Winter 2017

THE IDENTIFICATION OF MAJOR FACTORS IN THE DEPLOYMENT OF A SCIENCE DMZ AT SMALL INSTITUTIONS

Scott A. Valcourt
University of New Hampshire, Durham

THE IDENTIFICATION OF MAJOR FACTORS IN THE DEPLOYMENT OF A SCIENCE DMZ AT SMALL INSTITUTIONS

BY

SCOTT A. VALCOURT
B.A., Saint Anselm College, 1992
M.S., University of New Hampshire, 1999

DISSERTATION

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy

in

Engineering: Systems Design

December, 2017

This dissertation has been examined and approved in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Engineering: Systems Design by:

Dissertation Director, Dr. Radim Bartoš
Professor and Chair
Department of Computer Science

Dr. R. Daniel Bergeron
Professor
Department of Computer Science

Dr. Robert D. Russell
Associate Professor
Department of Computer Science

Dr. John R. LaCourse
Professor
Department of Electrical and Computer Engineering

Dr. Henry J. Neeman
Associate Professor
Gallogly College of Engineering
University of Oklahoma

On November 17, 2017

Original approval signatures are on file with the University of New Hampshire Graduate School.

DEDICATION

To Gregory, Elizabeth, and Michelle:

For being patient with me

and for helping me to reach the goal.

ACKNOWLEDGEMENTS

A long-time developed piece of work, such as this dissertation, has a very large collection of individuals who have woven in and out of the three topic changes that I encountered along the way. Yet, a few individuals have been my strong advocates and cheerleaders, urging me to continue, and in some cases, offering the "tough love" speeches that I needed to hear at the right times. A few of the countless individuals to which I am indebted for this work:

Dr. Radim Bartoš
Dr. Kent Chamberlin
Dr. Robert Russell
Dr. W. Thomas Miller
Dr. Dan Bergeron
Dr. John E. Pike
Dr. Phil Hatcher
Dr. Tom Franke
Dr. Bill Lenharth
Joanna Young
Barry Reinhold
Dr. Stan Waddell
Dr. Henry Neeman
UNH Information Technology
Dr. John LaCourse
Dr. Paul Kenison
UNH Computer Science Department
Dr. Roy Turner and Dr. Elise H. Turner
UNH Research Computing Center
Karla Vogel and Dr. Mehaela Sabin
UNH InterOperability Laboratory
Kevin Thompson and Anita Nicolich

And to my Mom and Dad for their patience and understanding with me.

This work has been partially funded through grants from the National Science Foundation including grant numbers ACI-1340972, ACI-1440661, and ACI-1659377.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

THE IDENTIFICATION OF MAJOR FACTORS
IN THE DEPLOYMENT OF A SCIENCE DMZ
AT SMALL INSTITUTIONS

by

Scott A. Valcourt

University of New Hampshire, December, 2017

The Science DMZ is a network research tool offering superior large science data transmission between two locations. Through a network design that places the Science DMZ at the edge of the campus network, the Science DMZ defines a network path that avoids packet inspecting devices (firewalls, packet shapers) and produces near line-rate transmission results for large data sets between institutions. Small institutions of higher education (public and private small colleges) seeking to participate in data exchange with other institutions are inhibited in the construction of Science DMZs due to the high costs of deployment. While the National Science Foundation made 18 awards in the Campus Cyberinfrastructure program to investigate the designs, methods, costs, and results of deploying Science DMZs at small institutions, there lacks a cohort view of the success factors and options that must be considered in developing the most impactful solution for any given small institution environment. This research examined the decisions and results of the 18 NSF Science DMZ projects, recording a series of major factors in the small institution deployments, and establishing the Science DMZ Capital Framework (SCF), a model to be considered prior to starting a small institution Science DMZ project.

CHAPTER 1

INTRODUCTION AND MOTIVATION

The Science DMZ was born out of the need to create a network tool that would allow better transmission of large science data between two locations. [3] The traditional network model seeks to protect and limit data flows into and outside of the enterprise network, which inhibits the reception and transmission of large science data due to packet inspection through the firewall traversal, bulky Transmission Control Protocol/Internet Protocol (TCP/IP) protocol overhead, [37, 38] and security practices that decrease packet movement. The Science DMZ design and installation bypasses the campus firewall, resides as close to the edge of the campus network as possible, and generally is located as a link off the wide-area network (WAN) border router, allowing uninhibited flows of data as configured in the border router. These flows are typically relegated to Layer 2 protocol transmissions reliant on VLANs and data flows (such as those provided by Software-Defined Networks (SDN)) [34] as well as access control lists (ACL) and disabled system logins to protect devices connected to the Science DMZ from unauthorized access. Protocols recommended for Science DMZ data transfers include GridFTP [2, 33] through a web-based interface such as Globus, [21, 22] minimizing the need to require login access to DTNs. Research outcomes demonstrate that this model produces near line-rate results when attempting to move large data sets from one site to another.

The typical data scientist is processing a raw data set using software tools enhanced by high performance computing (HPC) systems to perform as many computations as possible in the shortest period of time. While tuning the algorithms and models to apply to the data, a challenging element of this research is the acquisition of the data set from the origin or source of the data. This may be a scientific instrument or device on a network, or the data may be stored in a repository or electronic storage device accessible via network services. Moving and acquiring the data from its resting location, especially across multiple network devices and across distance, can be time consuming and difficult, which can be even more frustrating on a slow or unreliable link. If the data is stored on a data transfer node (DTN) within a Science DMZ network and the pathway between the DTN source and the data researcher's destination location are all located on Science DMZs, many of the barriers between these two locations are removed and data is tuned to flow easily and near line rate as compared to commodity network connections, regardless of institution size.

Small institutions of higher education (small colleges and community colleges, both private and public, with student and faculty totals below 3000) have research partnership needs that encourage or require small institution researchers to participate in the exchange of data with researchers at other institutions. Many newly-awarded PhD researchers who choose to teach at small institutions still require access to and use of research equipment and datasets that are housed at a previous institution, like the institution from which a terminal degree was granted. Other researchers at teaching-focused small institutions desire to enhance their teaching through a research experience engagement for their classroom students. However, the costs to deploy high research

university-class Science DMZs are often far beyond the means available to small institutions, with many large Science DMZ deployments having capital and operating expenses higher than the existing enterprise networks at small institutions. Smaller institutions tend to have a more difficult task in moving data to and from the campus network because small institutions are often focused on meeting the network and data needs of the majority of the campus students and not focusing on the specific, high throughput needs of the small institution campus researcher. In order to allow a small institution to participate in the deployment of a Science DMZ, special arrangements need to be considered including network design, impacted users, capital and operating costs, service levels and other factors.

While the National Science Foundation (NSF) has provided 18 awards in the Campus Cyberinfrastructure (CC*) program (including Infrastructure, Innovation and Engineering CC-IIE and Data, Networking and Innovation CC*DNI) [6, 7, 8, 9, 10, 11] to investigate the designs, methods, costs, and results of deploying Science DMZs at small institutions, there lacks a cohort view of those investigated factors and the options that must be considered in developing the most effective Science DMZ solution for any given small institution. This research documents the decisions and results of the 18 Science DMZ deployments across the country, identifies and highlights a series of factors to be considered in the deployment of a Science DMZ in a small institution, and designs a collection of campus capitals that can be assessed prior to the deployment of a Science DMZ to verify that the key factors associated with a successful Science DMZ deployment will be addressed. While the focus is on the small institution Science DMZ, these same factors exist in the large institution Science DMZ too, yet are often muted or

muffled by the other factors and resources at representative large institutions which do not happen in the small institution environment.

Consequently, we believe that not only is the design and installation of a Science DMZ beneficial for the researcher at the small institution, there are several specific design and implementation methods and factors that should be considered prior to embarking on the project that could offer benefits to an entire local network community of users. While this work has focused on the small institution, all institutions of every size stand to gain from this research, as the Science DMZ, deployed universally, can be a unifying and equalizing research tool for all disciplines that require the transfer of digital data between locations.

CHAPTER 2

BACKGROUND AND RELATED WORK

2.1 The Science DMZ Model

As scientific researchers continue to expand their data collections at a scale that consumes their available digital storage resources, those active researchers require an equally robust network to move that data from one location to another anywhere in the world to support collaboration and further computation on that data. Even though high speed networking has been actively deployed in major pathways across the world for several years, researchers continue to be inhibited with a lack of the appropriate cyberinfrastructure to transfer very large data sets from one location to another across these networks. Large and expensive scientific equipment falls into a similar category when a particular researcher or research institution serves as the installed home for this specialized equipment and, in spite of its networked capabilities, the pathways are not large and robust enough to support remote control, data staging, and data retrieval from these devices. This inability to share and move data delays forward progress in addressing science research goals.

The Science DMZ model was developed to increase the overall realized bandwidth rates between scientific data sources and research analysis equipment. [14] This paradigm outlines a collection of network design elements that increases the performance of high speed networks through end-to-end tuning to support data-intensive science applications. All components of the network are impacted through this design

paradigm: storage, compute nodes, local-area networks (LAN), and wide-area networks (WAN). While the less-than-ten-year-old Science DMZ model is considered young in comparison to the traditional enterprise network design, the benefits that have been realized by scientists and researchers have made the model an indispensable tool in the advancement of all data-intensive research. As defined, the Science DMZ integrates four key unified concepts together to serve as a foundational model:

- A network architecture explicitly designed for high-performance applications,
- A dedicated system of components focused on data transfer,
- Performance measurement and network testing systems (such as perfSONAR)[23, 47] regularly used and available for troubleshooting, and
- Security policies tailored for high performance science environments.

By observing these four components in the development of a Science DMZ, the network should have the greatest impact on the advancement of data-intensive research.

The Science DMZ term describes the computer subsystem (storage, computers, and networks) that is constructed to increase performance levels while remaining secure. Key in the design is the absence of a packet-scanning firewall, as networking performance studies have shown that packet loss and transport delays in high-speed networks are inhibited by the required passing of all data, including scientific data, through firewalls. [28, 29] The separate scientific data transport network, that is the Science DMZ, is constructed to handle the high volume data transfers typically associated with high-performance computing (HPC) and scientific data computation. The Science DMZ is solely focused on the work of a moderate number of high-speed

science flows, and is typically installed at or near the network perimeter so as to maximize the network data throughput.

The Science DMZ is designed to offer a network infrastructure that is scalable, extensible, and free from packet loss that results in poor TCP performance. [32] Inherent in the design model are mechanisms to measure and test the performance of the Science DMZ to ensure that network performance is consistent throughout the end-to-end flow between Science DMZs at local and remote locations. Science DMZs offer an increased access to WAN services for local scientific instruments and data, utilizing appropriate usage policies that minimize constraints that could hamper the results of high-performance applications.

When considering the potential delay points of the network pathway between compute and storage resources, between source and sink, the WAN is typically already in an optimized position. Its very purpose is to deliver as much bandwidth as possible across the WAN medium, usually fiber optic cable. The network traffic restrictions appear on the LAN, as the typical LAN was designed to deliver email, websites and other multi-purpose data traffic, not considering very large data flows associated with scientific research. Further, multi-purpose LANs employ firewalls and other security mechanisms to protect the wide variety of data that might traverse the LAN such as financial, personnel, medical and other protected data. Expecting a multi-purpose LAN to support the restrictive flow of a variety of data types while giving priority flow access for scientific data is infeasible, as one infrastructure to support all of this widely-varied traffic cannot exist. As a result, Science DMZs provide a parallel network environment optimized for scientific data flows outside of the enterprise network definition.

After the introduction of the Science DMZ model in 2011, many of the largest research institutions in the scientific community began to deploy Science DMZs on their campuses. The National Science Foundation (NSF), observing the best practice impact that Science DMZ deployment was having on the advancement of all science, adopted the model as a key investment eligible for funding and included the design and deployment as an area in the Campus Cyberinfrastructure-Network Infrastructure and Engineering (CC-NIE) program in 2012. The University of New Hampshire received a grant in 2013 (ACI-1340972) to deploy a Science DMZ on the UNH campus among a group of nearly 100 other institutions over the last four years doing the same. By deploying Science DMZs across campuses to decrease the transmission barriers for data-intensive research, a nationwide collection of research instruments sits ready to connect to other instruments to advance science.

2.2 Key Design Elements

The Science DMZ design consists of a well-defined alternative set of network configurations that describe how deployment and configuration should be set in delivering a network that supports advanced scientific discovery. The Science DMZ originated as a design model defined by ES.NET, the US Department of Energy (DoE) network services research team that maintains the nationwide network of DoE research devices and data. ES.NET continues to maintain the core documentation around the design, development, and maintenance of the Science DMZ model. The design has four repeatable areas that have been extensively shown to offer the removal of barriers that

have inhibited data transmission on traditional communications networks. These four design areas include

1) the proper location of Science DMZs to reduce complexity,
2) the inclusion of a dedicated data transfer node (DTN),
3) the performance monitoring of the network using perfSONAR, and
4) the application of appropriate security on the network and its attached devices.

We examine the details of each of these design areas to determine if any of these areas impact small institutions differently than large institutions.

2.2.1 Location Selection

The physical installation of the Science DMZ must be as close to the perimeter of the campus network as possible. Too far from the campus LAN network boundary and the Science DMZ will be ineffective as the traffic will need to traverse too many other devices on its path off the campus network and onto the WAN. Figure 1 details the typical location of the Science DMZ network in the existing campus LAN. Typically, the Science DMZ would be connected to the border router or campus device directly connected to the WAN service provider of the institution. Campus firewalls are generally located downstream from the border router, and the Science DMZ core device would be outside of the influence of the firewall to ensure that any throughput decline that might be introduced by the firewall packet inspection service would be eliminated.

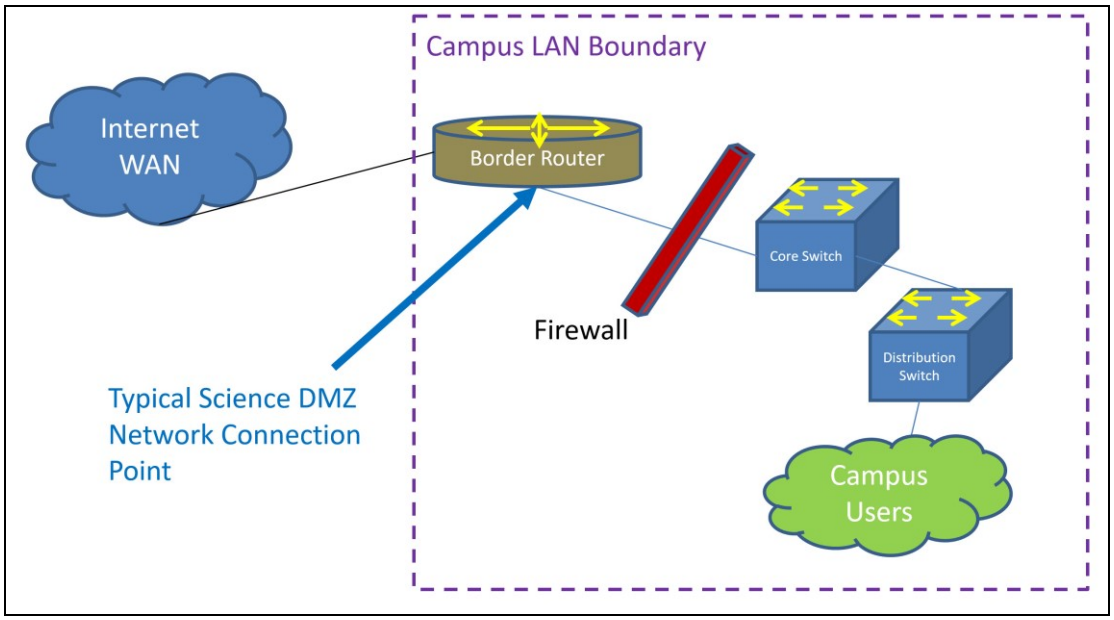

Figure 1: Typical Location of the Science DMZ in the Existing Campus LAN

When considering the end-to-end flow of research data from one endpoint device to another, the traffic flow, by nature, is going to include the transmission and reception capabilities of the endpoints and every network device on the path between those endpoints. The most widely used protocol on the worldwide Internet is Transmission Control Protocol (TCP), a protocol very susceptible to packet loss. WAN systems are designed to be large transmission pipes that deliver packets without regard to the nature of the packet. By positioning the Science DMZ as close logically to the WAN interface as possible, data transmission between end nodes become dependent on the configuration of the LAN interfaces at each end. By keeping those LANs simple and without many devices, the Science DMZ network, as the LAN, can undergo troubleshooting quickly to address end-to-end network performance issues without impacting the general-purpose network users on other segments off the WAN interface due to the separation of the science data flows from general user traffic.

2.2.2 Data Transfer Nodes (DTN)

Since the Science DMZ is a dedicated network for science data exchange, all of the services on the network should be tuned and dedicated to science data transfers. Therefore, each Science DMZ must have one or more data transfer nodes (DTN) to expressly move data from one location to another. DTNs are typically PC-based Linux servers [15] built with components that are specifically tuned to increase the performance associated with data transfers. The disk subsystem is high-speed, utilizing storage area networks (SAN) or high-speed parallel file systems such as GPFS [41] or Lustre. [31] Large amounts of random access memory (RAM) and cache memory augment the fast processors that make up the DTN in order to move stored data quickly onto the bus and out through the network interface to transfer data at near line rate of the network. DTNs are usually configured to utilize GridFTP and a user-oriented front-end interface called Globus. DTNs are most effective when configured with high-speed network interfaces, but those interfaces must align with the Science DMZ and WAN interface connections. When the Science DMZ and the WAN both support 10 Gbps interfaces, then the DTN should also utilize a 10 Gbps interface for the greatest potential throughput. A subsection of the ES.NET website is dedicated to the configuration of a commodity DTN design including hardware and configuration specifications for the standard device. [13]

Since DTNs are dedicated to the operation of data transfer, once tuned, they should have a regular, consistent performance which can be monitored, with system component failures resulting in clear performance failures. Since DTNs are single systems with a simple function (transferring data between end nodes), security policies for DTNs are applied to limit potential harm such as minimal direct login capability and

very basic service daemons remaining active. For DTNs to continue to have the greatest impact in the transfer of data, monitoring the utilization of the available storage space and building the DTN to flexibly grow with additional data storage is important for the Science DMZ to expand with additional research growth.

2.2.3 Performance Monitoring with perfSONAR

Scientists and researchers should not need to be network specialists to use the Science DMZ and the resources that it offers. There are times when the network environment might slow down or not perform at a level that the researcher will expect. In order to diagnose those failures, both locally and remotely, a key performance monitoring tool is deployed on all Science DMZs – perfSONAR. [23] A perfSONAR device monitors the local Science DMZ for performance levels as well as throughput connectivity with other perfSONAR devices remotely available across the WAN. Thus, simple multi-domain troubleshooting between Science DMZs can take place immediately upon the discovery of a data throughput error without having to deploy additional devices for testing. perfSONAR acts as a local test device that can be remotely consulted similar to having a network specialist on all of the network segments between the data source device and the data receiving device. Since perfSONAR devices are regularly testing network links to each other, the perfSONAR device maintains a grid-based dashboard (as demonstrated in Figure 2 below) [40] to indicate performance status between devices and networks and can inform researchers when a segment along the network pathway is failing or unavailable, which could impact the data transmission performance expected across the Science DMZ link segment. Since perfSONAR uses scheduled, periodic

testing packets to assess its connected network link health relative to a known set of perfSONAR locations at away sites, the overhead of operating a perfSONAR device is minimal. Using perfSONAR as a network monitoring device, the Science DMZ network can become a high-performance norm for data-intensive science experiments and increases the performance expectation for those researchers.

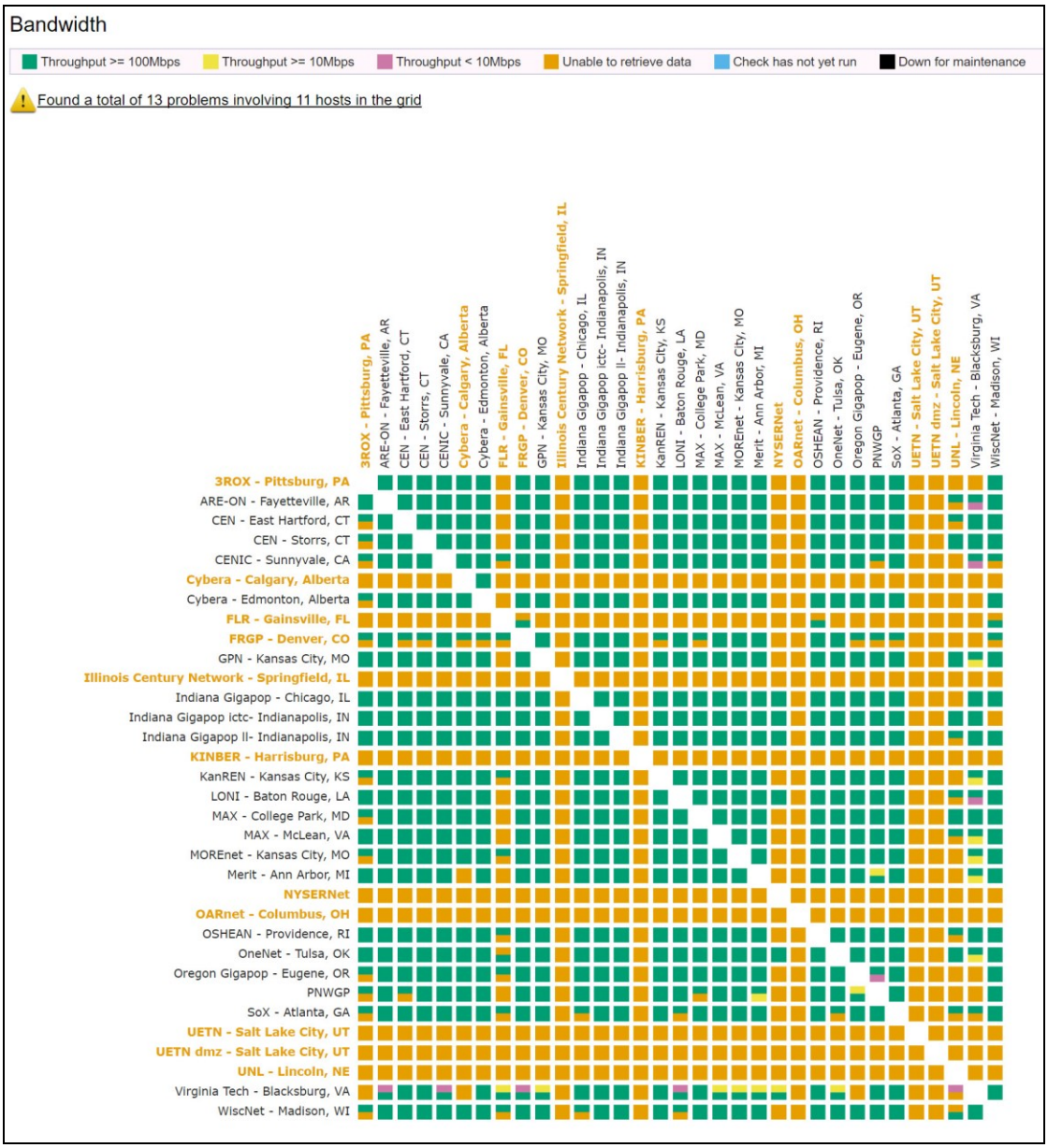

Figure 2: Sample perfSONAR Meshgrid from The Quilt

2.2.4 Security and Access Control Lists (ACL)

Since Science DMZs are located outside of the protection of firewalls, the importance of maintaining good security practices is elevated. Science DMZs are no different from general purpose networks in their need for good security, and in following the CIA (confidentiality, integrity, availability) concepts as outlined in the National Institute for Standards and Technology (NIST) FIPS-199 standard, [20] fear of network and device compromise should be no greater than that of any general network device. While these three concepts are key, a fourth concept, performance, must be considered when determining the most appropriate security policies to implement on the Science DMZ. Many of the modern security monitoring and testing tools used on general-purpose networks do not operate well in the Science DMZ high-speed environment. To assist in limiting the types of security monitoring that needs to be done on the Science DMZ, all non-science-related data traffic is forbidden on the network, including email, web traffic and all other general purpose connectivity. Security can be tailored for the data-intensive science environment of the Science DMZ, making this tool as effective as possible at moving data safely.

