perfSONAR

Introduction & Motivation

ASTRON perfSONAR training

Antoine Delvaux, PSNC, antoine.delvaux@man.poznan.pl
Szymon Trocha, PSNC, szymon.trocha@man.poznan.pl
24-26 September 2018

ESnet
ENERGY SCIENCES NETWORK

GÉANT

INDIANA UNIVERSITY

INTERNET2

UNIVERSITY OF MICHIGAN

Outline

- Problem Statement on Network Connectivity

- Supporting Scientific Users

- Network Performance & TCP Behaviors w/ Packet Loss

- What is perfSONAR

- Architecture and installation options

Problem Statement

- The global Research & Education network ecosystem is comprised of hundreds of international, national, regional and local-scale networks.


CAIDA's IPv4 & IPv6 AS Core
AS-level INTERNET Graph
Archipelago January 2014
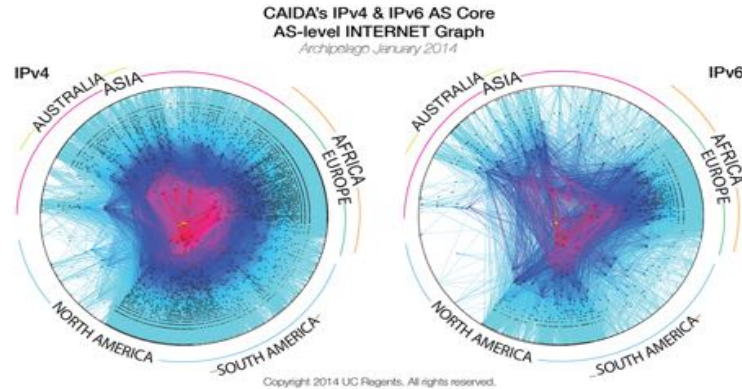
Problem Statement

- While these networks all interconnect, each network is owned and operated by separate organizations (called "domains") with different policies, customers, funding models, hardware, bandwidth and configurations.

The R&E Community

- The global Research & Education network ecosystem is comprised of hundreds of international, national, regional and local-scale resources – each independently owned and operated.

- This complex, heterogeneous set of networks ***must*** operate seamlessly from "end to end" to support science and research collaborations that are distributed globally.
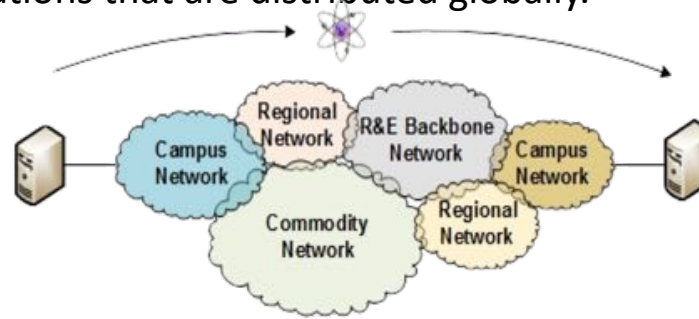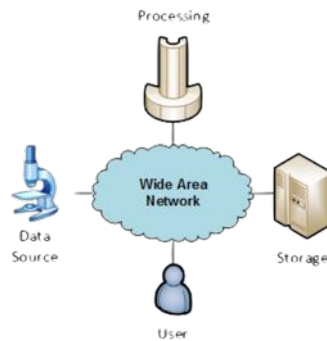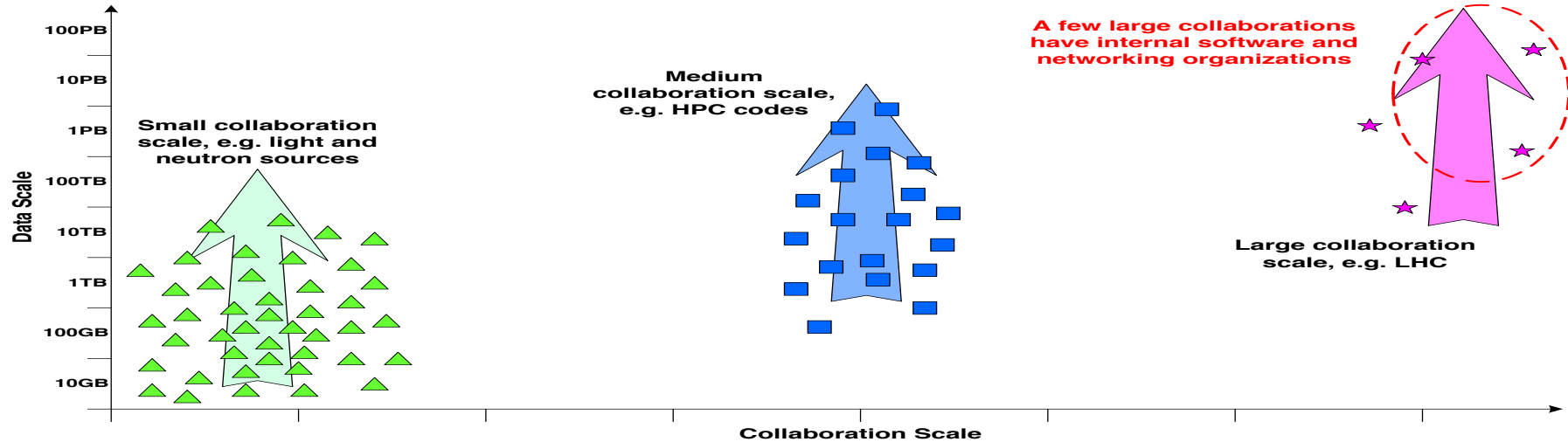

