ESnet

ENERGY SCIENCES NETWORK

The Science DMZ

With apologies to...

Eli Dart, Network Engineer

ESnet Science Engagement

Lawrence Berkeley National Laboratory

Building the Modern Research Data Portal

Yale University

October 12, 2016

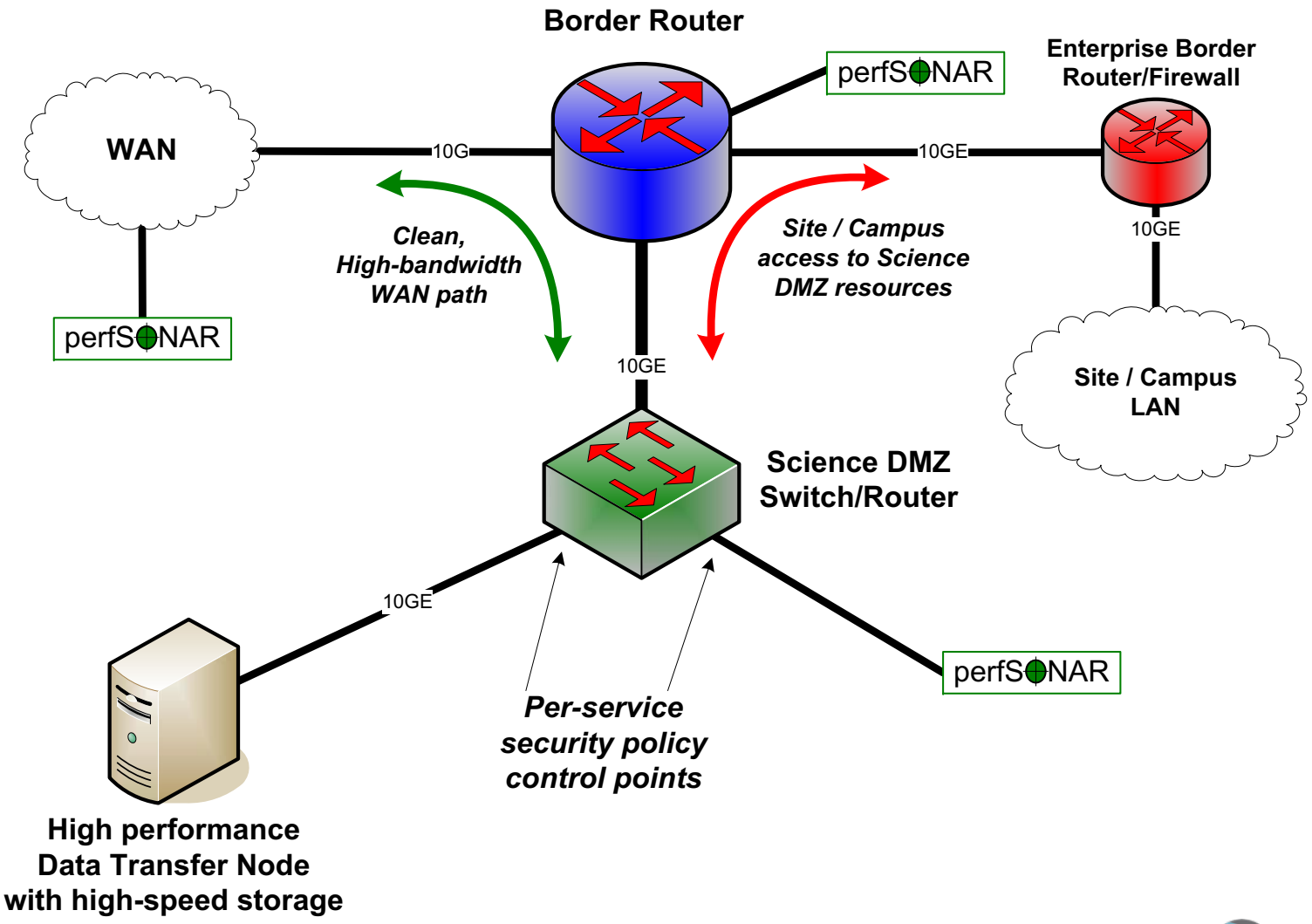
Outline

- Science DMZ in brief
- Context – Science DMZ in the community
- Science DMZ and Data Portals

