Security Challenges for PRACE High Throughput Data Transfers

WISE COMMUNITY

NSF Cyber Security Summit 2018, Alexandria, Virginia, US

August 21st, 2018

Ralph Niederberger, Jülich JSC

Acknowledgements

- Tim Chown who brought up this topic into WISE and provided a lot of input/slides to this presentation
- ESnet Fasterdata Knowledge Base for a very informative overview on Science DMZ at

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

The BIG DATA challenge on Security

Today increasing number of devices, sensors and people generate, share, and access data.

Data volumes have become so large, that conventional processing methods do not scale.

Nowadays we see decreasing storage costs, better storage solutions and algorithms.

Big Data technologies can be defined as

> *„New generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis.“*
>
> *Source: EMC/IDC, „The Digital Universe“ Study 2014*

Fundamentally, those technologies must include secure high throughput data transfers.

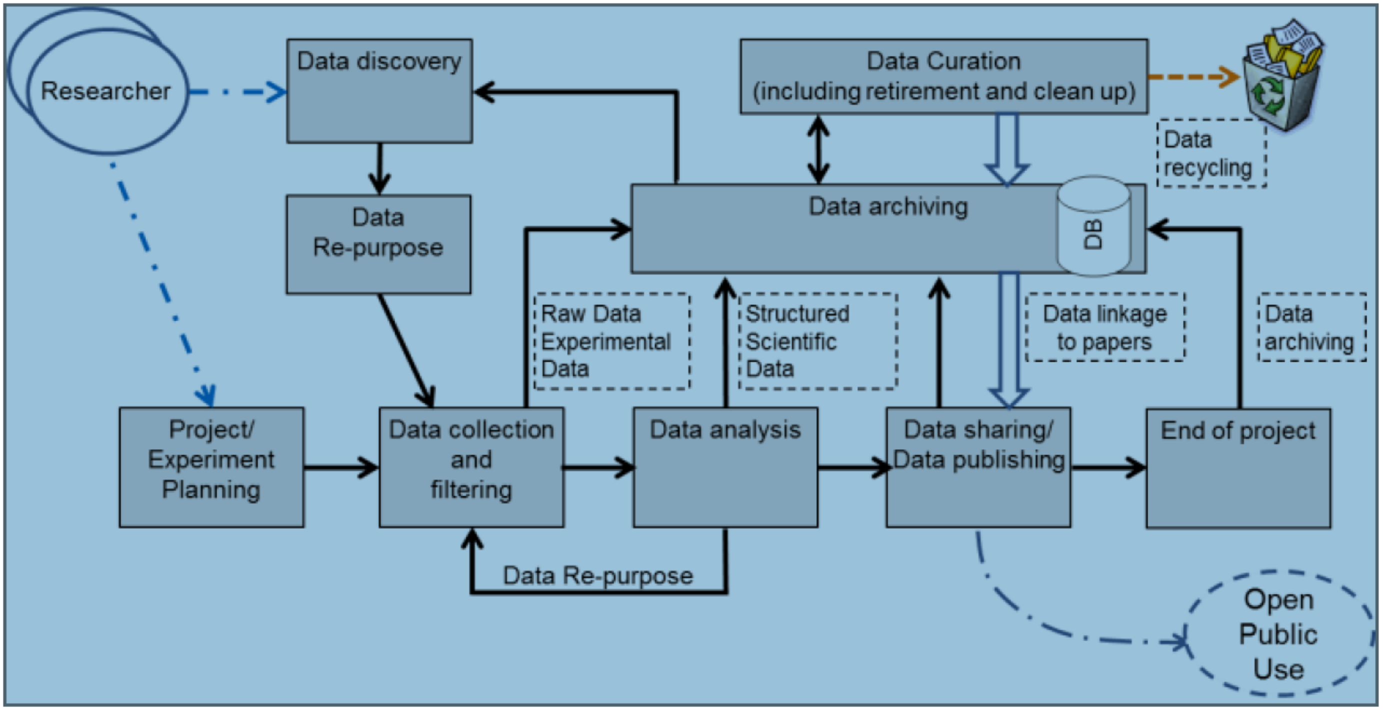
These transfers are dependent on the software stacks used within source and destination systems as well as security devices on the communication path.

