Improving Data Mobility Through Modern Security Infrastructure

Jason Zurawski, Ezra Kissel, George Robb
Energy Sciences Network (ESnet)
Lawrence Berkeley National Laboratory (LBNL)
Berkeley, CA, USA
{zurawski, kissel, grobb3}es.net

Corey Eichelberger, Nathaniel Mendoza
Texas Advanced Computing Center (TACC)
University of Texas at Austin
Austin, TX, USA
{ceichelberger, nmendoza}tacc.utexas.edu

***Abstract*—Modern scientific use cases require effective mechanisms to transfer data between sources, analysis facilities, collaborators, and long-term archives. Data transfer is most efficient when dedicated machines, on purpose built scientific networks, are deployed to support the activity; a Science DMZ is one example of an approach to ensure high-performance use cases are supported. It remains the case that "converged network" design, e.g., general purpose infrastructure to support all use cases, features security infrastructure that impedes efficient data transfer which impacts performance of TCP data movement tools.**

This paper will investigate the current state of performance degradation for data movement when deployed through network security infrastructure. A set of tests are being prepared for deployment at the SC24 conference being held in November 2024 in Atlanta GA, USA, that will evaluate the performance of data transfer tools through a heterogeneous network environment: SCinet. These results will help guide future design and deployment of network infrastructure to support scientific activities.

***Index Terms*—Network architectures, Network protocols, Data management systems**

I. Introduction

[3], [10] cite a number of patterns that represent scientific domains that have implemented low-latency, near-real-time workflows, spanning multiple domains. These use cases leverage access to high-throughput network connections to join together experiments, computing, and their end-user communities seamlessly. Designs like this are now routine, and will continue to depend on the underlying technologies (e.g., high-capacity networks, cutting edge computing hardware to support AI/ML, and large amounts of data storage) to execute on their scientific missions. A common requirement to deliver this vision exists in the form of predictable and routine data transfer between the participating entities.

A Science DMZ [1] is a portion of a network, built at or near a campus local network perimeter that is designed such that the equipment, configuration, and security policies, are optimized for high-performance workflows and large data sets [4]. This environment is mostly free from competing traffic flows and complex security middleware, such as firewalls or Intrusion Detection Systems (IDSs), that may impede data transfer performance. High performance servers, called Data Transfer Nodes (DTNs) are connected directly to this infrastructure, handling all data ingest/export tasks.

Over the past 20 years, a number of universities and organizations have adopted this model for their research support, but it remains the case that older "converged" network designs, featuring incapable infrastructure components that may impede efficient data transfer, may be present. Slow firewalls, under-buffered network devices, or other technology in the network path that are not able to handle high speed flows have a predictably negative impact the performance of TCP data movement tools, and on the overall throughput of a scientific workflow. [12] found that a Science DMZ design exhibits lower latency, higher throughput, and lower jitter behaviors.

SCinet [2], the dedicated network infrastructure that supports the SC Conference, temporarily becomes the most powerful and advanced network on Earth each November. This endeavor connects the SC community to the world. SCinet volunteers who deliver advanced networking each year are setting an ambitious goal of deploying a series of modern firewall architectures at SC24. While the necessary technology is widely available, and understood, the implications of deployment to support more than 18,000 users, each with multiple devices of different operating environments and ages, presents a unique technology and policy challenge.

The authors, working with the SCinet team, designed and implemented a set of tests to measure network performance through evolving security hardware. Results were collected during the operation of SCinet to show the impacts of security infrastructure on high speed networks, and the impacts they may have on scientific workflows.

II. Data Mobility

Because modern scientific use cases require efficient mechanisms to transfer data, modern cyberinfrastructure must be constructed on a robust set of hardware and software services that seamlessly enables productive and predictable outcomes. Efficient data movement enables distributed research by facilitating large volumes of data to travel between experimental facilities, analysis centers, and long-term storage. Added friction in network design can impact the overall effectiveness of these workflows: choice of network hardware, security infrastructure, server architecture, and data mobility software

all play a key role in ensuring success. Data transfer is most effective when dedicated machines on purpose built scientific networks, using advanced tools, are deployed to support the activity.

