ACCESS | Advancing Innovation

# STEP-1: Network Security
# May 23, 2024

*David Wheeler, NCSA / University of Illinois*

OpenCI Award: OCI 21-38296

NSF

# Outline

- Security Framework
- Assessing Risk
- Firewalls
- Strategic Security

# Security Framework

# Science DMZ as a Security Architecture

- **Allows for better segmentation of risks, more granular application of controls to those segmented risks**
  - Limit risk profile for high-performance data transfer applications
  - Apply specific controls to data transfer hosts
  - 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

# Science DMZ Security

- **Goal**: Disentangle security policy and enforcement for science flows from security for business systems
- **Rationale**
  - Science data traffic is simple from a security perspective
  - Narrow application set on Science DMZ
    - Data transfer, data streaming packages
    - No printers, document readers, web browsers, building control systems, financial databases, staff desktops, etc.
  - Security controls that are typically implemented to protect business resources often cause performance problems
- **Separation allows each to be optimized**

# Performance is a Core Requirement

- Core information security principles
  - Confidentiality, Integrity, Availability (CIA)
- In data-intensive science, **performance** is an additional core mission requirement: CIA -> PICA
  - CIA principles are important, but ***if the performance isn't there the science mission fails***
  - This isn't about "how much" security you have, but how the security is implemented
  - Need to appropriately secure systems without performance compromises

# Motivation

- **The big myth:** The main goal of the Science DMZ is to avoid firewalls and other security controls.
  - Leads to all sorts of odd (and wrong) claims like:
    - “Our whole backbone is a Science DMZ because there is no firewall in front of the backbone.”
    - “The Science DMZ doesn’t allow for **any** security controls.”
    - “The Science DMZ requires a default-permit policy.”
- **The reality:** The Science DMZ emphasizes reducing degrees-of-freedom, reducing the number of network devices (including middleboxes) in the path, eliminating devices that can’t perform, and ensuring that the devices that remain in the path are capable of large-scale data-transfer caliber performance

# Motivation

- Goal is to break down this myth by viewing the Science DMZ *as a security architecture.*
- That is, by thinking about Science DMZ as a form of security *control,* not just something that needs to be controlled.
- At the same time, Science DMZ enables us to do a better job of risk-based security through segmentation.

# Risk-based vs. Control-based Security

- Risk-based (ideal form):
  - Identify risks (impact and likelihood over a period of time)
  - Identify and/or create controls that are specifically designed to mitigate those risks
  - Apply controls as necessary
- Control-based (ideal form):
  - Select controls from a checklist or standard
  - Controls are, or at one point were, believed to mitigate a general set of risks.
  - Apply controls (more controls==better security)
- So why do we still practice control-based security in many instances?
  - Risk based security is actually pretty hard.

# Assessing Risk - Data

- **Simple goal:** Protecting data from destructive forces, and from the unwanted actions of unauthorized users
  - Know what you want to protect
    - PII (personally identifiable info)
      - *Information that can be used on its own or with other information to identify, contact, or locate a single person, or to identify an individual in context*
    - Master class of things like PHI (health information)
- All data types shouldn't be treated in the same manner
  - Not all research data is created equal – talk to the researchers about this.

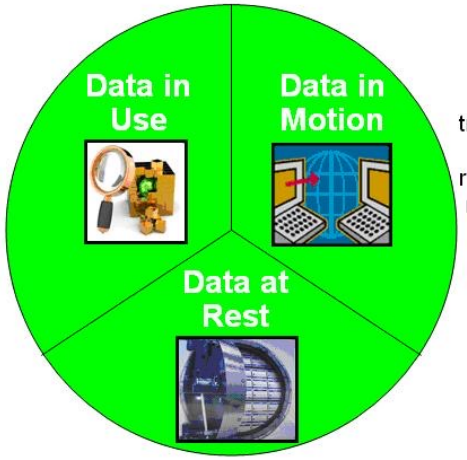
# Network Segmentation

- Think about residence hall networks, business application networks, and the networks that are primarily in research areas:
- The risk profiles are clearly different, so it makes sense to segment along these lines.
- Your institution may already be doing this for things like HIPAA and PCI-DSS. Why? *Because of the controls!*
- The Science DMZ follows the same concept, from a security perspective.
- Using a Science DMZ to segment research traffic (especially traffic from specialized research instruments) can actually *improve* the campus security posture.