Communication environments

- But are they all the same?
- No, they are not. We face:
  - business applications and human resources or more generell administrative tasks applications with well defined communication protocols, which are handled by any kind of firewalls (eg. smtp, http(s), H.323)
  - scientific data intensive applications with uncommon or at least rarely used specialized applications (e.g. gridftp, uftp, …) . Those are mostly unknown to firewalls
  - and of course, we see any kind of scenarios in between

- Why handle them all the same?

Scientific Data Lifecycle Management



Researcher
Data discovery
Data Re-purpose
Data Curation (including retirement and clean up)
Data recycling
Data archiving
DB
Raw Data Experimental Data
Structured Scientific Data
Data linkage to papers
Data archiving
Project/ Experiment Planning
Data collection and filtering
Data analysis
Data sharing/ Data publishing
End of project
Data Re-purpose
Open Public Use

Source: Demchenko Y., Ngo C., de Laat C., Membrey P., Gordijenko D. (2014) Big Security for Big Data: Addressing Security Challenges for the Big Data Infrastructure. In: Jonker W., Petković M. (eds) Secure Data Management. SDM 2013. Lecture Notes in Computer Science, vol 8425. Springer, Cham

Security impacts and challenges

With the Scientific Data Lifecycle we face

- Collaborating institutions, research projects, ad hoc communities,
- Generation systems, networks, data processing systems, storage systems
- Word wide distribution of instruments, HPCs, clouds, networks
- with diverse infrastructure set ups, security policies, national laws

Data set volumes are increasing

- 100 TB is no longer ‘large’
- Moving 100 TB takes 10Gbps of throughput for 24 hours
- How do we do this securely, AND with the necessary performance?

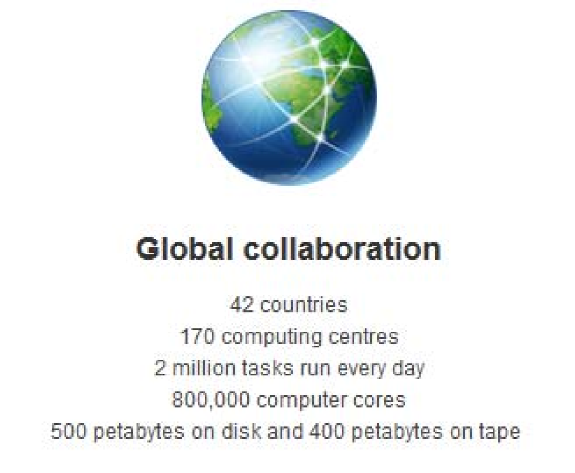
We need

- Secure and fast data transmission technologies, adequate infrastructures and corresponding security policies

Some examples: LHC – Large Hadron Collider

Generates 15 PB of data annually

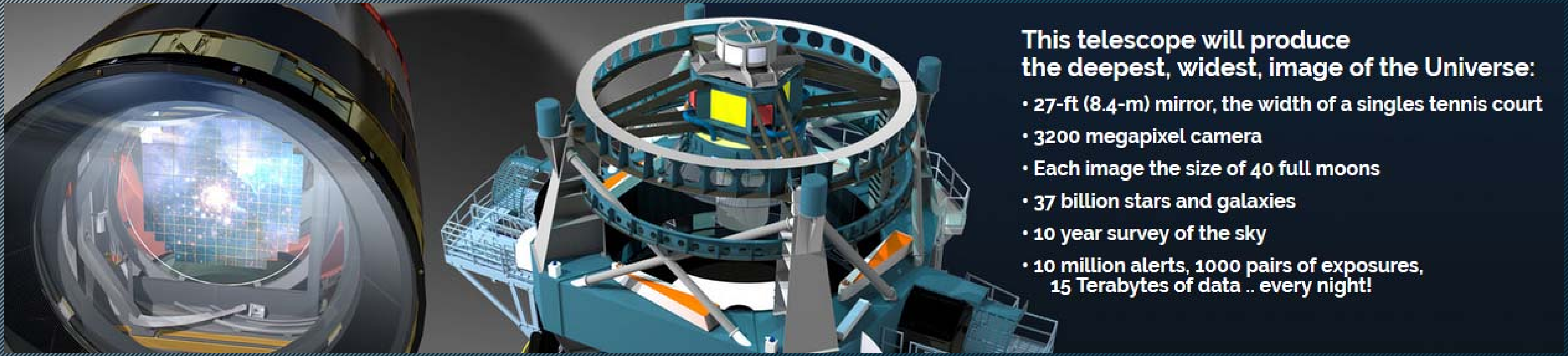
- ➢ 8000 physicists with near real-time access to LHC data,
- ➢ links thousands of computers and storage systems in over 170 centres across 42 countries
- After preprocessing transfer to those centers in Europe, Asia and US (Tier 1 sites) needed
- Postprocessing at Tier 2 sites