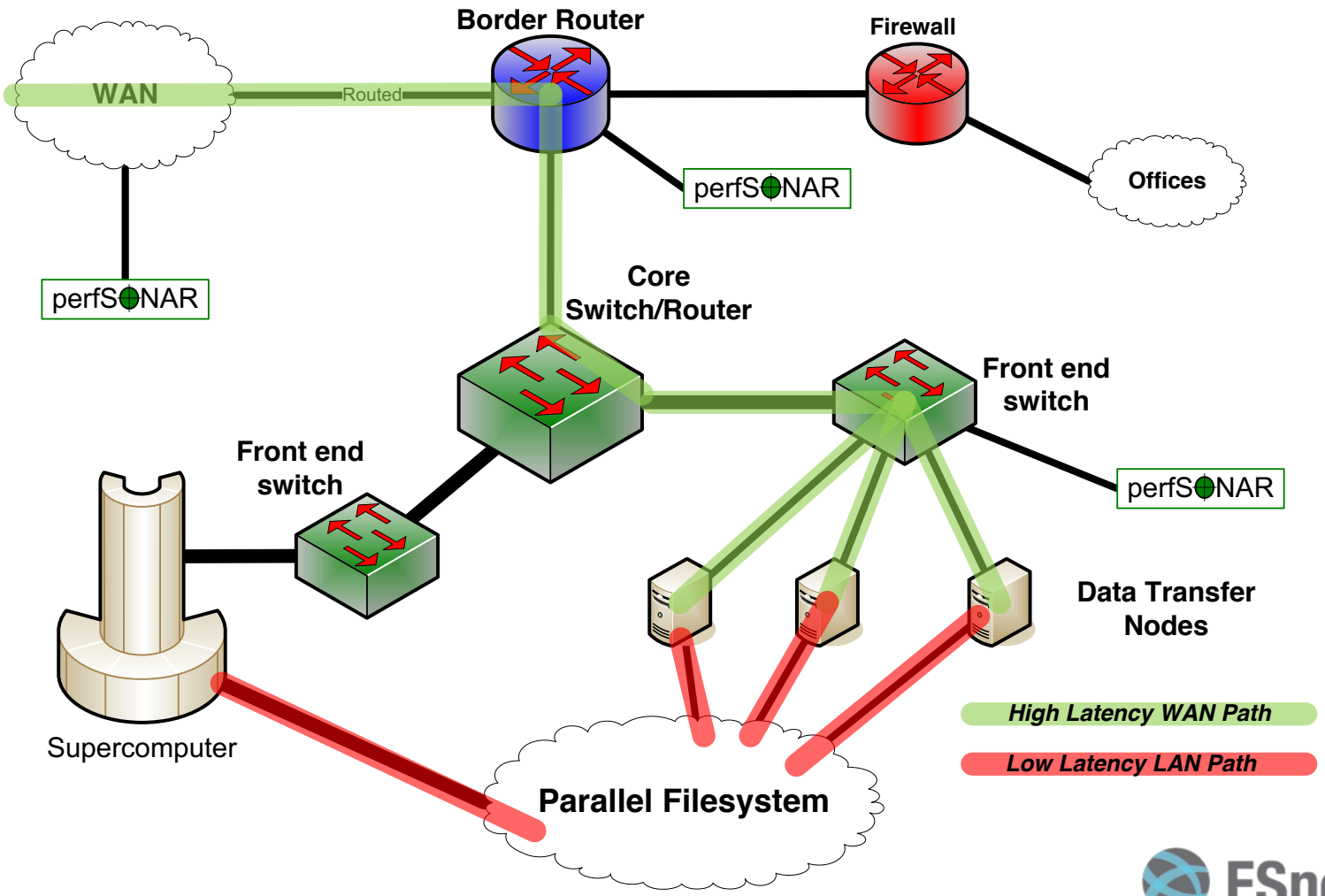
http://fasterdata.es.net/science-dmz/

ESnet

Science DMZ Design Pattern (Abstract)



