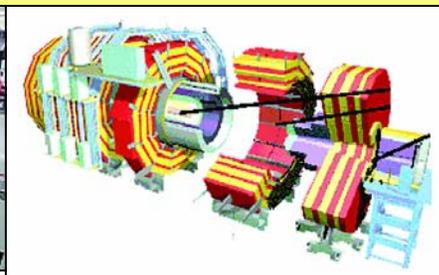
Compute Services

Application Servers

Asynchrony Support

Virtual Organization

Cybersecurity Protections





Distributed Science Reality



- Collaborations include as many as 1000's of scientists
- Collaborators located all over the world
- Many users never visit the site
- Virtual organization involved in managing the resources
 - Include multiple sites and countries
 - Distributed data storage
 - Distributed compute resources
 - Shared resources
- Do not control the computers users are accessing resources from
- High performance computing, networking, and data transfers are core capabilities needed
- Authentication, authorization, accounting, monitoring, logging, resource management, etc built into middleware
- ***These new science paradigms rely on robust secure high-performance distributed science infrastructure***



Current Research Middleware Reality wrt Cybersecurity



- Distributed Science Infrastructure is developed independent of operational cybersecurity considerations
 - Implications of site mechanisms
 - Protections from malicious code
 - Vulnerability testing
 - Interoperability with site cybersecurity mechanisms
 - Not commercial software
- Typically there is a long process of debugging prototype deployments
 - Negotiating ports and protocols with each site's cybersecurity group
 - Debugging unexpected behaviors
 - Debugging middleware security mechanisms
 - Identifying causes of performance problems
- ***This is a cross-agency and international issue***



Threats



- Viruses
- Worms
- Malicious software downloads
- Spyware
- Stolen credentials
- Insider Threat
- Denial of service
- Root kits
- Session hijacking
- Agent hijacking
- Man-in-the-middle
- Network spoofing
- Back doors
- Exploitation of buffer overflows and other software flaws
- Phishing
- Audits / Policy / Compliance
- ??????



Threats



- Viruses
- Worms
- Malicious software downloads
- Spyware
- Stolen credentials
- Insider Threat
- Denial of service
- Root kits
- Session hijacking
- Agent hijacking
- Man-in-the-middle
- Network spoofing
- Back doors
- Exploitation of buffer overflows and other software flaws
- Phishing
- Audits / Policy / Compliance
- ??????



Example - Credential Theft



- Widespread compromises
 - Over 20++ sites
 - Over 3000+ computers
 - Unknown # of accounts
 - Very similar to unresolved compromises from 2003
- Common Modus Operandi
 - Acquire legitimate username/password via keyboard sniffers and/or trojaned clients and servers
 - Log into system as legitimate user and do reconnaissance
 - Use “off the shelf” rootkits to acquire root
 - Install sniffers and compromise services, modify ssh-keys
 - Leverage data gathered to move to next system
- ***The largest compromises in recent memory (in terms of # hosts and sites)***



Cybersecurity Trend - Reactive



- Firewall everything – only allow through vetted applications with strong business need
- Users never have administrator privileges
- All software installed by administrators
- ***All systems running automated central configuration management and central protection management***
- ***Background checks for ALL government employees, contractors, and users with physical presence for issuance of HSPD-12 cards (PIV)***
- ***No access from untrusted networks***
- ***Conformance and compliance driven***
- ***It is a war***





Science is on the Front Lines



- The techniques needed to protect the open science environment today are needed by other environments tomorrow – Past examples
 - Network intrusion detection
 - Insider threat
 - Defense in depth
 - High performance network intrusion detection
- A next set of concerns
 - Reducing credential theft opportunities
 - Detection of insider attacks
 - Communication and coordination between components to recognize and react to attacks in real time
 - Tools which address day zero-1 vulnerabilities
 - Improved analysis techniques – data mining and semantic level searches
 - Prevention and detection of session hi-jacking



Current Operational Reality



- Cybersecurity group
 - Protect border
 - Protect network
 - Some host protections
 - Control access patterns
- System Administrators
 - Protect hosts
 - Authorize users
 - Define access capabilities
- Applications and software
 - Authenticate users
 - Authorize users
 - Open ports/connect to servers/transfer data
- Virtual Organizations
 - ????