Rather than developing complex security systems, such as firewalls and complex router rules, the Science DMZ utilizes access control lists (ACL) with a "deny all" configuration to limit access to scientific equipment. Ensuring all devices on the Science DMZ maintain good hygiene (regular operating system patches, updated virus signatures, and regular virus scanning) as well as limiting direct user login access to devices on the Science DMZ assists in maintaining a secure and clean environment for the movement of research data.

Another security mechanism applied in the Science DMZs design is the utilization of Layer 2 network configuration for device access. Devices on the worldwide Internet use the TCP/IP Layers 4/3 model for access and addressing, which makes reaching a device on any network possible from any other location worldwide, requiring packet filters and firewalls to protect devices from unwanted connectivity. By requiring devices on the Science DMZ to be accessible only via Layer 2, the configuration required to allow connectivity between two devices around the world would force the establishing of a virtual LAN, or VLAN, between those two devices, with those two devices being the only two communicating on that VLAN. By limiting the logical and physical pathways available to users of the Science DMZ network, an alternative security mechanism can be deployed that doesn't diminish the potential bandwidth capacity of the network.

Security is an open question in the Science DMZ community, not only for small institutions under our study. Therefore, we list security of Science DMZs as a potential area of future work.

2.3 Science DMZ Design Formats

The creation of the Science DMZ is merely a deployment of an alternate set of network configurations that meet the design principles associated with good network design. Consequently, there are multitudes of variations that could be designed and deployed to implement a Science DMZ in practice. There are some key designs that have been identified by ES.NET, each building upon the base model of the Science DMZ and its associated the four design principles. The three examples outlined next include a

simple Science DMZ deployment, a Science DMZ for a supercomputer center, and a Science DMZ for a big data site.

2.3.1 Simple Design

Common to most Science DMZ designs are a set of essential components. These include dedicated access to high-performance WAN connections, high-performance networking equipment, DTNs, and a means for the monitoring of the network, typically via perfSONAR. Internet traffic destined for the Science DMZ from the Internet passes through the core campus router and into the Science DMZ switch, and if allowed, traverses directly to the science devices on the Science DMZ, which could be science equipment, a data transfer node (DTN), or some other network-connected device. Non-Science DMZ traffic is passed through the core campus router through the campus firewall into the campus local area network (LAN) as it always has, unaware of the existence of the Science DMZ. Figure 3 shows the simple Science DMZ design that is deployed in many research institution networks.

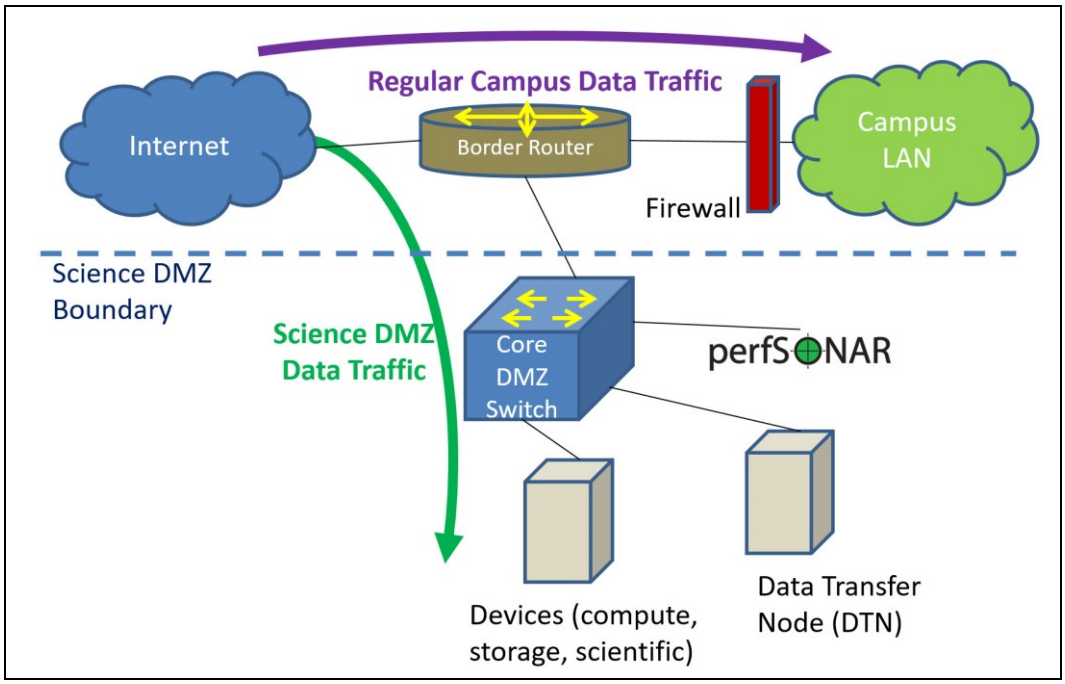

Figure 3: Simple Science DMZ Design Diagram

2.3.2 Supercomputer Center Design

Science DMZ designs that seek to maximize the utilization of large supercomputers have a slight variation to the basic Science DMZ design. To support rapid access to the data storage necessary during computation on the supercomputer, additional switches may be placed within the Science DMZ to increase throughput performance to multiple DTNs. Also, the core Science DMZ switch may be connected directly to the outside network, bypassing the need to move any research data through the core border router. Non-Science DMZ traffic, as always, is passed through the core campus router through the campus firewall into the campus LAN. Figure 4 shows the supercomputer center Science DMZ design that is deployed at large research data centers.

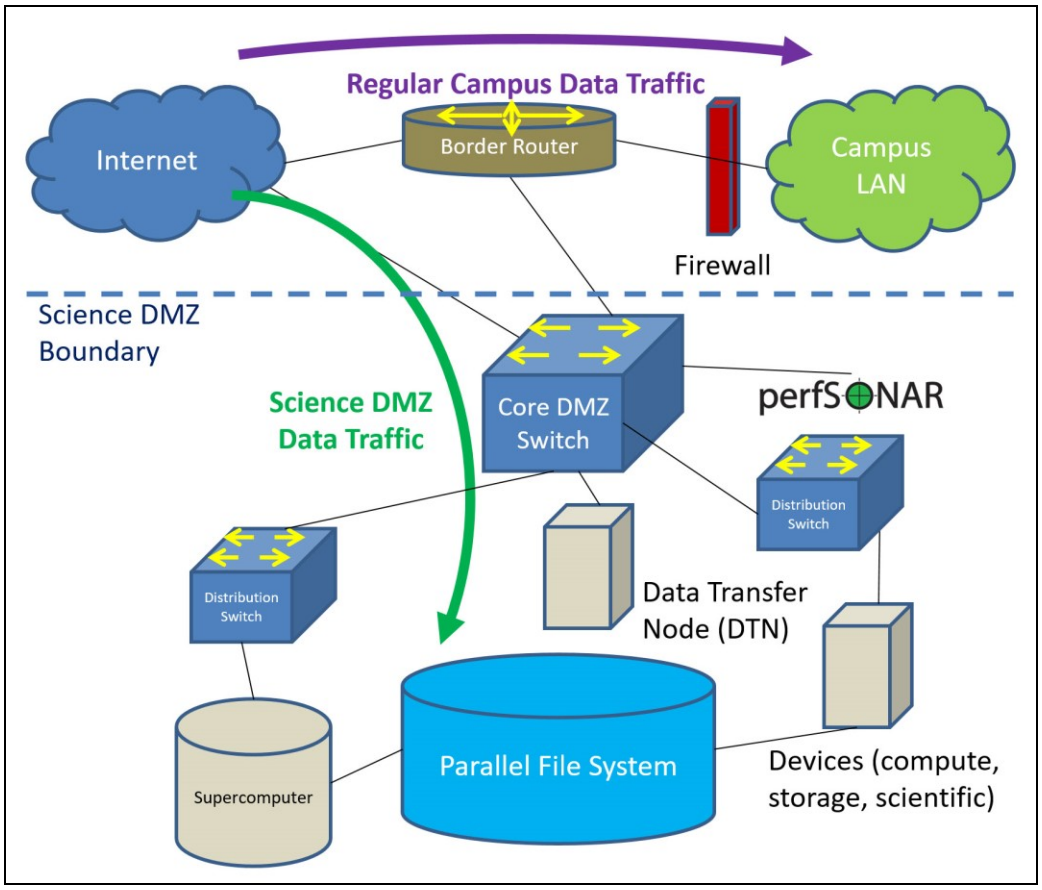

Figure 4: Supercomputer Center Science DMZ Design Diagram

2.3.3 Big Data Site

The implementation of a Science DMZ may be very complex in order to provide the best potential throughput for the connected equipment. The design in Figure 5 shows a meshed network design supporting a very large data cluster. In this configuration, large amounts of data can flow between the outside network to the campus LAN or to the data transfer cluster to maximize the flow of traffic.

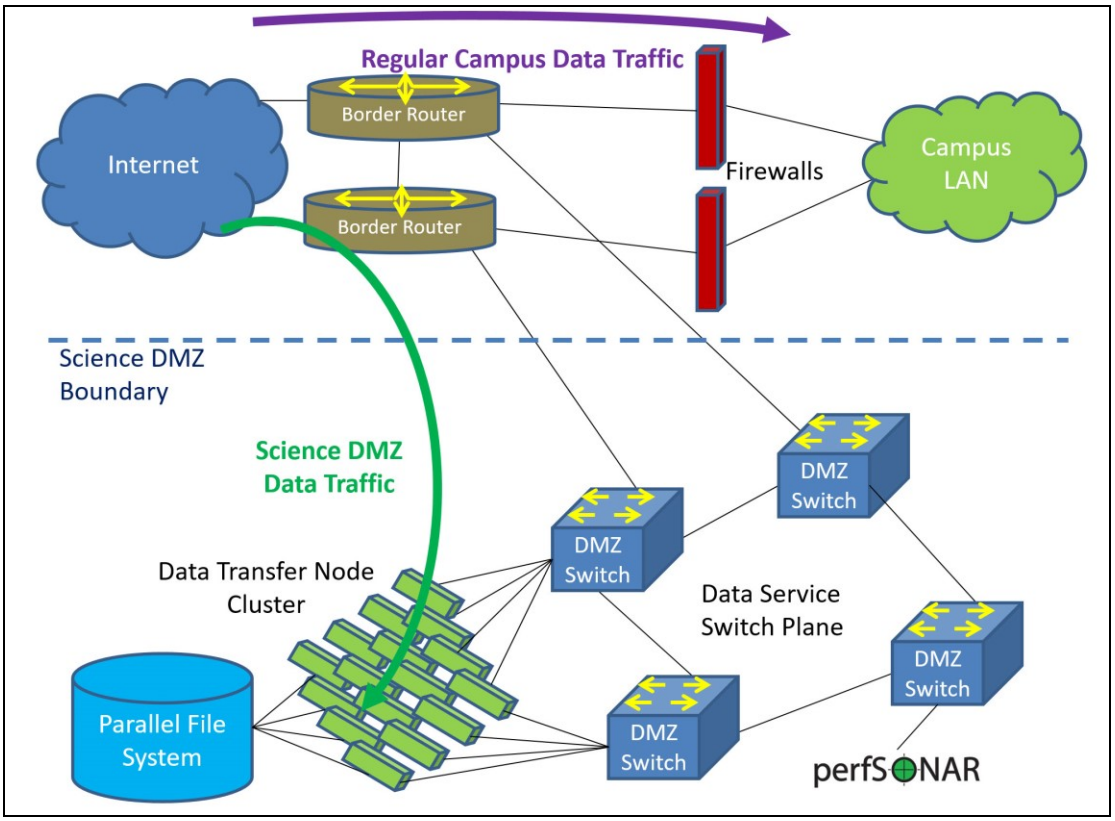

Figure 5: Extreme Data Cluster Science DMZ Design Diagram

Regardless of the format, each of these Science DMZ designs maintain a common set of aspects that define the essence of a Science DMZ: a means of data flow outside of the campus network firewall boundary (a defined network path outside of the campus border router), a means for connecting science devices to the non-campus network (a Science DMZ core device), and a means for sending and receiving data from the institution's network (DTN). We consider all three of these minimal aspects when examining the small institution designs proposed by the 18 NSF CC* awards.

CHAPTER 3

SMALL INSTITUTIONS

Nearly all of the literature and documentation on Science DMZs has focused on the large research institution. [16, 17, 42] Large institutions have a concentration of researchers with a common set of communications needs, which makes them natural places in which to place Science DMZs and impact science in a big way. However, there is a group of higher education institutions that are unable to participate in the development of the Science DMZ and in advanced data-enabled research itself. These schools can be broadly characterized as small based on their student population size (usually between hundreds and thousands of students) and their research activity (usually a few grant and contract awards to a handful of researchers). Many of these schools are teaching-focused, where research activity is limited to activity that supports the teaching mission of the institution.

These institutions are rapidly becoming more involved in advanced data-enabled science and research for two reasons: enhancing the undergraduate educational experience requires engaging in meaningful research as part of the curriculum, and many newly-minted PhD faculty are seeking the opportunity to teach undergraduates at small institutions without having to give up access to large research institution resources that can now be made available through network communication links. By developing Science DMZs for the small institution, researchers at small schools can continue to collaborate with colleagues at larger institutions without the significant expense of

traveling to those schools to use equipment, access or acquire data, or limit the students who can be involved in achieving research goals. All of those core functions can be completed across a network link that supports data-intensive transfers.

The National Science Foundation, recognizing this gap in meeting the needs of the whole research community, began offering in 2012 a set of funding to support the design and deployment of Science DMZs at research institutions to enhance data-driven science and research. Through an encouragement of small institutions to apply for the grant awards, as well as to partner with larger institutions already with Science DMZs online, the NSF provided grant awards starting in 2013 that funded the design of Science DMZs in the first year of the award and, after review and approval, funded the implementation of the design on the small institution campus in the second year. Figure 6 graphically illustrates the location of the 18 small institution Science DMZ awards made in 2014 and 2015 by the NSF. A more detailed table of these institutions is located in Table B1 in the Appendix.

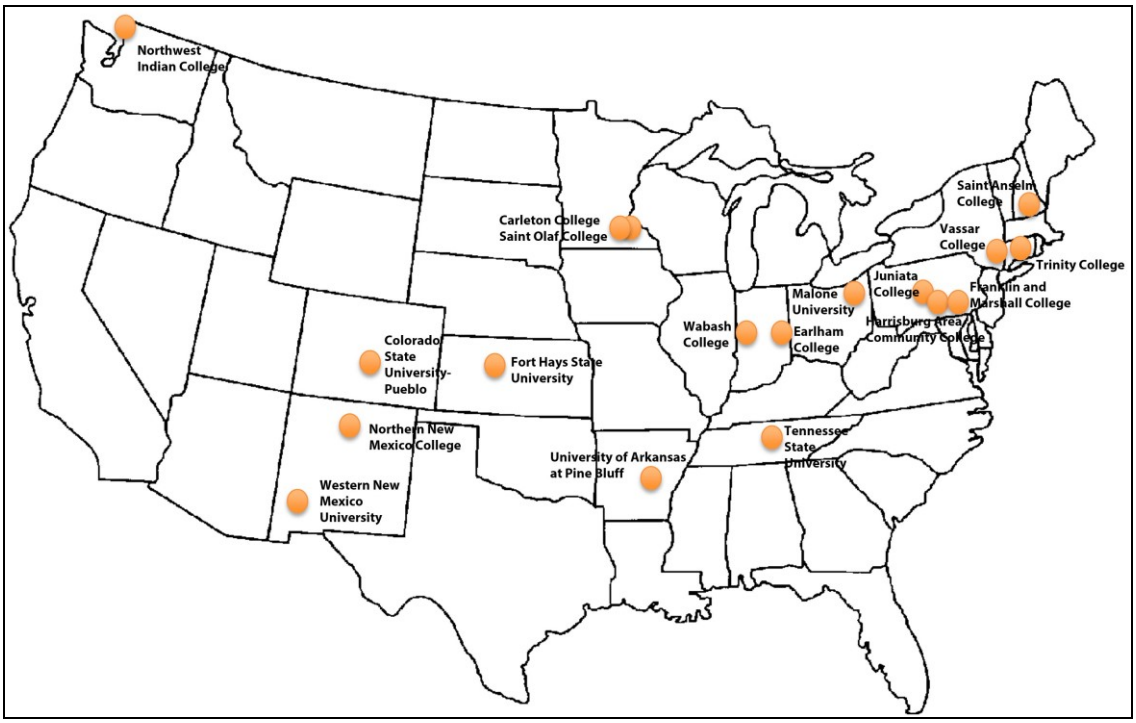

Figure 6: NSF-Funded Small Institution Science DMZ Deployments 2014-2015

What makes this research so important is that the researcher is key in the development of the Science DMZ. Regardless of institution size, researchers are not network specialists and should not have to be in order to accomplish their research. Yet, researchers are in the best position to advocate for a Science DMZ to advance their studies. By knowing the key factors that go into the development of a Science DMZ, researchers can assist in building the best case for a Science DMZ on their local campus, find the best partners to join them in the advocacy for a Science DMZ, and detail the key technical functions that will lead to the advancement of research for themselves and their worldwide research partners.

CHAPTER 4

RESEARCH APPROACHES

While much of the Science DMZ development has been taking place at large research institutions and government laboratory facilities, the transmission capabilities and data engagement that supports collaborative research takes place at all institutions. Colleges and universities that are classified as small to medium-sized have researchers that earned their terminal degrees at large institutions and want to continue that research at the small institution in conjunction with the desire to increase research opportunities for undergraduates as part of the teaching process. The need for a Science DMZ is increasingly important for small and medium-sized institutions as a means for their participation in the advancement of research, as a tool for teaching students using the most current methods, tools, data, and collaborative resources available, and as a recruiting and retention tool for quality faculty to conduct research and teach students in the small institution setting.

Science DMZ development for the small institution began to be a priority with the release of the National Science Foundation (NSF) program in the Campus Cyberinfrastructure (CC*) Network Integration and Engineering (CC-NIE) program area to support the Campus Design for Small Institutions. Significant efforts were made to encourage small institutions to apply for the program, including the offering of physical and virtual workshops on how to apply for CC* funds. Since 2014, a total of 18 awards have been made to institutions of higher education that self-identify as small to medium

sized and are in need of a Science DMZ based on the science research that is taking place on their campus. Just like the regular Science DMZ development program that has awarded over 100 Science DMZs to large institutions, the small institution awards are also scattered across the country, and are often led by key campus Information Technology leaders – Directors of Research Computing, Chief Information Officers, Provosts, and Vice Presidents of Research. While most of these awards have been successfully completed or are in progress presently, the results of this develpment work have not been published beyond the annual and final reports submitted to the NSF. A listing of the 18 NSF CC* awards, their NSF award numbers, titles, and institutions awarded, is detailed in Table B1 in the Appendix.

In 2014, Saint Anselm College, a small, Catholic, liberal arts college in Manchester, New Hampshire, in partnership with the University of New Hampshire (UNH), was a recipient of a NSF CC* Campus Design for Small Institutions award (ACI-1440661). This award was specifically given to Saint Anselm College because of the small institution partnership with a large institution (UNH) that had already completed the design and most of the implementation of a Science DMZ in one of the other areas of the NSF program. The Principal Investigator on this grant was the Chief Information Officer for Saint Anselm College, while the author of this dissertation served as the Co-Principal Investigator. Other Senior Personnel involved in the project were the Network Architect for UNH and the Network Manager for Saint Anselm College.

Entering into the two-year grant award, the first year was to be devoted to determining the best design implementation of the Science DMZ on the small institution campus. All of the researchers and technologists engaged in the design tasks were

expecting to design a Science DMZ that was of a smaller scale than that of the design of UNH's Science DMZ. The group also expected to have much smaller costs associated with the design because less equipment would be required to support a smaller number of research projects identified in the grant proposal. However, a significant number of unknown factors were discovered along the way that changed the nature of the design project and adjusted the list of assumptions that the team needed to consider. In the course of completing the Saint Anselm College Science DMZ project, we wondered how many of the other small institutions that embarked upon the creation of a Science DMZ encountered any of the same problems the our project team had seen, or whether other Science DMZ projects encountered different issues that our project overlooked or avoided.

Beyond cost and scale, the Saint Anselm College project and informal conversations with other NSF CC* awardees have identified other factors that warrant investigation and detail, including:

- the integration of a Science DMZ design project into an existing small campus LAN design,
- the final Science DMZ designs deployed on small campuses at the end of these projects and how those final designs varied from their original designs,
- the type of references used to propose the awarded NSF project,
- the type of references used during the design phase of the NSF project,

- the sustainability costs to maintain and operate Science DMZ equipment when its use is anticipated to serve only one or two active research projects,
- the need to upgrade campus WAN connectivity to support Science DMZ deployment,
- the necessary on-campus knowledge and expertise that understands Science DMZ network equipment and its operation and maintenance, especially if the small institution technology staff is already minimal,
- the political landscape surrounding the deployment of a Science DMZ that doesn't serve the entire campus community,
- the connection of the small campus Science DMZ to other research data sources, both on-campus and off-campus, and
- the most effective way to move data to and from the small institution in support of data-enabled science.

Through panel presentations at the last three NSF-sponsored CC* Principal Investigator (PI) meetings on September 29, 2015 in Austin, TX, October 19, 2016 in Philadelphia, PA, and on October 3-4, 2017 in Albuquerque, NM, [19] we have come to learn of some of the successes and challenges that small institutions have faced when developing a Science DMZ for their campus. While we are reminded that small institutions cannot build full Science DMZ configurations in exactly the manner that ES.NET describes on their website, we find that most small institutions do their utmost to be as close to the large institution deployment as possible. However, there are many

more factors involved in the construction of the Science DMZ on the campus of a small institution beyond those that we have already encountered.

The literature is very sparse on the topic of Science DMZs in a general sense, with most papers pointing back to the definitive design paper from ES.NET. [15, 16, 17] Other than a few papers that describe some of the proposed modifications to the base design particularly focused on security enhancements, not much is written on the extension of the Science DMZ as a research tool. [34, 43] Nothing appears in the literature on the design modifications, challenges and direction options that small institution Science DMZ designs need to consider. Through anecdotal conversations, we knew that the design deployed for Saint Anselm College was not the same model as that of Vassar College in New York. We also knew that the design choices of Franklin and Marshall College of Pennsylvania are different from the previous two institutions and warrant an examination of the differences as well as the reasons for making those choices.

We began this research by collecting the network designs and implementation details on the 18 small institution Science DMZs deployed across the US. Through the review of NSF CC* awarded proposals, and other written materials such as annual reports, outcomes documents, and final reports, as well as through personal interviews and presentations of Principal Investigators and key personnel on some of the grant awards, we collected and categorized the design decisions that have been deployed thus far, including summaries on the architecture, management scheme, best practices, policies, security, sustainability costs, and other details. We found that many of the designs follow conceptually the Science DMZ model that ES.NET has proposed. Yet, we

also observed design decisions that align with the small institution's network challenges. We recognize that the small institution's factors are not entirely the same as those of the large institution. We expected to see many of the same factors observed in the design and deployment of the Saint Anselm College Science DMZ to be common among the 18 design awards, which, to a degree, was observed.