Global collaboration

42 countries
170 computing centres
2 million tasks run every day
800,000 computer cores
500 petabytes on disk and 400 petabytes on tape

Some examples:
LSST – Large Synoptic Survey Telescope



This telescope will produce the deepest, widest, image of the Universe:

- 27-ft (8.4-m) mirror, the width of a singles tennis court
- 3200 megapixel camera
- Each image the size of 40 full moons
- 37 billion stars and galaxies
- 10 year survey of the sky
- 10 million alerts, 1000 pairs of exposures, 15 Terabytes of data .. every night!

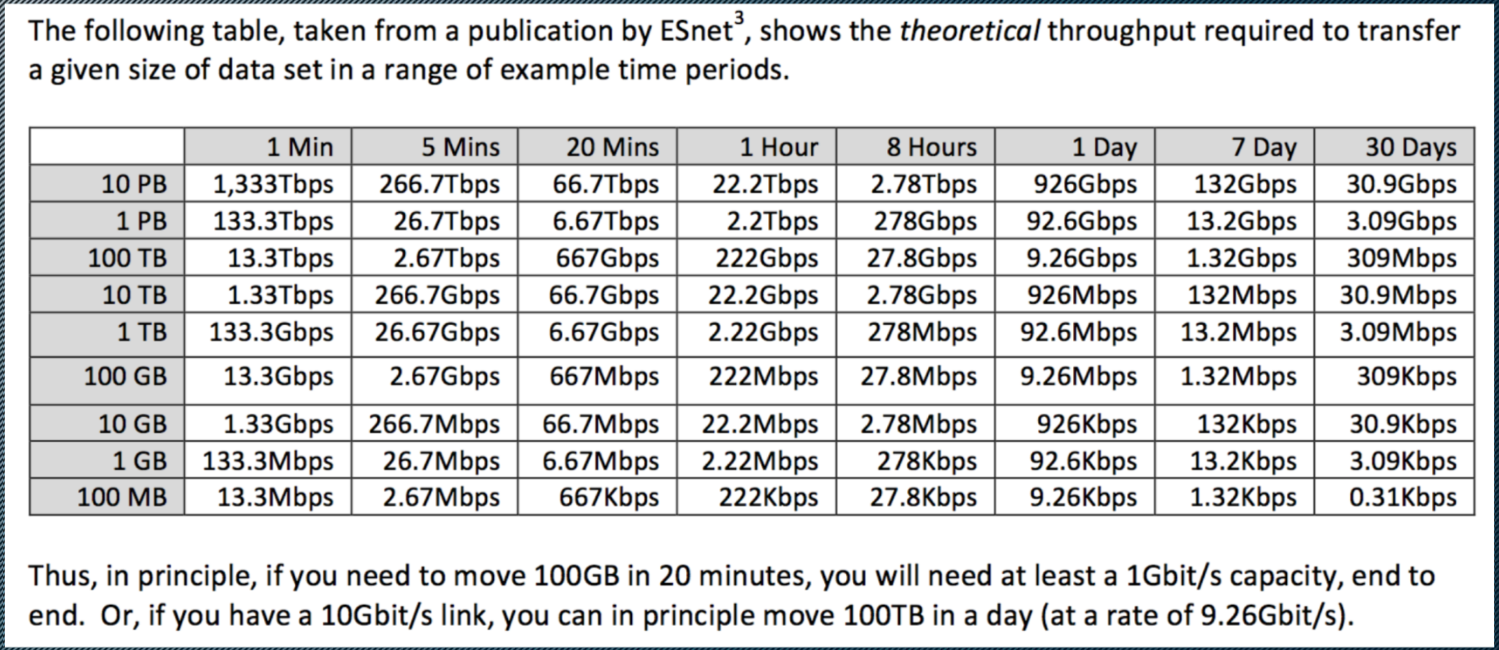
Source: https://www.lsst.org/

- Data Mining: more than 100.000 TB of data expected till end of project
- Shared with the public