Protecting High Performance Distributed Science



- Coordination between cybersecurity components
 - Border intrusion detection mechanisms
 - Network intrusion detection mechanisms
 - Host security mechanisms
 - Software authentication and authorization mechanisms
- Authentication mechanisms for users who never physically visit the site
- Analysis of data particularly in high-performance environments
- Efficient forensics information gathering
- Cybersecurity as an integral consideration in building middleware
- Proxy mechanisms
- Continuous data collection and data correlation
- Forensics collection including middleware
- Improved recovery capabilities – it is currently weeks to recover a supercomputer
- ***A new operations oriented Cybersecurity R&D effort is needed to help protect open science***



Example Advantages of Research and Operations Working Together



● Bro – network intrusion detection

- Introduced layered approach to high-speed intrusion detection
- Protocol awareness allowed detection of anomalous behavior at the protocol level
- Developed policy language and interpreter to describe policy
- Research platform for investigation of new approaches and events
- Implemented and deployed through teaming with operations
- Developments based on experience with real traffic and the operational environment
- Currently leveraging the Bro communication capabilities to add decryption of encrypted traffic streams



Example2: One-time Password

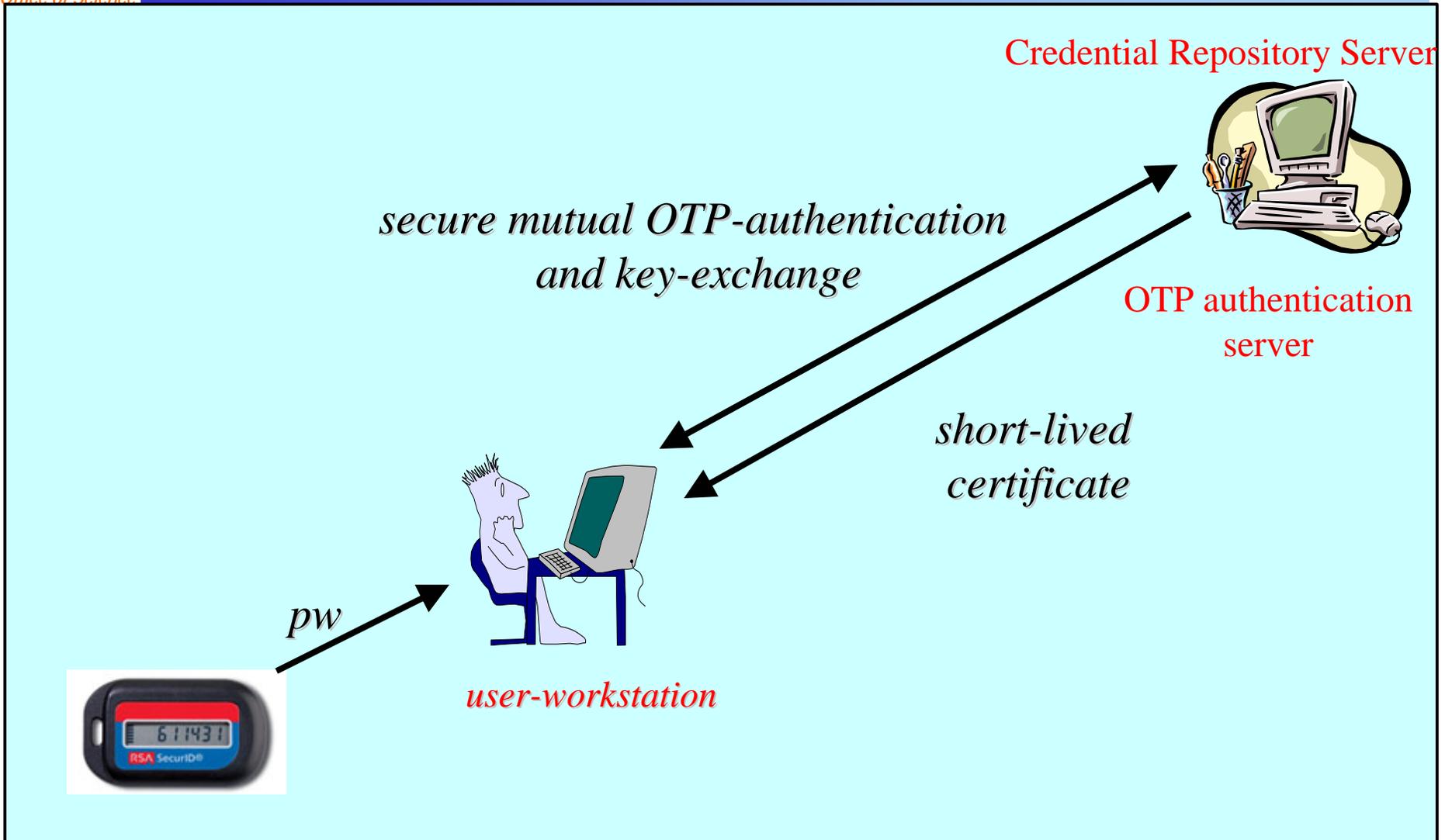


- Deploying at many sites and facilities to combat credential theft
- Many products out there on the market
- 1-factor, 2-factor, cards, software-based, etc
- Federation an important issue to reduce cost and the number of tokens a user must carry – must be secure to avoid creating cross-site propagation vectors