# Assessing Risk - Data

- The data lives in 3 major states:

**Data in Use:**
Active data under constant change stored physically in databases, data warehouses, spreadsheets etc.




**Data in Motion:**
Data that is traversing a network or temporarily residing in computer memory to be read or updated.

**Data at Rest:**
Inactive data stored physically in databases, data warehouses, spreadsheets, archives, tapes, off-site backups etc.

- How does the motion impact the security approach?

# Assessing Risk - Data

- Research Data
  - Imagine a climate scientist – they download a lot of observation data from collaborators. When they are done, they normally delete it. What needs to be secured?
  - The data itself has 2 portions:
    - Metadata: information on where/when it was gathered. Could be PII (e.g., you should ask)
    - Data: time/value pairs. Without the metadata, this is basically worthless. Is there a way to tie it back? Is the metadata PII?
  - Protection strategy:
    - Protect the metadata at rest/in motion
    - Data can be given a pass on protections depending on how it relates back to metadata

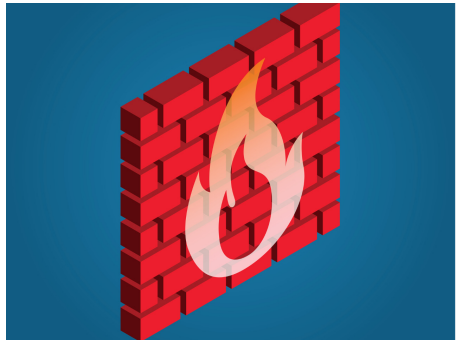
# Firewalls

# Science DMZ Placement Outside the Enterprise Firewall

- Why? For performance reasons
  - Specifically: **Science DMZ traffic does not traverse the firewall data plane**
  - This has nothing to do with whether packet filtering is part of the security enforcement toolkit
- Lots of heartburn over this, especially from the perspective of a conventional firewall manager
  - Organizational policy directives can **mandate** firewalls
  - Firewalls are designed to protect converged enterprise networks
  - Why would you put critical assets outside the firewall??
- **The answer:** Firewalls are typically a poor fit for high-performance science applications

# Typical Firewall Internals

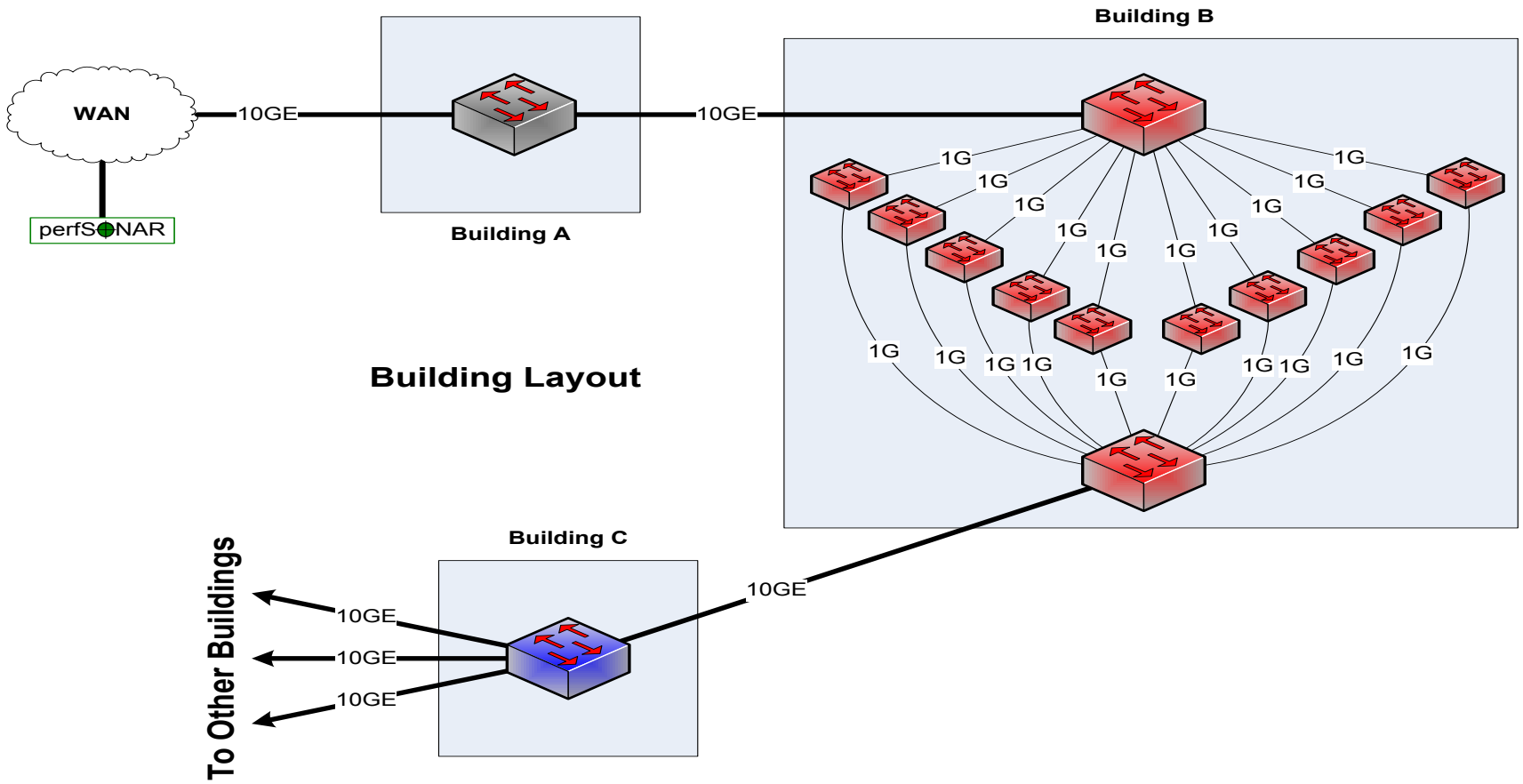
- Composed of a set of processors which inspect traffic in parallel
  - Traffic distributed among processors so all traffic for a particular connection goes to the same processor
  - Simplifies state management
  - Parallelization scales deep analysis
- Excellent fit for enterprise traffic profile
  - High connection count, low per-connection data rate
  - Complex protocols with embedded threats
- Each processor is a fraction of firewall link speed
  - Significant limitation for data-intensive science applications
  - Overload causes packet loss – performance crashes