We gathered this information through an IRB-approved protocol for data collection (UNH #6598) including requests for NSF proposals, NSF annual reports, NSF final reports, NSF outcome reports through a Freedom of Information Act (FOIA) request (#17-116F) to the NSF, along with general surveys sent to 18 PIs followed by specific interview questions made to PIs and other identified senior personnel on those grants. Through the review of the pre-project proposal and those projects that had already completed outcomes reports, we were able to determine some initial assumptions and design variance that the project teams encountered during their deployments. However, our detailed, project-specific personal interview questioning led to identifying and confirming the major factors that determined the Science DMZ design that was finally deployed to fulfill the NSF grant project award. In order to ensure that the data collected doesn't compromise the operational integrity of the 18 institutions studied, we have randomized their results and refer to specific institutions in this document via a coded name, such as College A through College R.

We expected that the data collected by this research endeavor would synthesize into a few major network design models that either parallel the ES.NET Science DMZ basic designs or diverge entirely from them, really not fitting the basic design model, but rather turning towards a means of upgrading small institution enterprise networks for the

benefit of all users, not specifically targeting data-enabled science. While this approach does violate some of the major assumptions that a Science DMZ should support no email or general Internet traffic, and a restriction on the access by general users to the Science DMZ, small institutions need to balance the need for restricted science-focused data with everyday data traffic flows for the predominant majority of the campus. Understanding how that alternative design model achieves success for data-enabled science will have a potential to impact the deployment of all small institution Science DMZ designs.

We also anticipated that we would uncover a few factors that we had yet to see in our informal conversations and initial research work with Saint Anselm College. Those factors previously unidentified were examined and weighed to determine how they could have a place in a standard Science DMZ design for small institutions.

We presented some of our preliminary results via a peer-reviewed poster session at the annual Practice & Experience in Advanced Research Computing (PEARC17) conference in New Orleans, LA in July, 2017. [36] Our early results from data gathering and research synthesis led towards a litany of colleagues across the country, from small and large institutions, offering confirmation and validation of the results we reported in the poster. With so little documented on the design and development of Science DMZs, our conversations confirmed that the academic research community and the networking practitioner community are craving good approaches towards addressing the hard problem of building the best possible Science DMZ network in the midst of a collection of challenges that face small institutions. We captured our collected data and presented our findings as a set of guidance tools for small institutions nationwide that may be considering the design and deployment of Science DMZs. That collection of tools would

best be published alongside the support documents that ES.NET publishes on Science DMZs for large institutions, thereby adding to the knowledgebase of Science DMZs nationally. The community guidance on the development of a Science DMZ, in its present form, poses a significant burden for small institutions to participate in the Science DMZ community because the cost of the Science DMZ alone often dwarfs many of the present-day small institution campus operating networks and the guidance is all written from the perspective of the large institution and small institutions need to translate the designs to fit their environment. [15, 16, 17]

While this research is more applied than basic, we are uncovering many questions that others have asked and have received answers that are not shared with the larger community. By systematically approaching the discovery of questions and answers, we believe that the results that follow will lead towards better tools to support digital data exchange for small and large institutions alike.

CHAPTER 5

SMALL INSTITUTION SCIENCE DMZ FACTORS

5.1 Introduction

The need for a Science DMZ is driven by the need to connect science data and devices with other resources across the campus or across the world. While the technological aspects rooted in network theory and design and modern information technology account for the Science DMZ hardware choices, there are several other aspects that go into the decision making of a Science DMZ that are not solely technical. Our research uncovers several factors that are already recognized by small institutions in the design of their campus Science DMZs. In a review of the 18 initial project descriptions developed by the NSF CC* PIs, several factors that influenced the design and implementation of the Science DMZ became evident and consistently surfaced as influences on the small institution Science DMZ designs at the onset of the projects.

When we consider the motivations that drive any institution, large or small, to pursue the development of a Science DMZ, we observe some fundamental flaws in the design preparation. Researchers at large institutions have long complained about slow data transfer rates between institutions. Several TCP/IP network modifications and network operation models have developed over the years to help move large data sets from one location to another. [4] Applications that adjust the TCP window size and European Council for Nuclear Research's (CERN) multi-tier data distribution grid network [1, 44] are two examples of many innovations that have been attempted to

maximize the line rate available across the physical layer of the network protocol stack. While the Science DMZ was born out of this necessity of moving data between sites quickly, the funding motivation to build these Science DMZs across the county did not support the true motivation.

The National Science Foundation, through the solicitation process within the Campus Cyberinfrastructure program, asked researchers to draft funding proposals based on two primary tasks – 1) build a Science DMZ based upon the ES.NET model, and 2) identify a collection of on-campus science drivers that require the uninhibited movement of data that would benefit from the Science DMZ. Peer review panels judged the proposals to be awarded based primarily on these two criteria, though the requisite NSF language around the need for a solid Data Management Plan (DMP) and a reasonable sustainability plan of the once operational Science DMZ was also required of all proposals. Small institution awards were judged with even less restrictions as the process required the small institution to spend the first year of the award designing a Science DMZ environment that would fit their environment, making the science drivers at a small institution the most compelling reason for funding the construction of a Science DMZ. Small institutions, in most cases, considered the true impact of the Science DMZ deployment and operation only after the grant award of funding was received.

In reviewing the small institution Science DMZ designs and conversing with some of the Principal Investigators (PI) of the funded projects, several prevalent factors surfaced in the planning and operation of those Science DMZ projects. These factors are identified and detailed below.

5.2 Cost

One adage that our research team is fond of repeating is, “A free kitten isn’t always free,” and the same can be said about a NSF-funded Science DMZ network. Small institutions are typically constrained in their financial investments. Limited capital resources compete with a collection of possible investments that may have higher priority than science research. Since the capital costs of a Science DMZ network are covered by external sources, like a NSF grant, the small institution is delighted to be awarded these grant funds to build the Science DMZ network. Having a sense of the real costs of the Science DMZ equipment, prior to the application for grant funds, is a key consideration to achieve success. While the NSF solicitation is clear that small institutions had one (1) full year to design the right Science DMZ to fit their unique campus while following the ES.NET model, nearly all of the proposals had considered at least one possible approach towards the design and installation of the Science DMZ in the existing campus network. From those initial designs, PIs considered the potential costs that would be required to redesign the campus network and include the Science DMZ hardware, even if the final design is not well aligned with the four components of a Science DMZ. Some proposals were up-front about wanting to redesign the entire campus network with new fiber and hardware for the campus network in order to support a Science DMZ.

Since the application-to-award period for most of the projects reviewed averaged nine months, and most of the projects required nearly the entire first year of the award to craft a solid Science DMZ design in light of the existing network, the retail costs for the equipment dropped and the technological capabilities of that equipment increased as often happens when seeking to deploy rapidly-evolving technological assets. Several

projects that used maximum cost estimates 1.5 years earlier were pleasantly able to secure the same, if not better, equipment for lower prices, making the thoughtful delay of designing the Science DMZ financially worthwhile.

Even if the proposal covered most, if not all, of the costs for the Science DMZ equipment, the small institution has the obligation to consider the long-term operational costs of the equipment and what to do once the equipment reaches the end of its useful life. We address this question in greater depth around sustainability further in this chapter. We noticed that many of the proposals did not specifically speak to the detailed operational costs such as hardware maintenance, equipment repair, and patch installation that have real costs, either paid to the vendor or expended by the small institution in staff labor. Many institutions simply consider the Science DMZ, once installed, to be another element of their existing network that must be maintained to support the institution. Knowing and understanding the real costs of designing and building a Science DMZ is an important first step in building a Science DMZ for any size institution.

5.3 Design

We devote the entire Chapter 6 to an examination of the eighteen (18) Science DMZ projects and the general design aspects that we observed to have been developed by small institutions. Dropping a Science DMZ into an existing small institution local area network (LAN) isn't a design. Besides the typical overshadowing cost of a Science DMZ relative to the rest of the LAN, we noticed that many of the NSF projects studied don't have LAN designs that are Science DMZ-ready. Most required some form of adjustment to prepare for a Science DMZ, including upgrading or creating a central campus router

that could route Science DMZ traffic away from the campus LAN and the firewall, switching technology to segment Science DMZ traffic efficiently, or WAN connection upgrades that allow the campus network to operate with wide area network (WAN) access of 1 Gbps or better. Several campus LAN networks required the installation of either single-mode fiber optic (SMF) cable or multi-mode fiber optic (MMF) cable to upgrade the existing LAN to Gigabit speeds higher than their existing Megabit speeds over copper. These upgrades were included in the cost estimates to upgrade the network to support the Science DMZ design and installation and could be viewed as network replacements rather than network modifications to support Science DMZs. With these campus network upgrades in place, nearly all of the small institution campus networks were prepared to design and install a Science DMZ network that conformed to the general Science DMZ design.

5.4 Capabilities

Science DMZ networks have a variety of components that could be included in the overall design. Data Transfer Nodes (DTN), perfSONAR and other monitoring tools, intrusion detection systems (IDS), high-speed routing and switching, and other capabilities are being developed, particularly at larger institutions, to support specialized equipment on the Science DMZ. Not all of the components need be part of the small institution design, and a careful matching of the science drivers, their critical needs, and their long-term desires must be weighed carefully in the overall Science DMZ design. A few small institutions required connectivity to Internet2 [25] services to interconnect their Science DMZ networks to other remote sites important for the continuing work of the

science drivers. Internet2's Advanced Layer 2 Switching (AL2S) [26] offers the definition of a layer 2 pathway to be established from the small institution's GigaPOP (Point-of-Presence) to the remote connection's GigaPOP to maximize throughput across the WAN between these two sites. The collection of Internet2 services are attractive benefits available to small institutions that register and connect with Internet2, and the Science DMZ design motivates having an Internet2 connection to support data exchange. PIs who design Science DMZs to offer every capability, though, will quickly find that the costs and the sustainability of such a Science DMZ could have negative consequences on the small institution campus. Therefore, the key capabilities that designers of Science DMZs for small institutions need to include, besides a properly-configured network, are perfSONAR, a DTN, and a high-speed network pathway to off-campus.

5.5 Sustainability

The sustainability costs to maintain and operate Science DMZ equipment vary greatly for small institutions as compared to large institutions. A large institution with a Science DMZ that will be serving 1% of its research population may result in 7 to 10 projects maintaining the resource. A small institution serving the same 1% of its research population may result in less than 1 faculty member using the resource. As a consequence, the dedicated bandwidth required to follow the Science DMZ model is unsustainable by a single researcher, potentially making the small institution Science DMZ more costly to maintain than the entire campus LAN. Developing creative means for meshing the resources needed to sustain a small institution Science DMZ requires

designers to create a new network configuration that varies from the traditional Science DMZ design.

A couple of the approaches that small institution designers have taken is through the integration of the Science DMZ with the regular campus network. Network operational personnel will turn down the bandwidth rates of the Science DMZ switch and its subsequent network connection when not in use, allowing the core campus network users to take advantage of the WAN bandwidth off campus. If a researcher has a specific data transfer need, then that researcher will schedule a window of time to move data to a DTN while the network personnel turn up the bandwidth rate to the Science DMZ switch and allocate WAN bandwidth of a sufficient level to support the transfer needs of the researcher while not starving the campus users' network needs. Network administrators need to consider the typical campus network use patterns before embarking on this practice, however, as reallocating significant enough bandwidth to support data movement may need to occur when the regular campus network is in a low use period, such as very early in the morning or on the weekends. In using this operating pattern, the researchers have advanced bandwidth access while the campus network personnel can satisfy the campus network demand without needing to maintain two dedicated WAN routes, in all cases increasing overall bandwidth levels while decreasing per-Megabit bandwidth costs, allowing the overall campus network service to benefit from the increased connectivity required by the Science DMZ design. This approach is a good start for serving the researchers at the small institution, but the manual operation required for this service may be too expensive for the small institution IT staff to continually change the campus WAN traffic flow and could put the proper functioning of the whole

network in jeopardy should some misconfiguration occur. Implementing automation to support network pathway management or allocating a regular, repeating window of data transfer time each day may serve as effective approaches that small institution Science DMZ designers should consider if there are resources available for investment.

The long-term sustainability of the Science DMZ, regardless of the design and implementation, is a key factor in the deployment success of this research tool. The benefits achieved by researchers who use the Science DMZ justify the costs associated with the expenses required to keep the tool in long-term operation. However, small institutions need to be prepared to understand the full scope of those costs associated with their Science DMZ network designs.

5.6 Upgrade Requirements

As described above, the small institution WAN configuration may not be under the control of the campus network personnel at the start of the project. Some campuses receive their network WAN service from commercial providers who maintain the campus border router from within the service provider network. Other WAN configurations offer border routers from within the small institution campus network, but the service provider might be maintaining the campus border router and not the local campus network personnel. Most small institutions we studied maintain their own campus border router, but their depth of knowledgeable personnel may be limited to only one person. With the ability to bypass firewalls and gain access to WAN resources that are nearest to the campus LAN border, clear access to a campus border router is a factor in the design of the Science DMZ for small institutions that is not readily considered at the project onset.

Many campus designs relied on the partnership of larger organizations assisting the small institution in the overall design and deployment of the Science DMZ network. These larger institutions took the form of neighboring universities in the state with experience in Science DMZ design and operation, or of regional network service providers that were either serving the small institutions already or were the logical entities to eventually serve the small institutions due to proximity and long-time experience in addressing campus networking needs. In all cases, the expertise of these large institutions was welcomed and instrumental in the creation of designs that satisfied the requirements and integrated well with upstream WAN resources.

Since Science DMZ deployment on the small institution network may require changes and upgrades to the existing campus network, designers need to take into consideration those changes that are necessary to meet the science missions associated with the Science DMZ installation. External partners may provide significant insight into existing and past practices that may lead to the best changes that should be made as compared to those changes in the campus network that could be ignored for the initial implementation of the Science DMZ.

5.7 Local Knowledge

With a demand for research resource access through a Science DMZ growing among small institution faculty, the need for on-campus knowledge and expertise that understands Science DMZ network equipment and its operation and maintenance grows as well. Most small institution network management staff is at the minimum, with some campuses having less than one full time staff member to maintain a campus network that

serves a community of 2000 faculty, staff and students. We found few Science DMZ design projects in our dataset that embarked with full consideration of the knowledge required to incorporate this resource into the current environment, with many PIs invoking comments like, "we had no idea what we were really getting ourselves into," or "we learned far more than we ever expected to learn by embarking on this project." Most PIs have the opinion that the knowledge required to maintain the equipment will be gained through vendor training and live problem solving on the network, as an external vendor can always be hired to address major issues with the equipment.

Those small institutions that partnered with large institutions to design and deploy a Science DMZ found that partnership to be much more than a design resource. Examples of operations training, configuration decision making, and best practice transfer were detailed throughout the construction and implementation phases of the projects. While many of the science driver research partnerships that justified the Science DMZ funding spanned multiple institutions, the network technicians and expertise to operate the Science DMZs spanned across similar institutions and have led to a regional community that regularly learns from each other. While Science DMZs are designed to connect research assets together for the benefit of advancing science, the side benefit of connecting knowledgeable personnel is a key factor in helping small institutions acquire the knowledge they lack at the start of a project.

5.8 Politics

The political landscape surrounding the deployment of a Science DMZ that doesn't serve the entire campus community can destroy the effort from the start. Small

institutions have resource constraints and campus leadership must make daily decisions to apply scarce resources where they will provide the greatest return for the benefit of the whole institution. A single researcher at a small institution is unlikely to have the political capital to convince the small institution leadership to invest in the development and maintenance of a Science DMZ. Many of the research projects we examined were led by campus leaders with the political capital necessary to drive a Science DMZ to a sustainable future. The titles of these PIs include Chief Information Officer, Provost, Senior Provost, and one institution has even had its PI promoted to becoming College President after the award was made, presumably not a result of the NSF award. Many of the projects have PI and technical personnel teams that are a combination of IT leadership and science researchers working together to design and deploy the Science DMZ on the small institution campus. PIs reported that the interchange between IT and the researchers had long-term impacts beyond the Science DMZ and its use. The group found that they had an easier engagement when having to address other IT or network issues involving the research community, and also observed a greater awareness of the needs for research on their campuses, where in the past, the needs of research were simply assumed to be embedded within the needs of the academic users.

Somewhat related to the political environment of the small institution is the public image that is conveyed by the institution and the research work done there. Very few of the small institutions made any attempt to publicize their NSF award or the Science DMZ project that they were undertaking. The few small institutions that did release articles and press statements saw very substantial interest from their campus leadership when the capital funding levels and day-to-day impacts where outlined for the public to digest.

The campus PIs suspect that influential alumni and donors viewed the federal funds inflow to the small institution as a badge of honor that was achieved by the institution, and that recognition was anticipated to draw additional funding from benefactors.

Therefore, to meet the political factor associated with the Science DMZ design and development, small institutions need to leverage the whole community, not just the science community, in encouraging investment in the Science DMZ research tool. Communicating the development and actual deployment of the Science DMZ on the small campus network will rally public support for this new and interesting device. Announcing any external funding awarded to the small institution in conjunction with the use of the Science DMZ will strengthen the meaningful purpose of the tool for all of the researchers who depend on the tool.

5.9 External Contacts

The connection of the small campus Science DMZ to other research data sources is a key consideration in the design and deployment project phases. Remote data sites, either for the reception or origination of data relative to the small institution, must be capable of compatibility with the elements of the Science DMZ. DTNs and dedicated network pathways between the two sites need to be determined if near-line-rate transmission is to be effective. Small institutions with commercial network providers are not directly served by Internet2 resources, which may make throughput difficult when transmitting between academic or research sites. Commercial providers may have traffic shaping in place to limit the maximum throughput levels that can be achieved by any one node on the campus side of the provider's network offering. These upstream network

considerations are key decision points when designing the Science DMZ for the greatest possible throughput. Internet2 does not impose any in-network restrictions to the WAN data flow and is the best choice for Science DMZ data transmission.

Most of the projects reviewed had some external partner that was guiding the small institution in the development of their Science DMZ. While many of these partners are other large institutions that have already built Science DMZs on their campuses, their insight and guidance has been reported to be highly valuable in navigating the correct choices to make in the Science DMZ design and in the choice of vendor equipment. A few PIs reported that they did not have an external partner in the design of their Science DMZ. By reviewing the Science DMZ literature and working with their Internet Service Provider (ISP), they were able to achieve the result they were seeking, yet in hindsight, would have found value in consulting with another party that had experience with building a Science DMZ.

As identified above, Internet2 connectivity was a key goal for all of the reviewed projects that did not already have some mechanism for Internet2 access. The national network community associated with ES.NET staff has served as a strong influencer in the development of Science DMZ networks at all sizes of institution. Through workshops, both in-person and remotely-delivered, guidance on how to build Science DMZs and draft grant proposals that would win funding has been well-received by the small institution community. Knowing and leveraging the external community is a key factor in designing small institution Science DMZs, as many of the projects we reviewed took advantage of those relationships. Without working closely with that external community, it is difficult to consider the long-term purpose of the Science DMZ other than being a

local data store as the access and movement of digital data requires external partners to reach a successful outcome.

5.10 Best Practices

Even after reviewing the previous factors and considerations when choosing to design and deploy a Science DMZ, some small institutions should begin with a basic process approach to the movement of digital data between locations. The most effective way to move data to and from the small institution in support of data-enabled science may not be through a Science DMZ. Data repositories, distribution networks, and other DTNs can and do exist on networks that are not Science DMZs. The dedication of WAN resources towards the movement of research data is a critical component of the Science DMZ design. Small institutions need to understand this factor before embarking on a project to invest in a Science DMZ.

Beyond the factors identified at the project onset as detailed in this chapter and via the conversations with many of the 18 NSF CC* grant awardees, we recognize two key tasks that need to be done by Science DMZ project leaders at small institutions before proposing a Science DMZ project – design the right Science DMZ for the campus network that exists or should exist, and assess the campus environment to ensure that the Science DMZ designed will have a possibility of succeeding and thriving after the Science DMZ deployment project is complete. Our research has led us to develop the following two chapters that outline the observed Science DMZ models that were deployed by the 18 NSF projects (Chapter 6) and propose a framework to be used to capture the existing campus environment (Chapter 7) relative to the observed factors

above to support an effective Science DMZ deployment for small institutions and the long-term impact of those investments on campus networks.

CHAPTER 6

SMALL INSTITUTION SCIENCE DMZ DESIGNS

6.1 Introduction

Beyond an examination of the major factors that have led to the design decisions of the Science DMZ on small campuses, we encountered a more fundamental challenge that stands before the small institution in the deployment of the Science DMZ. Unlike the large institution campus network design, with several switched or routed network segments to isolate campus academic, campus residential, campus research, and campus administrative computing traffic, small institutions have relatively flat, single segment networks to address all network traffic to be managed.

The integration of Science DMZ designs into existing small campus LAN designs often starts in one direction and ends in a completely different approach. With no examples of small institution Science DMZ designs on which to base a working model, designers have had to rely on translating the design models for large institutions described above into models that would fit the small institution. Every campus network design is different, and we reviewed 18 different initial campus configurations, 18 different proposed Science DMZ network configurations to satisfy the grant requirements put forth by the NSF, and 18 different implementations with most completed or near completed.

In this chapter, we detail the 18 campus network configurations we observed, both at the start of the project and the final configuration produced, noting any significant

deviations that were made during the course of the design implementation and explaining why those deviations occurred. We further record any interesting project facts that relate to the development of the Science DMZ at each small institution. At the end of the chapter, we summarize the major findings that specifically impact the design options of the Science DMZ on the small institution network. As a reminder, we randomized and neutered the specifics of the institution's design and implementation, both to satisfy our human subjects research and to focus less on the specifics of the small institution and more on the design approach that has transferrable qualities to other small institutions nationwide. Consequently, we address each design labeled as College A through College R.

6.2 College A

The small campus network of College A was remarkably robust for embarking on the Science DMZ design, as their initial border router was connected to their Internet2 and commodity Internet provider via a 1 Gbps fiber connection. The border router was mesh-networked to four segment routers via 10 Gbps links that pass through firewall ports prior to traffic distribution along 10 Gbps uplinked switches with 1 Gbps downlinks to user ports. A substantially built-out network in its own right, College A simply sought to build a Science DMZ network segment off the border router, already in place with 10 Gbps traffic capability, attach the segment to the Science DMZ switch, and install a perfSONAR node and a DTN. In addition, their network design included the expansion of the WAN port from 1 Gbps to 10 Gbps to accommodate the increased traffic

bandwidth required to exchange data with a large institution within their state, who was also a key partner in assisting with their WAN upgrade.

Figure 7 shows the network configuration before the Science DMZ build and Figure 8 details the College A network after the project is completed.

