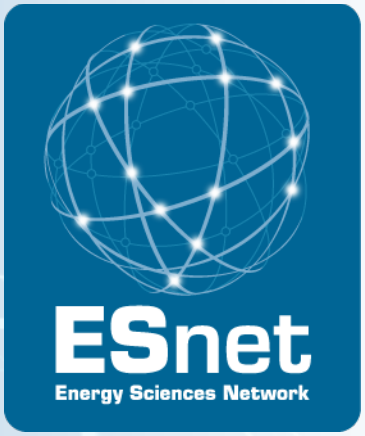


A Completely Serious Overview of Network Performance for Scientific Networking

Jason Zurawski, Science Engagement Engineer

ESnet

Focused Technical Workshop

Lawrence Berkeley National Lab, Berkeley, CA

July 18th 2013

Outline (Completely Serious)

- Endgame – Efficient networks to support science
- Use Case
- Expectations & Realities
- Problem Definition
- Solution Space
- Conclusions

Endgame = Scientific Network Use

- What we have seen so far yesterday, and today:
  - Network Infrastructure
  - Storage and Processing
  - Workflow
  - Applications
- What we are missing:
  - Preparing for the worst, e.g. sometimes it just doesn't work
  - Addressing problems (deliberate and non-deliberate)
  - "Festivus"
    - e.g. data movement/network design strategies for the rest of us – may involve feats of strength and airing of grievances

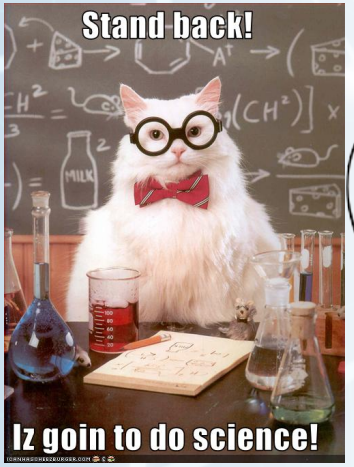
Outline

- Endgame
- Use Case – Meet our actors and action
- Expectations & Realities
- Problem Definition
- Solution Space
- Conclusions

Use Case = End to End Exchange

- Alice & Bob are collaborators
  - Experts in their field
  - Physically separated (common)
  - Rely on networks, but are not IT experts (common & expected)
  - They know their local IT staff. May also have an adversarial relationship with them (e.g. Alice and Bob are 'troublemakers' since they use the network, and expect it to work)
- Alice & Bob want to embark on a new project
  - Instrumentation @ one end, processing/analysis @ the other
  - Keep in mind they know about the science, not about the technology in the middle
  - Use infrastructure they are comfortable with, perhaps cobbled together by local support staff

Meet Alice & Bob - They are Science Cats



MOAR
DATA 4 U

INTERNETS

HAI, I CAN
HAZ DATA?

K, THX
BAI

Outline

- Endgame
- Use Case
- Expectations & Realities – What could possibly go wrong?
- Problem Definition
- Solution Space
- Conclusions

Expectations & Realities

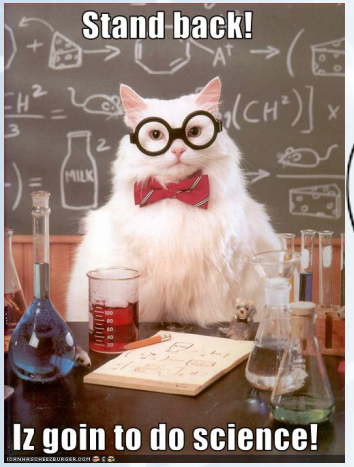
"In any large system, there is always something broken."

Jon Postel

© Dog Shaming 2012

- Modern networks are large and complicated
- Many users will encounter unforeseen (and therefore challenging) situations:
  - Upgrading networks breaks them (loss of configuration, etc.)
  - Synergy between the new and the old
  - Statistical anomalies, e.g. that 7 year old interface will stop working eventually…
- Mitigating the risk can be done in a number of ways:
  - Analysis and alteration to architecture
  - Careful thought to security/data policies in target areas
  - Integration of software designed to exercise the network, and alert/visualize

Meet Alice & Bob; Sad Reality



“The Network Is Working Fine”

- Trouble prompted a call to support (e.g. know when you are in too deep)
- The response can be frustrating (we all have been there…)
  - The engineers look at magic (and secret) monitoring tools, and tell you, firmly, that things are as they should be.
- In fairness – the network ****may**** be fine locally. e.g. We don’t see the signs of stress for the pieces we control
- In reality we need ‘end to end’ visibility, its what you experienced after all