# Thought Experiment

- We're going to do a thought experiment
- Consider a network between three buildings – A, B, and C
  - This is supposedly a 10Gbps network end to end (look at the links on the buildings)
  - Building A houses the border router – not much goes on there except the external connectivity
  - Lots of work happens in building B – so much so that the processing is done with multiple processors to spread the load in an affordable way, and aggregate the results afterwards
  - Building C is where we branch out to other buildings
- Every link between buildings is 10Gbps – this is a 10Gbps network, right???

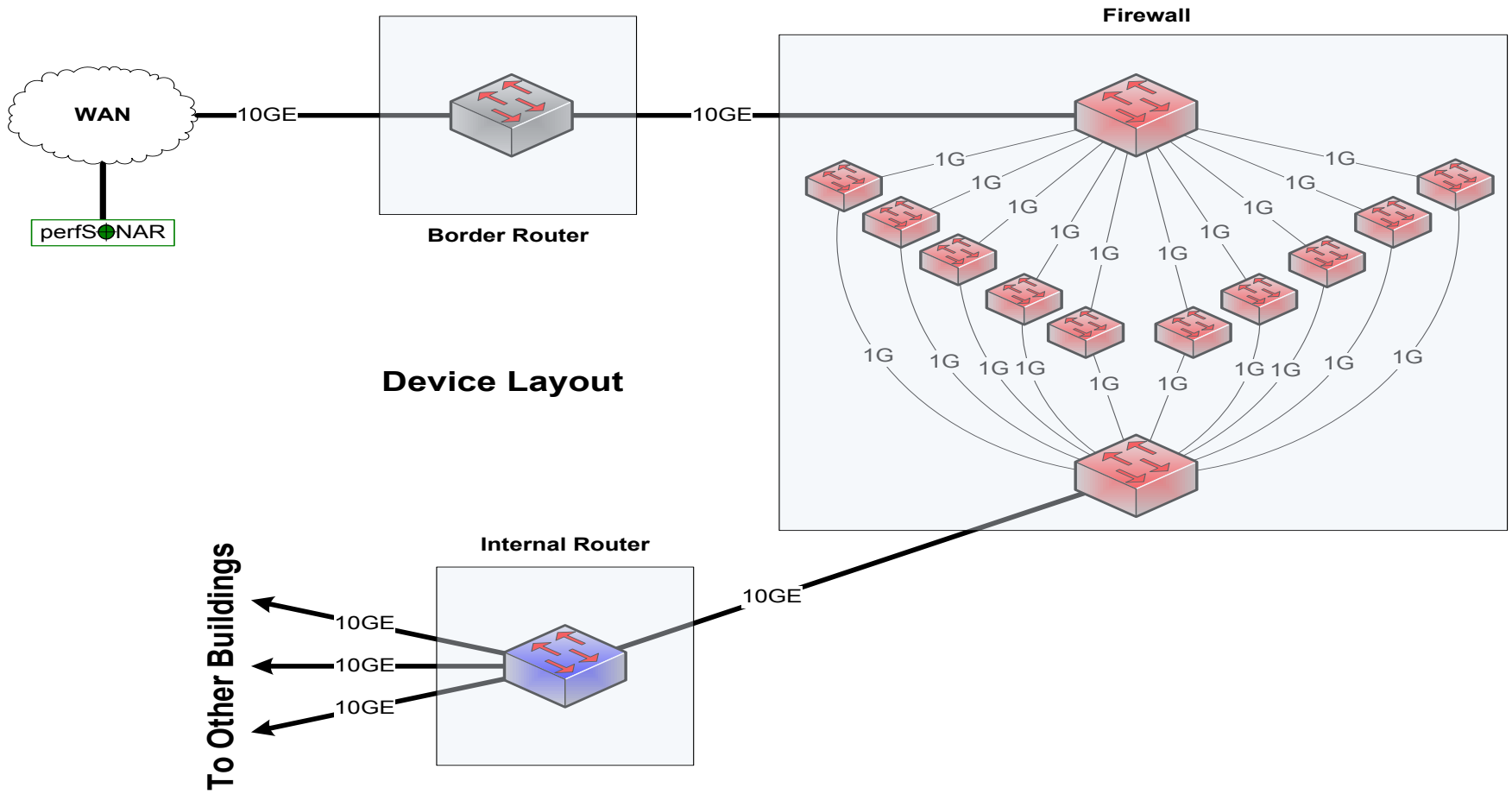
# Notional 10G Network Between Buildings