- Data mobility is required; there is no liquid market for HPC resources (people use what they can get – DOE, XSEDE, NOAA, etc. etc.)
  - To stay competitive, we must learn the use patterns, and support them
  - This may mean making sure your network, and the networks of others, are functional

Outline

- Problem Statement on Network Connectivity

- Supporting Scientific Users

- Network Performance & TCP Behaviors w/ Packet Loss

- What is perfSONAR

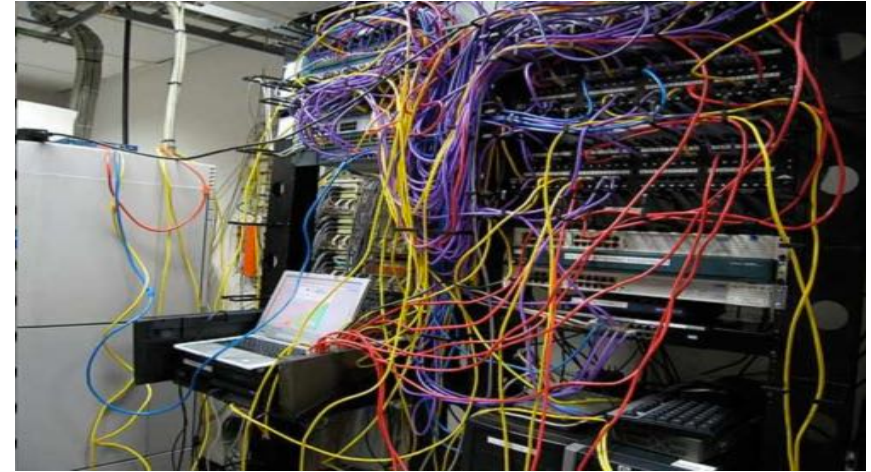- Architecture and installation options

Understanding Data Trends

perfSONAR

A few large collaborations have internal software and networking organizations

Small collaboration scale, e.g. light and neutron sources

Medium collaboration scale, e.g. HPC codes

Large collaboration scale, e.g. LHC

Data Scale

100PB, 10PB, 1PB, 100TB, 10TB, 1TB, 100GB, 10GB

Collaboration Scale

http://www.es.net/science-engagement/science-requirements-reviews/

ESnet ENERGY SCIENCES NETWORK   GÉANT   INDIANA UNIVERSITY   INTERNET2   UNIVERSITY OF MICHIGAN

Challenges to Network Adoption

- Causes of performance issues are complicated for users.

- Lack of communication and collaboration between the CIO's office and researchers on campus.

- Lack of IT expertise within a science collaboration or experimental facility

- User's performance expectations are low ("The network is too slow", "I tried it and it didn't work").

- Cultural change is hard ("we've always shipped disks!").

- Scientists want to do science not IT support

ESnet ENERGY SCIENCES NETWORK   GÉANT   IU INDIANA UNIVERSITY   INTERNET2®   M UNIVERSITY OF MICHIGAN

Outline

- Problem Statement on Network Connectivity

- Supporting Scientific Users

- Network Performance & TCP Behaviors w/ Packet Loss

- What is perfSONAR

- Architecture and installation options
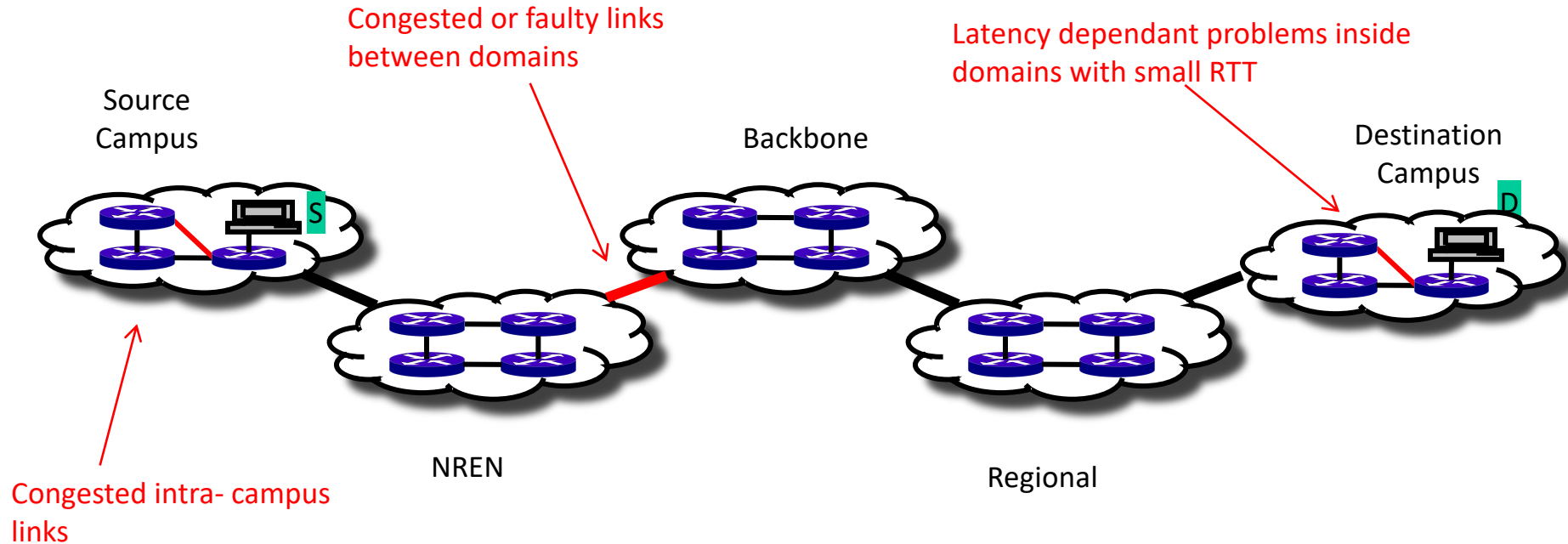
Lets Talk Performance ...