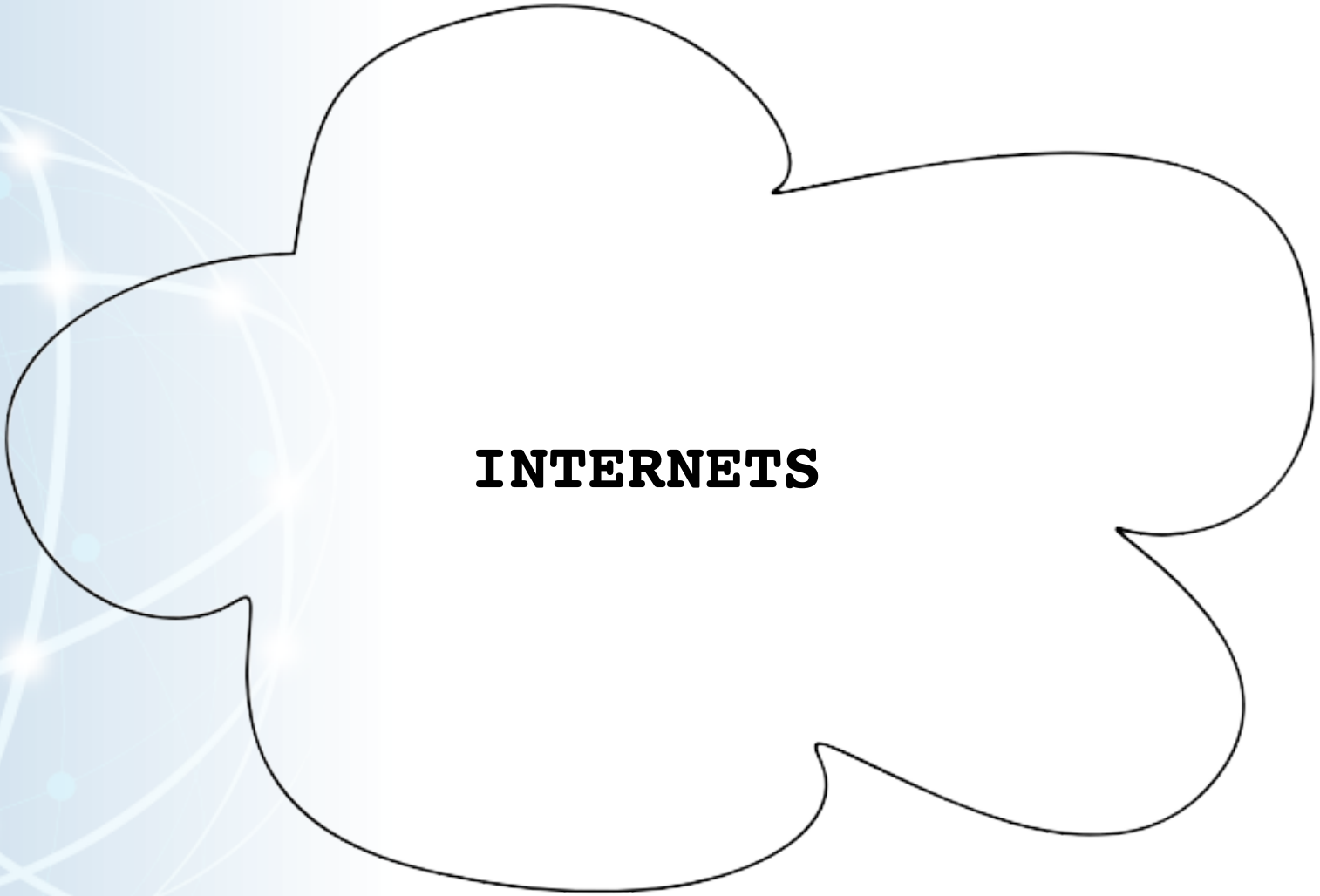
Identification of Problems

- Use Case = End to End
  - All collaboration is multi-domain, which implies you need to care about how others are running their networks as well
  - Many problems may fall between the cracks as the RTT increases (N.B. work by many, including M. Mathis)
- Two behaviors that need adjustment:
  - “Not My Network, Not My Problem”
  - Patience of Researcher