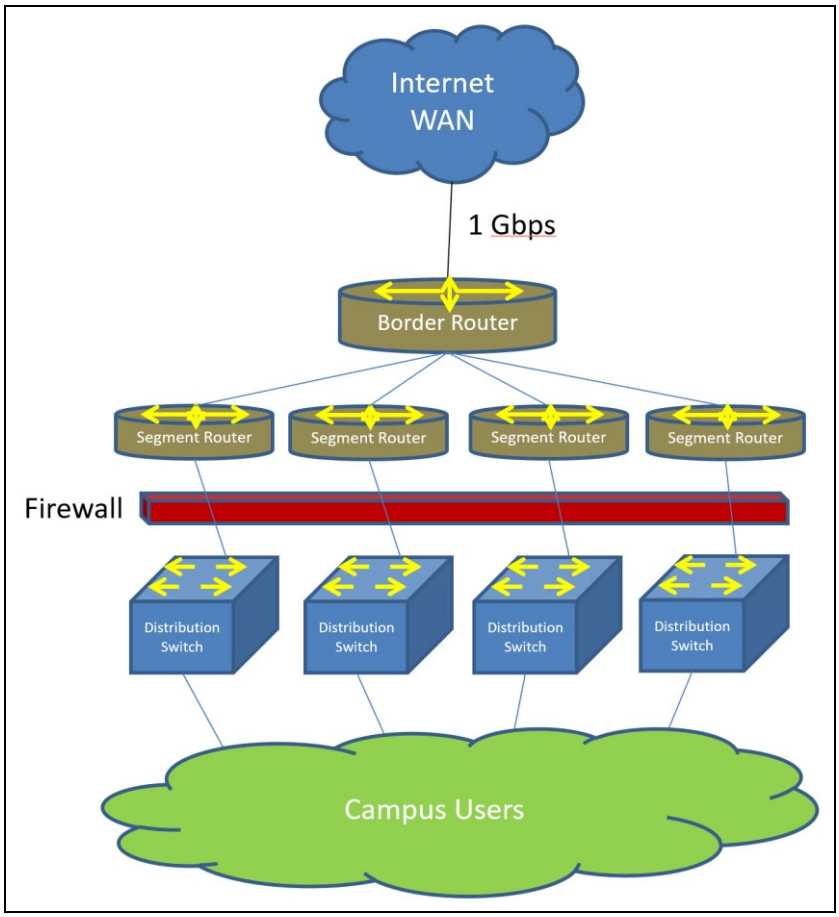

Figure 7: College A Campus Network Before the Science DMZ Project

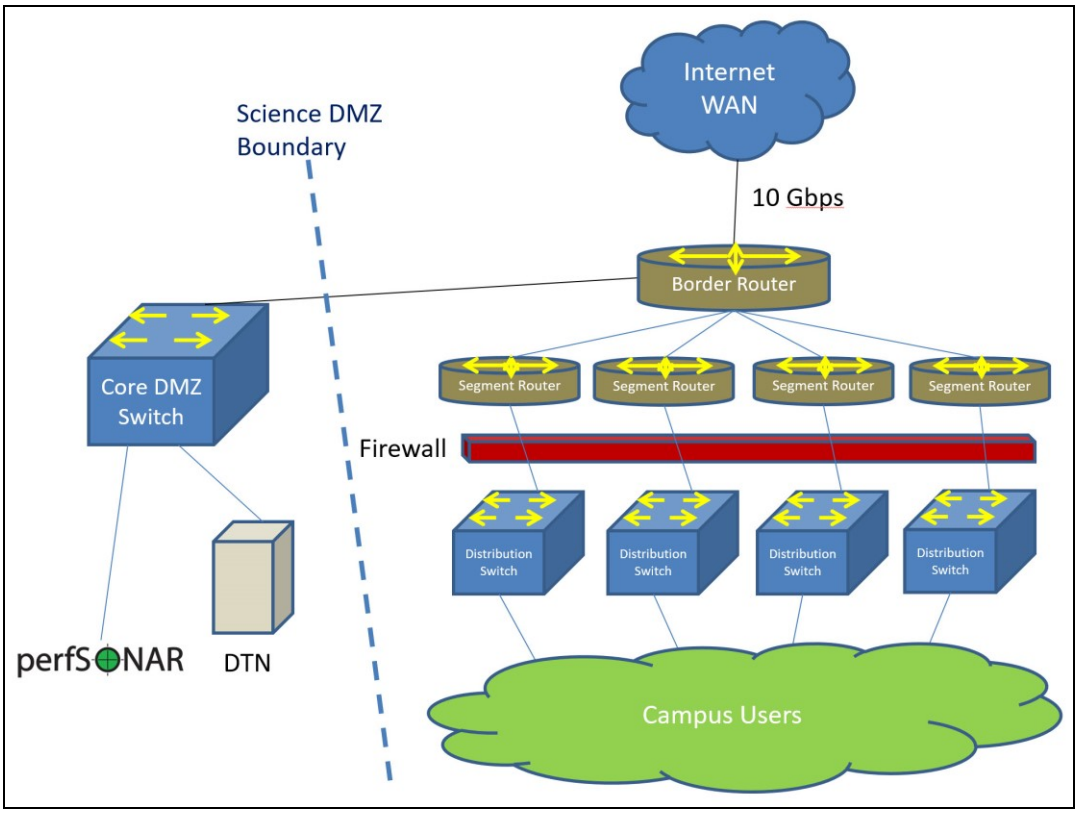

Figure 8: College A Campus Network After the Science DMZ Project

College A is a classic Science DMZ network configuration nearly identical to the model designed by ES.NET. This small institution had prepared for the Science DMZ expansion in advance of the grant project to support expanded network access for their enterprise users while upgrading the border router to handle 10 Gbps network flows both in and out of the campus. Pre-grant investments in the campus infrastructure, which overshadow the grant budget by nearly a factor of ten, made installation of the Science DMZ a useful experiment and a valuable investment in the campus environment.

6.3 College B

Small institutions are not synonymous with a lack of financial resources, and College B is another example of a small institution network that was already prepared for

a Science DMZ network. The existing campus network offered a border router connected to the WAN via a 1 Gbps link which, after passing through a firewall, served a core campus switch with 1 Gbps links both upstream and downstream to 1 Gbps-serving distribution switches at the users' ports. Unlike the ES.NET design model, College B chose to upgrade all of the network links from 1 Gbps to 10 Gbps, including the WAN link, and to upgrade the core switch on the inside of the campus network with one that supports 40 Gbps uplinks. That switch is connected via a direct 40 Gbps link to the WAN provider for science traffic, and VLANs within the campus switching environment allow the campus networking staff to isolate science equipment onto a high bandwidth science network that is both outside of the firewall boundary and on the high-speed link off-campus.

The design and final implementation was not clear in identifying some form of network monitoring, such as perfSONAR, as well as any deployment of a DTN as a Science DMZ support tool. The design philosophy presumes that researchers could be on any port of the campus network and, with a VLAN configuration, allow for data transfer from any port on the campus network through the core Science DMZ switch via the 40 Gbps WAN connection.

Figure 9 illustrates the College B network configuration before the Science DMZ implementation and Figure 10 details the network after the project is completed.

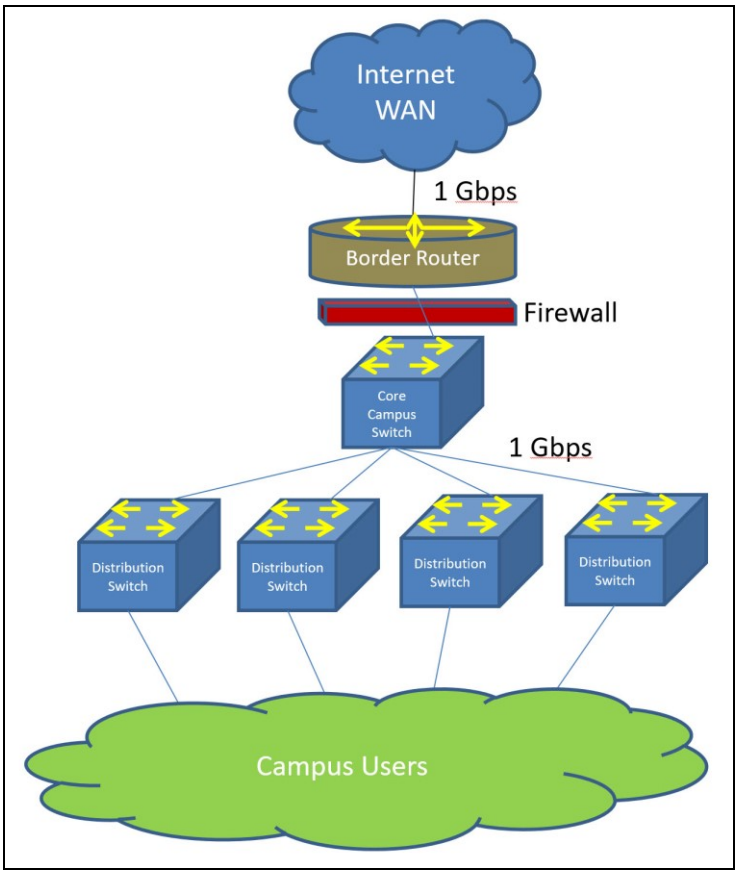

Figure 9: College B Campus Network Before the Science DMZ Project

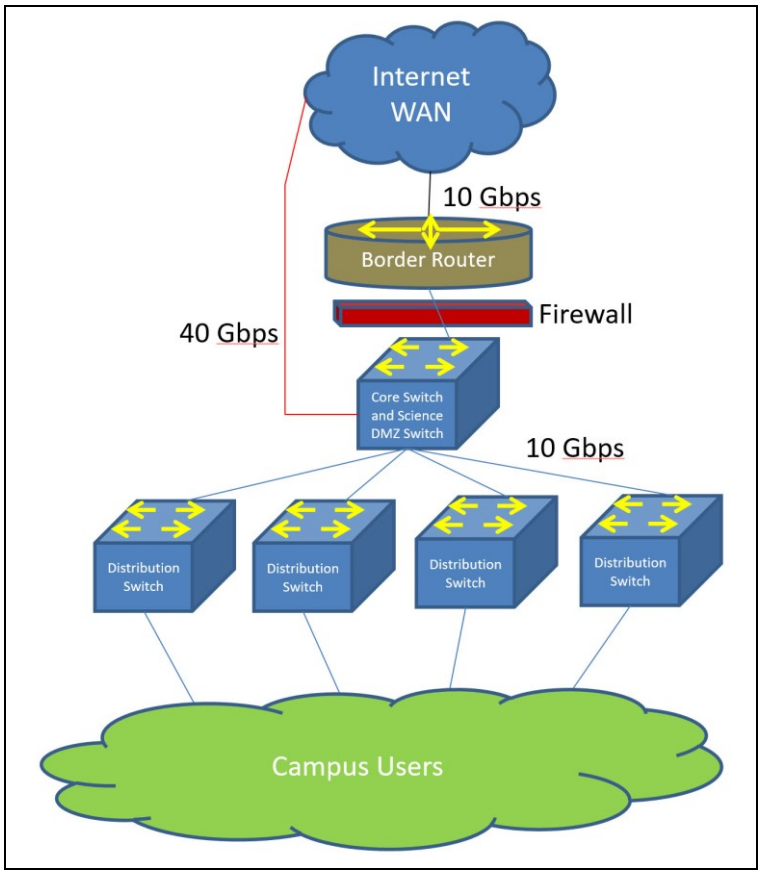

Figure 10: College B Campus Network After the Science DMZ Project

Following the ES.NET definition of a Science DMZ, we are challenged to classify this network upgrade as a true and complete implementation of the four major components of a Science DMZ. Much of the investment of this project went into the upgrading of the WAN and LAN links across campus, and the upgrading of the core switch to support firewall-bypassed data traffic. While a set of increased pathways off and around the network will assist in the transfer of data, without a dedicated DTN or monitoring tools such as perfSONAR, we believe that this implementation will fall short of fully achieving the outcomes that the research community at other small institutions with Science DMZs will enjoy. Had this small institution had a larger research partner to assist in the design, the outcome may have been different.

6.4 College C

The initial campus design on the College C network was fairly robust for the general campus community. Having a border router connected to commodity Internet services at 1 Gbps, their core network was a mesh of 10 Gbps central switches that connected 1 Gbps uplinks delivering 10/100 Mbps ports to the campus users over Power over Ethernet (PoE) switches. With a perfSONAR monitoring device already on the network, this small institution had some initial data that demonstrated the need for greater transmission capability to support scientific research. The Science DMZ network upgrade focused on three aspects—the installation of a Science DMZ switch with a 10 Gbps uplink, a DTN, and an upgraded WAN connection that is through a Sponsored Educational Group Participant (SEGP) 100 Mbps connection to Internet2, which is 10 times less bandwidth than the commodity Internet connection.

The College C network configuration prior to the Science DMZ implementation is found in Figure 11, with Figure 12 illustrating the post-project network configuration.

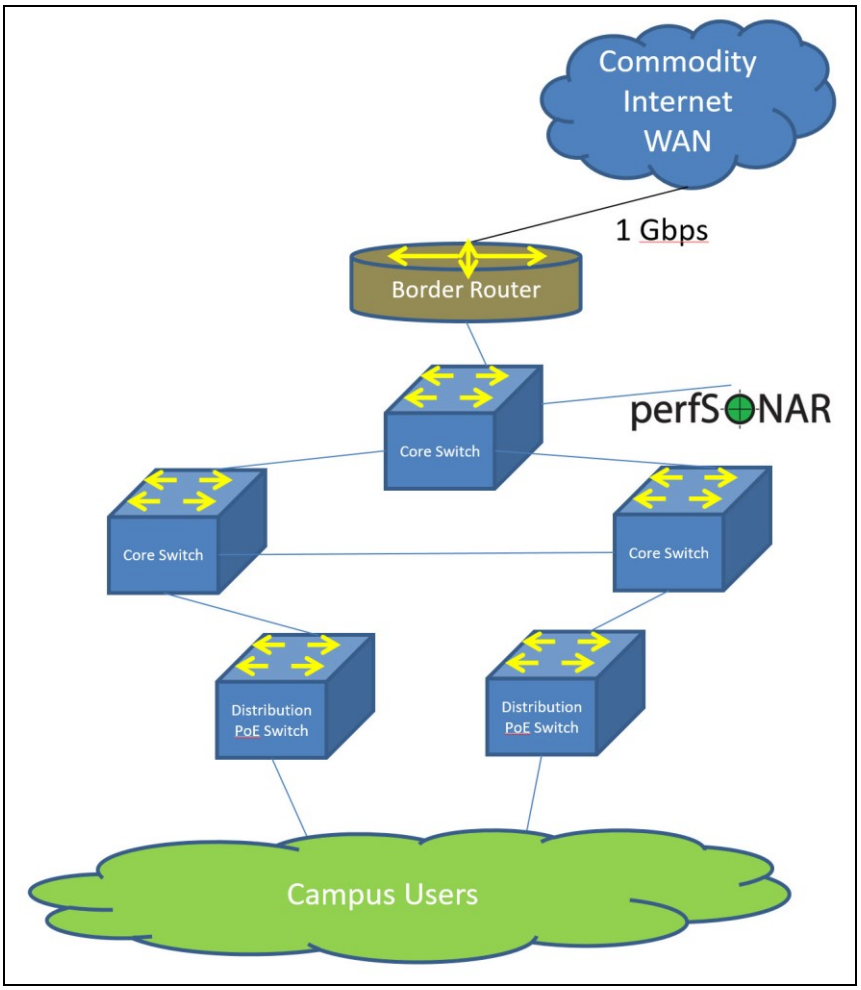

Figure 11: College C Campus Network Before the Science DMZ Project

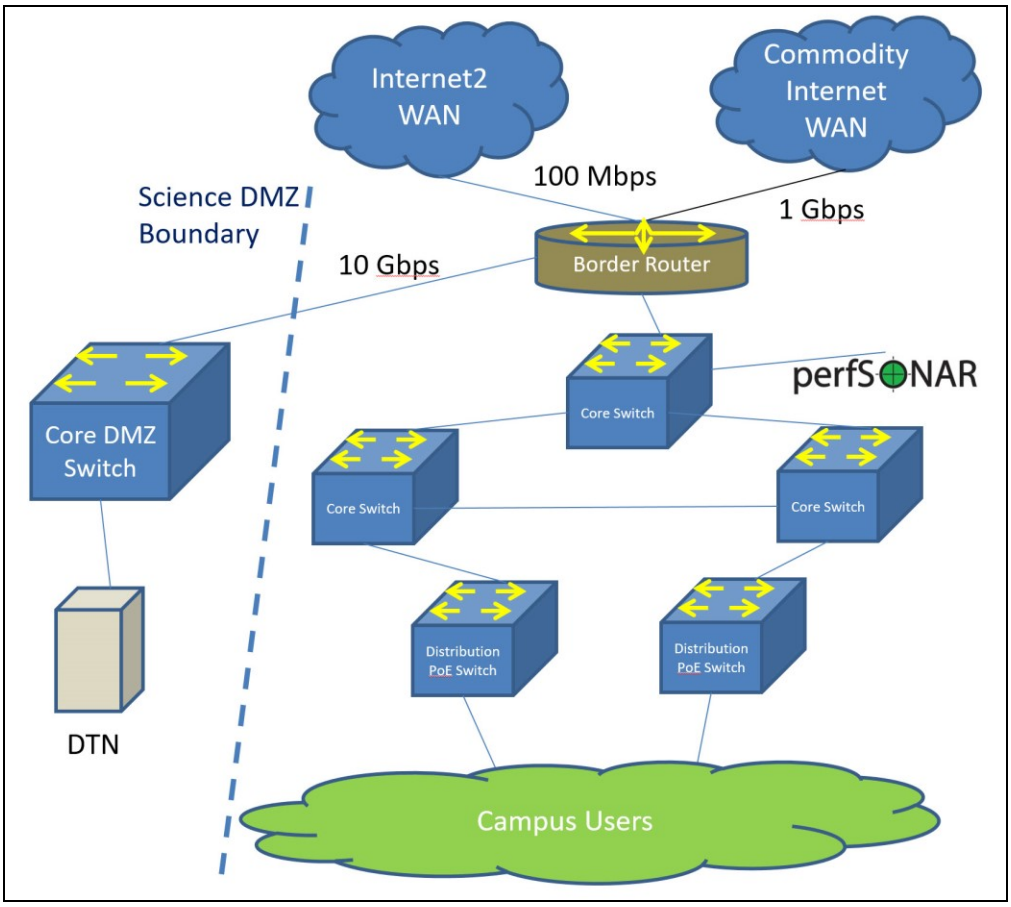

Figure 12: College C Campus Network After the Science DMZ Project

This network configuration is a standard Science DMZ configuration as compared to the ES.NET model. To have a small institution previously committed to the use of perfSONAR as a monitoring tool was a positive sign that the campus leadership was focused on ensuring that the campus network paths would operate well for all of the campus users. However, the 100 Mbps Internet2 connection seems to be a low research WAN connection given that the commodity Internet connection has already been upgraded to 1 Gbps. Yet, going from no Internet2 connectivity to even a small bandwidth level will impact positively the digital data transmission needs of the College C research community.

6.5 College D

The small institution campus network at College D was already served by two large border routers aggregating two Internet service providers with a combined bandwidth of 1.75 Gbps. With an existing campus network distribution plant of single mode fiber (SMF) between building locations, the College D campus user community enjoyed commodity Internet access for data transmission. Rather than building a Science DMZ, College D chose to invest in a 1 Gbps link to a local Internet2 Point-of-Presence (POP), and upgrade the border routers to support VLAN tagging and increased traffic over the campus network. College D did not install a perfSONAR network monitoring device, nor did they install a DTN to support the research community and make this network follow the Science DMZ network design.

Figure 13 details the College D network configuration prior to the project implementation with the post-project network upgrades revising the campus network map as presented in Figure 14.

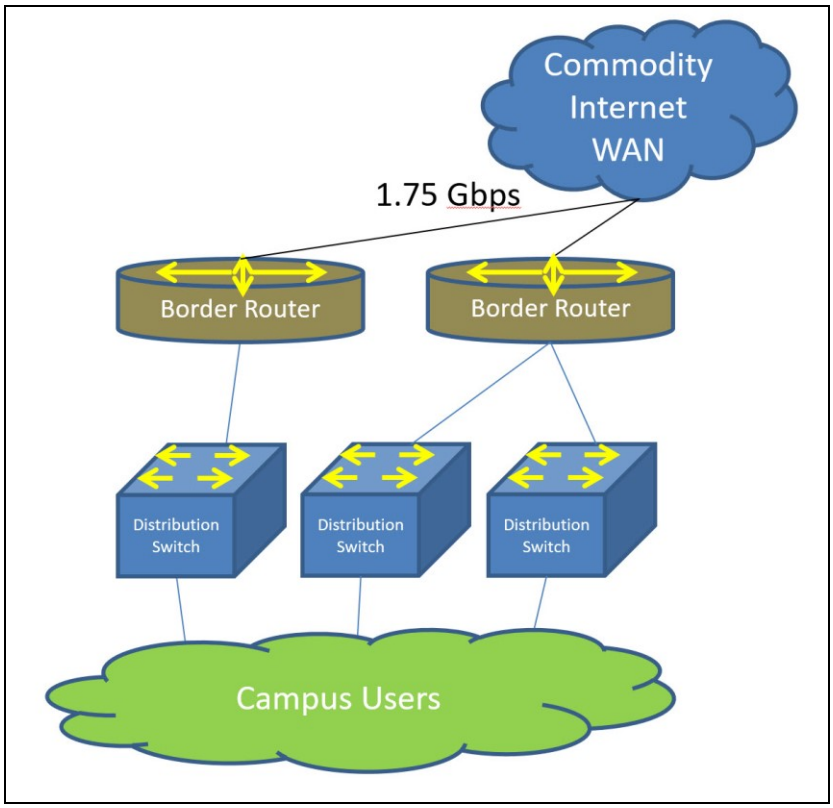

Figure 13: College D Campus Network Before the Science DMZ Project

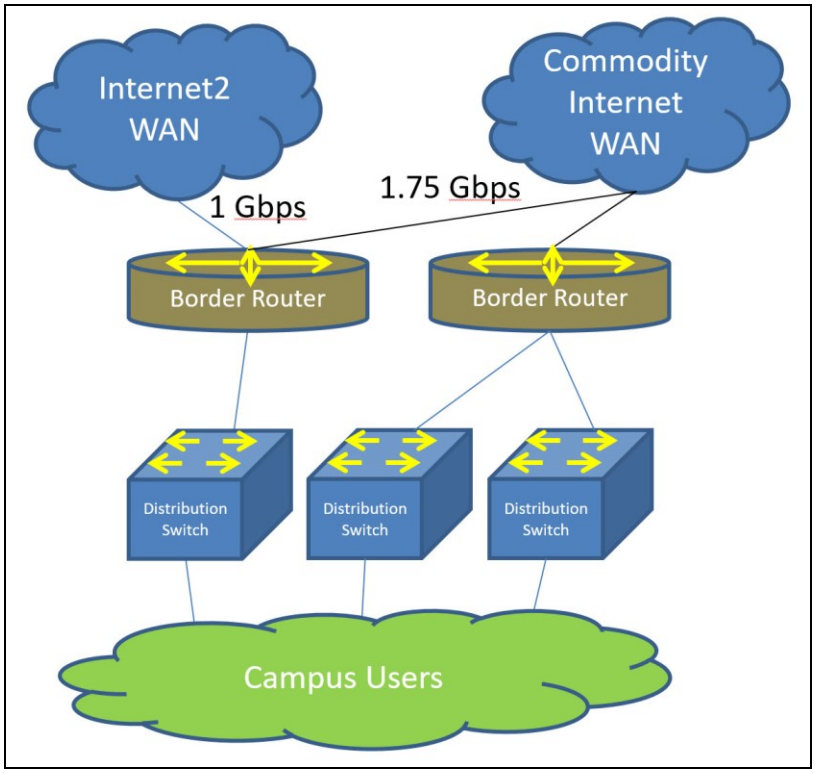

Figure 14: College D Campus Network After the Science DMZ Project

With such a well-distributed campus network environment already in place, College D realized the benefits of digital data transmission across campus with single-mode fiber optic cabling (SMF) in place. Like other campus networks elsewhere, we expect that the Internet2 connection will entertain the majority of the major data transmission for research, while the commodity Internet connections serve the regular campus community's need for communications exchange elsewhere. However, the lack of a basic DTN coupled with the absence of a network monitoring device such as perfSONAR prevents this design from achieving the full bandwidth benefits that a Science DMZ gains.