Some examples:
SKA - Square Kilometre Array

- Simulation of a huge radio telescope (> 1000 antenna) one square kilometre
- The dishes of the SKA will produce 10 times the global internet traffic
- Approximately 160 Gb/s of data transmitted from each radio dish to a central processor
- The SKA central computer will have the processing power of about one hundred million PCs.
- The aperture arrays in the SKA could produce more than 100 times the global internet traffic
- SKA generates more than 960 PB of data annually
- The processed data will be used by astronomers worldwide requiring connections to HPC facilities and enormous archive capability to store the data.
- Connectivity challenge (hundreds of Gb/s) on international connection systems and local networks. This will require network infrastructure to surpass the global internet by a huge factor in terms of the amounts of data being sent globally.
- Signal processing has never witnessed anything on this scale.

Theoretical throughput



The following table, taken from a publication by ESnet[3], shows the *theoretical* throughput required to transfer a given size of data set in a range of example time periods.

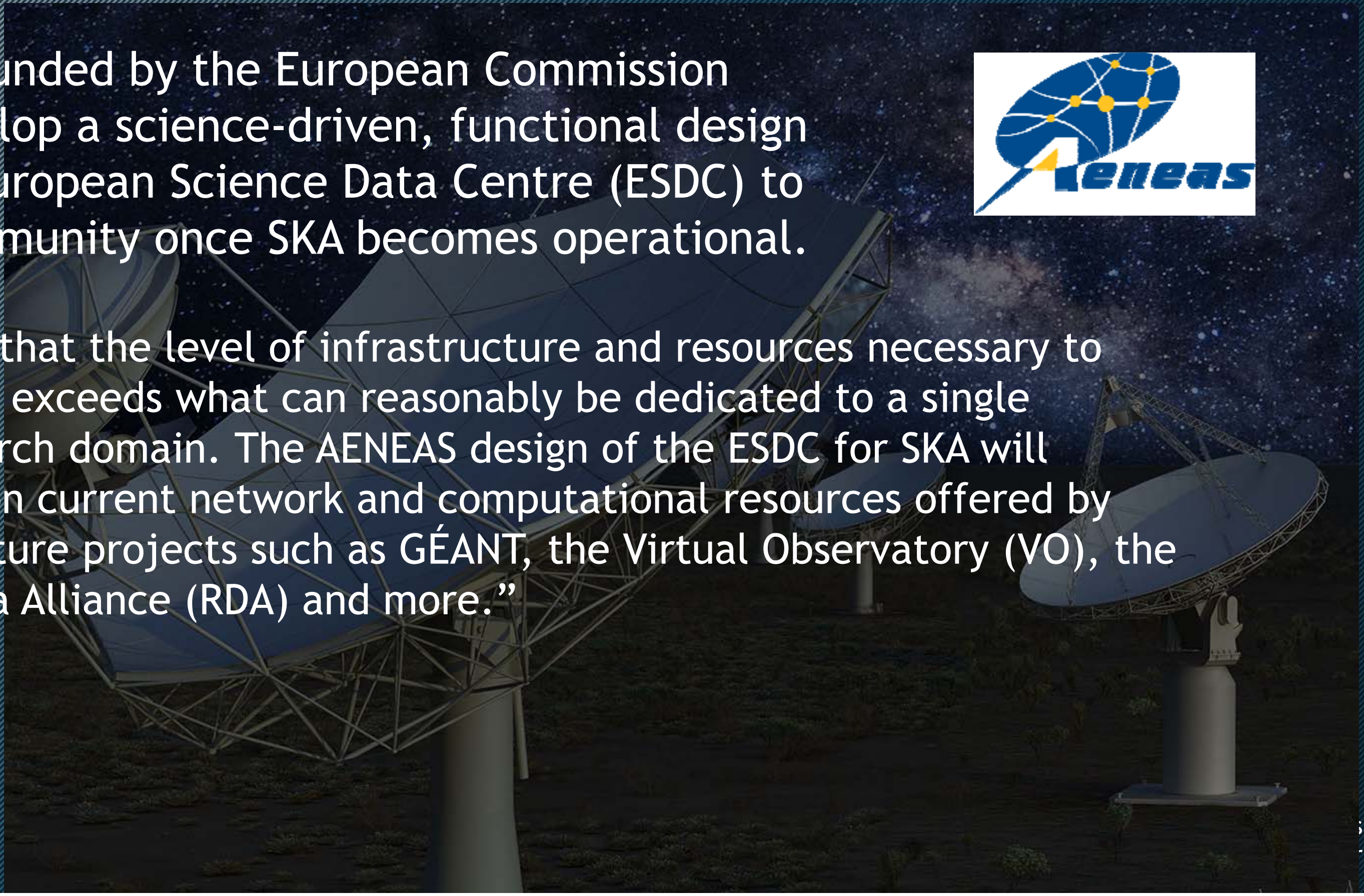
	1 Min	5 Mins	20 Mins	1 Hour	8 Hours	1 Day	7 Day	30 Days
10 PB	1,333Tbps	266.7Tbps	66.7Tbps	22.2Tbps	2.78Tbps	926Gbps	132Gbps	30.9Gbps
1 PB	133.3Tbps	26.7Tbps	6.67Tbps	2.2Tbps	278Gbps	92.6Gbps	13.2Gbps	3.09Gbps
100 TB	13.3Tbps	2.67Tbps	667Gbps	222Gbps	27.8Gbps	9.26Gbps	1.32Gbps	309Mbps
10 TB	1.33Tbps	266.7Gbps	66.7Gbps	22.2Gbps	2.78Gbps	926Mbps	132Mbps	30.9Mbps
1 TB	133.3Gbps	26.67Gbps	6.67Gbps	2.22Gbps	278Mbps	92.6Mbps	13.2Mbps	3.09Mbps
100 GB	13.3Gbps	2.67Gbps	667Mbps	222Mbps	27.8Mbps	9.26Mbps	1.32Mbps	309Kbps
10 GB	1.33Gbps	266.7Mbps	66.7Mbps	22.2Mbps	2.78Mbps	926Kbps	132Kbps	30.9Kbps
1 GB	133.3Mbps	26.7Mbps	6.67Mbps	2.22Mbps	278Kbps	92.6Kbps	13.2Kbps	3.09Kbps
100 MB	13.3Mbps	2.67Mbps	667Kbps	222Kbps	27.8Kbps	9.26Kbps	1.32Kbps	0.31Kbps