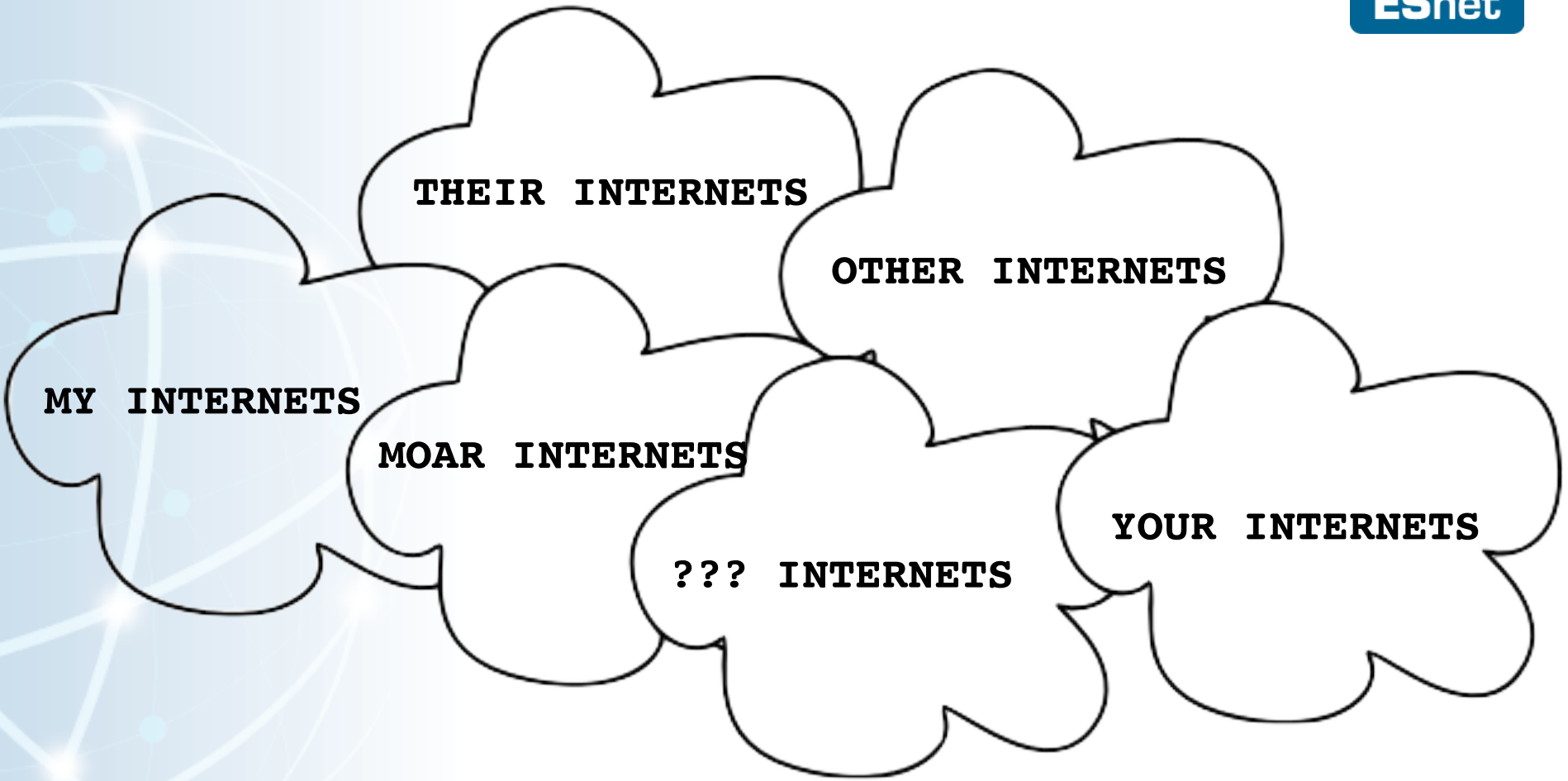
Outline

- Endgame – Efficient Use of the Network to Support Science
- Use Case
- Expectations & Realities
- Problem Definition – More than meets the eye
- Solution Space
- Conclusions