6.6 College E

The small institution network for College E originated with a 250 Mbps WAN link to a border router that was upgraded to support 1 Gbps prior to the start of designing a Science DMZ in conjunction with a NSF grant. Most of the core campus network was served by multi-mode fiber optic (MMF) cabling with 1 Gbps uplinks from the campus distribution switches that aggregate into a single campus core switch, which, in turn, served the science building with a 10 Gbps SMF distribution switch. The Science DMZ project for College E replaced all of the MMF with SMF across the campus, upgraded the core switch into a dual redundant switch, all with 10 Gbps capability, and improved the WAN connection with Internet2 connectivity at 10 Gbps. A Science DMZ switch was attached to the core router for scientific traffic to a newly-installed DTN and perfSONAR monitoring device.

The College E campus network connection before the revisions are located in Figure 15, with the Science DMZ implementation and campus network modifications identified in Figure 16.

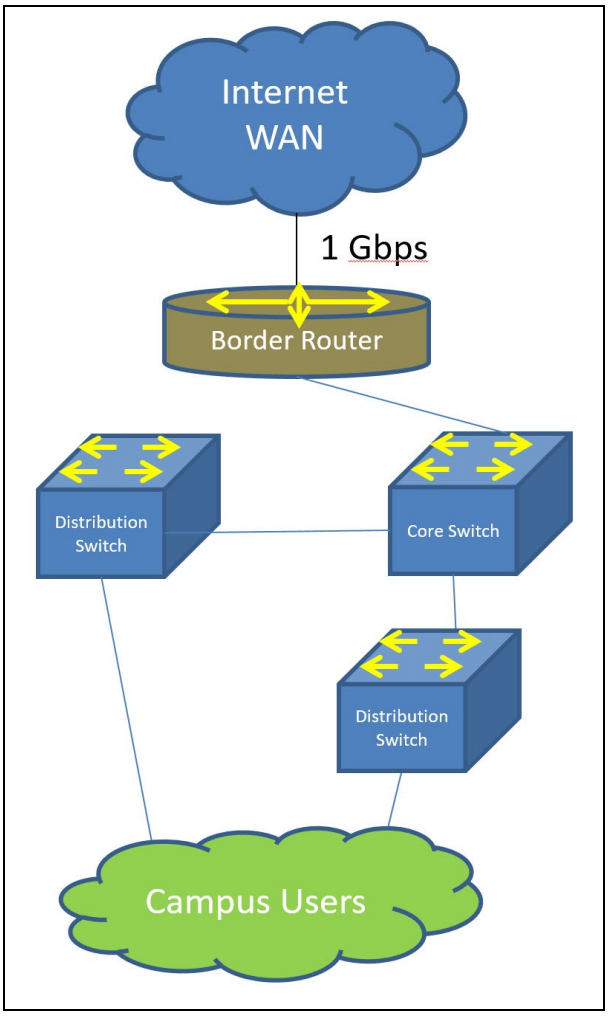

Figure 15: College E Campus Network Before the Science DMZ Project

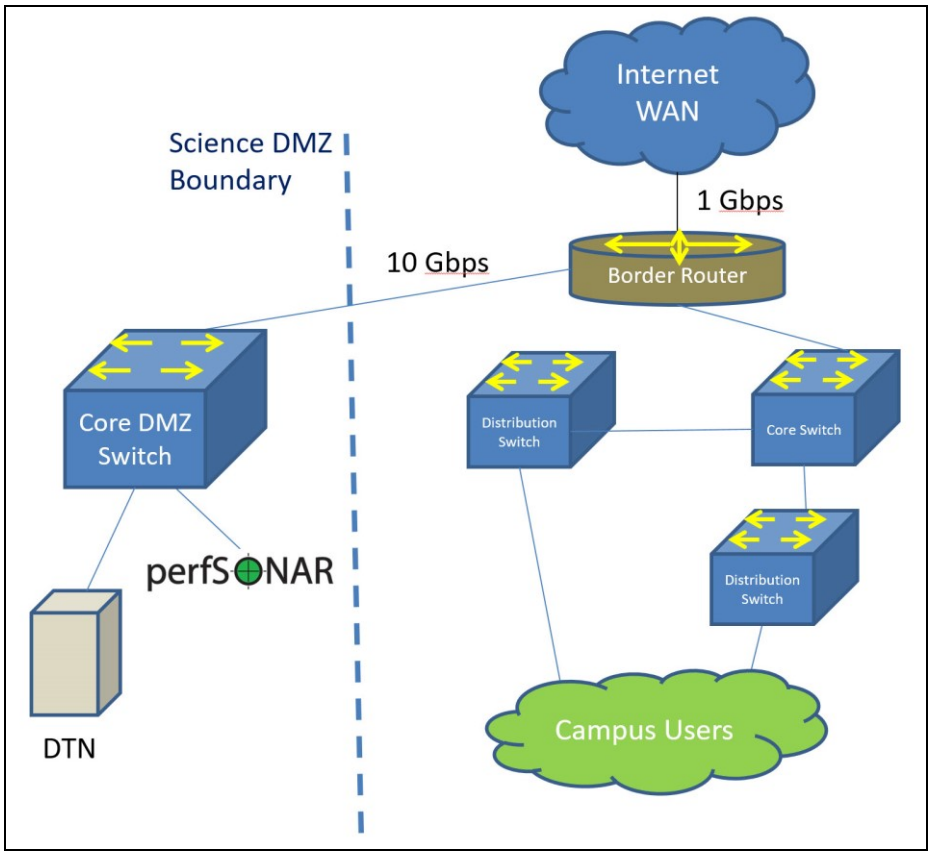

Figure 16: College E Campus Network After the Science DMZ Project

College E already seemed to be robust in their network service for science data transfers, but their Science DMZ implementation completed the full opportunity for researchers. The 10 Gbps connection to their WAN provider is a growing opportunity for both parties, as the WAN provider is a key partner in assisting College E with the upgrade of their campus network to support data-driven research.

6.7 College F

Similar to other small institution network upgrades funded by the NSF, College F has built a Science DMZ network in a familiar pattern. Receiving commodity Internet at 250 Mbps and Internet2 connectivity at 1 Gbps, College F was in a great position to support the research community. The border router offered a 1 Gbps link through the

campus firewall and the packet shapers to distribute network services via 1 Gbps MMF links to key buildings including the computer science (CS) department and a recently-developed science complex. The Science DMZ upgrade project for College F focused on upgrading both the commodity Internet connection to 500 Mbps and the Internet2 WAN link to 10 Gbps, as well as upgrading the entire campus network with SMF. The border router, core switch, and distribution switches in the CS department and science complex building were upgraded to support 10 Gbps links, as well as VLAN tagging to directly move traffic flows from the router to the CS department and Science DMZ network switches. A perfSONAR device was connected in the science complex to monitor traffic to and from the College F network.

The network configuration for College F at the beginning of the Science DMZ upgrade project can be found in Figure 17. A post-network installation campus map is located in Figure 18 below.

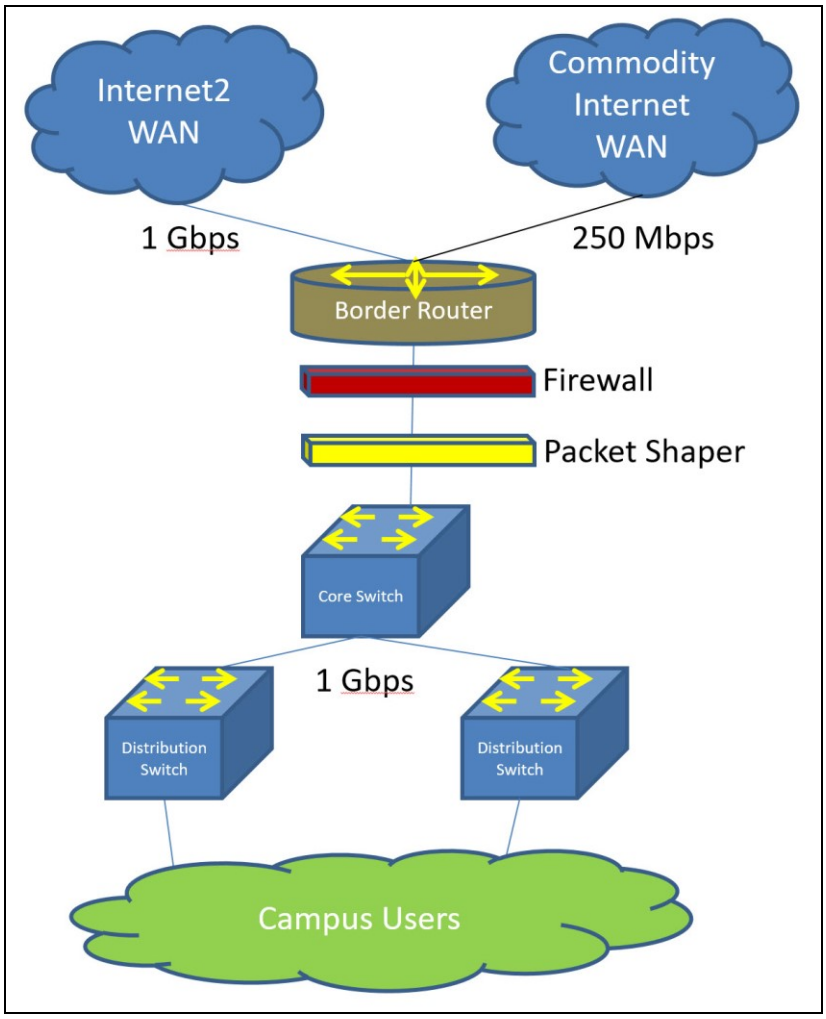

Figure 17: College F Campus Network Before the Science DMZ Project

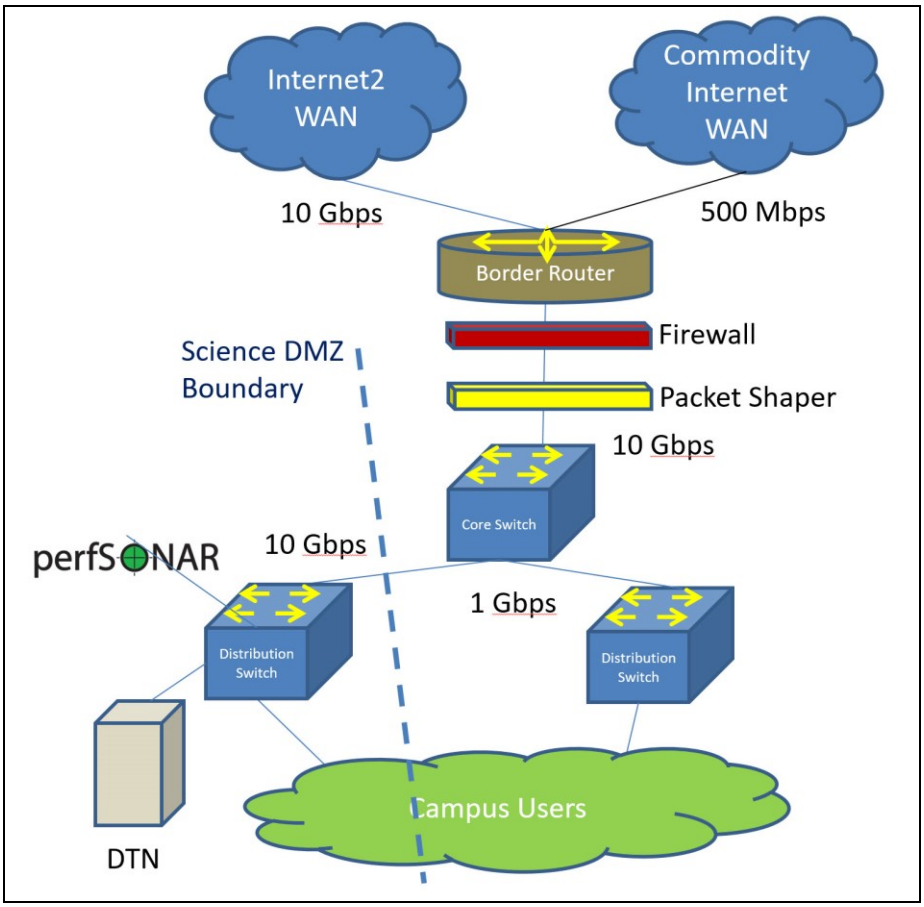

Figure 18: College F Campus Network After the Science DMZ Project

The College F network uses VLANs to define a pathway through the firewall and the packet shapers to pass traffic directly to the Science DMZ switch in the science complex. The Internet2 service provider was a key partner in the completion of this project, and the enhanced 10 Gbps link propelled College F's network to one that can support large data flows.

6.8 College G

The College G network offers a simple campus network at the onset, supporting a 4 Gbps Internet2 WAN connection with a border router passing campus LAN traffic through a firewall to a 1 Gbps distribution switch. College G chose to install a Science

DMZ switch with 10 Gbps links from the border router to support science equipment. The Science DMZ is monitored by a perfSONAR box, and with additional funding, the campus distribution switch and the WAN connection were both upgraded to 10 Gbps links, lifting the whole campus with better connectivity through the investment in science data flows.

Figure 19 describes the network configuration initially in place at College G prior to the start of the Science DMZ upgrade project. Figure 20 outlines the College G network with the Science DMZ in place.

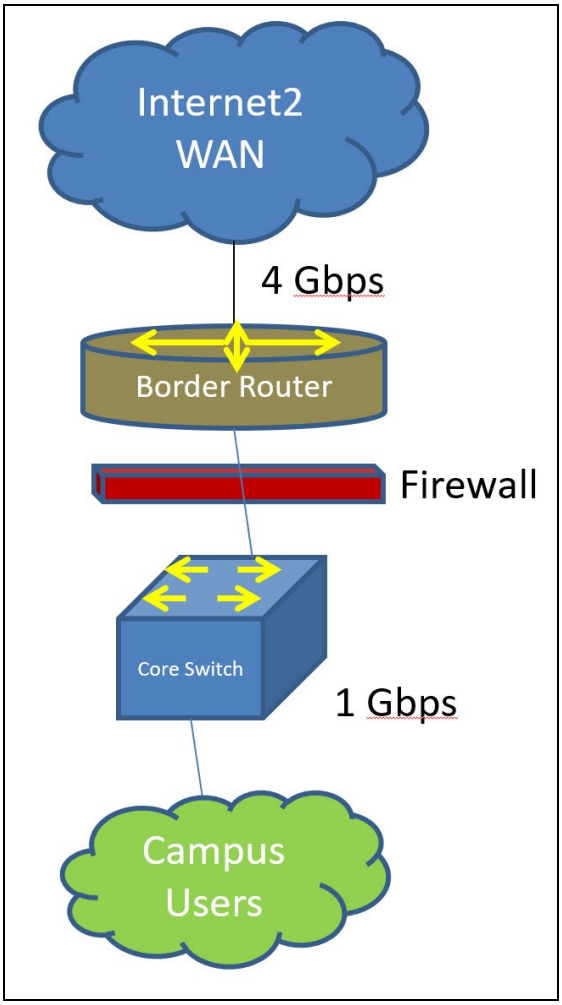

Figure 19: College G Campus Network Before the Science DMZ Project

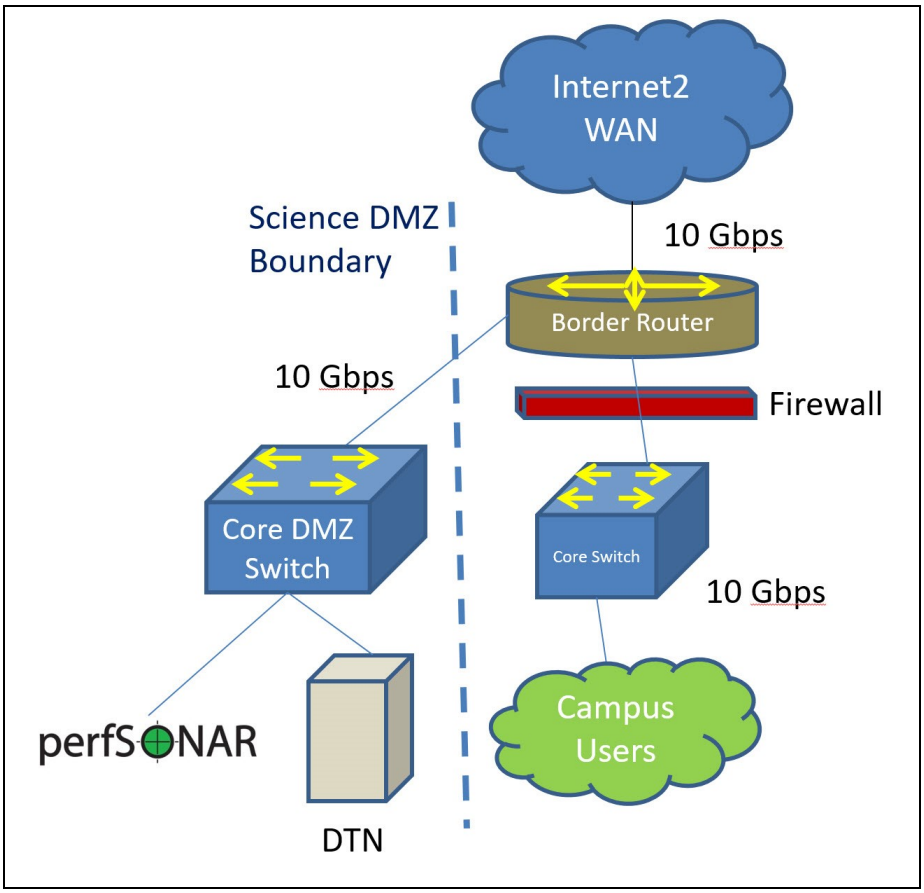

Figure 20: College G Campus Network After the Science DMZ Project

College G approached the design and implementation of the Science DMZ entirely on their own as there were no large institutions nearby that could assist with this project, though they relied on conceptual assistance from the documentation found at the ES.NET website and archived on the development of large institution Science DMZ networks.

6.9 College H

We observe a similar network design for College H as we have seen on other small campuses in this study. The Internet2 WAN provider for College H presented 800 Mbps to the campus border router, which delivered between 10 Gbps and 40 Gbps,

depending on the distribution switches attached to the network, which in turn delivered between 100 Mbps and 1 Gbps to the campus user community. This basic network design served the users with commodity Internet and Internet2 WAN service for multi-purpose digital communication. The Science DMZ project enhances this network with a traditional Science DMZ core switch with 10 Gbps links to support a DTN, perfSONAR, and research equipment. The Internet2 WAN connection is also upgraded to support 6 Gbps, with the ability to burst 10 Gbps when needed.

The starting network configuration for College H is recorded in Figure 21 below. After the Science DMZ project was completed, the final network configuration for College H is found in Figure 22.

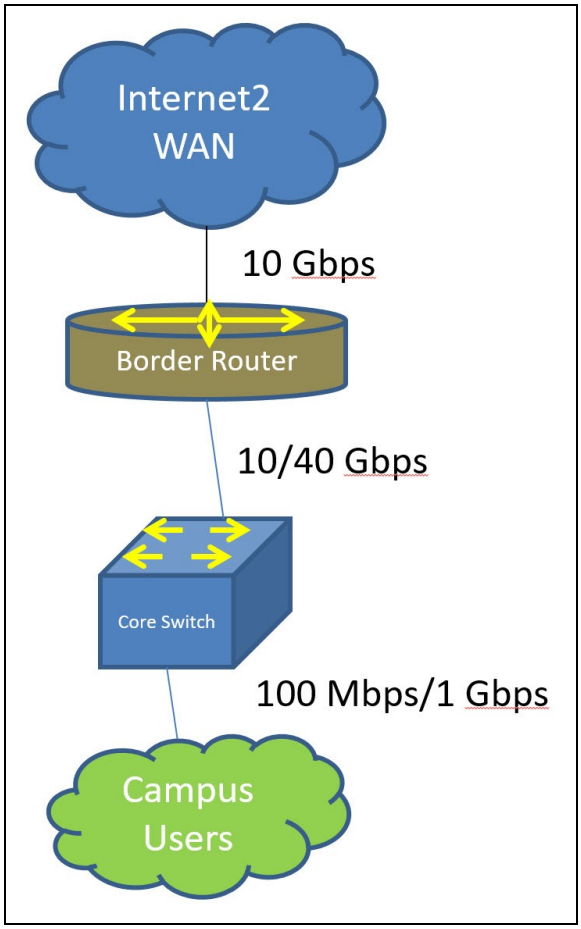

Figure 21: College H Campus Network Before the Science DMZ Project

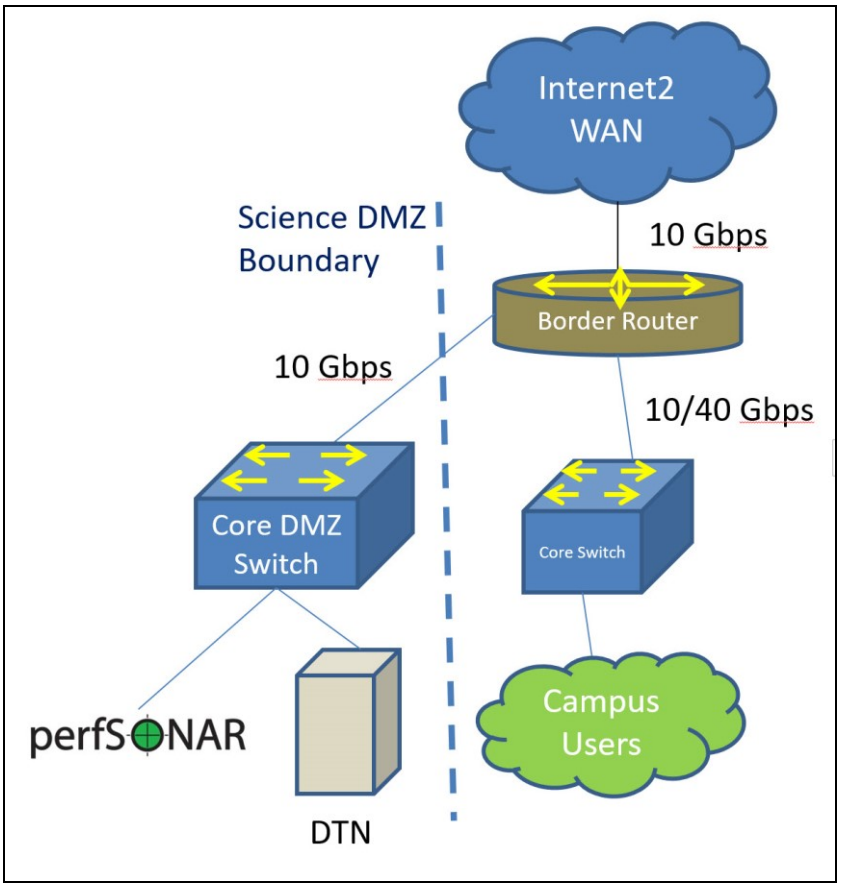

Figure 22: College H Campus Network After the Science DMZ Project

Besides the experience that the technology leadership from College H had in-house, College H project leadership reached out to another small institution Science DMZ grant award recipient for guidance and feedback on their emerging designs, which could be a factor in the stark similarities between College H and their consulting partner.

6.10 College I

Another basic small campus network design is visible at College I. The WAN service at 250 Mbps delivered commodity Internet to the campus border router, passing through the firewall, to the core switch. The distribution switches connected to the core switch had 1 Gbps uplinks and served each campus port with 100 Mbps for users. After

the implementation of the Science DMZ, which College I did by installing a Science DMZ core switch with 10 Gbps links between the research equipment and the campus core switch, the network devices were configured to use VLANs to transmit traffic between the WAN port and the Science DMZ network. In addition, all of the campus hardware was replaced with equipment that supports between 10 Gbps and 20 Gbps as uplinks, while campus user ports can serve up to 1 Gbps of bandwidth. To support all of the additional network traffic required, the WAN port was upgraded to 2 Gbps with the capacity to grow to 10 Gbps before requiring a hardware upgrade. perfSONAR was installed on the core Science DMZ switch to monitor the network. In addition, the campus was connected to Internet2 through the campus ISP, who was a critical partner in supporting the Science DMZ design and build.