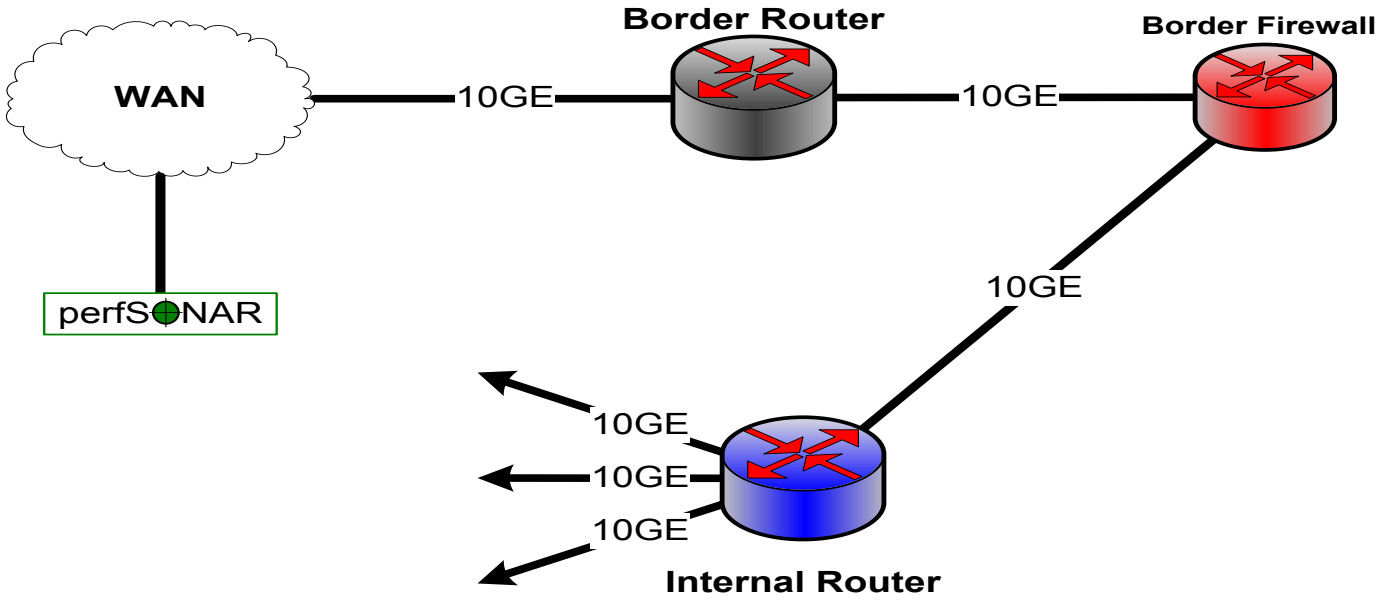
# Clearly Not A 10Gbps Network

- If you look at the inside of Building B, it is obvious from a network engineering perspective that this is not a 10Gbps network
  - Clearly the maximum per-flow data rate is 1Gbps, not 10Gbps
  - However, if you convert the buildings into network elements while keeping their internals intact, you get routers and firewalls
  - What firewall did the organization buy?  What's inside it?
  - Those little 1G "switches" are firewall processors
- This parallel firewall architecture has been in use for years
  - Slower processors are cheaper
  - Typically fine for a commodity traffic load
  - Therefore, this design is cost competitive and common