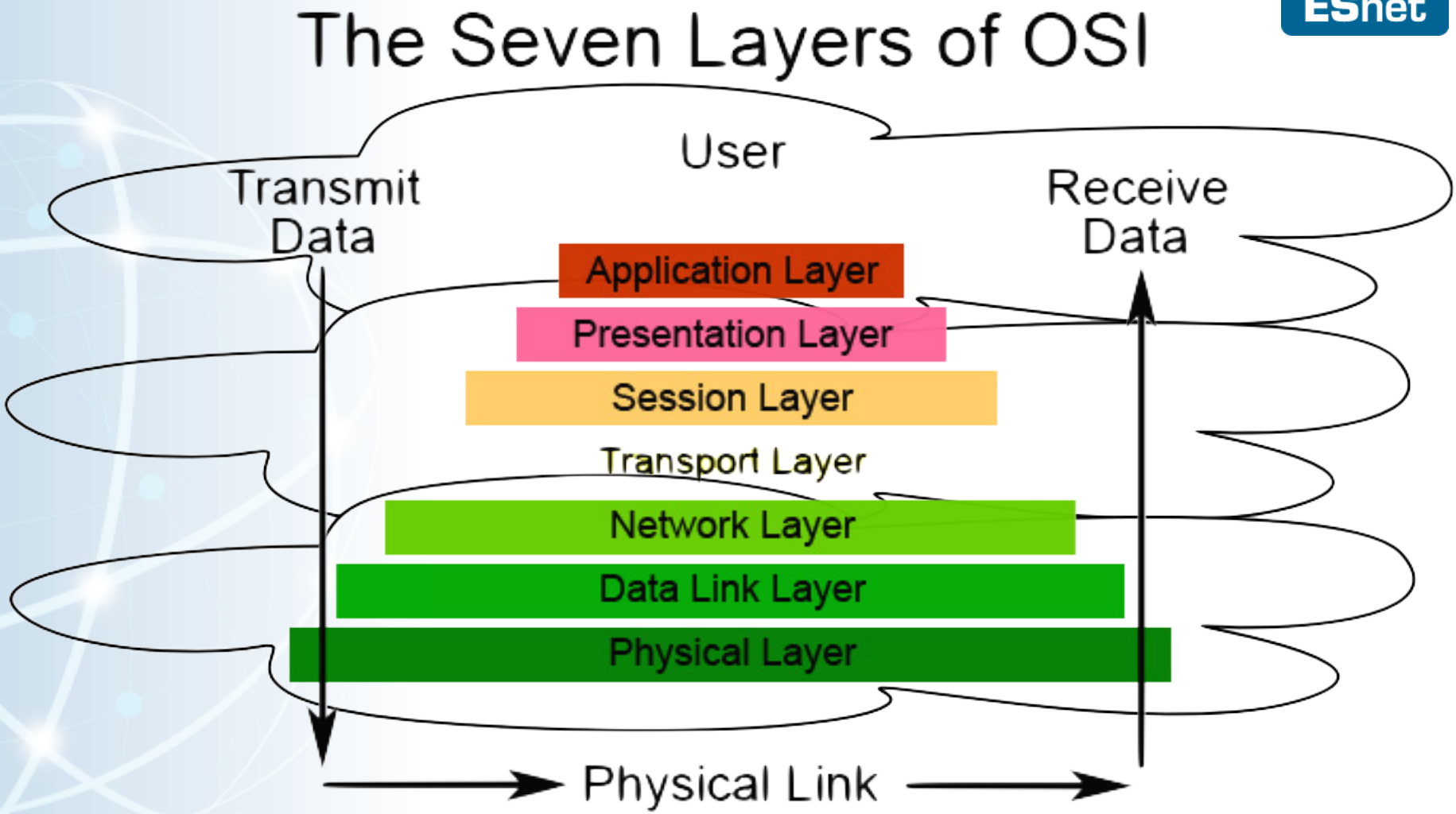
Abstraction Helps & Hurts



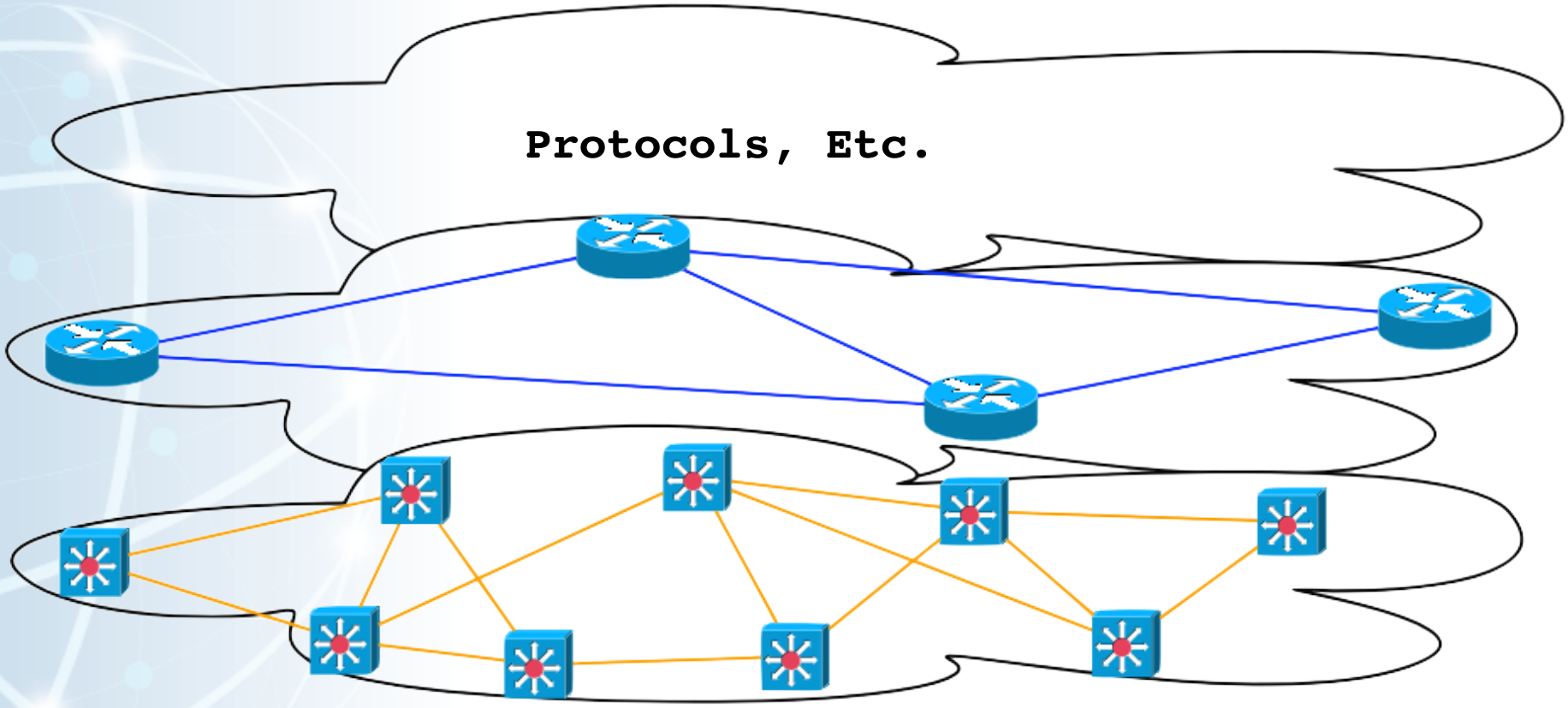
Abstraction Helps & Hurts



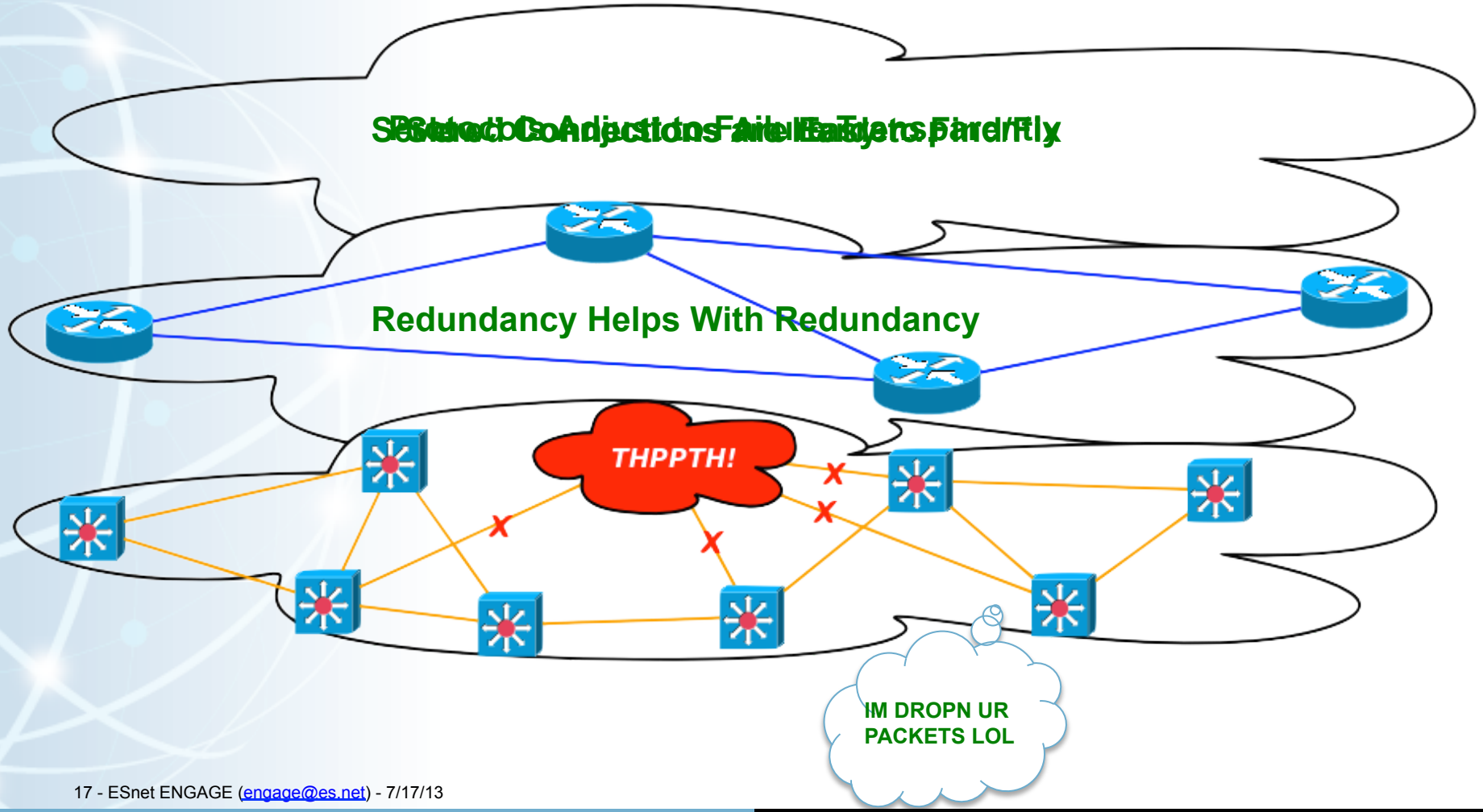
Abstraction Helps & Hurts



Abstraction Helps & Hurts



Failure in the Layers



Outline

- Endgame – Efficient Use of the Network to Support Science
- Use Case
- Expectations & Realities
- Problem Definition
- Solution Space – It's not as bleak as it seems
- Conclusions