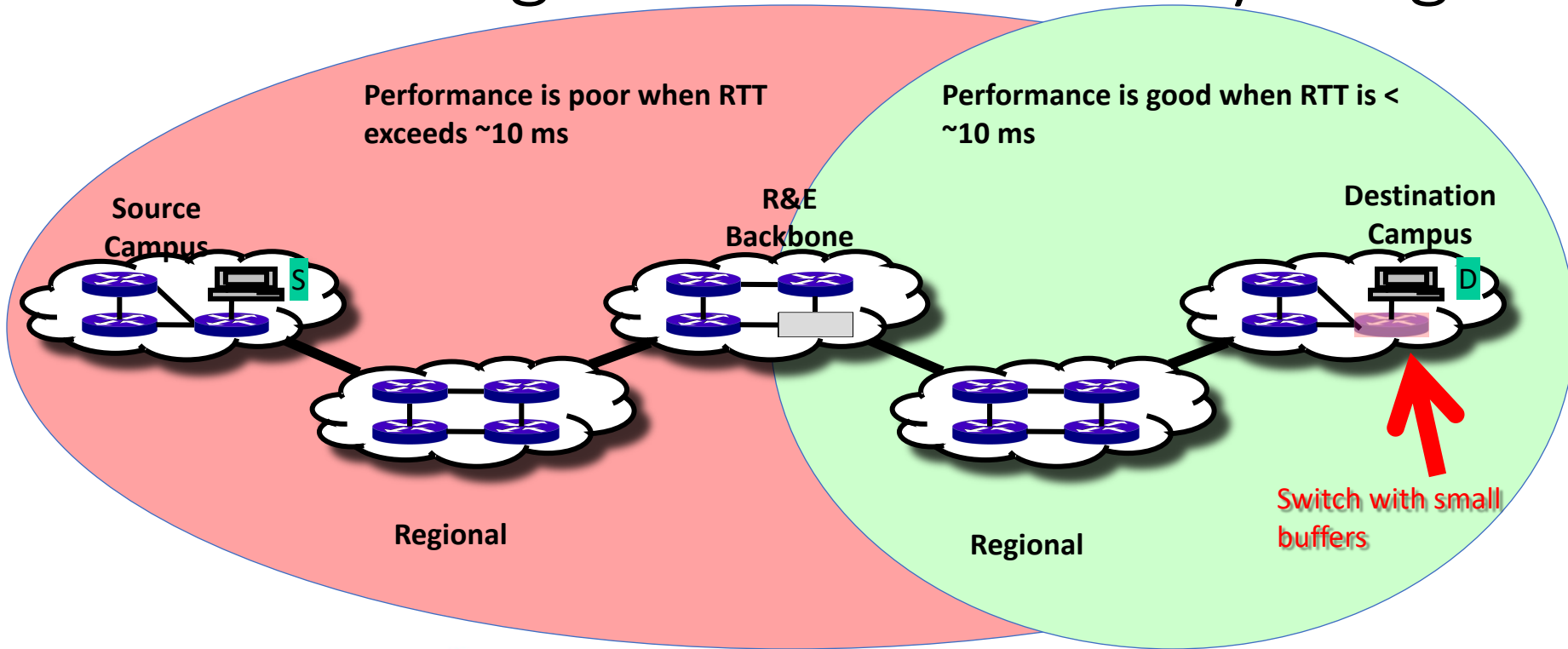"In any large system, there is always something broken."