- Analysis from a cryptographic perspective of the various tools identified important short-comings
- Needs to be integrated with distributed science infrastructure to be fully realized





Using OPKeyX in Grid environments





Proposed Cybersecurity R&D Program



- Coordination of distributed science software infrastructure with cybersecurity mechanisms
 - Authentication, authorization, and encryption in the middleware can coordinate with the cybersecurity systems to open temporary ports etc
- Coordination between cybersecurity components
 - Significantly improve detection of attacks
 - Notify broadly of attacks as they are identified
 - Help recognize insider attacks
 - Improve handling of encrypted sessions
- Improved risk- and mission-based cybersecurity decisions
 - Research and development of methodologies for cyber assessment
- Tools for the high-performance computing environment
 - Analysis tools which can efficiently ingest and analyze large quantities of data
 - Semantic level investigation of data
 - Security tools for high bandwidth reserved paths
- Improved data collection, forensics, recovery
- ***Focus on practical solutions, integrating middleware security, and working with operations personnel during the development and testing***



ROI Model



\$\$\$

- Starts with a review of cyber incidents to determine actual damage in dollars
- Depends on the best thinking and estimates of those responsible for protecting cyber resources
- Requires the cooperation and teaming with the resource owners
- Calculates risk avoided and return on investment for protective measures



Example Cost Based Analysis



- The next few slides show an example of using a cost-based methodology to
 - determine the nominal, probable, and possible damage of different cyber incidents
 - calculate the cyber damage avoided
 - evaluate the cost effectiveness of individual protective measures