Solution Space

- Basic idea:
  - Architectural changes
  - Solution for Monitoring/Emulation of User Behavior
  - Workflow Analysis/Adoption of New Tools (*but we have heard that part already*)
- Architecture
  - Split out enterprise concerns from data intensive ones
  - Directed security policies, instead of blanket enforcement
  - E.g. the Science DMZ
- Monitoring:
  - Dedicated resources at different vantage points in the network
  - Running some standard and useful types of measurement
  - Integrated with tools that allow you to see/hear when a problem arises

Science DMZ Overview

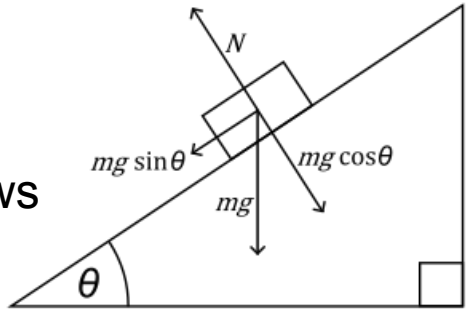
- Significant commonality in issues encountered with science collaborations … and similar solution set
  - The causes of poor data transfer performance fit into a few categories with similar solutions
    - Un-tuned/under-powered hosts
    - Packet loss issues
    - Security devices
  - A successful model has emerged – the Science DMZ
    - This model successfully in use by CMS/ATLAS, ESG, NERSC, ORNL, ALS, and others
- The Science DMZ is a ***design pattern*** for network design.
  - Not all implementations look the same, but share common features
  - Some choices don't make sense for everyone, caveat emptor

The Science DMZ in 1 Slide

Consists of **three key components**, all required:

"Friction free" network path

- Highly capable network devices (wire-speed, deep queues)
- Virtual circuit connectivity option
- Security policy and enforcement specific to science workflows
- Located at or near site perimeter if possible



Dedicated, high-performance Data Transfer Nodes (DTNs)

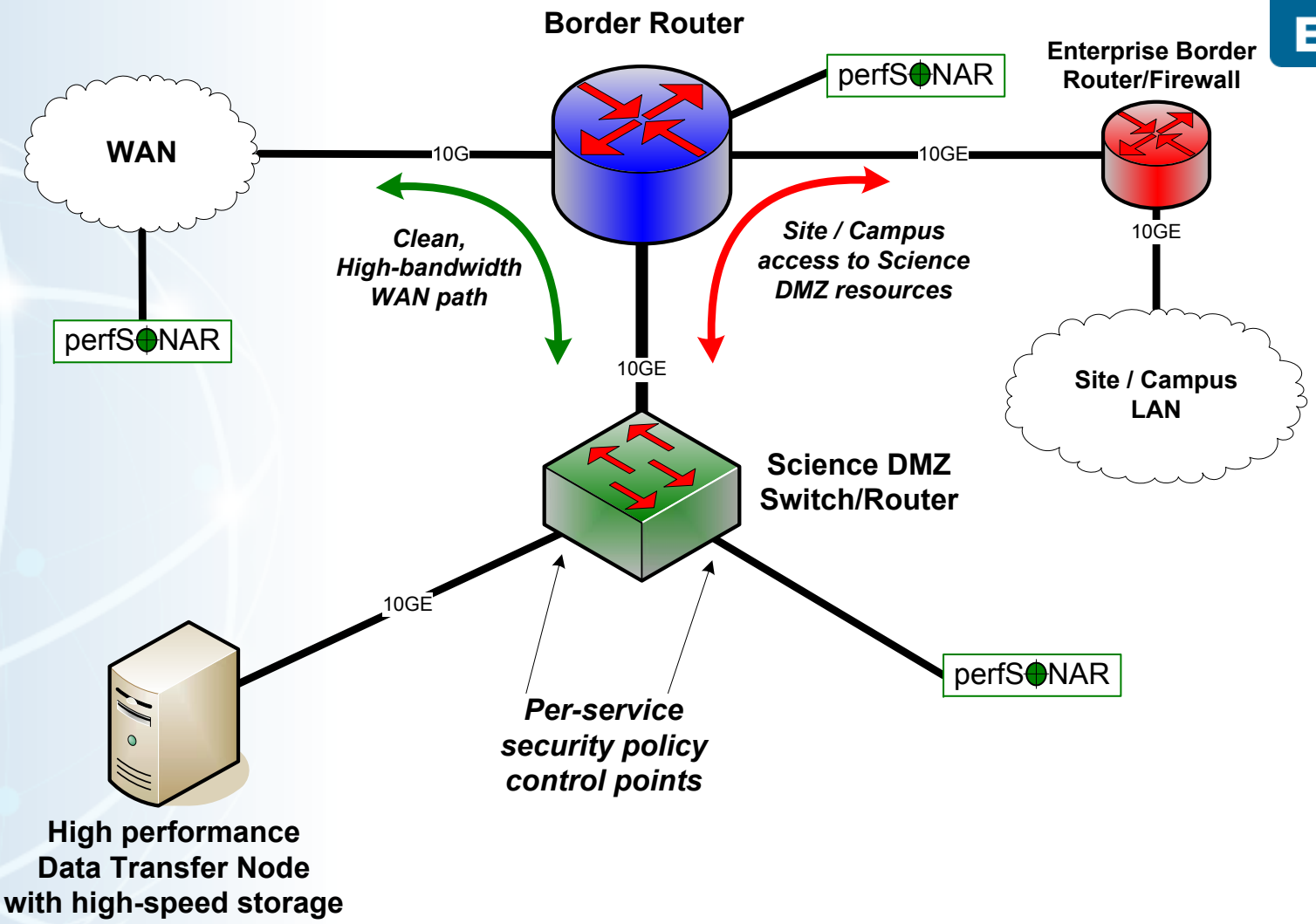
- Hardware, operating system, libraries all optimized for transfer
- Includes optimized data transfer tools such as Globus Online and GridFTP