Jon Postel

- Modern networks are occasionally designed to be *one-size-fits-most*
- e.g. if you have ever heard the phrase "converged network", the design is to facilitate CIA (Confidentiality, Integrity, Availability)
  - This is not bad for protecting the HVAC system from hackers.
- Its all TCP
  - Bulk data movement is a common thread (move the data from the microscope, to the storage, to the processing, to the people – and they are all sitting in different facilities)
  - This fails when TCP suffers due to path problems (***ANYWHERE*** in the path)
  - its easier to work with TCP than to fix it (20+ years of trying…)
- TCP suffers the most from unpredictability; Packet loss/delays are the enemy
  - Small buffers on the network gear and hosts
  - Incorrect application choice
  - Packet disruption caused by overzealous security
  - Congestion from herds of mice
- It all starts with knowing your users, and knowing your network
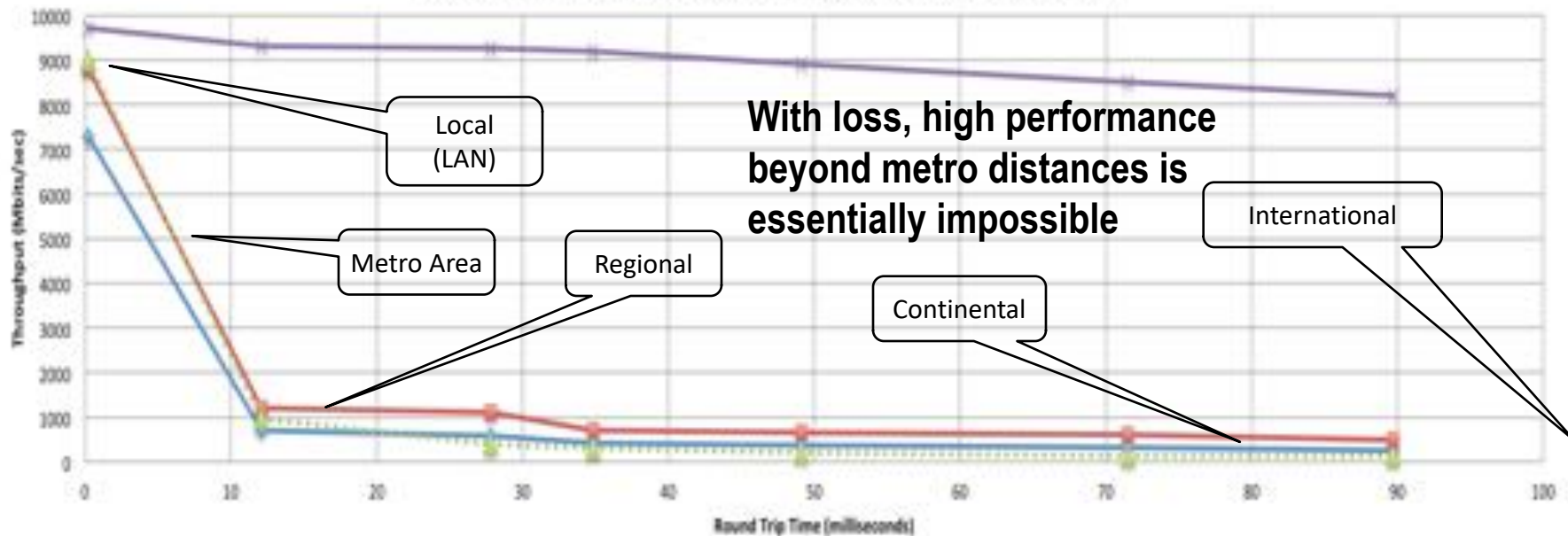
Where Are The Problems?

perfSONAR

Congested or faulty links between domains

Latency dependant problems inside domains with small RTT

Source Campus

Backbone

Destination Campus

S

D

NREN

Regional

Congested intra- campus links

ESnet
ENERGY SCIENCES NETWORK

GÉANT

INDIANA UNIVERSITY

INTERNET2

UNIVERSITY OF MICHIGAN

Local Testing Will Not Find Everything

Performance is poor when RTT exceeds ~10 ms

Performance is good when RTT is < ~10 ms

Source Campus

R&E Backbone

Destination Campus

Regional

Regional

Switch with small buffers

Soft Failures Cause Packet Loss and Degraded TCP Performance

Throughput vs. Increasing Latency with .0046% Packet Loss

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

Problem Statement:
Hard vs. Soft Failures

- "Hard failures" are the kind of problems every organization understands
  - Fiber cut
  - Power failure takes down routers
  - Hardware ceases to function
- Classic monitoring systems are good at alerting hard failures
  - i.e., NOC sees something turn red on their screen
  - Engineers paged by monitoring systems
- "Soft failures" are different and often go undetected
  - Basic connectivity (ping, traceroute, web pages, email) works
  - Performance is just poor
- How much should we care about soft failures?

Causes of Packet Loss

- Network Congestion
  - Easy to confirm via SNMP, easy to fix with $$
  - This is not a 'soft failure', but just a network capacity issue
  - Often people assume congestion is the issue when it fact it is not.

- Under-buffered switch dropping packets
  - Hard to confirm

- Under-powered firewall dropping packets
  - Hard to confirm

- Dirty fibers or connectors, failing optics/light levels
  - Sometimes easy to confirm by looking at error counters in the routers

- Overloaded or slow receive host dropping packets
  - Easy to confirm by looking at CPU load on the host

Outline

- Problem Statement on Network Connectivity

- Supporting Scientific Users

- Network Performance & TCP Behaviors w/ Packet Loss

- What is perfSONAR

- Architecture and installation options

But ... It's Not Just the Network

- Perhaps you are saying to yourself "I have no control over parts of my campus, let alone the 5 networks that sit between me and my collaborators"
  - Significant gains are possible in isolated areas of the OSI Stack

- Things "you" control:
  - Choice of data movement applications (say no to SCP and RSYNC)
  - Configuration of local gear (hosts, network devices)
  - Placement and configuration of diagnostic tools, e.g. *perfSONAR*
  - Use of the diagnostic tools