Figure 23 describes the network configuration initially in place at College I prior to the start of the Science DMZ project. Figure 24 outlines the College I network with the Science DMZ in place.

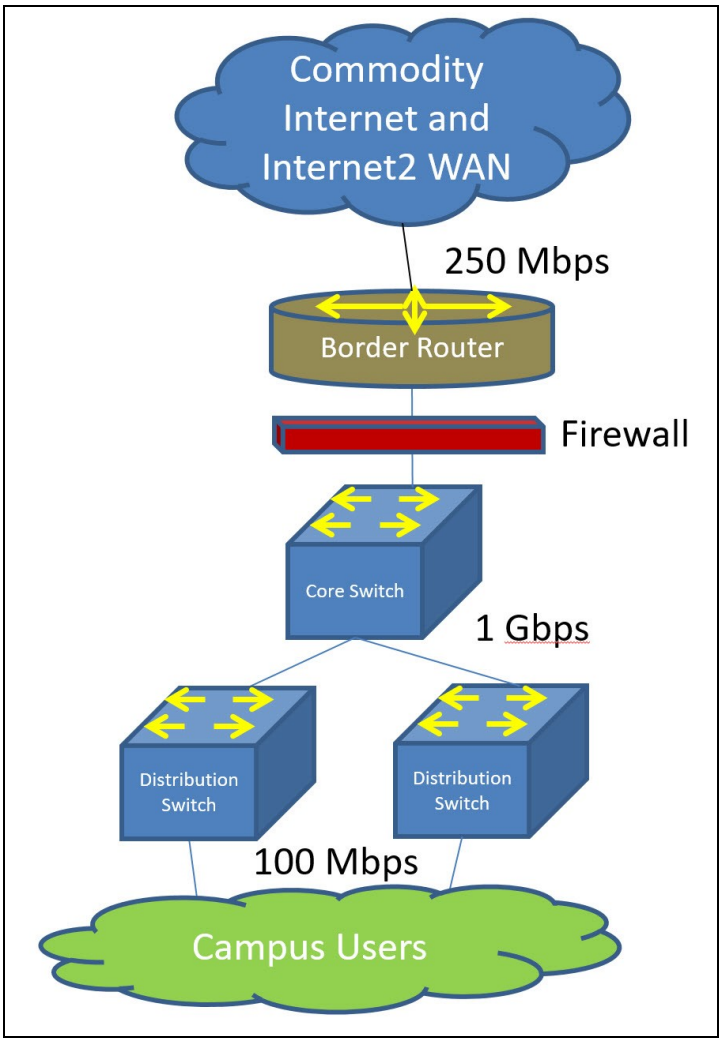

Figure 23: College I Campus Network Before the Science DMZ Project

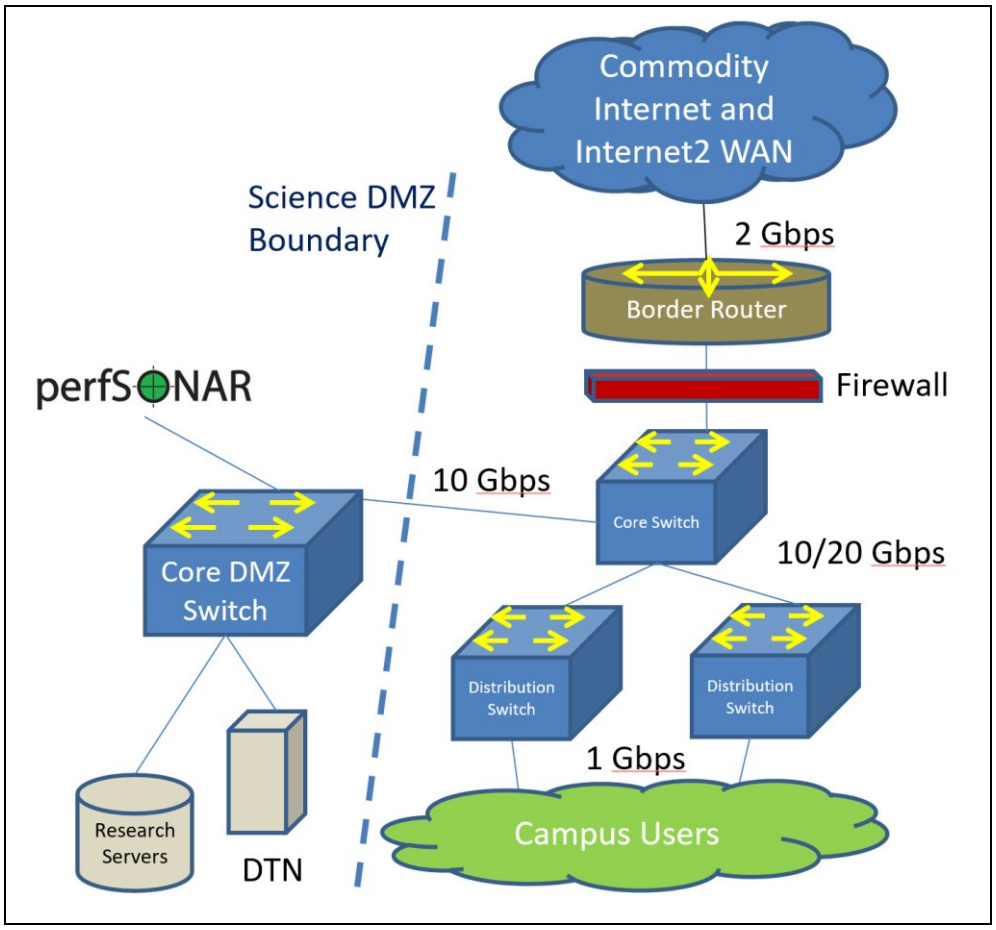

Figure 24: College I Campus Network After the Science DMZ Project

College I, like many small institutions, took advantage of the opportunity when upgrading the network to support a Science DMZ, to also upgrade the general campus user community with better networking services. This Science DMZ installation that follows the ES.NET model will be in a position to grow as demand requires additional bandwidth and resources.

6.11 College J

With new leadership arriving at the start time of the Science DMZ network design and installation project, College J took the opportunity to consider what would make the greatest impact on the researchers who would use the Science DMZ. College J's existing

network had a 1 Gbps Internet2 WAN connection to their ISP, who was a key partner in the design and installation of the campus network upgrades. Eight border routers in a mesh environment supported both College J and a downstream educational institution, while campus-destined traffic passed through a firewall before being distributed to the general user community. College J, unlike others, began their project by formally surveying the research community through electronic surveys and in-person focus groups to identify the key requirements that they had relative to data transfer. Their final installation included a Science DMZ core switch connected to the border routers via 10 Gbps, monitoring via perfSONAR, connecting a DTN to the network to support data exchange, and upgrading the Internet2 WAN connection to 10 Gbps.

Figure 25 describes the network configuration initially in place at College J prior to the start of the Science DMZ upgrade project. Figure 26 outlines the College J network with the Science DMZ in place.

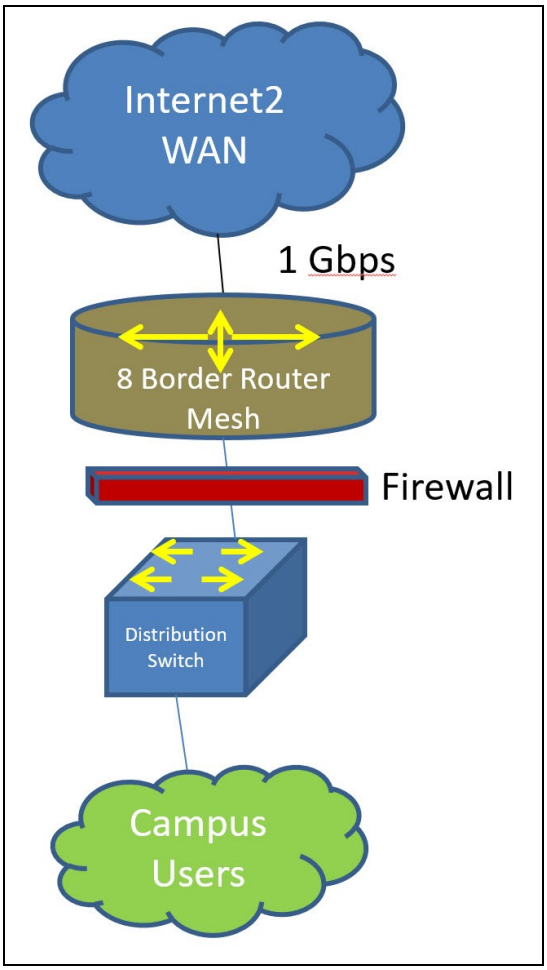

Figure 25: College J Campus Network Before the Science DMZ Project

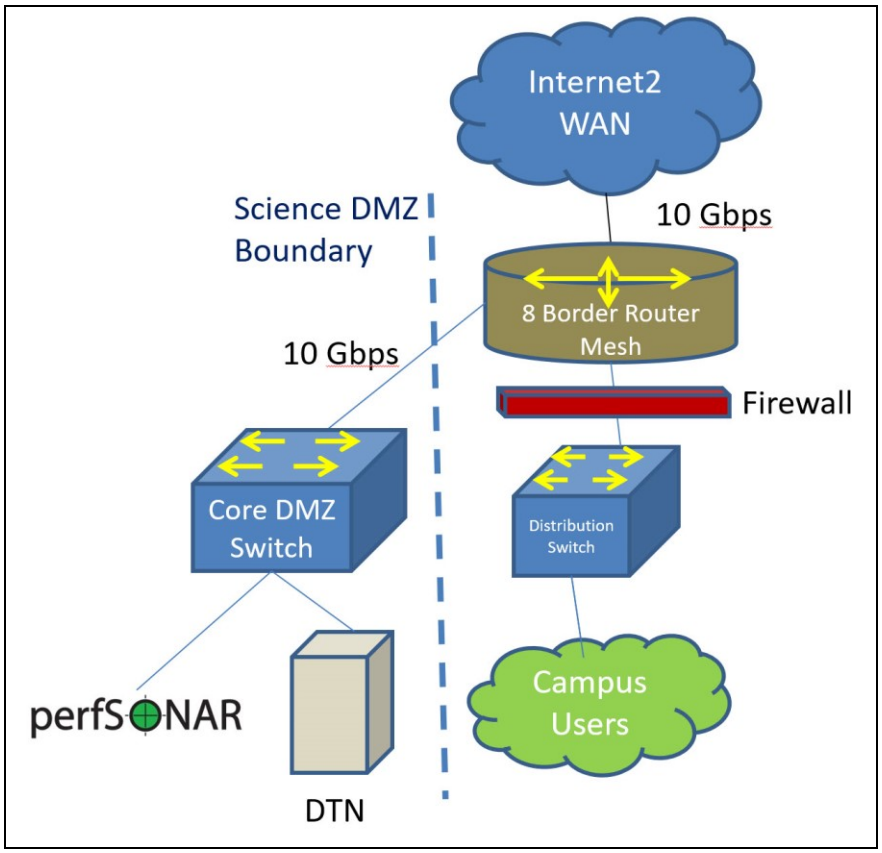

Figure 26: College J Campus Network After the Science DMZ Project

The philosophy of College J's network investments and design is to attempt to match as best as possible those resources that would be installed at a large institution, in spite of being a small institution. The science drivers that led to the justification for the College J Science DMZ has encouraged the transfer of digital data with other institutions beyond their local campus network.

6.12 College K

The pre-existing campus network for College K, like other small institutions, consists of a distributed MMF network serving the campus community with network services. The border router of College K had a WAN connection that had been recently upgraded from 40 Mbps to 200 Mbps. The Science DMZ network project award allowed

for the installation of a Science DMZ core switch, an upgrade of the border router to support 10 Gbps traffic to and from the Science DMZ, the installation of a DTN for traffic exchange with off-site researchers, as well as a perfSONAR monitoring device to maintain visibility into the network. The WAN connection was upgraded to 1 Gbps, and the entire campus network was upgraded to SMF to offer up to 10 Gbps to every network port across College K's campus network.

The pre-existing network configuration for College K prior to the Science DMZ project is detailed in Figure 27, while the post-Science DMZ network deployed for College K is highlighted in Figure 28.

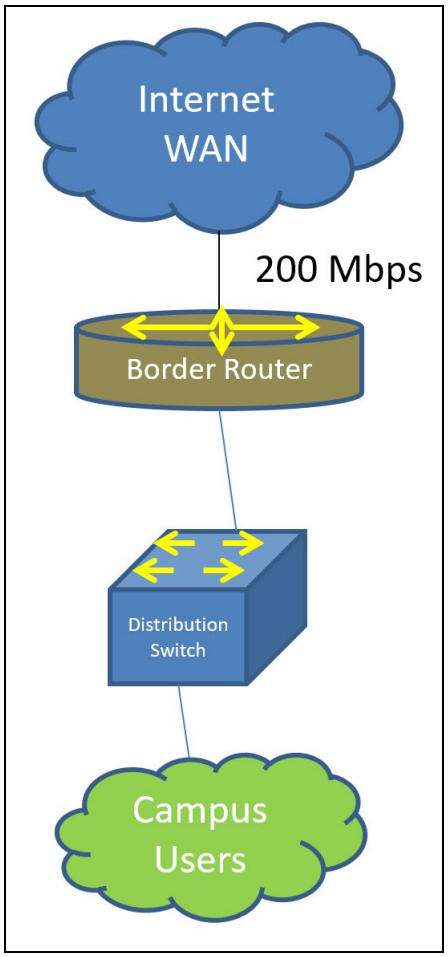

Figure 27: College K Campus Network Before the Science DMZ Project

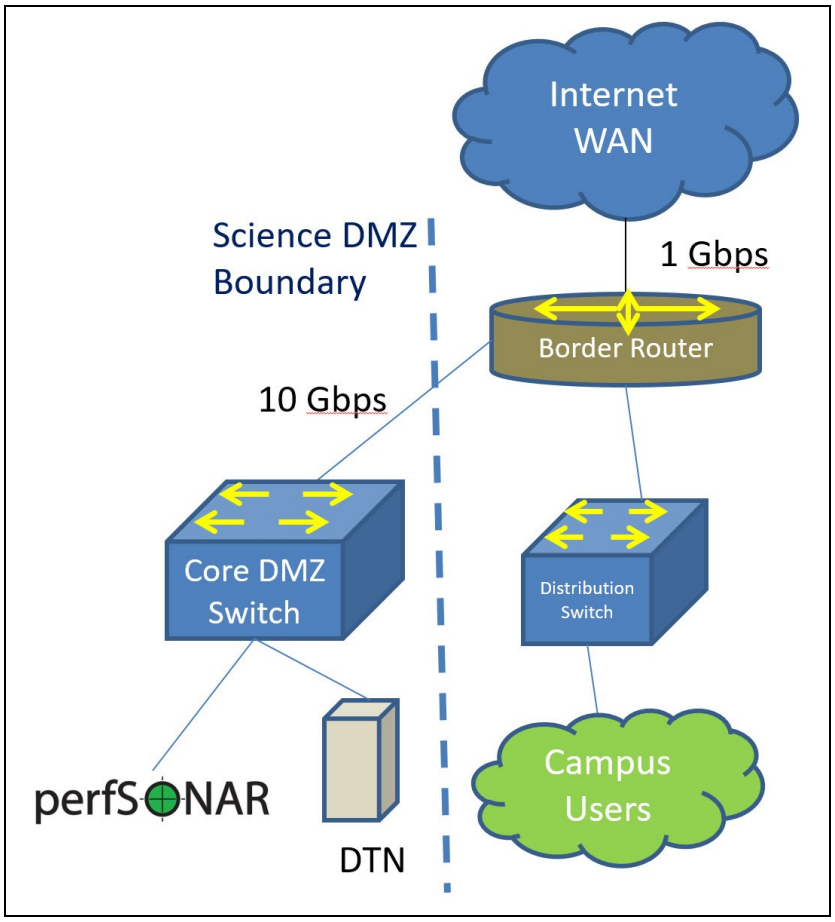

Figure 28: College K Campus Network After the Science DMZ Project

Working in collaboration with a large institution within the state, the College K network is prepared to connect to other research institutions across Internet2 using the basic Science DMZ network model and an increased WAN connection.

6.13 College L

The network configuration of College L is quite different from other small institution configurations, as College L is a combination of seven locations in in one area. Each area is served by a distribution switch connected via a border router managed by the WAN provider offering 51 Mbps of commodity Internet access. Working closely with the WAN provider, the overall network environment was upgraded to 100 Mbps

Internet2 WAN connectivity with a new Science DMZ core switch added to one of the sites providing a DTN and perfSONAR. The increase in WAN capacity allows the campus network to take advantage of a remotely-accessible Learning Management System and a private cloud storage facility, though that access is out of the scope of the Science DMZ specifically.

Figure 29 describes the seven-site campus network configuration in place prior to the Science DMZ project at College L. The revised network configuration, including the Science DMZ elements, is detailed in Figure 30.

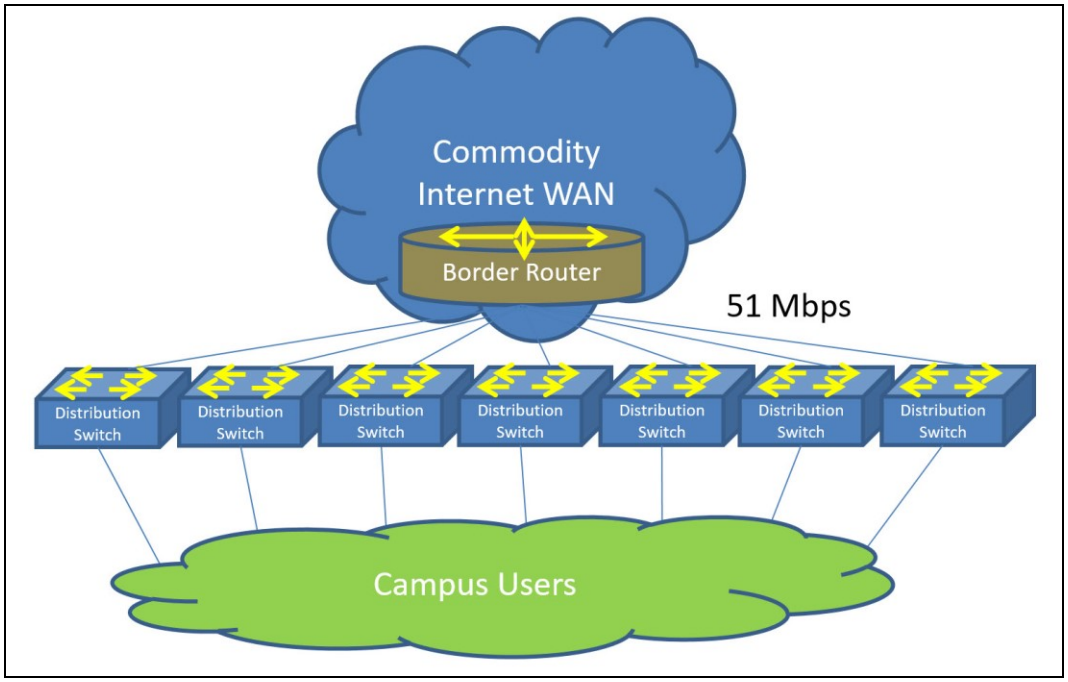

Figure 29: College L Campus Network Before the Science DMZ Project

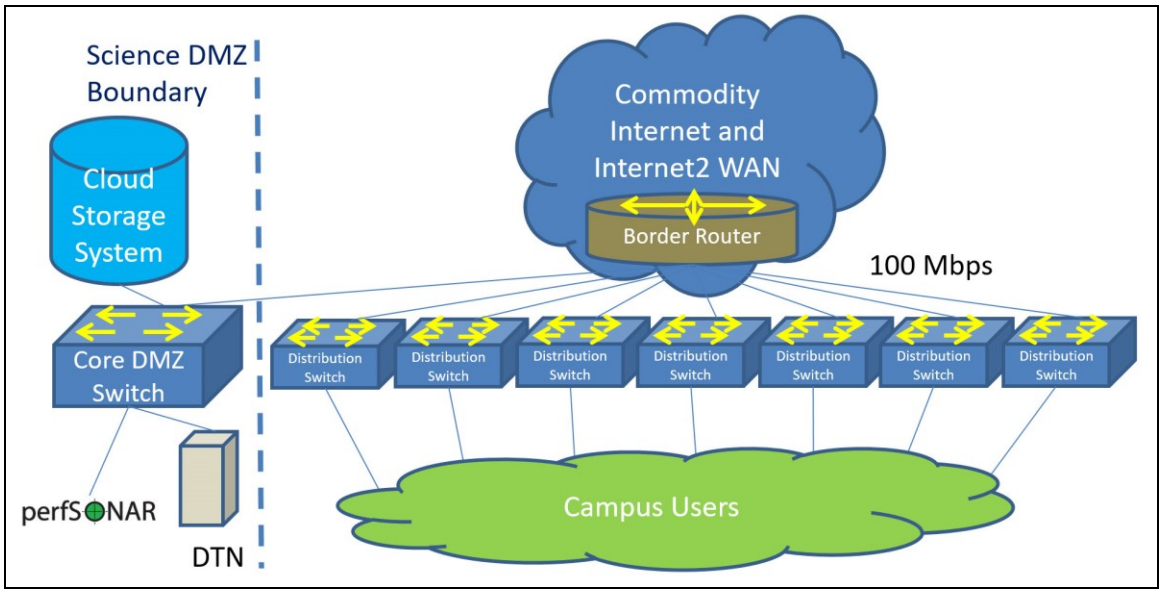

Figure 30: College L Campus Network After the Science DMZ Project

While the initial network configuration doesn't appear to follow other campus designs, the Science DMZ installation at College L aligns very well with the ES.NET model, and serves this small institution with the additional bandwidth and technology tools to advance the transmission of digital research data between the College L Science DMZ and with external research partners.

6.14 College M

The physical location of College M to its research partners exceeds 80 miles and traverses 2 network hops introducing a 50 msec latency on their initial 100 Mbps WAN connection. Outside of that distinctive motivation for building a better WAN pathway, the campus network appeared similar to most other small institutions, with a border router and firewall protecting the campus core and distribution network with 1 Gbps uplinks offering 100 Mbps links to campus users' desks. In partnership with both the WAN provider and the research partners at the other end of the WAN, College M built a

data center in which the Science DMZ core switch with a perfSONAR monitoring device was installed, in addition to upgrading the border router and core campus switch with 10 Gbps connectivity. All of the campus upgrades were primarily to support the construction and development of a 10 Gbps WAN uplink to Internet2 that bypassed the original physical pathway and offers 5 msec of latency, making data transmission across the WAN significantly available to researchers with datacenter connectivity.

A network design of College M's environment prior to the Science DMZ project is in Figure 31, with Figure 32 offering the layout of the College M network with the Science DMZ and other components installed.