Thus, in principle, if you need to move 100GB in 20 minutes, you will need at least a 1Gbit/s capacity, end to end. Or, if you have a 10Gbit/s link, you can in principle move 100TB in a day (at a rate of 9.26Gbit/s).

The question to consider here is how do we achieve high throughput on the end-to-end data transfer path, while applying appropriate security measures to the traffic in flight

Welcome to AENEAS

Advanced European Network of E-infrastructures for Astronomy with the SKA



AENEAS is a 3 year initiative funded by the European Commission Horizon 2020 program to develop a science-driven, functional design for a distributed, federated European Science Data Centre (ESDC) to Support the astronomical community once SKA becomes operational.

“The AENEAS project recognizes that the level of infrastructure and resources necessary to enable SKA science within Europe exceeds what can reasonably be dedicated to a single instrument or even a single research domain. The AENEAS design of the ESDC for SKA will therefore integrate and build upon current network and computational resources offered by European and Global e-infrastructure projects such as GÉANT, the Virtual Observatory (VO), the EGI federation, the Research Data Alliance (RDA) and more.”

Jülich Supercomputing Centre (JSC)

The Jülich Supercomputing Centre operates supercomputers of the highest performance class.

It enables scientists and engineers to solve their highly complex problems by simulations.

Currently, we run one of the fastest supercomputers in Europe. So we are part of several EU projects like PRACE, HBP, EOSC-Hub and a lot of others all related to HPC.

Jülich Supercomputing Center is also part of the AENEAS project and especially highly interested in secure communications.

Jülich/Frankfurt, 25 June 2018 - When it comes to developing innovative supercomputer architectures, Europe is about to take the lead. A striking example of this is the new supercomputer that started operation in Jülich. JUWELS is a milestone on the road to a new generation of ultraflexible modular Supercomputers. With its first module alone, JUWELS qualified as the best German computer for the TOP500 List of the fastest supercomputers in the world published today.