- Things that need some help:
  - Configuration of remote gear
  - Addressing issues when the diagnostic tools alarm
  - Getting someone to "care"

Network Monitoring

- All networks do some form monitoring.
  - 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

perfSONAR

- All the previous Science DMZ network diagrams have little perfSONAR boxes everywhere
  - The reason for this is that consistent behavior requires correctness
  - Correctness requires the ability to find and fix problems

    - ***You can't fix what you can't find***
    - ***You can't find what you can't see***
    - ***perfSONAR lets you see***

- Especially important when deploying high performance services
  - If there is a problem with the infrastructure, need to fix it
  - If the problem is not with your stuff, need to prove it
    - Many players in an end to end path
   - Ability to show correct behavior aids in problem localization
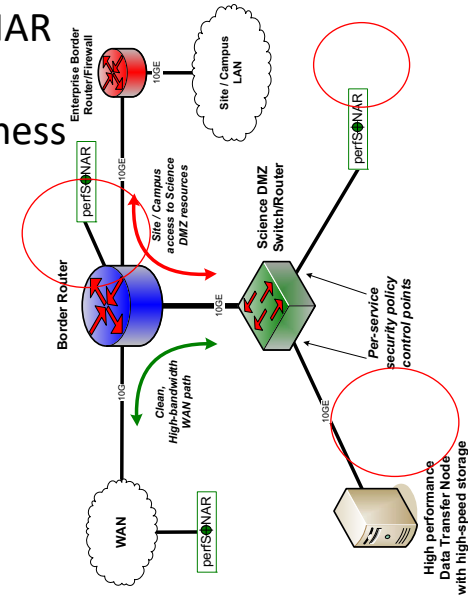
What is perfSONAR?

- 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

perfSONAR HISTORY TIMELINE
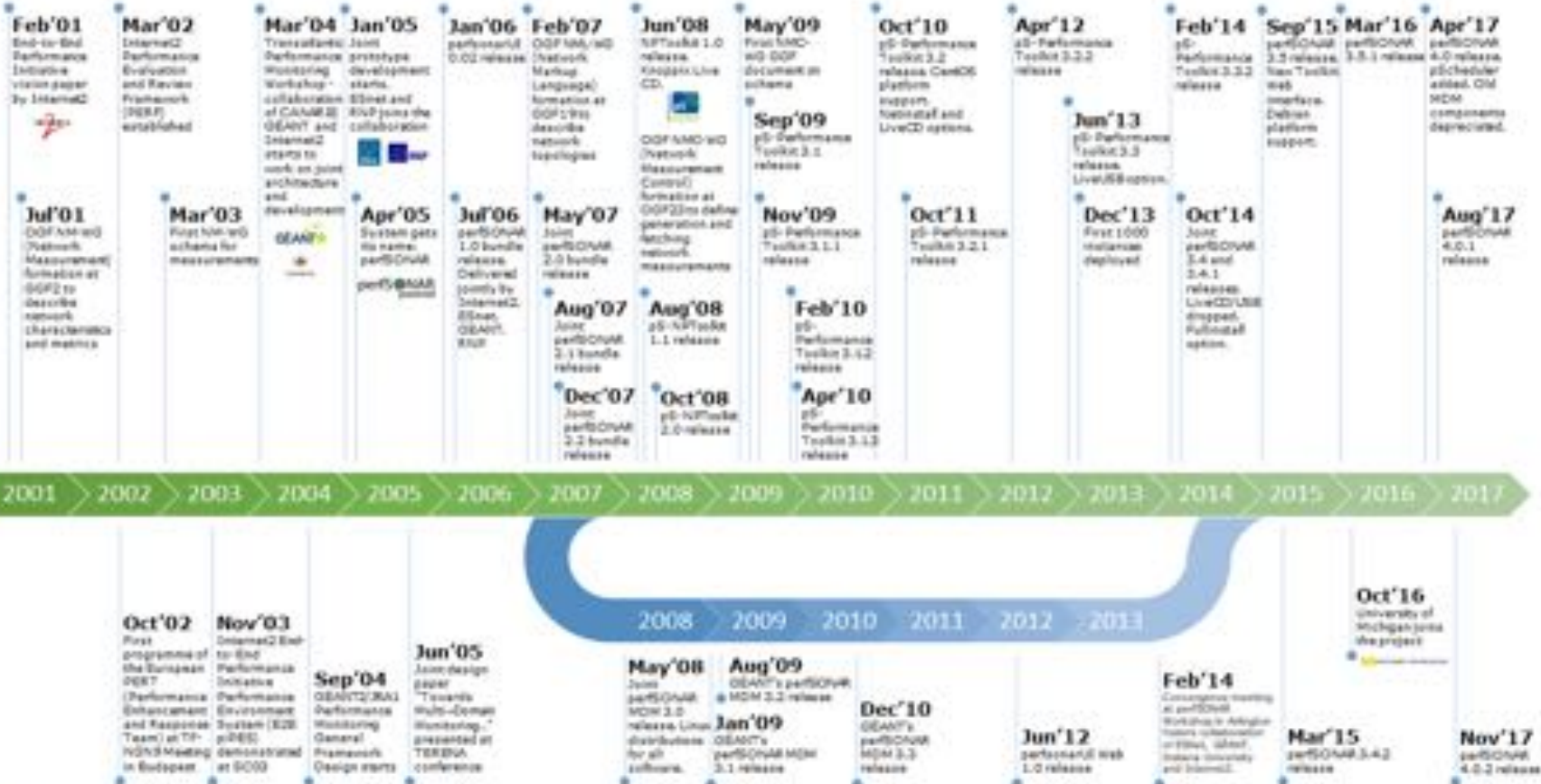
perfSONAR History

- perfSONAR can trace its origin to the Internet2 "End 2 End performance Initiative" from the year 2000.