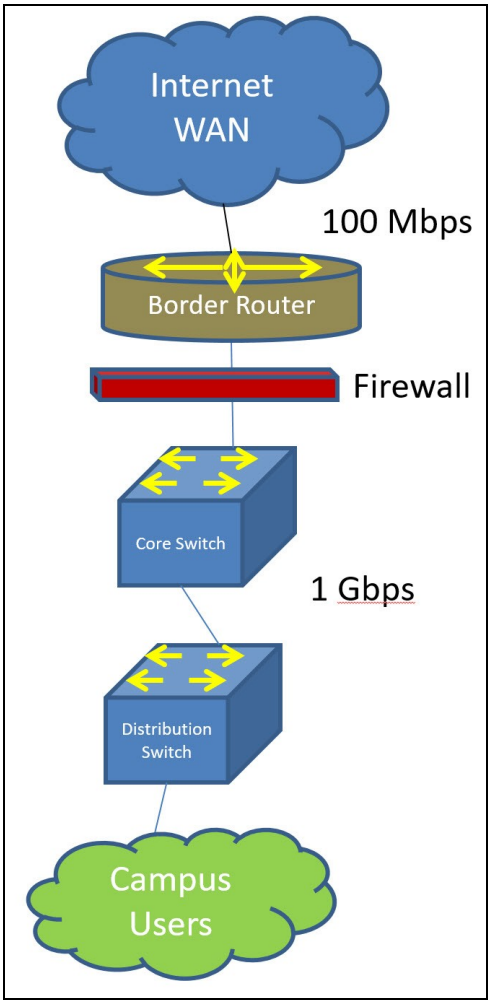

Figure 31: College M Campus Network Before the Science DMZ Project

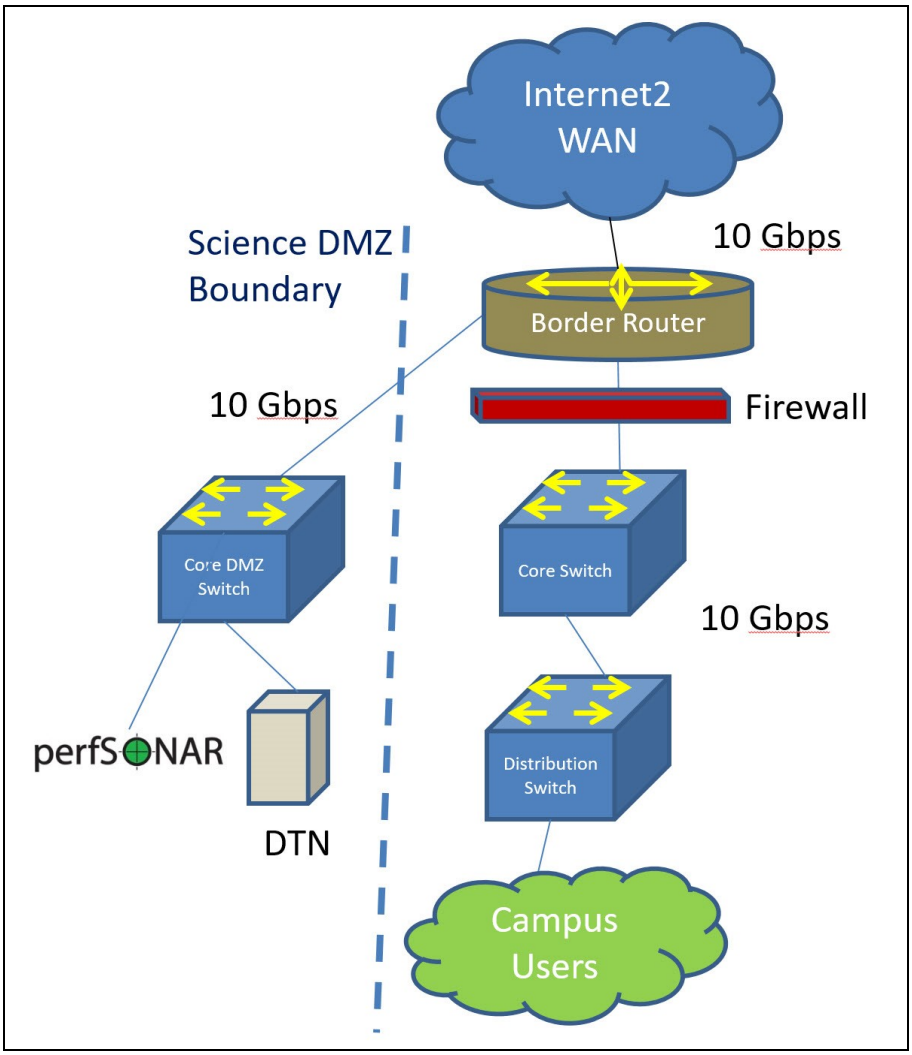

Figure 32: College M Campus Network After the Science DMZ Project

The College M opportunity to decrease latency through a revised WAN connection offered a significant benefit to the research community that further upgrading of the Science DMZ network equipment would be in a position to address.

6.15 College N

While not specifically submitted as a collaborative project, College N and College O are closely related because they share the same research and WAN service provider

partner and are connected together via a private fiber link between the two campuses, though any upgrade of the link is not specifically part of College N's Science DMZ project. College N had configured the network with a core switch ahead of the campus router. Besides their fiber connection, College N and College O both have been granted a no-cost extension (NCE) from the NSF for an additional year to complete their respective projects, so much of the design work that has been proposed is still not completed as of this research project.

The switch and router ports were 10 Gbps, and the core switch has a WAN connection with a 1 Gbps service level. The campus is further served in two places: a data center with a distribution switch offering 10 Gbps connectivity to a 1.5 PB NetApp Storage Area Network (SAN), and the rest of campus with a distribution switch with a 1 Gbps uplink due to a campus cable plant of MMF. Since this project is still in development and not yet completed, the current working Science DMZ design plan included modification of the existing data center with a 10 Gbps switch, which is expected have a DTN connected to it. The WAN connection is expected to be upgraded to 10 Gbps, and the link from the core router to the campus distribution switch is slated to be upgraded with SMF and link connections at 10 Gbps.

Figure 33 describes the network configuration in place today at College N, without the link to the peer college. Figure 34 is the current working design plan for the College N Science DMZ network upgrade.

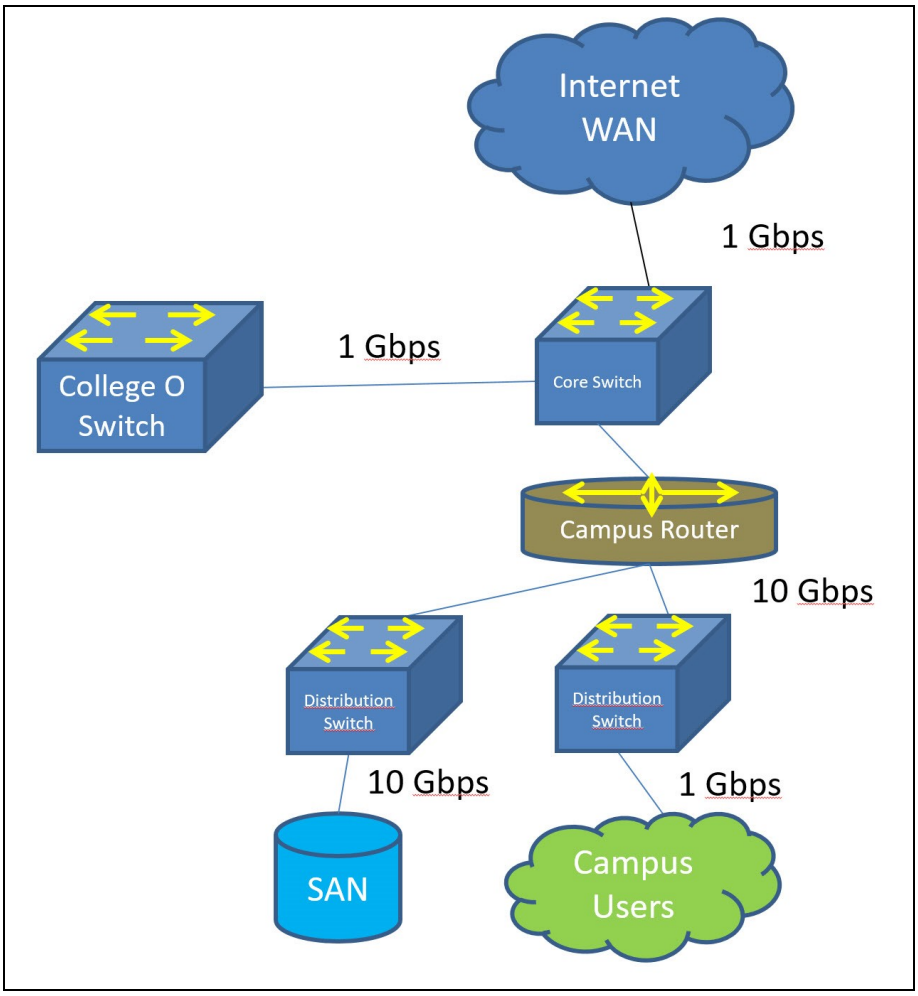

Figure 33: College N Campus Network Before the Science DMZ Project

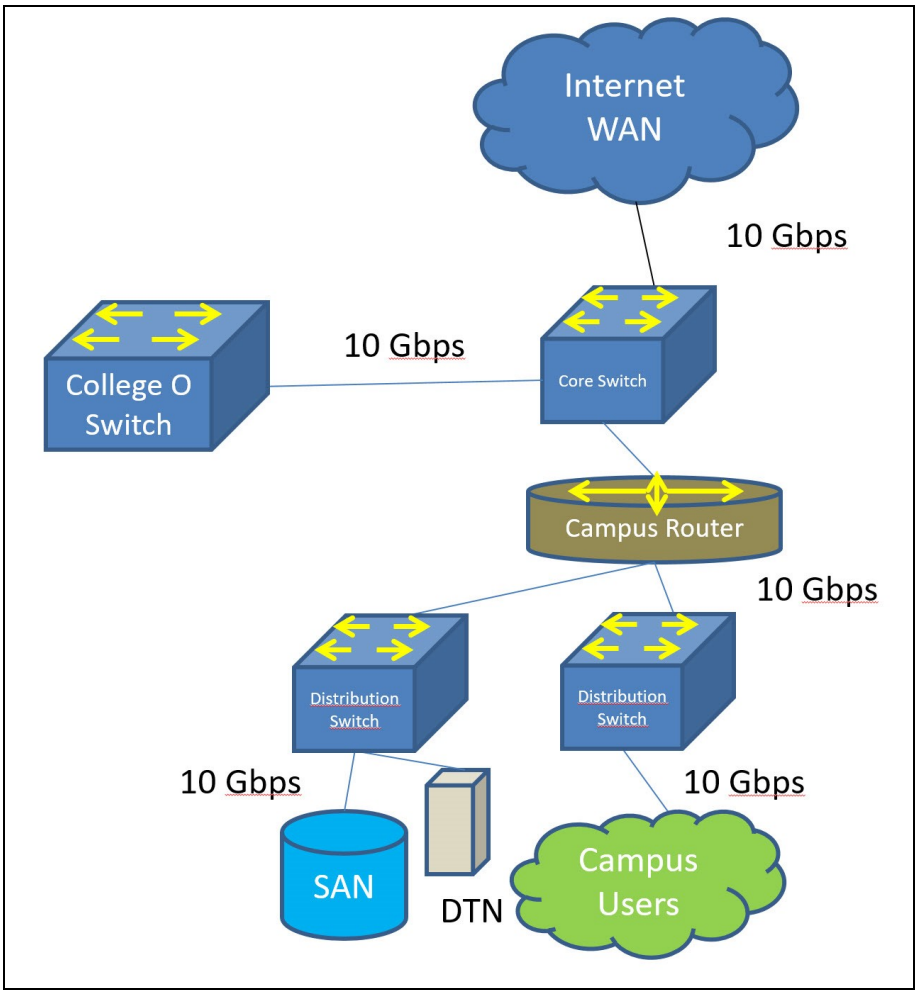

Figure 34: College N Campus Network After the Science DMZ Project

This Science DMZ project has been granted a no-cost extension (NCE) to complete in 2018, so the final implementation is still ongoing, though the design has been completed and is on schedule to complete during the extension period.

6.16 College O

College O is the other side of a separate, yet collaborative, project with College N that serves as a research partner, shares the same WAN provider that is also collaborating on the Science DMZ project, and operates a private fiber link between College O and

College N with a 1 Gbps bandwidth rate. Additionally, this project has been granted a NCE from the NSF, so an additional year will allow this project to be completed.

The College O network consists of a data center housing a border router and core switch connected to a 500 Mbps Internet2 WAN connection. Nearly one third of the data center houses research equipment on a network segment separated by a distribution switch with 1 Gbps links in both the upward and downward directions. Off the core switch and passing through a firewall, the campus distribution switch network has 1 Gbps connections to serve the user community, though poor fiber across the campus LAN limits the aggregated bandwidth levels that can be traversed between the core and distribution switches. Defined in the Science DMZ project plan, the link to College N, the Internet2 WAN link to the ISP, and the switch links in the research side of the data center will be upgraded to 10 Gbps and the research distribution switch will act as a Science DMZ switch, serving high-bandwidth connectivity to research equipment in the data center. perfSONAR will be installed on the Science DMZ switch to monitor traffic across the network.

The design map for the existing College O network is featured in Figure 35, with the proposed College O Science DMZ network design detailed in Figure 36.

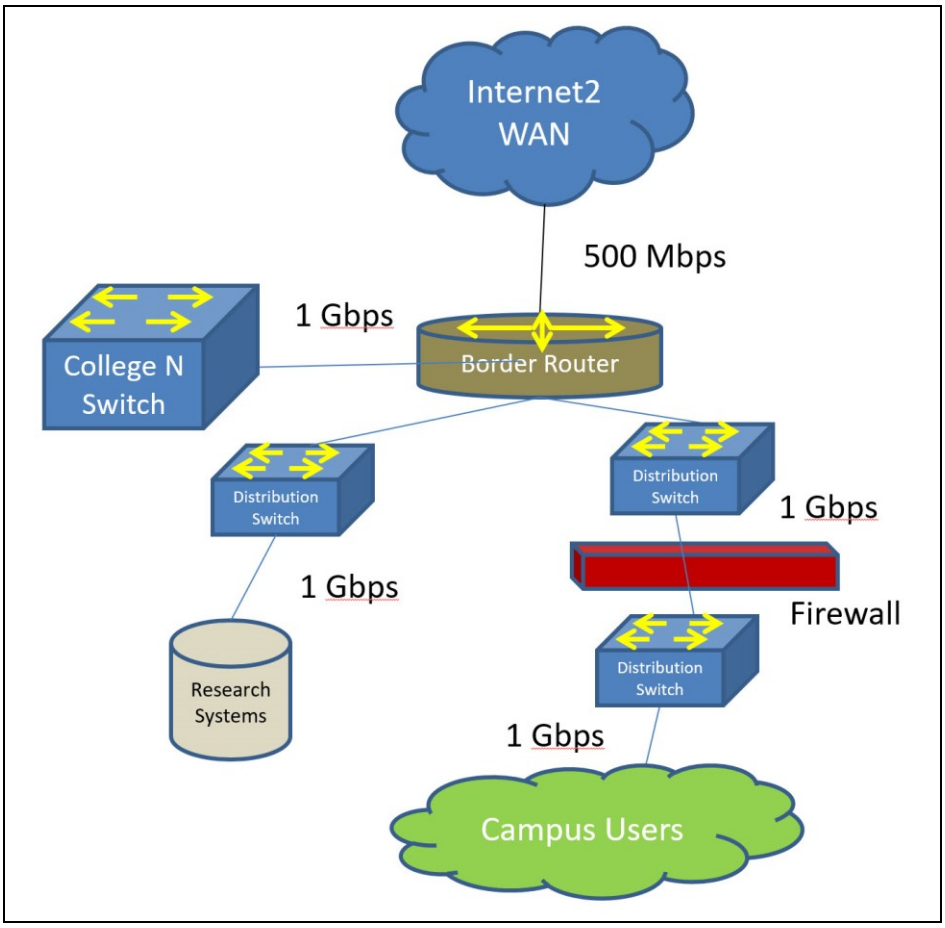

Figure 35: College O Campus Network Before the Science DMZ Project

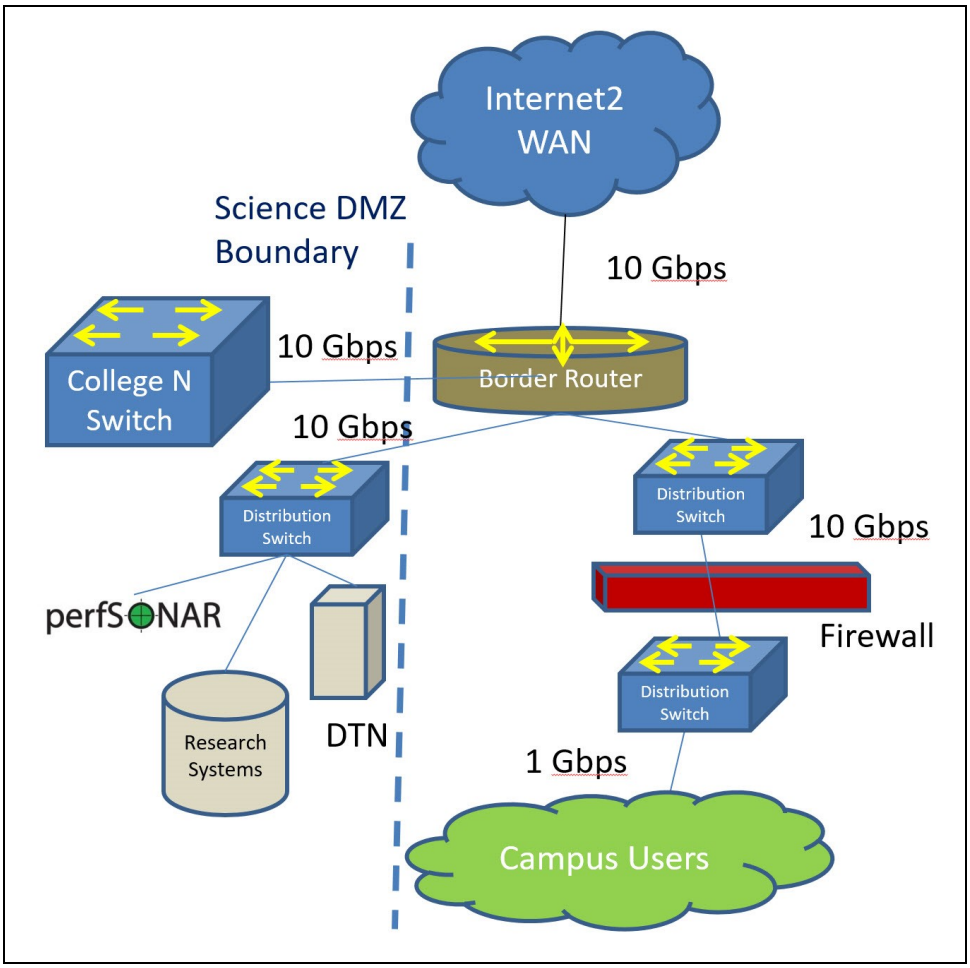

Figure 36: College O Campus Network After the Science DMZ Project

Like its partner institution, College N, the College O Science DMZ project has requested and been granted an extension to complete the project elements in 2018. With so many parts of the network already prepared for the Science DMZ upgrade, from our observation, this project should be able to complete the proposed design by the end of the extension period.

6.17 College P

The initial network configuration for College P is typical for many small institutions. The college doesn't have a large enough technology team to operate all of

the network services, so they hired their WAN ISP to operate a border router for the college. The core switch on the college campus had a 1 Gbps connection to the WAN, and distributed to all of the campus buildings a 1 Gbps uplink to offer 1 Gbps ports for campus network users. College P partners with a nearby large institution as part of the Science DMZ grant award. This large institution now serves as the Internet2 SEGP WAN connection, offering a 3 Gbps bandwidth rate that is burstable to 10 Gbps with the installation of a new border router that has been brought onto the campus and managed by College P network staff. A new Science DMZ switch, DTN, and perfSONAR device were installed in the science complex, all operating at 10 Gbps.

The College P network configuration, before the start of the Science DMZ design and installation project can be found in Figure 37 below, while the post-installation network configuration for College P's Science DMZ network is located in Figure 38 below.

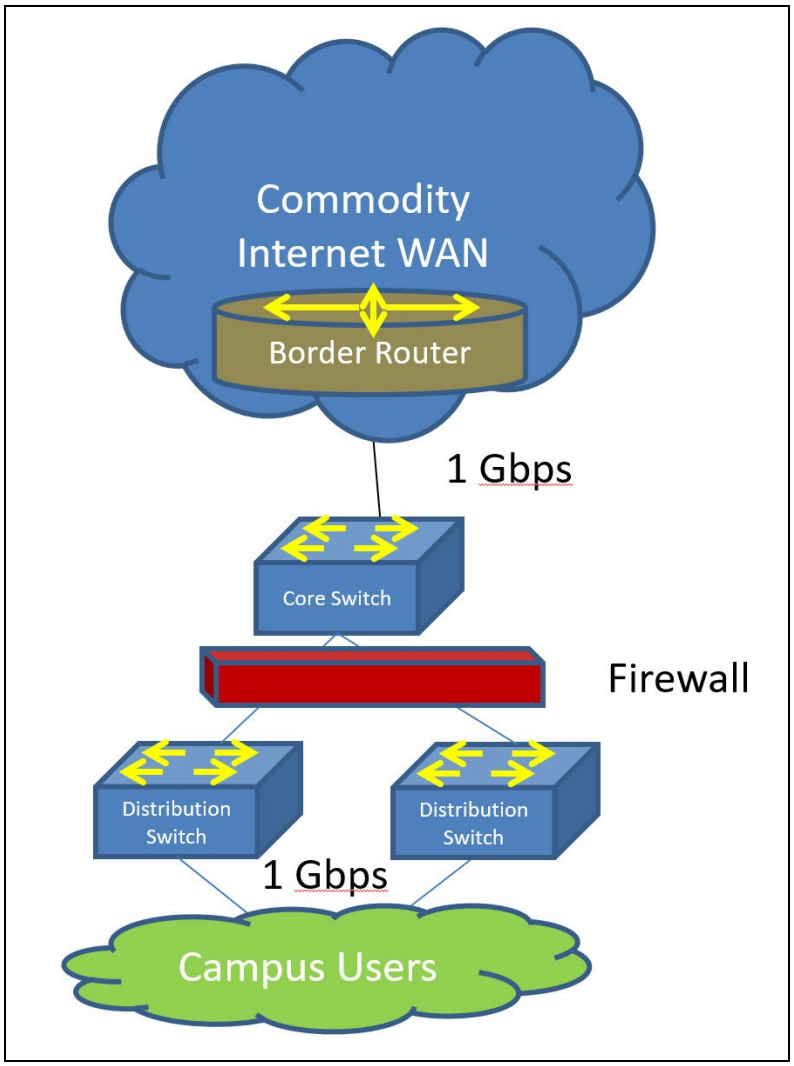

Figure 37: College P Campus Network Before the Science DMZ Project

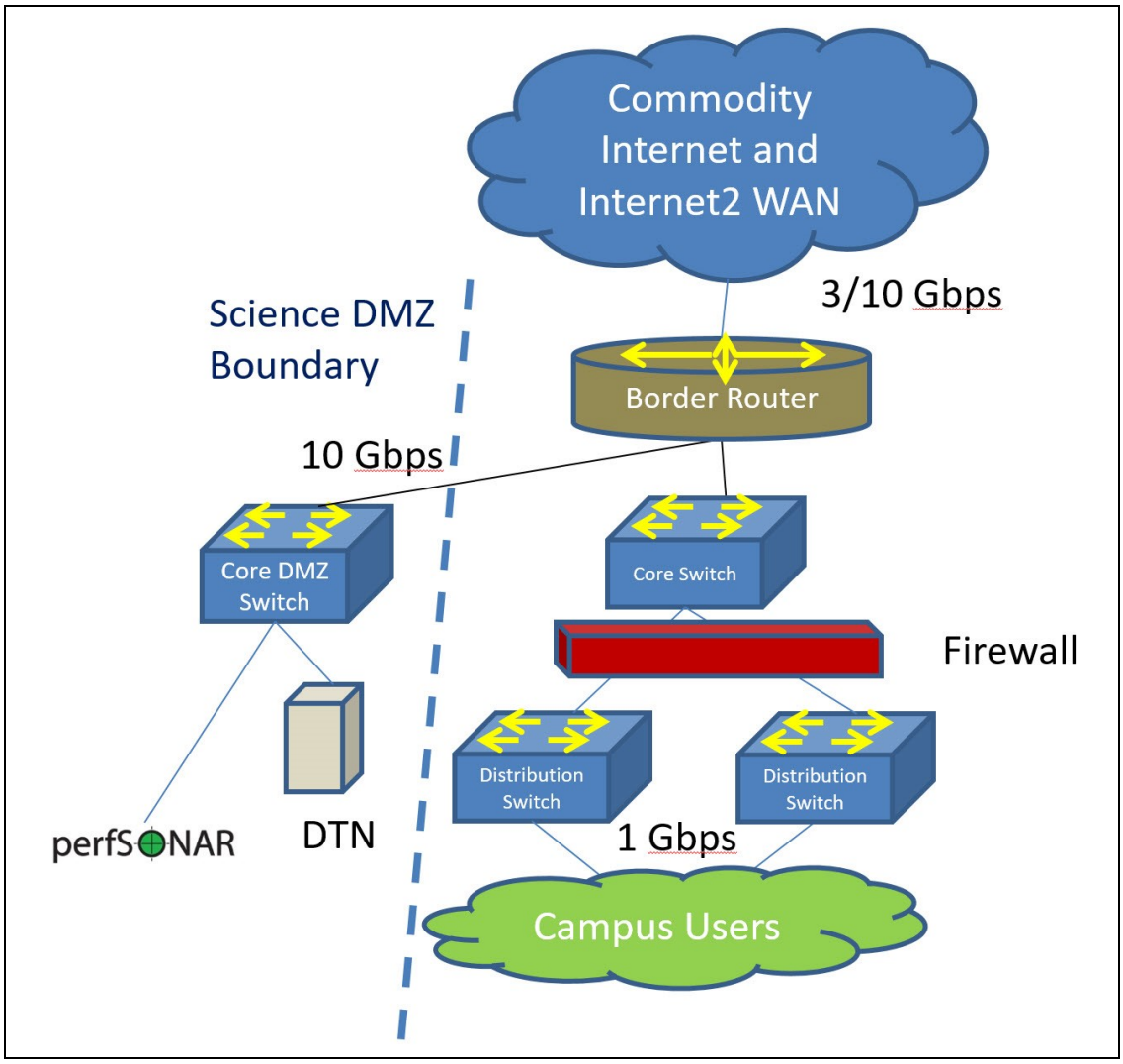

Figure 38: College P Campus Network After the Science DMZ Project

College P was prepared with a SMF cable plant installed a few years prior to the start of the Science DMZ project. However, not having a block of IP addresses took a long time to broker a range in order to properly assign the Autonomous System Number (ASN) for the border router that was to be housed on campus. The large partnering institution played a significant role in teaching College P's staff, as well as assisting in the design of the network.