From Jim Rothfuss's Security Tutorial, LBNL



Nominal Cost Estimates

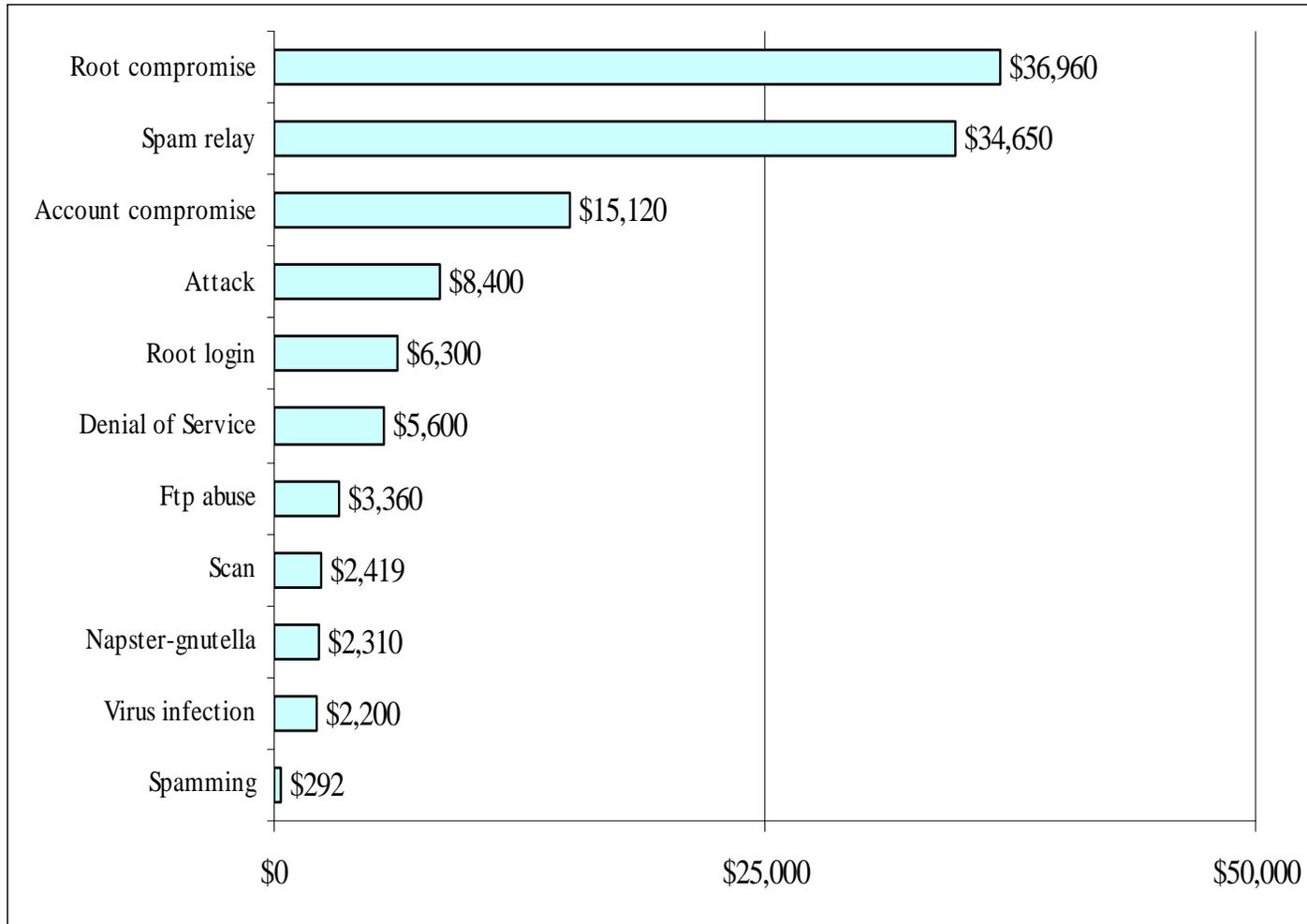


Incident type	Damage Source					Totals	
<i>Name</i>	<i>Diagnostic Effort</i>	<i>Legal Effort</i>	<i>Public Relations Effort</i>	<i>Repairs</i>	<i>Reporting Effort</i>	<i>Total Person Days</i>	<i>Nominal Damage Per Hit</i>
Account compromise	0.75	0.1	0.3	1	0.25	2.4	\$1,680
Attack	0.75				0.25	1	\$700
Denial of Service	1.75			2	0.25	4	\$2,800
File damaged or destroyed		0.01	0.01	1		1.02	\$714
Ftp abuse	0.75	0.1	0.3	1	0.25	2.4	\$1,680
Inappropriate use	0.75	1	0.3	1	0.25	3.3	\$2,310
Root compromise	1.75	0.1	0.3	2	0.25	4.4	\$3,080
Root login	0.25			0.5	0.25	1	\$700
Scan	0.003125				0.003125	0.00625	\$4
Spam relay	0.75		0.3	2	0.25	3.3	\$2,310
Spamming	0.000347			0.000347		0.000694	\$0.80
Virus/Worm				0.125		0.125	\$88