- What has changed since 2000?
  - The Good News:
    - TCP is much less fragile; Cubic is the default CC alg, autotuning is and larger TCP buffers are everywhere
    - Reliable parallel transfers via tools like Globus Online
    - High-performance UDP-based commercial tools like Aspera
  - The Bad News:
    - The *wizard gap* is still large
    - Jumbo frame use is still small
    - Under-buffered and switches and routers are still common
    - Under-powered/misconfigured firewalls are common
    - Soft failures still go undetected for months
    - User performance expectations are still too low

Simulating Performance

- It's infeasible to perform at-scale data movement all the time – as we see in other forms of science, we need to rely on simulations

- Network performance comes down to a couple of key metrics:
  - Throughput (e.g. "how much can I get out of the network")
  - Latency (time it takes to get to/from a destination)
  - Packet loss/duplication/ordering (for some sampling of packets, do they all make it to the other side without serious abnormalities occurring?)
  - Network utilization (the opposite of "throughput" for a moment in time)

- We can get many of these from a selection of measurement tools – enter the perfSONAR Toolkit

ESnet
ENERGY SCIENCES NETWORK

GÉANT

INDIANA UNIVERSITY

INTERNET2®

UNIVERSITY OF MICHIGAN

© 2018, http://www.perfsonar.net

perfSONAR Toolkit

- The "perfSONAR Toolkit" is an open source implementation and packaging of the perfSONAR measurement infrastructure and protocols
  - http://docs.perfsonar.net/install_getting.html
- All components are available as RPMs, DEBs, and bundled as CentOS 7, Debian 8,9 or Ubuntu 14,16,18 -based packages (as for perfSONAR v. 4.1.2)
  - perfSONAR tools are much more accurate if run on a dedicated perfSONAR host
- Very easy to install and configure
  - Usually takes less than 30 minutes