PRACE in a few words

PRACE Mission

The mission of PRACE (Partnership for Advanced Computing in Europe) is to enable high-impact scientific discovery and engineering research and development across all disciplines to enhance European competitiveness for the benefit of society. PRACE seeks to realize this mission by offering world class computing and data management resources and services through a peer review process.

PRACE also seeks to strengthen the European users of HPC in industry through various initiatives. PRACE has a strong interest in improving energy efficiency of computing systems and reducing their environmental impact.

PRACE in a few words (2)

PRACE Research Infrastructure (RI)

PRACE is established as an international not-for-profit association (aisbl) with its seat in Brussels. It has 26 member countries creating a pan-European HPC infrastructure, providing access to computing resources and services for large-scale scientific and engineering applications at the highest performance level.

The HPC systems and their operations accessible through PRACE are provided by 5 PRACE members (BSC (Spain), CINECA (Italy), ETH Zurich/CSCS (Switzerland), GCS (HLRS, LRZ, Jülich; Germany), and GENCI (France). In pace with the needs of the scientific communities and technical developments, systems deployed by PRACE are continuously updated and upgraded to be at the apex of HPC technology.

The PRACE project received and is still receiving EC funding under the PRACE Preparatory and Implementation Phase Projects (PRACE-1IP, 2010-2012, RI-261557 | PRACE-2IP, 2011-2013, RI-283493 | PRACE-3IP, 2012-2017, RI-312763 | PRACE-4IP, 2015-2017, 653838 | PRACE-5IP, 2017-2019, 730913). The total funding of the PRACE Projects amounts to €132M over 9 years (2010 – 2019) of which €97M is provided by the European Commission (EC).

For more info see: http://www.prace-ri.eug

PRACE in a few words (3)

PRACE HPC Access

PRACE systems are available to scientists and researchers from academia and industry from around the world through 2 forms of access:

Preparatory Access is intended for short-term access to resources, for code-enabling and porting, required to prepare proposals for Project Access and to demonstrate the scalability of codes.

Applications for Preparatory Access are accepted at any time, with a cut-off date every 3 months.

Project Access is intended for individual researchers and research groups including multi-national research groups and can be used for 1-year production runs, as well as for 2-year or 3-year (Multi-Year Access) production runs.

Project Access is subject to the PRACE Peer Review Process, which includes technical and scientific review. Technical experts and leading scientists evaluate the proposals submitted in response to the bi-annual calls. Applications for Preparatory Access undergo technical review only.

Getting things together

- Prace partners are connected to each other via a MD-VPN provided by GÉANT allowing fast access between HPC systems.
- Firewalls may be implemented in between, but „Net of trust“ idea doesn‘t necessitate this. It is already a Science DMZ?
- But what about Input/Output data?
- Researchers need fast access from outside the PRACE MD-VPN to the PRACE HPC systems. One example will be AENEAS now and the related SKA project in the future. Another one is the Human Brain Project (HBP) and there are a lot of others in the future.

Factors affecting E2E

Achieving optimal end-to-end performance is a multi-faceted problem.

It includes:

- Appropriate network capacity provisioning between the end sites
- Properties of the local campus network (at each end), including capacity of the external connectivity, internal LAN design, the performance of firewall / IDS devices, and the configuration of other devices on the path
- End system configuration and tuning; network stack buffer sizes, disk I/O, …
- The choice of tools used to transfer data, e.g. scp, Globus, rsync, Aspera, …

To optimise end-to-end performance, you need to address each aspect

Nevertheless, there will inevitably be a bottleneck somewhere

How to design the local campus network?