Performance measurement/test node

- perfSONAR

Details at http://fasterdata.es.net/science-dmz/

Science DMZ – Simple Abstract Cartoon



Perf-what?

Network Monitoring

- E.g. everyone has some form on their network. Addresses needs of local staff for understanding state of the network
  - Would this information be useful to external users?
  - Can these tools function on a multi-domain basis?
- Beyond passive methods, there are active tools.
  - E.g. often we want a 'throughput' number. Can we automate that idea?
  - Wouldn't it be nice to get some sort of plot of performance over the course of a day? Week? Year? Multiple endpoints?

perfSONAR = Measurement Middleware

What is perfSONAR?

http://psps.perfsonar.net

perfSONAR is a tool to:

- Set network performance expectations
- Find network problems (“soft failures”)
- Help fix these problems

All in multi-domain environments

- These problems are all harder when multiple networks are involved

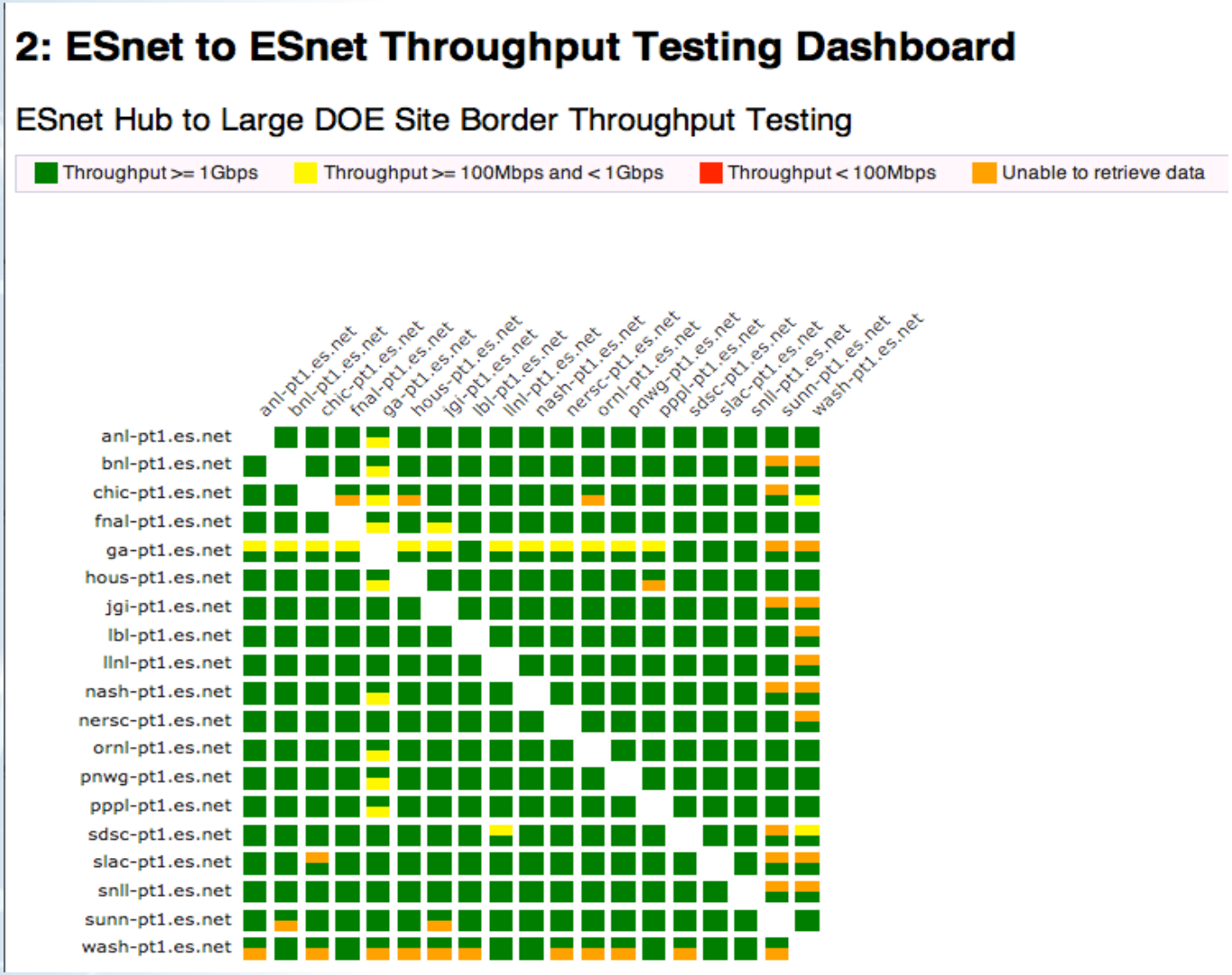
perfSONAR is provides a standard way to publish active and passive monitoring data