The basic Science DMZ model has been successfully implemented in numerous scenarios; and these efforts have been notably recognized by the National Science Foundation, which has awarded multiple rounds of funding (as Campus Cyberinfrastructure (CC*) programs [11]) to U.S. academic institutions to construct Science DMZ environments on their campuses to support research at scale.

A. Background

The capabilities required to effectively deploy and support high-performance science applications are based on having access to networks that support high bandwidth operation and emphasizes operational soundness and a focus on information security. This infrastructure must not compromise on expected performance baselines, or it becomes a hinderance to the scientific workflows it is designed to support. Security requirements will emerge from the need to ensure correctness, prevent misuse, and to avoid embarrassment or other negative publicity that can compromise the reputation of the site or the science.

A Science DMZ provides an environment mostly free from competing traffic flows and complex security middleware, and features high performance servers designed to handle all data ingest/export tasks. While the DMZ model benefits researchers, the benefits are not automatic - careful network tuning based on specific use cases is still required.

B. Science DMZ Architecture

The Science DMZ architecture meets these needs by instantiating a simple and scalable network enclave that explicitly accommodates high-performance science applications, while explicitly excluding general-purpose computing and the additional complexities that go with it. Ideally, the Science DMZ is connected directly to a network border, in order to minimize the number of devices that must be configured to support high-performance data transfer and other scientific applications. Achieving high performance is very difficult to do with system and network device configuration defaults, and the location of the Science DMZ at the site perimeter simplifies the system and network tuning processes.

A simple Science DMZ has several essential components. These include dedicated access to high-performance wide area networks and advanced services infrastructures, high-performance network equipment, and dedicated science resources such as Data Transfer Nodes. Fig. 1 shows a notional diagram of a simple Science DMZ showing these components, along with data paths.

A simple Science DMZ has several essential components. These include dedicated access to high-performance wide area networks and advanced services infrastructures, high-performance network equipment, test and measurement infrastructure provided by perfSONAR [13], [14], and dedicated

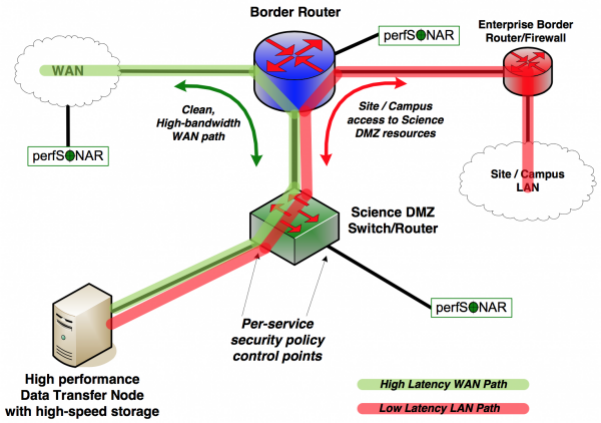

Fig. 1. Science DMZ Architecture

science resources such as DTNs. Fig 1 shows a notional diagram of a simple Science DMZ showing these components, along with data paths.

The Science DMZ architecture encourages a security posture that is implemented by policies and techniques that should avoid the use of a network firewalls; an approach that is accepted and accepted throughout the world when implementing a converged network design. Rather than relying on a single device (e.g., the firewall) to address all security needs, servers in the Science DMZ are protected by several straightforward cybersecurity concepts and mechanisms that can be tailored to fit the institution's needs: reduction in services exposed, removal of "degrees-of-freedom" on the network (e.g., implementation of IP address filters within the host and the network switching devices), and ongoing monitoring to ensure expected behavior. Placing these high bandwidth transfers outside the general-purpose network also has the benefit of reducing the load on enterprise network devices, and the ability to support a wide range of security postures [15].

C. Data Transfer Capabilities

The computer systems used for wide area data transfers perform far better if they are purpose-built and dedicated to the function of wide area data transfer. These DTNs are typically PC-based Linux servers built with high-quality components and configured specifically for wide area data transfer. The DTN can also have access to local storage, whether it is a local high-speed disk subsystem, a connection to a local storage infrastructure such as a Storage Area Network (SAN), or the direct mount of a high-speed parallel filesystem such as Lustre or GPFS. To mitigate security risks, no general-purpose computing tasks are allowed on the DTN; e.g., only tools required to manage data mobility.