HPC Center Data Path



Context: Science DMZ Adoption

- Initially deployed within DOE National Laboratories
- Growing adoption among institutions of all sizes
- NSF CC* programs have funded many Science DMZs
- Other US agencies, e.g. NIH, USDA
- International, e.g. Australia, Brazil, UK

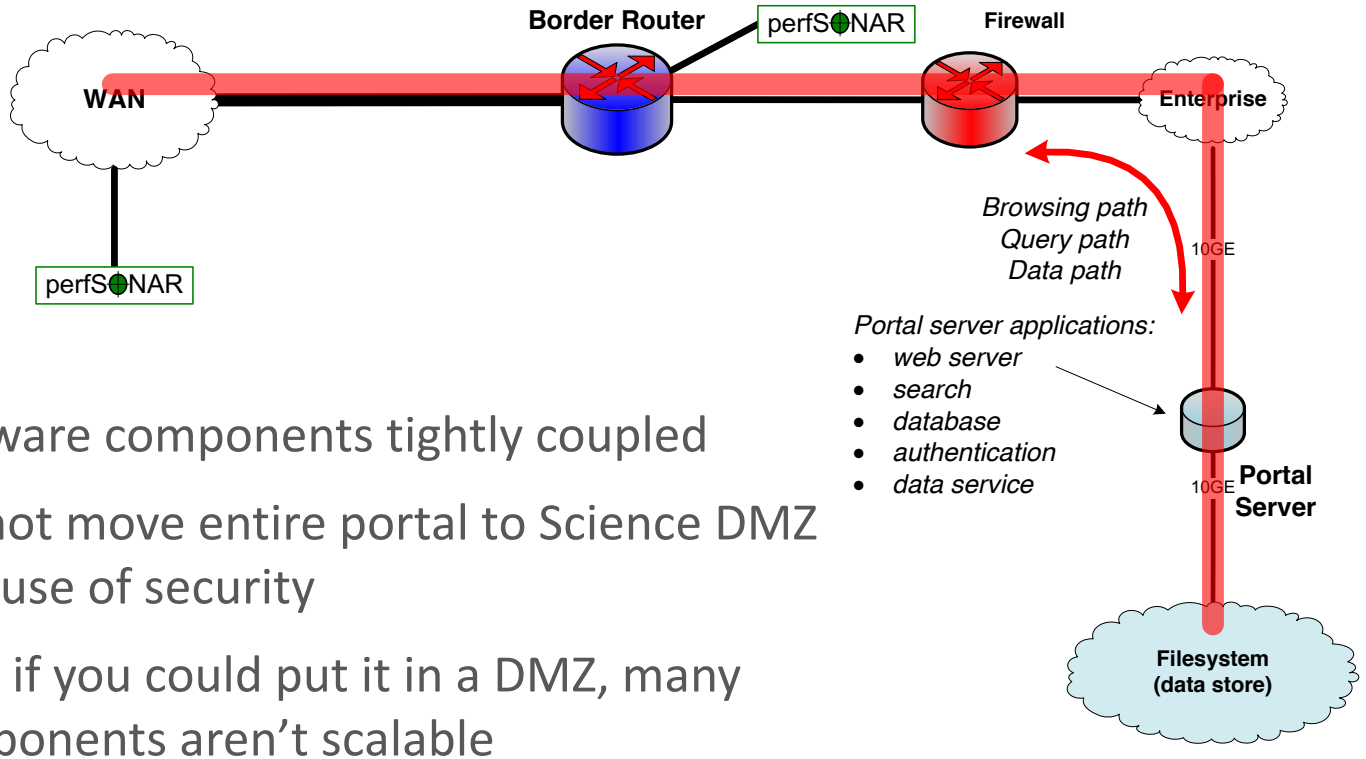
ESnet

Strategic Impacts

- We are undergoing significant cyberinfrastructure upgrades...
- ...but enterprise networks need not be unduly perturbed ☺
- Significantly enhanced capabilities compared to 3 years ago
  - Terabyte-scale data movement is much easier
  - Petabyte-scale data movement possible outside the LHC experiments
    - ~3.1Gbps = 1PB/month
    - ~14Gbps = 1PB/week
  - Widely-deployed tools are much better (e.g. Globus)
- Metcalfe's Law of Network Utility
  - Value of Science DMZ proportional to the number of DMZs
    - n^2 or $n(\log_n)$ doesn't matter – the effect is real
  - Cyberinfrastructure value increases as we all upgrade

ESnet

Legacy Portal Design