From Jim Rothfuss's Security Tutorial, LBNL



Nominal Damage From Cyber Incidents



From Jim Rothfuss's Security Tutorial, LBNL



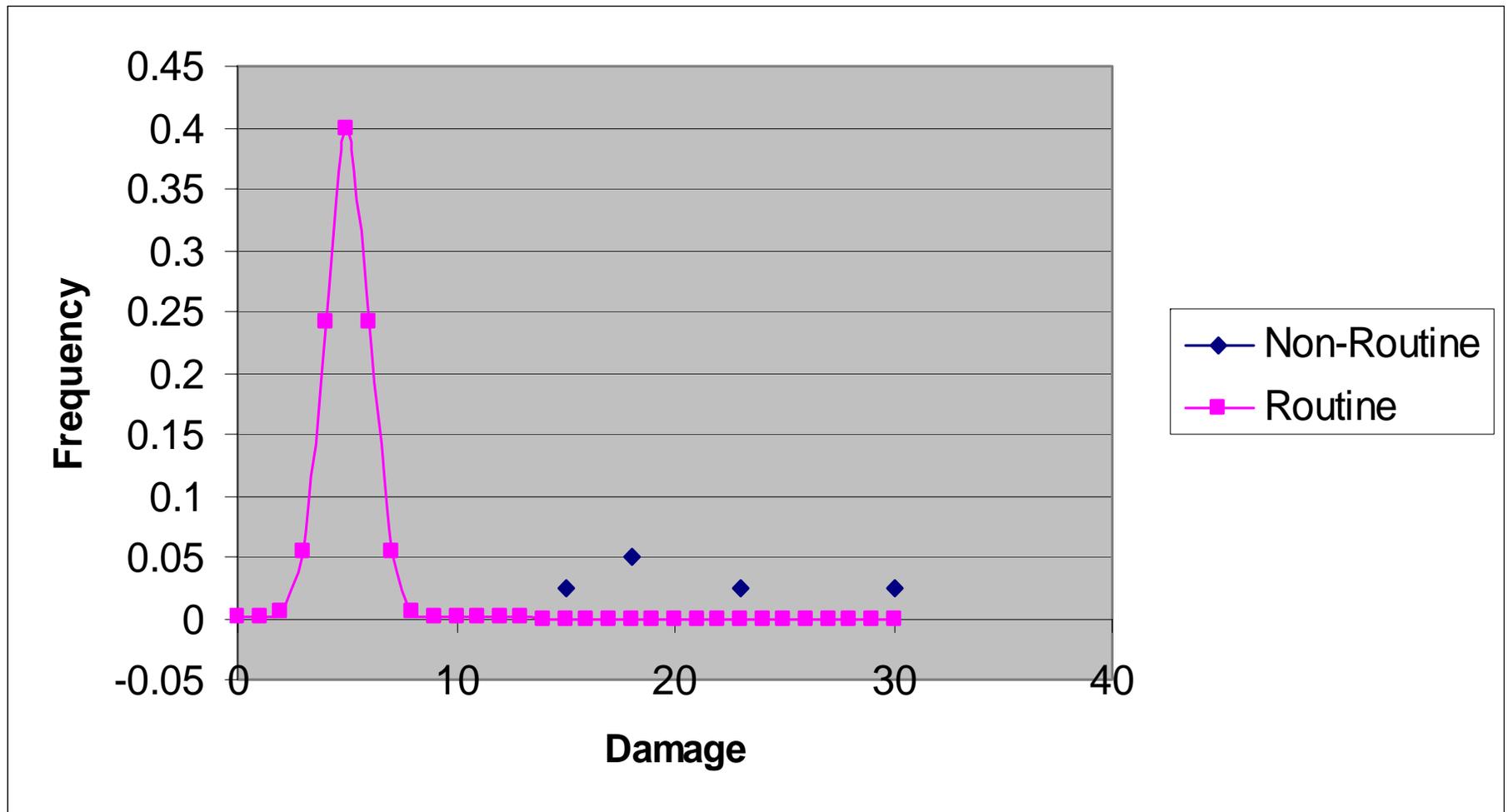
Probable Damage Associated With Incidents



- **Probable Damage** includes a factor for *non routine* incidents
 - Assumed non-routine incidents do not exceed nominal damage by more than a factor of 1000
 - Calculated using probability of incurring costs of ten, one hundred, and one thousand times nominal damage.
 - Essentially a scale factor on Nominal Damage



Non-Routine Incidents



From Jim Rothfuss's Security Tutorial, LBNL



Probable Damage Estimate



Name	Nominal Damage Per Hit	P10	P100	P1000	Probable damage per hit
Account compromise	\$1,680	0.05	0.01	0.001	\$5,778
Attack	\$700	0.05	0.01	0.001	\$2,407
Denial of Service	\$2,800	0.05	0.01	0.001	\$9,629
File damaged or destroyed	\$714	0.05	0.01	0.001	\$2,455
Ftp abuse	\$1,680	0.05	0.01	0.001	\$5,778
Inappropriate Use	\$2,310	0.05	0.01	0.001	\$7,944
Root compromise	\$3,080	0.05	0.01	0.001	\$10,592
Root login	\$700	0.05	0.01	0.001	\$2,407
Scan	\$4	0.05	0.01	0.001	\$15
Spam relay	\$2,310	0.05	0.01	0.001	\$7,944
Spamming	\$0	0.05	0.01	0.001	\$2
Virus/Worm	\$88	0.3	0.05	0	\$757