In the most general use case, the DTN is connected directly to a high-performance Science DMZ network infrastructure, which in turn is connected directly to the border router. The DTN's job is to efficiently move science data, and must be properly designed and tuned for this task. The steps

required to tune a DTN for optimal performance can prove challenging. Considerations such as hardware configuration (e.g., CPU clock speed, main memory quantity and speed, network capacity), software choice (e.g., Globus, XRootd, or versions of default data movement tools such as SSH/SCP with high-performance patches applied [5]–[7]), and overall system configuration (e.g., choice of TCP algorithms, allocation of system memory, software-based based "pacing" that can be used to better match the bottleneck network speed), must be considered during the deployment and validation steps.

D. D. Friction form Security Infrastructure

To better understand the reasoning behind the aforementioned steps to design a Science DMZ, it's necessary to explore the root cause of performance abnormalities on networks via the protocols they use to communicate. The Transmission Control Protocol (TCP) [16] of the TCP/IP protocol suite is the primary transport protocol used for the reliable transfer of data between applications. TCP is robust in many respects; in particular it has sophisticated capabilities for providing reliable data delivery in the face of packet loss, network outages, and network congestion. However, the very mechanisms that make TCP so reliable also make it perform poorly when network conditions are not ideal.

TCP interprets packet loss as network congestion, and reduces its sending rate when loss is detected. In practice, even a tiny amount of packet loss is enough to dramatically reduce TCP performance, and thus increase the overall data transfer time. When applied to large tasks, this can mean the difference between a scientist completing a transfer in days rather than hours or minutes. Because TCP interprets the loss as network congestion, it reacts by rapidly reducing the overall sending rate. The sending rate then slowly recovers due to the dynamic behavior of the control algorithms. Network performance can be negatively impacted at any point during the data transfer due to changing conditions in the network. This problem is exacerbated as the latency increases between communicating hosts. This is often the case when research collaborations sharing data are geographically distributed [1].

Two very common causes of TCP packet loss are firewalls and aggregation devices with inadequate buffering. An important note regarding TCP-based flows are that they rarely runs at an observed, or "average", speed; TCP flows are composed of bursts and pauses. These bursts are can be very close to the maximum data rate for the sending host's interface.

Firewalls are often built with an internal architecture that relies on a set of lower-speed processors to achieve an aggregate throughput. This architecture works well when the traffic is composed of a large number of low-speed flows (e.g., a typical converged network traffic pattern). However, this causes a problem when a host with a network interface that is faster than the firewall's internal processors emerges. Since the firewall must buffer the traffic bursts sent by the data transfer host until it can process all the packets in the burst, input buffer size is critical. Firewalls often have small input buffers because that is typically adequate for the traffic profile of a business network. If the firewall's input buffers are too small to hold the bursts from the science data transfer host, the user will suffer severe performance problems caused by packet loss.

III. SCinet

SCinet is a global collaboration of networking experts who provide the fastest and most powerful volunteer-built network in the world for the SC Conference. Designed and created from new technology requirements each year, the SCinet network brings together experts who provide a platform that connects attendees and exhibitors to the world [2].

A. Background

SCinet has become more than a research network. It provides wired and wireless network connectivity to all conference attendees while in the host city's convention center. Thousands of attendees and presenters, each bringing numerous devices, expect and depend on SCinet to provide a reliable, high-speed, open network infrastructure.

B. Network Architecture

The SCinet Network Architecture is designed to address two core use cases:

- Operational network that supports connectivity for approximately 18,000 attendees, volunteers, and staff
- Research-oriented network that supports high-performance demonstrations around the world

The SCinet infrastructure relies on optical transport provided by six wide area network (WAN) providers, delivered over four different transportation systems. This heterogeneity of technology is a core strength of SCinet and something the volunteers take pride in yearly: interoperability across platforms helps build understanding of how each will operate in a non-conference scenario.

C. SC24 Achievements

At SC24 in Atlanta, GA, SCinet was comprised of more than 200 volunteers hailing from 9 countries, 34 states, and 114 institutions. The SCinet teams installed nearly 13 miles of fiber, over 450 wireless access points, and delivered a WAN capacity of 8.42 terabits per second (Tbps). All of this was accomplished following the SCinet creed: one year to design, one month to build, one week to operate, and one day to tear down.