- Software components tightly coupled
- Cannot move entire portal to Science DMZ because of security
- Even if you could put it in a DMZ, many components aren't scalable
- Performance improvement requires architectural change

ESnet

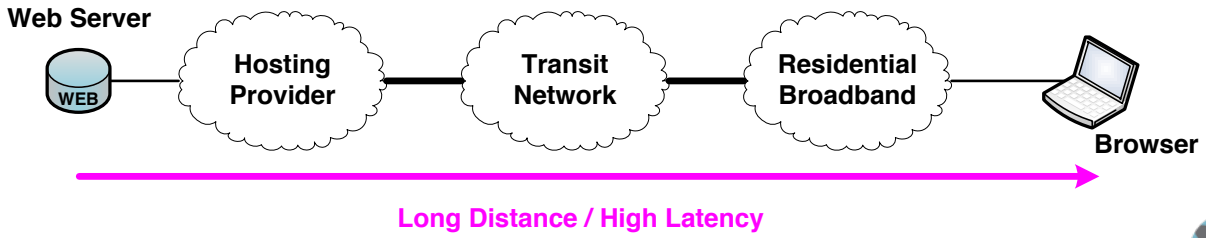
Example of Architectural Change – CDN

- CDNs are a well-deployed design pattern
- Inherent in Internet architecture (e.g. Netflix, Amazon)
- Store static and dynamic content in separate locations
  - Static content is simple (but often BIG)
  - Application dynamics are complex (stateful, synchronous)
- Separation of application and data service allows each to be optimized

ESnet

Classical Web Server Model

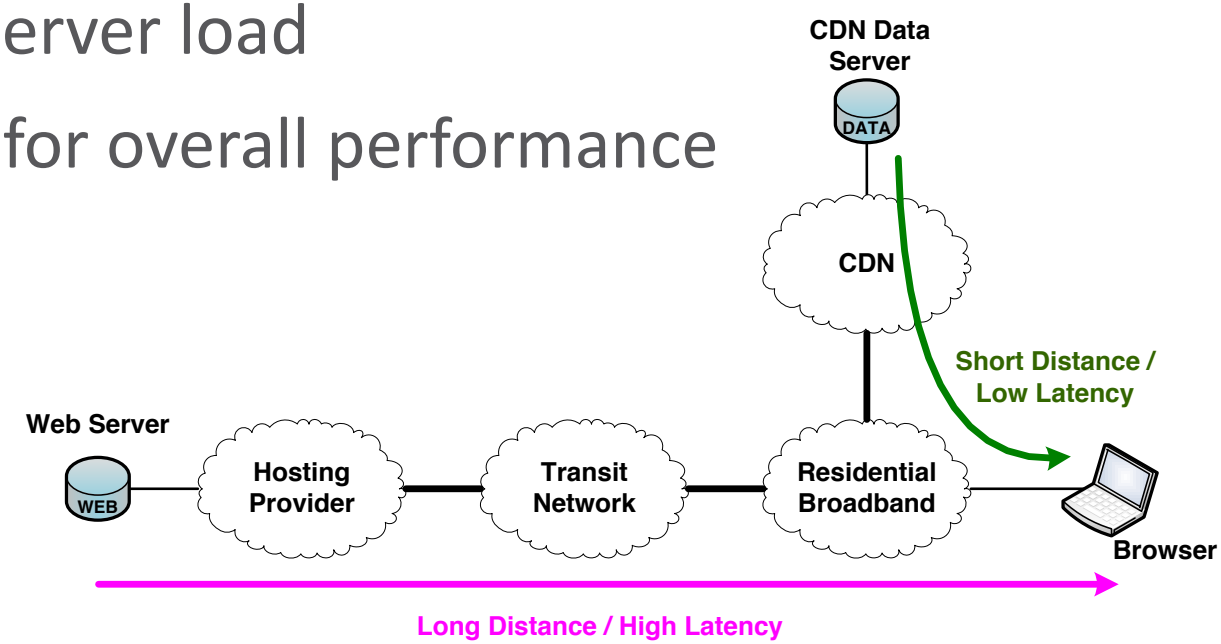
- Web browser fetches pages from web server
- All content stored on the web server
- Web applications run on the web server
- Web server sends data to client browser over the network
- Issue: Latency increases time to page render
- Issue: Packet loss + latency = problems for large static objects