- This data is interesting to network researchers as well as network operators

World-Wide perfSONAR-PS Deployments: 572 bwctl nodes, 552 owamp nodes as of Jun '13

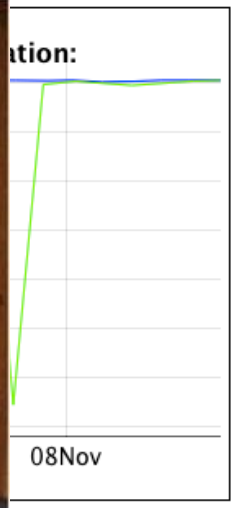
Visualizations & Alarms – Automation is the key

http://ps-dashboard.es.net



One motivation for Science DMZ & perfSONAR models: Soft Network Failures

- Soft failures are where basic connectivity functions, but high performance is not possible.
- TCP was intentionally designed to hide all transmission errors from the user:
  - "As long as the TCPs continue to function properly and the internet system does not become completely partitioned, no transmission errors will affect the users." (From IEN 129, RFC 716)
- Some soft failures only affect high bandwidth long RTT flows.
- Hard failures are easy to detect & fix
  - Soft failures can lie hidden for years!
- One network problem can often mask others



Graph Key

Src–Dst throughput

Dst–Src throughput

ESnet

© 2013 Clever Noob

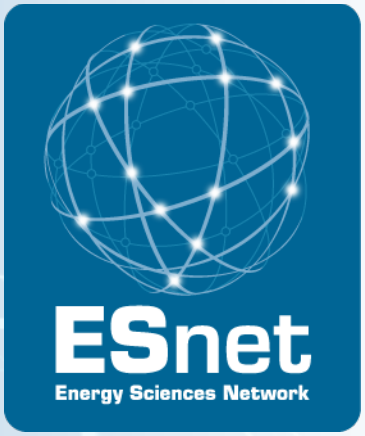
© 2013 REDBUBBLE

Outline

- Endgame – Efficient Use of the Network to Support Science
- Use Case
- Expectations & Realities
- Problem Definition
- Solution Space
- Conclusions

Conclusions

- Goals:
  - Efficient end to end use of the network, no matter the use case (bulk data movement to video)
  - 'Happy' users & operators
- Problems:
  - Network design may need work, sometimes things just break
  - How to find/fix the problems that are out there
- Solutions
  - Architectural changes
  - Adoption of new solutions
- Continued Conversation:
  - perfSONAR:
    - http://psps.perfsonar.net
    - perfsonar-node-users@internet2.edu
    - https://lists.internet2.edu/sympa/info/performance-node-users
  - Science DMZ:
    - http://fasterdata.es.net
    - sciencedmz@es.net
    - https://gab.es.net/mailman/listinfo/sciencedmz



A Completely Serious Overview of Network Performance for Scientific Networking

Questions?

Jason Zurawski – zurawski@es.net

http://www.es.net/

http://fasterdata.es.net/