IV. Experimental Results

For SC24, several demonstrations will focus on real-world use cases for data mobility; some of which will feature experiments that try to deliver realistic and performant use cases for migration of data between SC24, and collaborators worldwide. The objective of experimentation is to simulate the common components of a scientific workflow, and evaluate the impacts of friction-inducing network components in the path. In doing so, we are attempting to characterize:

- How security friction impacts performance

- Ways security friction can be mitigated through non-disruptive means (e.g., configuration changes, new approaches to data mobility)

Demonstrators constructed a set of tests between a well-connected DTNs internal, and external, to SCinet. Baselines were established to understand the un-impacted ideal performance expectation, and compared against the same tests run through a set of three network security devices. Special attention will be paid to ways that the network security devices behave when faced with different protocols (IPv4, IPv6), and the impacts of packet size (1500 versus 9000 Byte MTUs).

A. SC24 Architecture

The SC24 architecture was finalized in October of 2024 One notable change from previous generations of SCinet is the construction of a dedicated Science DMZ enclave, and the allocation of experimental hardware to create the "testbed" environment to evaluate network device performance (pictured on the left side).

Fig. 2 shows the final experimental setup of a set of security devices that will be evaluated. These are located in the "Firewall Sandbox" and consist of:

- Cisco Secure Firewall 4245 (e.g., "fw_A")
- Palo Alto PA-7500 (e.g., "fw_B")
- Fortinet FortiGate 4801F (e.g., "fw_C")

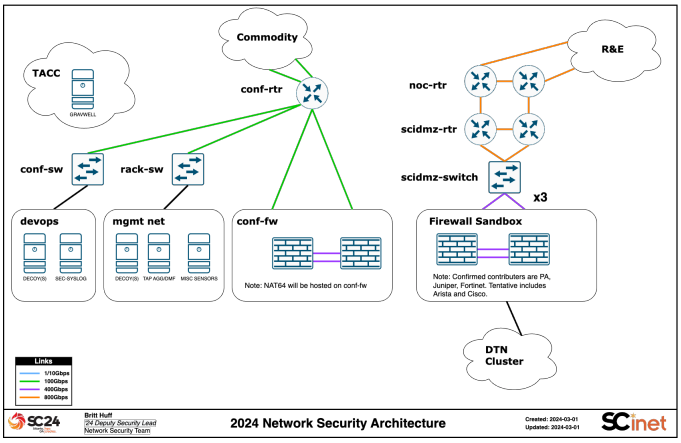

Fig. 2. Final SC24 Firewall architecture, and experimental setup.

A DTN was positioned behind these three security devices, along with a set of switches and routers, before they are able to communicate with the outside world. The DTNs used for testing:

- 2x AMD EPYC Milan 73F3 (16 cores each) with 3.5Ghz clock speed
- 256 GB RAM
- 25TB Data Disk: 10x Micron 9300 MAX 3.2TB U.2/2.5" NVME
- NVIDIA MCX613106A-VDAT ConnectX-6 EN Adapter Card 200GbE
- Ubuntu LTS server (release 24.04)

A set of tests was designed to evaluate the performance of traffic from the outside world destined for SCinet. All connections in this testbed are a minimum of 100Gbps, with some uplinks being multiples of 100Gbps, or 400Gbps. Due to the nature of SCinet, cross-traffic was always present and could disrupt TCP performance during experimentation.

B. Test Procedure

A set of tests was designed to evaluate the performance between Chicago, IL and SCinet in Atlanta, GA. The latency on the path between these 100Gbps capable DTN resources was found to be 30ms, and traversed the ESnet backbone network: a 400Gbps+ capable infrastructure [19]. Tests were performed using version 3.17 of the iperf3 tool [17]. Tests were designed to exercise the following for the control situation, as well as each of the 3 firewalls. Note that customized Layer 2 paths were configured through the SCinet architecture to isolate each of the test scenarios.