ESnet

Solution: Place Large Static Objects Near Client

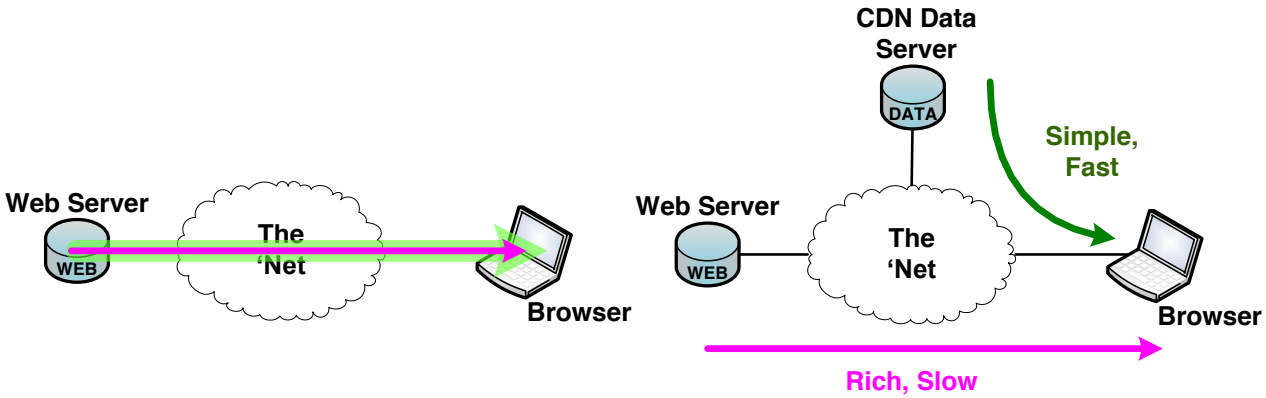
- Reduced latency
  - Faster page rendering
  - Faster static content delivery
- Reduced web server load
- Significant win for overall performance



ESnet

Client Simply Sees Increased Performance

- Client doesn't see the CDN as separate entity
- Web content is all still viewed in a browser
  - Browser fetches what the page tells it to fetch
  - Different content comes from different places
  - User doesn't know/care
- CDNs provide architectural solution to performance