Outline

- Problem Statement on Network Connectivity

- Supporting Scientific Users

- Network Performance & TCP Behaviors w/ Packet Loss

- What is perfSONAR

- Architecture and installation options

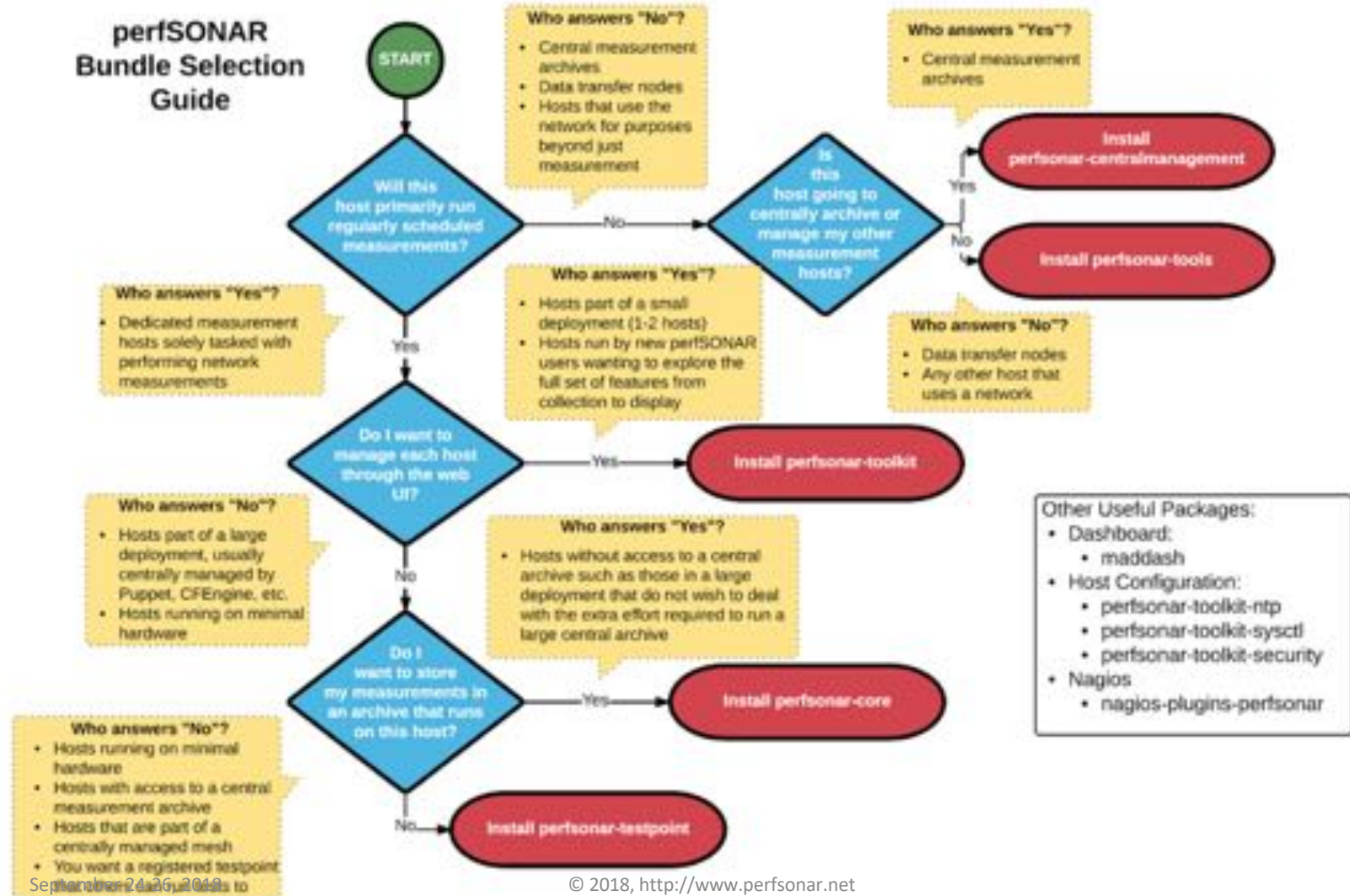# Architecture

© 2018, http://www.perfsonar.net
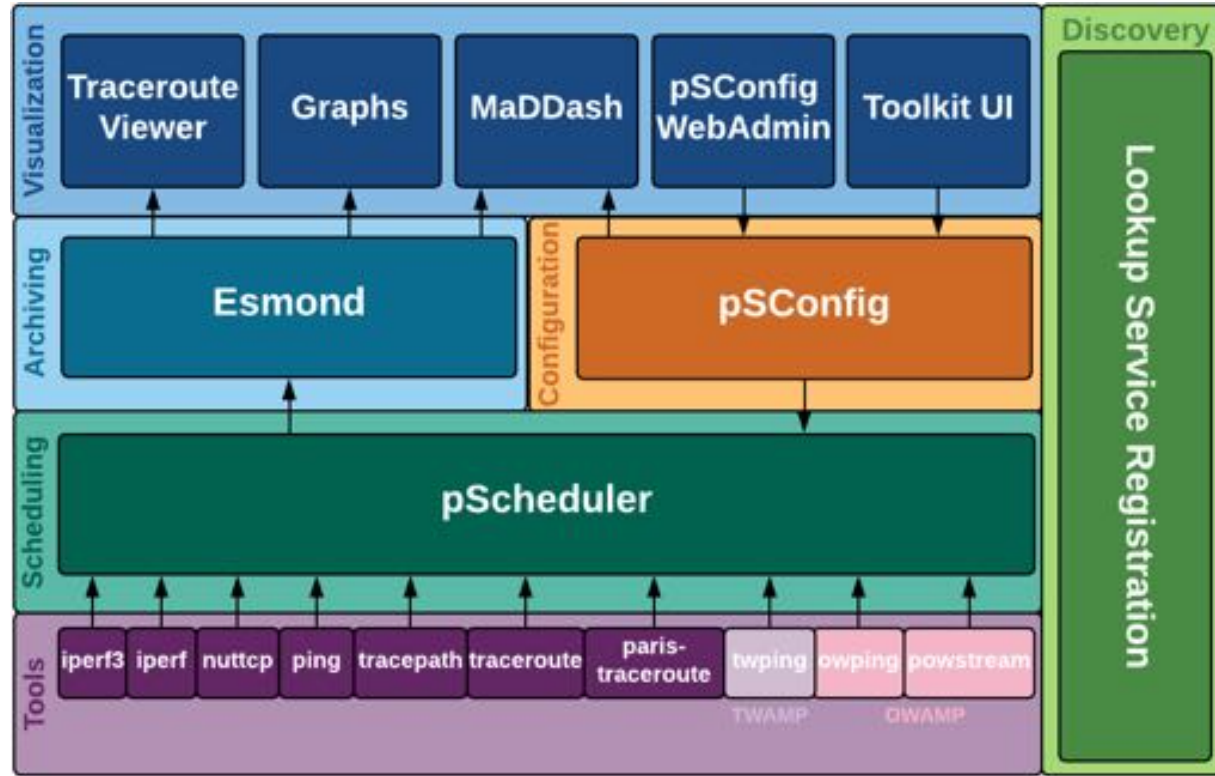
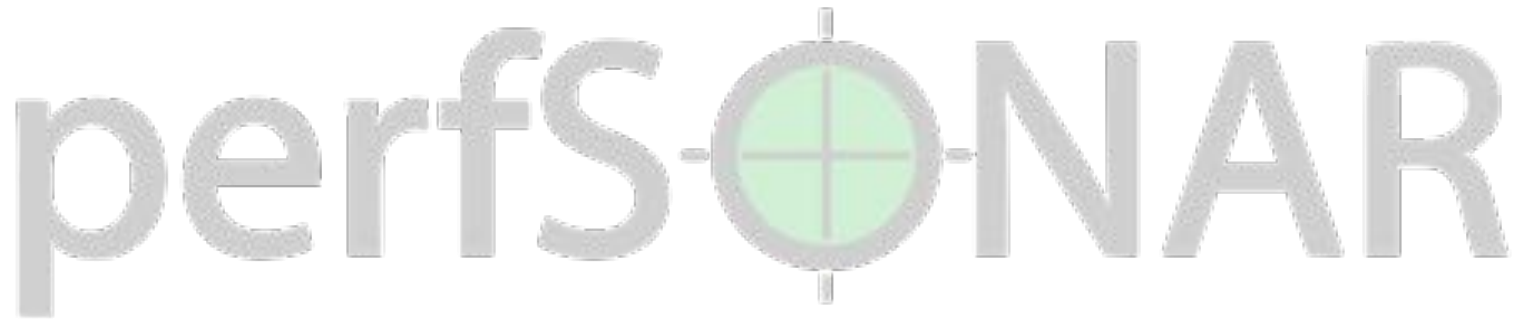# Install Options: Classic or Advanced