- A set of 5 sequential tests, that were to run for 5 minutes in duration each
- 8 parallel TCP streams per test, with a maximum TCP pacing of 15Gbps per stream.
- The source DTN was always configured to be Chicago, IL
- The destination DTN was always configured to be Atlanta, GA
- One set of tests where MTU was configured to be 1500 Bytes, and a second set with "Jumbo Frames" (e.g., 9000 Bytes) enabled.
- Testing using both IPv4 and IPv6 addressing and routing between source and destination DTNs.
- Default system tuning for each DTN that follows ESnet's recommendations [18].
- Firewall configurations were default, except for operational considerations for the SCinet environment.

Fig. 3 showcases the results for the 1500 Byte testing. Due to a configuration issue with the fw_A hardware at the time of experimentation, it was not possible to gather IPv6 test results, and the IPv4 results were abnormally lower than expected. It was not possible to run additional tests for this scenario due to the time constraints of the SCinet environment.

The results for each of the other firewalls (and the control) were consistent between the protocol tests, with IPv6 traffic performing slightly better across the board (by a factor of 6.4 to 8.7% versus the control). In all of these cases, the firewalls performed very close to the expected control value.

Fig. 4 showcases the results for the 9000 Bytes, and features a more chaotic set of outcomes. The performance of fw_B remained consistent and incredibly similar to the control for both protocols. The performance of fw_A and fw_C was significantly lower for both the IPv4 and IPv6 traffic situations when the larger MTU was configured. In the IPv6 case in particular, performance was nearly 65% worse than the control for fw_A and 30% for fw_C.

Due to limited operational time, it was not possible to dig too deeply into the root causes of the performance for the 9000

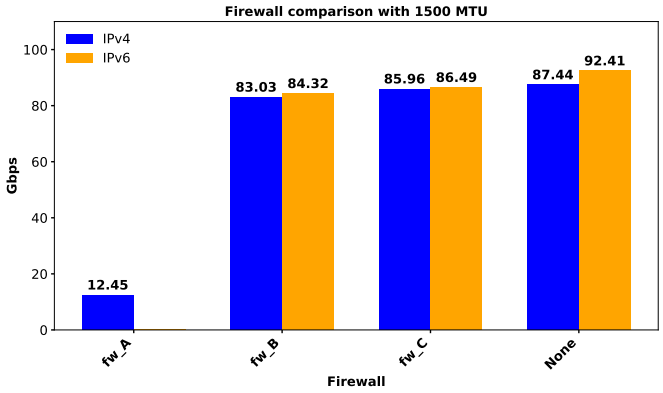

Fig. 3. Experimental results when MTU is set to 1500 Byte MTU.

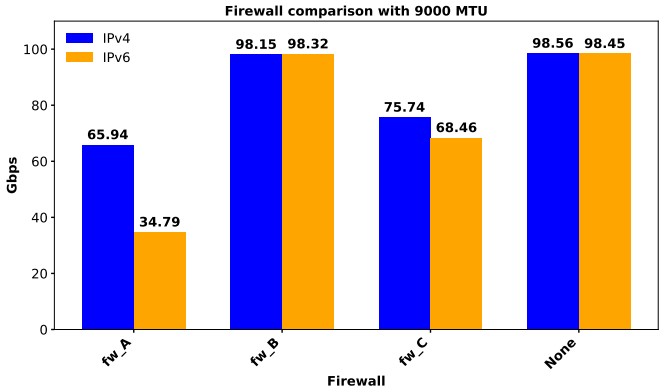

Fig. 4. Experimental results when MTU is set to 9000 Byte MTU.

Byte testing scenario. The authors believe that both device configuration could play an important factor in improving performance, as often default settings are designed to offer base functionality versus peak efficiency. 9000 Byte MTUs are still uncommon in enterprise environments, but are incredibly popular on high-performance networks.

Given the inconclusive nature of the results, it is not possible to say with certainty that modern firewalls are incapable of meeting the demands of scientific workflows; but it can be said that configuration, testing, and tuning of the network environment will remain a factor now and into the future to ensure proper performance behavior.