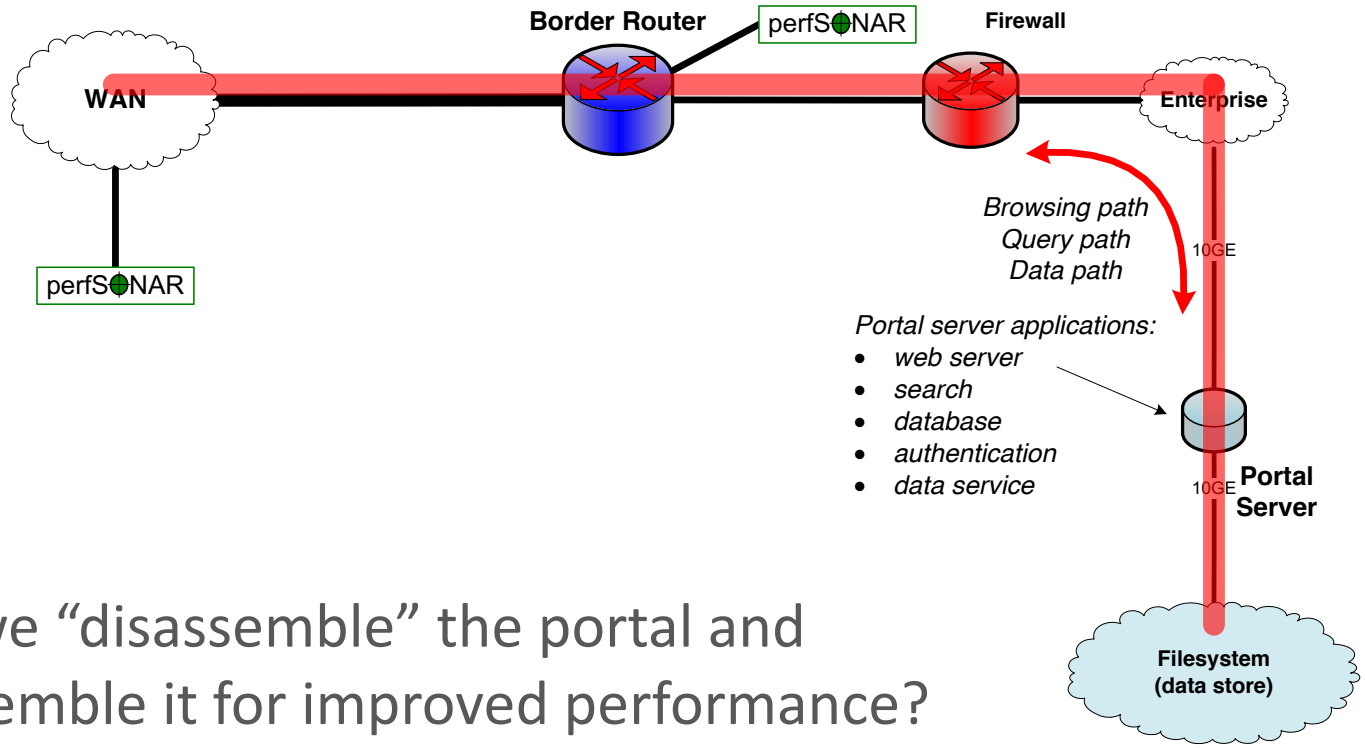


Architectural Examination of Data Portals

- Common data portal functions
  - Search/query/discovery
  - Data download method for data access
  - GUI for browsing by humans
  - API (ideally incorporates search/query/download)
- Performance issues primarily in data download
  - Rapid increase in data scale eclipsed legacy software stack
  - Portal servers often stuck in enterprise network