# Notional 10G Network Between Devices



# Notional Network Logical Diagram



# Firewall Capabilities and Science Traffic

- Firewalls have a lot of sophistication in an enterprise setting
  - Application layer protocol analysis (HTTP, POP, MSRPC, etc.)
  - Built-in VPN servers
  - User awareness
- Data-intensive science flows don't match this profile
  - Common case – data on filesystem A needs to be on filesystem Z
    - Data transfer tool verifies credentials over an encrypted channel
    - Then open a socket or set of sockets, and send data until done (1TB, 10TB, 100TB, …)
  - One workflow can use 10% to 50% or more of a 10G network link
- Do we have to use a firewall?

# Firewalls as Router Access Control Lists

- When you ask a firewall administrator to allow data transfers through the firewall, what do they ask for?
  - IP address of your host
  - IP address of the remote host
  - Port range
  - That looks like an ACL to me – I can do that on the router!
- Firewalls make expensive, low-performance ACL filters compared to the ACL capabilities are typically built into the router
- Router ACLs do not drop traffic permitted by policy, while enterprise firewalls can (and often do)

# Security Without Enterprise Firewalls

- Data intensive science traffic interacts poorly with firewalls
- Does this mean we ignore security? NO!
  - We must protect our systems
  - We just need to find a way to do security that does not prevent us from getting the scienc done
- Key point – security policies and mechanisms that protect the Science DMZ should be implemented so that they do not compromise performance

# Strategic Security

# Systems View Of Science Infrastructure

- **Security is a component, not a gatekeeper**
- *Think about workflows*
- Think about the interfaces to data (tools, applications)
  - How do collaborators access data?
  - How could they access data if the architecture were different?
- **Cost/benefit**
  - What is a new cancer breakthrough worth?
  - $30k for a few DTNs – what is that in context?
- **Risks**
  - What risks do specific technologies mitigate?
  - What are the opportunity costs of poor performance?

# Other Security Mechanisms: ACLs

- **Aggressive access lists**
  - More useful with project-specific DTNs
  - Exchanging data with a small set of remote collaborators = ACL is fairly easy to manage
  - Large-scale data distribution servers = difficult/time-consuming to handle
    (but then, the firewall ruleset for such a service would be, too)
- **Limitation of the application set**
  - Makes it easier to protect
  - Keeps unnecessary applications off the DTN (but watch for them anyway using a host Intrusion Detection System – take violations seriously)

# Other Security Mechanisms: Network IDS

- Intrusion Detection Systems (IDS)
  - One example is zeek (formerly bro)
    - http://zeek.org/
  - Zeek is high-performance and battle-tested
    - Zeek protects several high-performance national assets
    - Zeek can be scaled with clustering:
      - https://docs.zeek.org/en/master/cluster-setup.html
  - Other IDS solutions are available also

# Other Security Mechanisms: Host IDS

- Using a Host IDS is recommended for hosts in a Science DMZ

- Several open-source solutions have been recommended:
  - OSSec: http://www.ossec.net/
  - Rkhunter: http://rkhunter.sourceforge.net (rootkit detection + FIM)
  - chkrootkit: http://chkrootkit.org/
  - Logcheck: http://logcheck.org (log monitoring)
  - Fail2ban: http://www.fail2ban.org/wiki/index.php/Main_Page
  - denyhosts: http://denyhosts.sourceforge.net/

# Collaboration Within The Organization

- **All stakeholders should collaborate on Science DMZ design, policy, and enforcement**
- **The security people have to be on board**
  - Political cover for security officers
  - If the deployment of a Science DMZ is going to jeopardize the job of the security officer, expect pushback
- **The Science DMZ is a strategic asset, and should be understood by the strategic thinkers in the organization**
  - Changes in security models
  - Changes in operational models
  - Enhanced ability to compete for funding
  - Increased institutional capability – greater science output

# Summary

- **Think about what the Science DMZ is trying to do.**
  - Improve performance, both by removing impediments and improving the performance of the devices that must be in line
  - Ease troubleshooting
  - In general, reduce degrees of freedom from science networks
  - Maximize performance **and** security **and** resiliency

# Questions?