V. Future Work

The authors intend to prepare a follow-up set of experiments for SC25 in St. Louis, MO. Some factors that will be considered for the next round of testing include:

- Understanding the gap in performance between MTU settings
- Understanding IPv4 and IPv6 behaviors
- Adding additional firewall hardware
- Varying the destination (e.g., longer latency, shorter latency)

Acknowledgment

ESnet is stewarded by Lawrence Berkeley National Laboratory (Berkeley Lab), which is operated by the University of California for the U.S. Department of Energy, Office of Science, under contract DE-AC02-05CH11231.

EPOC [8] is a partnership between Lawrence Berkeley National Laboratory and the Texas Advanced Computing center, which is funded by the National Science Foundation grant number 2328479.

References

[1] Eli Dart, Lauren Rotman, Brian Tierney, Mary Hester, and Jason Zurawski. 2013. The Science DMZ: a network design pattern for data-intensive science. In Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis (SC '13). Association for Computing Machinery, New York, NY, USA, Article 85, 1–10. DOI:https://doi.org/10.1145/2503210.2503245

[2] The International Conference for High Performance Computing, Networking, Storage, and Analysis. 2024. SCinet. Retrieved August 1, 2024 from https://sc24.supercomputing.org/scinet/

[3] Eli Dart, Jason Zurawski, Carol Hawk, Benjamin Brown, and Inder Monga, ESnet Requirements Review Program Through the IRI Lens: A Meta-Analysis of Workflow Patterns Across DOE Office of Science Programs. United States, 2023. doi:10.2172/2008205.

[4] Nathan Hanford, Brian Tierney, and Dipak Ghosal. 2015. Optimizing data transfer nodes using packet pacing. In Proceedings of the Second Workshop on Innovating the Network for Data-Intensive Science (INDIS '15). Association for Computing Machinery, New York, NY, USA, Article 4, 1–8. https://doi.org/10.1145/2830318.2830322

[5] Globus, Retrieved August 1, 2024 from https://www.globus.org

[6] XROOTD project, Retrieved August 1, 2024 from https://xrootd.slac.stanford.edu

[7] HPN-SSH: High performance SSH/SCP, Retrieved August 1, 2024 from https://www.psc.edu/hpn-ssh-home/

[8] Engagement and Performance Operations Center (EPOC), Retrieved August 19, 2024 from https://epoc.global

[9] The International Conference for High Performance Computing, Networking, Storage, and Analysis. 2024. SCinet Architecture. Retrieved November 25, 2024 from https://sc24.supercomputing.org/scinet/scinet-technology/

[10] DOE's Integrated Research Infrastructure Program, Retreived November 25th, 2024 from https://iri.science/

[11] Campus Cyberinfrastructure (CC*), Retreived September 5th, 2024 from https://new.nsf.gov/funding/opportunities/campus-cyberinfrastructure-cc

[12] Mutter, E, and Shannigrahi, S. "Science DMZ Networks: How Different Are They Really?". Country unknown/Code not available: 2024 IEEE 50th Conference on Local Computer Networks (LCN). https://par.nsf.gov/biblio/10534241.

[13] perfSONAR, Retreived September 5th, 2024 from https://www.perfsonar.net

[14] Brian Tierney, Jeff Boote, Eric Boyd, Aaron Brown, Maxim Grigoriev, Joe Metzger, Martin Swany, Matt Zekauskas, and Jason Zurawski. perfSONAR: Instantiating a Global Network Measurement Framework. In SOSP Workshop on Real Overlays and Distributed Systems (ROADS '09), Big Sky, Montana, USA, October 2009. ACM.

[15] Ishan Abhinit, Hans Addleman, Kathy Benninger, Don DuRousseau, Mark Krenz, and Brenna Meade, 2022. Science DMZ: Secure High Performance Data Transfer. Center for Applied Cybersecurity Research, Bloomington, IN. DOI: https://doi.org/10.5967/f4aj-t870

[16] J. Postel. Transmission Control Protocol. Request for Comments (Standard) 793, Internet Engineering Task Force, September 1981.

[17] Iperf3, Retrieved December 1, 2024 from https://software.es.net/iperf/

[18] DTN Tuning, Retrieved December 1, 2024 from https://fasterdata.es.net/DTN/tuning/

[19] ESnet Network, Retrieved December 1, 2024 from https://www.es.net/engineering-services/the-network/