- CentOS 7 ISO image
  - Full toolkit install
  - Easy, all contained

- Want more control? Bundle of packages
  - perfsonar-tools
  - perfsonar-testpoint
  - perfsonar-core
  - perfsonar-toolkit
  - perfsonar-centralmanagement
  - + optional packages

Package bundles structure
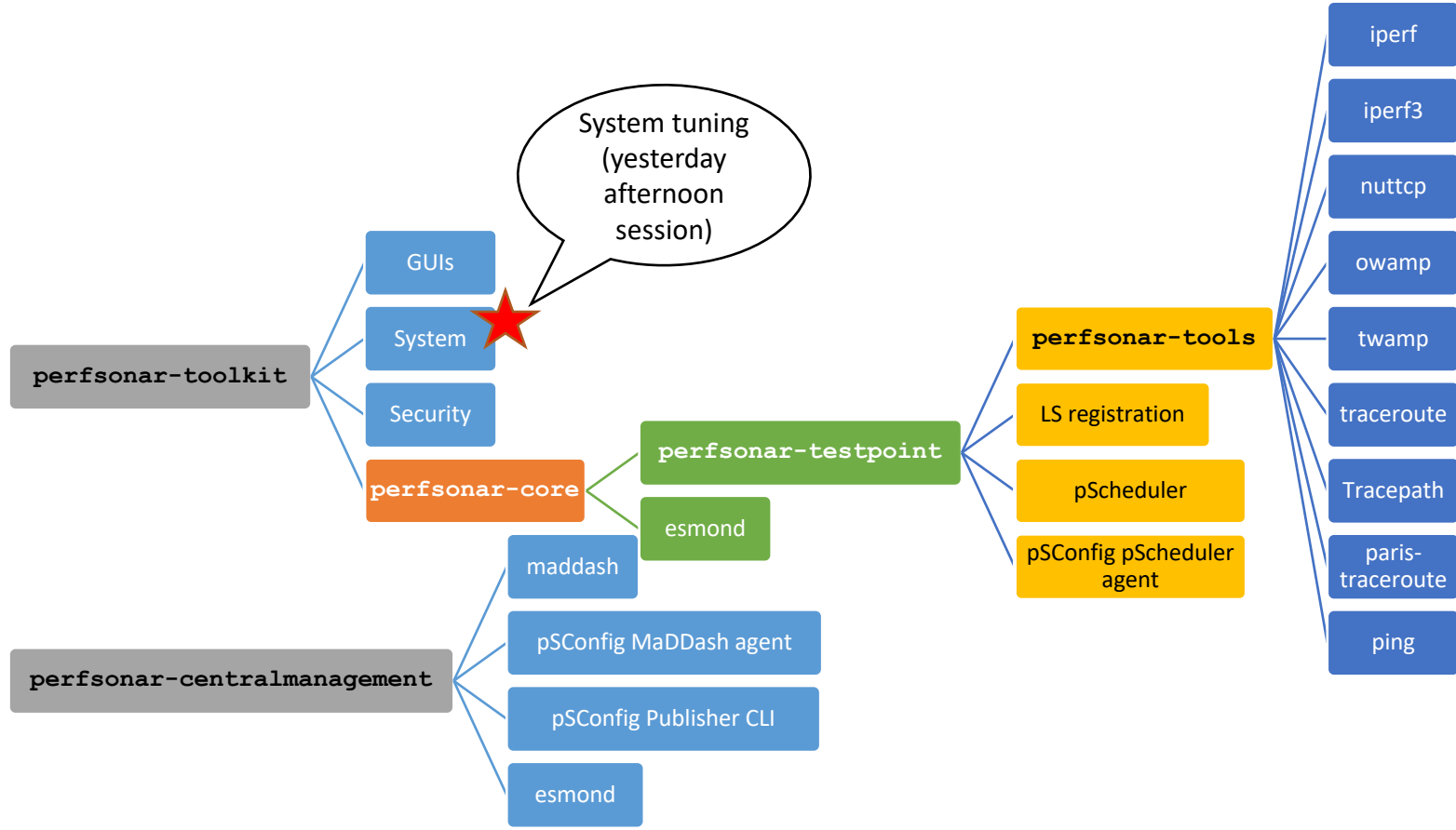
perfSONAR Bundle Selection Guide

© 2018, http://www.perfsonar.net

41

Introduction & Motivation

ASTRON perfSONAR training

Antoine Delvaux, PSNC, antoine.delvaux@man.poznan.pl
Szymon Trocha, PSNC, szymon.trocha@man.poznan.pl
24-26 September 2018