An application using TCP will see its performance degrade if packets are lost, with more degradation the higher the path's RTT

- Very small loss can have a surprisingly significant impact
- Therefore we need to engineer towards zero packet loss

Zero loss implies both sufficient capacity and performant network elements

The challenge is that many campus security appliances, esp. corporate firewall/IDS, are designed for 1000's of small flows, not tens of very large flows, and they can / will thus drop packets

There is already an answer to this: The Science DMZ

The Science DMZ

ESnet published the Science DMZ 'design pattern' in 2012/13

https://www.es.net/assets/pubs_presos/sc13sciDMZ-final.pdf

Three key elements:

- Network architecture improvements; avoiding local bottlenecks
- Network performance measurement
- Data transfer node (DTN) design and configuration

Also termed a "high speed on-ramp" to the campus storage

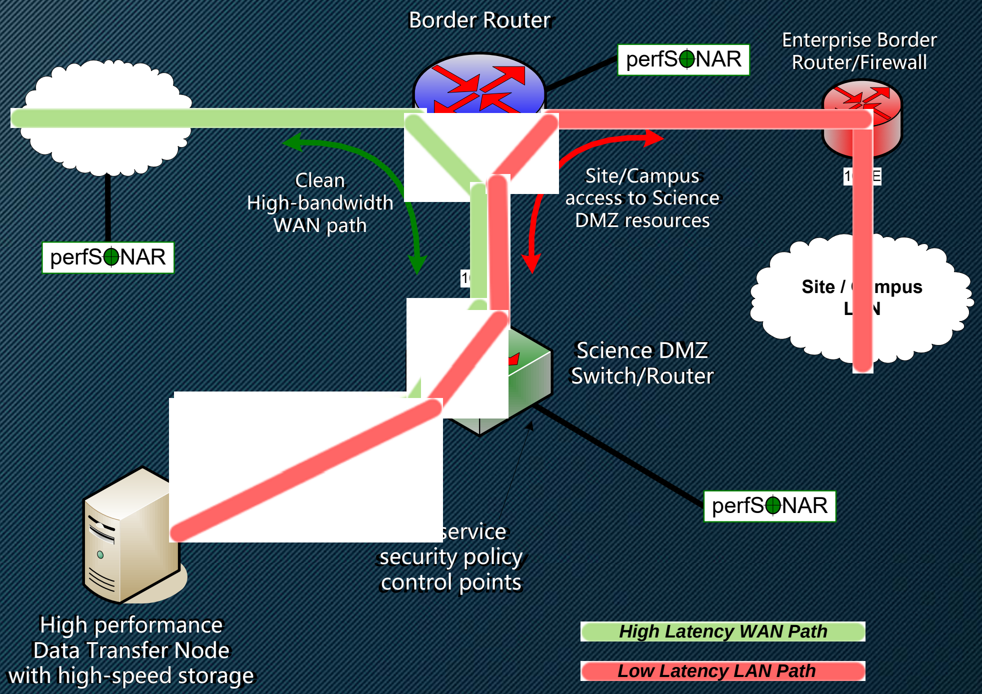
- Splits the internal and external latency domains

The NSF Cyberinfrastructure (CC*) Program funded this model in over 100 US universities:

See https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504748

Science DMZ Design Pattern

- There are a lot of examples of sites using some form of Science DMZ deployment
- In many cases the deployments were made without knowledge of the Science DMZ model!
- Science DMZ is just a set of good principles to follow, so it's not surprising that several sites were already doing it



Border Router
perfSONAR
Enterprise Border Router/Firewall
Clean High-bandwidth WAN path
Site/Campus access to Science DMZ resources
perfSONAR
1 E
1
Site / mpus L N
Science DMZ Switch/Router
perfSONAR
service security policy control points
High performance Data Transfer Node with high-speed storage
High Latency WAN Path
Low Latency LAN Path

Source: https://fasterdata.es.net/science-dmz/science-dmz-architecture/

The science DMZ principle

One fundamental remark:

- Often it is said, that with a Science DMZ you avoid using a firewall.
- This isn‘t true. You just split traffic into two classes.
  - those, which can be handled by „fast packet screens“,
    e.g. routers with access lists

  and

  - those, which need „deep packet inspection“

(Stateful) packet inspection

- So is a router right doing this ACLs? The answer is: Maybe
- Why isn‘t is Yes? Take GridFTP and you know the answer.
- GridFTP has a control connection and a lot of data connections using TCP, i.e. you have to open a port range for those data cons.
- If you use a packet screen only, how do you differentiate data connections started from outsite and any unknown connection started from inside, if you don‘t concentrate on TCP streams (stateless).
- So your security policy decides if the answer is „Yes“ or „No“.
- If you don‘t care on connections going out, it‘s fine.