From Jim Rothfuss's Security Tutorial, LBNL



Total Possible Damage



<i>Incident Type</i>	<i>Probable damage per hit</i>	<i>Total Unblocked Attacks</i>	<i>Total Blocked Attacks</i>	<i>Probable damage per year</i>	<i>Damage avoided per year</i>	<i>Total possible damage per</i>
Account Compromise	\$5,778	9	1490	\$51,000	\$8,607,000	\$8,658,000
Attack	\$2,407	12	633	\$28,000	\$1,524,000	\$1,552,000
DOS	\$9,629	2	283	\$19,000	\$2,724,000	\$2,743,000
Ftp Abuse	\$5,778	2	327	\$11,000	\$1,888,000	\$1,899,000
Inappropriate Use	\$7,944	1	70	\$7,000	\$552,000	\$559,000
Root Compromise	\$10,592	12	1987	\$127,000	\$21,041,000	\$21,168,000
Root Login	\$2,407	9	540	\$21,000	\$1,301,000	\$1,322,000
Scan	\$15	553	28078	\$8,000	\$422,000	\$430,000
Spam Relay	\$7,944	15	2058	\$119,000	\$16,351,000	\$16,470,000
Total				\$391,000	\$54,410,000	\$54,801,000

From Jim Rothfuss's Security Tutorial, LBNL



Protective Measures with Estimated Effectiveness



<i>CounterMeasure</i>	<i>Account compromise</i>	<i>Attack</i>	<i>Denial of Service</i>	<i>Ftp abuse</i>	<i>Napster-grnutella</i>	<i>Root compromise</i>	<i>Root login</i>	<i>Scan</i>	<i>Spam relay</i>
Warning banner program					2%				
Regular password cracking	10%			10%		10%			
Firewall programs	19%	19%	19%	19%	19%	19%	19%	19%	19%
Router control lists program	19%	19%	19%	19%	19%	19%		19%	19%
Network Connection Control.	18%		18%	18%	18%	18%	18%		18%
Level 1 vulnerability scanning	50%		50%	50%		50%			50%
Intrusion detection sensors and analysis infrastructure	95%	95%	95%	95%	95%	95%	95%	95%	95%
"Crown jewels" intrusion detection program	40%	40%	40%	40%	40%	40%	40%	40%	40%
Dial in service security program	4%	4%		4%		4%	4%		
New employee orientation; System Administrator training.	9%		9%	9%	9%	9%	9%		9%
VPN infrastructure	1%					1%	1%		
Web server security requirements	3%		3%	3%		3%	3%		



Risk Avoided and Return on Investment



<i>CounterMeasure</i>	<i>Operating Cost</i>	<i>Risk Avoided</i>	<i>ROI</i>
Intrusion detection sensors and analysis infrastructure	\$140,000	\$7,522,015	5273%
Level 1 vulnerability scanning program	\$35,000	\$332,359	850%
"Crown jewels" intrusion detection program	\$7,280	\$263,930	3525%
Firewall program	\$7,000	\$92,864	1227%
Router control lists program	\$7,000	\$87,495	1150%
Network Connection Control.	\$4,200	\$79,725	1798%
New employee orientation; System Administrator training.	\$4,200	\$35,908	755%
Regular password cracking program	\$5,000	\$21,274	325%
Dial in service security program	\$46,200	\$9,528	-79%



Conclusions



- Distributed science has become core to the conduct of science
- Robust, **secure**, and supported distributed science infrastructure is needed
- Attackers are getting more malicious and quicker to exploit vulnerabilities
- Need to set the example for protecting distributed infrastructure
- COTS is a key component of the solution but will not solve many aspects of the problem
- ***Need to partner cybersecurity operations, cybersecurity researchers, system administrators, and middleware developers***