ESnet

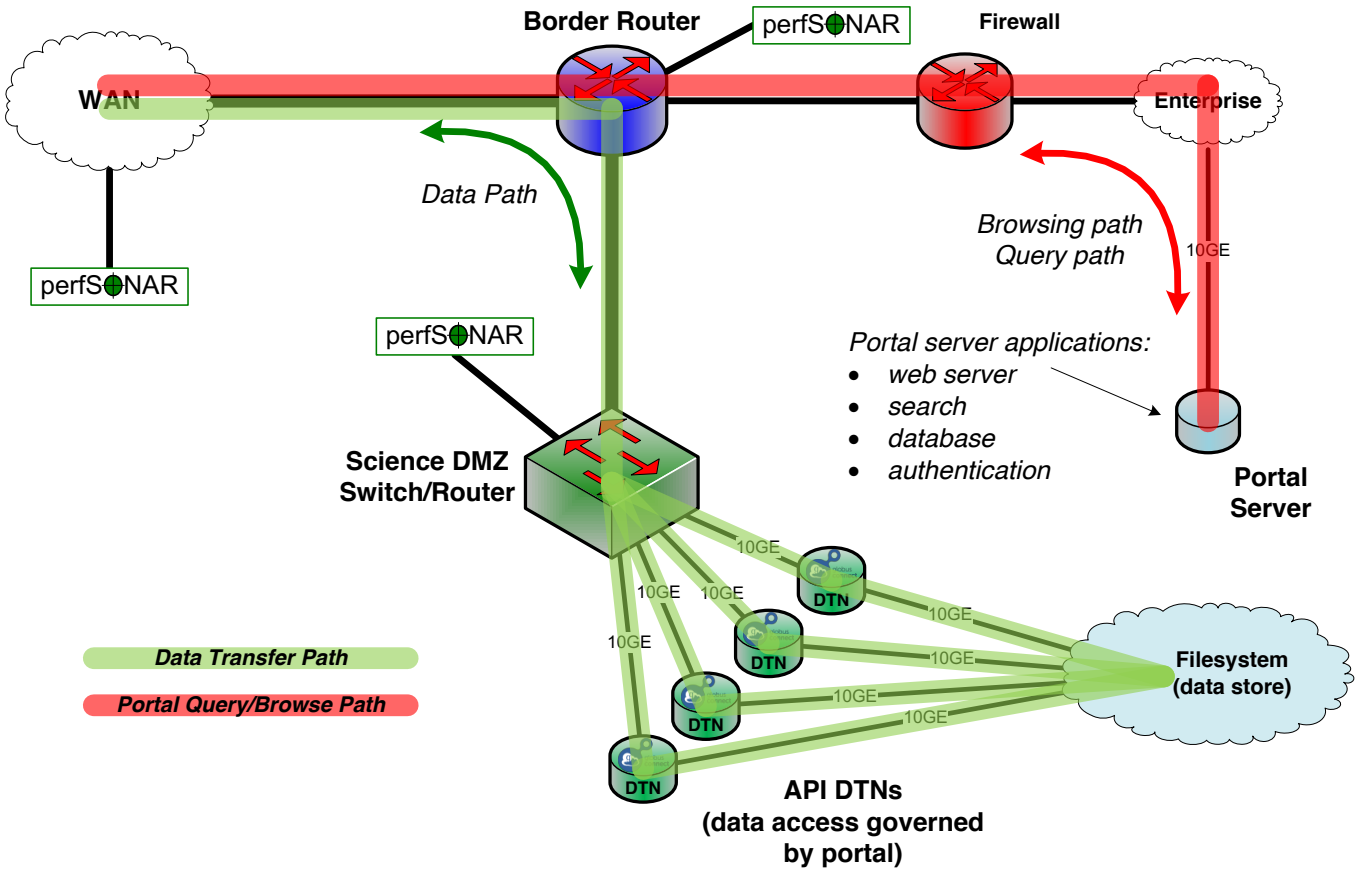
Legacy Portal Design



Can we "disassemble" the portal and reassemble it for improved performance?

- Use Science DMZ as a platform for the data piece
- Avoid placing complex software in the Science DMZ

Modern Data Portal Leverages Science DMZ



Separating data handling from portal logic

- Portal GUI, search, etc. all function as before
- Query returns pointers to data objects in Science DMZ
- Portal freed from ties to data servers
- Data handling is separate, and scalable
  - High-performance DTNs in the Science DMZ
  - Scale without modifying the portal software
- Outsource data handling to computing centers

ESnet

Links and Lists

- ESnet fasterdata knowledge base
  - http://fasterdata.es.net/
- Science DMZ paper
  - http://www.es.net/assets/pubs_presos/sc13sciDMZ-final.pdf
- Science DMZ email list
  - Send mail to sympa@lists.lbl.gov with subject "subscribe esnet-sciencedmz”
- perfSONAR
  - http://fasterdata.es.net/performance-testing/perfsonar/
  - http://www.perfsonar.net
- Globus
  - https://www.globus.org/

ESnet
ENERGY SCIENCES NETWORK

Thanks!

Eli Dart dart@es.net
Energy Sciences Network (ESnet)
Lawrence Berkeley National Laboratory

http://fasterdata.es.net/
http://my.es.net/
http://www.es.net/