Other examples of campus network engineering

Many sites split their external connectivity
- e.g., 40G total; 1x10G campus, 1x10G research science data, 2x10G resilience
- And then apply Science DMZ principles to the research data path
- Or employ Science DMZ in their data centre

The Worldwide LHC Computing Grid (WLCG) has used physical / virtual overlays
- LHCOPN (private optical network) / LHCONE (virtual network)
- LHCONE implicitly becomes a 'trusted' network

Similar did the DEISA and PRACE project
- Starting with a dedicated star like 10 Gb/s fiber wavelength network
- And now a PRACE-VPN operated on the GÉANT MDVPN backbone

But how should campuses cater for multiple data-intensive science disciplines?
- Would one new overlay network per research community scale?

Science DMZ as a Security Architecture

It allows for better segmentation of risks, and more granular application of controls to those segmented risks.

- Limit risk profile for high-performance data transfer applications
- Apply specific controls to data transfer nodes (DTNs)
- Avoid including unnecessary risks, unnecessary controls

Remove degrees of freedom – focus only on what is necessary

- Easier to secure
- Easier to achieve performance
- Easier to troubleshoot

Performance is a key requirement; e.g., use efficient ACLs

- See https://www.slideshare.net/JISC/science-dmz-security (Kate Mace)

It is not true, that a Science DMZ does not use a firewall.

They just use a stateless packet screen, i.e. a simple old fashioned firewall and it serves its needs.

And what about data transfer technologies

https://fasterdata.es.net/data-transfer-tools/say-no-to-scp/

Say no to SCP:

Sample Results Berkeley, CA to Argonne, IL (near Chicago).

RTT = 53 ms, network capacity = 10Gbps.

Tool	Throughput
scp	330 Mbps
HPN patched scp	1.2 Gbps
Gridftp, http (e.g.:wget), 1 stream	6 Gbps
Gridftp, 4 streams	8 Gbps (disk limited)

Further transfer technologies

FTP (File Transfer Protocol), PFTP (Parallel FTP), SFTP (Secure FTP), …
HTAR (High Performance Storage System Tape Archiver)
NFT (Network File Transport)
rsync
(open)ssh/scp, HPN SCP (from PSC), and there are others around
Uftp (Unicore FTP)
Gridftp
… and more and more

What is there impact on security? How do those scale in WAN environments?
And do we always need encrypted transfers? Maybe we only need encrypted authentication?

Why should Wise bother with all of this?
Implications on WISE activities

The classic 'Science DMZ' model has value; many did it anyway; Well-tuned DTNs with host-based security

But what are the security implications ->

1.) Analysis of implications of Science DMZs on IT security and risks

2.) bring in line Science DMZs and strict security policies

and what can help improving data transfer throughput ->

3.) (development of or) analysis of secure high speed data transfer protocols and

4.) fast en-/decryption technologies to speed up secure transmissions

First potential deliverable/whitepaper: „Recommendations for best practices in delivering high performance data transfers while maintaining appropriate security"

-> New WISE Group dealing with these aspects has been setup.

Join, engage and contribute!!

A draft Charter

Security is one of the main drivers in Information technology today. Making your local networks more and more secure adds overhead to all involved software and hardware components.

Since you will never get a 100 % secure network, the question arises, where to stop adding components. A controversial discussion has risen in the last years. Do we need all these security components for every action taking part in our networks? One of these discussions has led to the "Science DMZ" network architecture. Why doing deep inspection of packets on the communication path (Intrusion prevention), where risks are minimal or nearly zero, and using instead intrusion detection techniques (Bro clusters)?

This working group focuses on the security aspects of the "Science DMZ architecture" and its security risks. Are all components sketched in the concept needed, do we need more? What is missing, what are the drawbacks? Can we offer similar concepts fitting different scenarios?

The second question we would like to answer follows consequently. Why are we introducing "Science DMZs"? The answer is: To make transfers faster, but which tools are outside there and what are their advantages/disadvantages. Which should be used for optimal high speed transmissions? Which are easy to be setup and used?

S4HST intends to focus on those two aspects:

a.) Security level of Science DMZs and similar approaches and

b.) secure high speed data transmissions.

The WG will produce two whitepapers on the above security areas. Emails are the means of communication of the working group. S4HST will mainly meet via teleconferences, but face-to-face meetings (at WISE Work Shops) will also be considered and organised.

Questions and discussion

WISE COMMUNITY