6.18 College Q

The network design proposed and delivered for College Q doesn't have the specific elements of an ES.NET Science DMZ model, but a number of the components of a Science DMZ already were in place on the enterprise network of College Q prior to the project's start. Partnered with their Internet2 WAN provider, College Q had a 400 Mbps WAN connection that served their border router in one campus location and feeds a campus router at another campus location via a 100 Mbps link. Both campus locations are configured exactly the same with a 1 Gbps link from the router to a distribution switch offering campus users 1 Gbps links to the network and access to local data stores. The existing campus network uses NetFlow [12] to monitor the network's health rather than perfSONAR, and has a host of DTNs already distributed across the network in both campus locations. As a consequence, the only upgrades requested as part of this project were to upgrade the WAN link to 1 Gbps and to upgrade the router-to-router link to 1 Gbps as well.

Figure 39 describes the network configuration initially in place at College Q prior to the start of the link upgrade project. Figure 40 shows the same network configuration as nothing in terms of device hardware changed, other than the link rates between the major pathways between routers and the WAN connection.

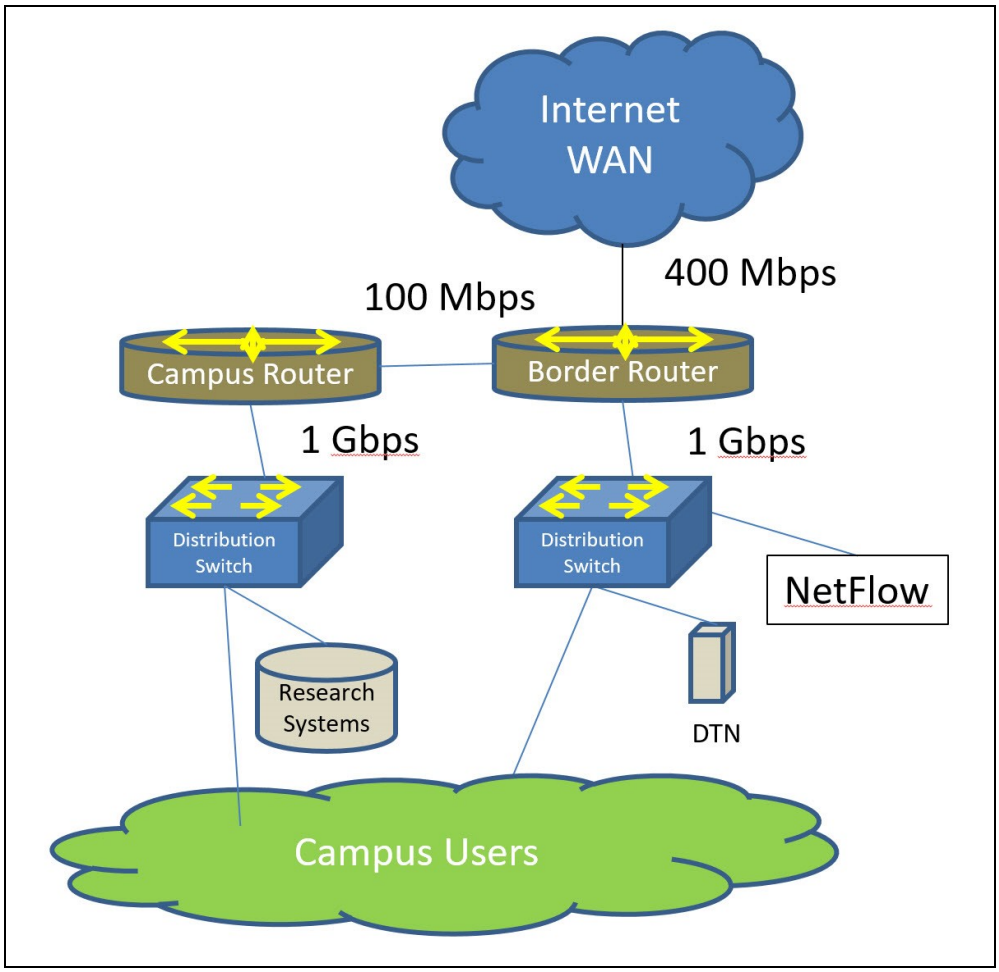

Figure 39: College Q Campus Network Before the Science DMZ Project

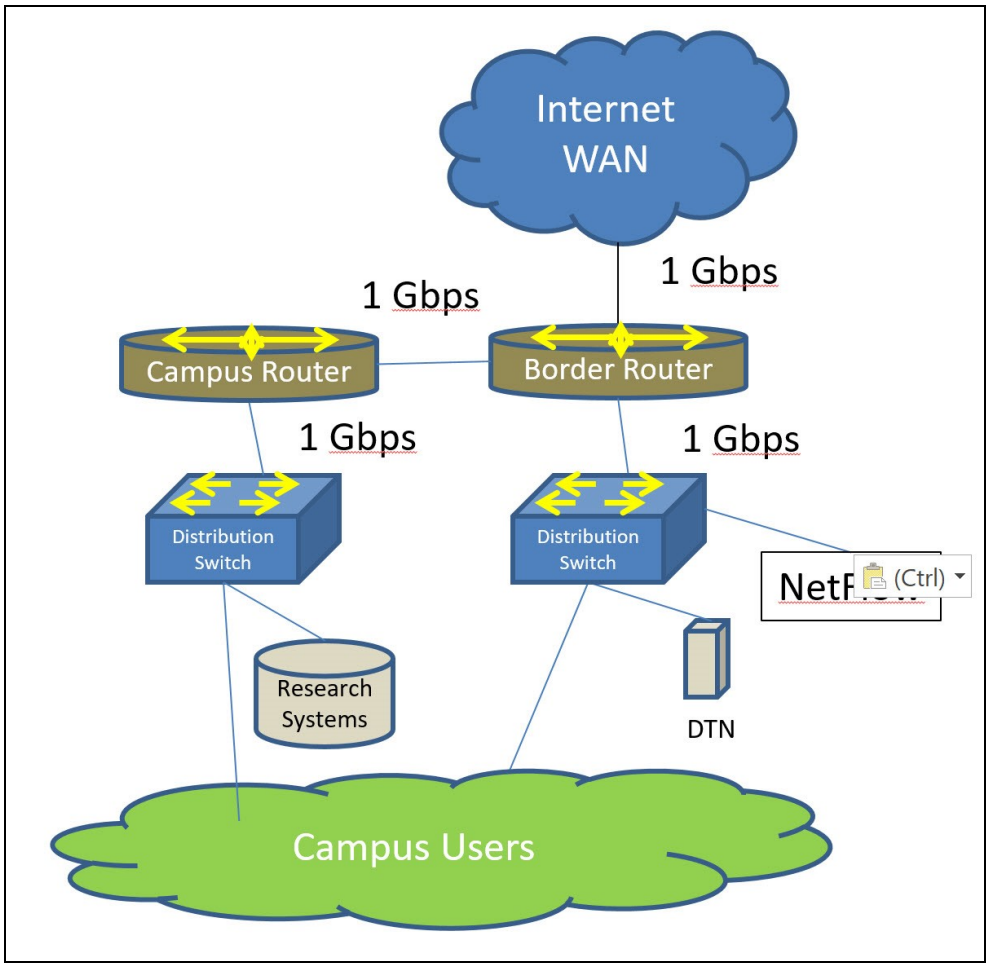

Figure 40: College Q Campus Network After the Science DMZ Project

We are challenged to refer to this project as a Science DMZ project as the only upgrades focused on two links that were elevated to 1 Gbps capacity. However, in spite of not having to upgrade the network hardware due to its initial installation at the start of the project, the major components of a Science DMZ appeared to have been in place prior to the start of the project.

6.19 College R

College R, for a small institution, had a robust initial network configuration that would rival many Carnegie-classified high research universities. Redundant core

switches connected redundant 1 Gbps Internet2 WAN connections. A mesh of building distribution switches served a series of edge switches to offer users campus network connectivity. Partnering with two large institutions on the design project and the research drivers, the Science DMZ network installation upgrades the existing 2 WAN connections with 10 Gbps links and adds a third 10 Gbps link for a Science DMZ border router and switch offering 10 Gbps ports for a DTN, a perfSONAR device, and access for a variety of research connection devices including digital storage and a networked telescope.

The initial College R network configuration is described in Figure 41 below. Figure 42 details the College R network after the Science DMZ network and WAN upgrades are installed.

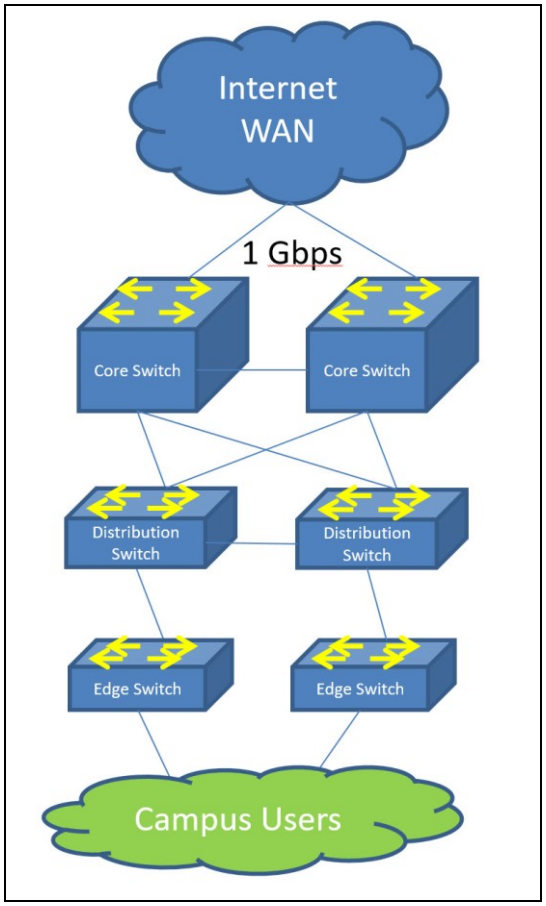

Figure 41: College R Campus Network Before the Science DMZ Project

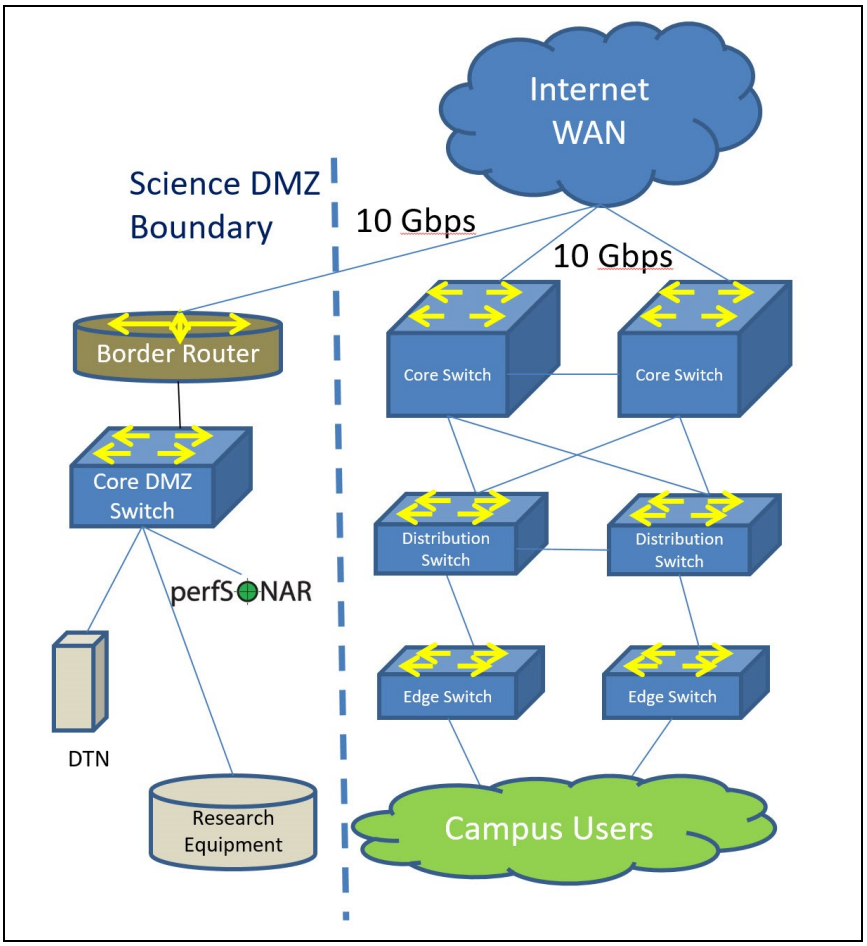

Figure 42: College R Campus Network After the Science DMZ Project

The College R network, after the Science DMZ project, follows the ES.NET model for a Science DMZ quite well, and has the WAN connection capacity to engage in significant amounts of digital data exchange.

6.20 Summary of Observations

Our review of the 18 NSF CC* award projects yielded much more than a picture of the network configurations, both before and after, associated with the projects. We were able to speak directly with personnel and leadership from 10 of the 18 projects to gain greater insight into the related aspects required for building Science DMZ networks

on small campuses. Many of the insights provided by project leadership led to the development of the Science DMZ Capital Framework (SCF) that we discuss in Chapter 7. However, a cohort view of the network designs of small institutions reveals some key points that any Science DMZ designer should consider prior to embarking on the task.

6.20.1 Robust Initial Networks

Nearly all of the campus networks observed in this study have networks that provide some service for their research and campus community. Even the worst-case initial network design of College L with a WAN connection of 51 Mbps shared between 7 campus locations still had a campus network design that served the users with some connectivity. Many of the existing small institution campus networks have received regular investments in upgrades and maintenance by their campuses. Between campus information technology (IT) leadership and faculty researchers, both taking leadership in serving the campus with robust learning tools, the initial network designs were in very good positions to add Science DMZ components going forward.

6.20.2 Traditional Science DMZ Model Deployment

After a review of the 18 projects in the study, 13 projects have a similar project deployment model where the traditional campus network remained in place while the Science DMZ network was designed to be a spur network off the border router or core network switch to serve high capacity network links for research. All of these Science DMZ switches are connected with 10 Gbps links as the hardware for such data rates is considered commodity and readily available in large, less-expensive quantities. DTNs

and perfSONAR nodes are connected to these Science DMZ switches and are scattered among the 13 designs, some with these tools and some without or not documented specifically, though one would presume that a scientific instrument that generated data on the Science DMZ would have to have a way to transfer that data to some other location across its network interface.

6.20.3 Flat Campus Networks with Border Routers

Small institutions often have only a few key active IP addresses, with the remainder of the campus network being served by a Network Address Translation (NAT) router on the campus side of the network. One campus network didn't have their own IP address block and used addresses as assigned by their WAN provider, requiring the campus network to be reconfigured each time they changed WAN providers. These configurations are not compatible with the traditional Science DMZ design and require a reconsideration of the campus network address configuration model. In two cases that we observed, the small institution does not operate its own border router, but receives network connectivity from its commercial provider. Only one of those two cases brought the router into the campus network, while the other partnered with their service provider to keep managing the external router and serving the Science DMZ link from the outside. Not all providers are willing to take on this new network design, but in close partnership, this can be a solution to address the strategic desire to not having to manage the network border router directly. Some small campus networks are often too small to warrant the expense of hiring large numbers of information technologists to maintain a network that doesn't change much from day-to-day. By flattening the campus network and assigning

live IP addresses to the DTN and other specific devices that should be connected beyond the Science DMZ network, small institutions will be able to effectively transfer digital data without the need to translate packets, which would delay significantly any packet movement through the Science DMZ.

6.20.4 Border Router and WAN Connection Upgrades

Nearly half of the proposed network upgrades included plans to purchase or upgrade a border router or core switch to support higher bandwidth traffic, and all of the projects included upgraded access to WAN services to support the Science DMZ design. Most projects that upgraded their WAN connection sought 10 Gbps rates, while a few sought rates between 1 Gbps and 10 Gbps. Those small institutions in rural areas that have few options for high-capacity bandwidth took conservative WAN connection upgrades that either doubled or tripled their existing rates, maintaining a "wait and see" attitude about how much demand their research community will place on the network environment. All project proposers recognized the critical need to offer higher rates to make the transmission of digital data reasonable in terms of total time to transfer large data files.

Since the investment in Science DMZ networking hardware was most likely to involve data rates that were not already on campus initially, we were not surprised to see every project seek to upgrade the WAN connection. We expected to see more border router and core switch upgrades than was observed, but this appears to be another example of key IT leadership planning for the future and installing important networking resources with future considerations in mind.

6.20.5 Multiple Locations

One campus has seven distributed locations that are all served by "household-grade connections," meaning that they have no border router onsite to serve their campus. Two campuses in the study connect to each other in a distributed fashion and face the same issues. The seven location network, College L, resorted to building a completely new and separate Science DMZ network with access to an upgraded WAN connection, which follows the Science DMZ design model. Not much of that build impacts the overall campus significantly, but the focus of the project was not to upgrade the campus anyway. The two location colleges, Colleges N and O, were already operating their networks like a Science DMZ-serving campus, so their network upgrades were focused on increasing their campus-to-campus connection and their collective connection to Internet2. With their project delayed until 2018, we do not know yet if this approach will yield the expected results, though we anticipate their success.

6.20.6 WAN Upgrades Only

There are two projects that only upgraded their campus WAN connections, Colleges D and Q, and are among the eight projects whose leadership we were unable to interview in person. In reviewing their campus network configurations, it is clear that they have taken great effort in ensuring that the campus network supports robust data movement across the campus network as well as through the Internet2 WAN connection. College Q already had a number of services in place prior to the WAN connection upgrade, so one could argue that they may have already had the makings of the Science DMZ before the award. College D is relying on a robust campus network environment

and an upgraded Internet2 WAN connection to move data to and from the campus network. We believe that College D may continue to be at a deficit as the network that has been designed doesn't meet the basic definition of the Science DMZ as there is no DTN, no network monitoring, and no clear digital data pathway that isn't in competition with data streams that are not associated with scientific discovery.

6.20.7 The Science DMZ Network of the Whole

College B displays the most interesting of the Science DMZ project designs in that the network was upgraded to such a level that any user on the network has a 10 Gbps port at the desktop and can transmit over a WAN connection that is 40 Gbps, the highest WAN connection rate of all of the projects. While not on the design map, we know there is a perfSONAR device and DTNs across the network, but they were not detailed. This network design, both at the start and finish, is an outlier among small institutions, as the majority do not have such robust on-campus networks nor the resources to subscribe to WAN connection rates as high as that of College B. Additionally, this design model seeks to simply increase performance in data movement rather than attempting to develop a separate network allocated to scientific equipment. By integrating the access of scientific equipment to the regular campus network with a higher capacity network in place, the security aspects of the Science DMZ design are ignored, calling into question whether this design can even be consider a Science DMZ network in the end. In the Science DMZ model, research devices need to be separate from the commodity network.

6.21 Design Summary

With so many interesting variations on the base model of the Science DMZ, we are challenged to identify specific design components that should be applied to the small institution campus network. Clearly, a key design element is the inclusion of a high-capacity device, such as a router or switch, which is capable of avoiding the delays that come from the traditional firewall and security scanning systems that are TCP/IP enabled. Data Transfer Nodes (DTN) with high-capacity network interface cards and with low-latency, high capacity storage are valuable instruments to include in the Science DMZ network. perfSONAR devices across all campus segments are useful tools for every user and researcher as the device collection aids in any network segment troubleshooting and identifies any segment slowdowns or breaks. We know that some form of WAN connection that can support large data flows is a key component in the external connectivity to remote data. However, the size of that WAN connection is a variable assignment. Those small institutions that have only a few researchers with a periodic pattern of data to transfer may be sufficiently served with a small WAN capacity around 1 Gbps. Small institutions with researchers who receive large amounts of data or data continuously over time should consider larger WAN capacity links in the 10 Gbps or greater level. While a dedicated WAN connection for the Science DMZ ensures an uninhibited pathway between the small institution and the remote institution where the data to be transferred is located, small institutions would be better served and may gain more campus leadership support if the WAN capacity increase benefitted both the research community via the Science DMZ and the overall campus network users. This

could be achieved through the careful configuration of routers and switches at the campus border.

The best design for a small institution must be aligned to the purpose and intent of the deployment of the Science DMZ. If the need for greater performance in the movement of digital data between two sites needs to be better than the present network by orders of magnitude, or simply faster than boxing hard drives and shipping them from site-to-site, then that model will be different in design emphasis than the model that focuses on the delivery of digital data in near-real-time.

The collection of network designs reviewed and recorded in this chapter serve as a series of examples of small institution network configurations before and after a Science DMZ project installation. By following those network patterns that appear similar to a new small institution installation, many of the difficulties in determining the best design for a small institution should be removed, allowing more focused attention to be paid to the other aspects of the upgrading of the small institution.

CHAPTER 7

SCIENCE DMZ CAPITAL FRAMEWORK

There exists a body of applied research in the area of economic development that provides a usable structure to assist in the design and development of Science DMZs for small institutions. The Community Capitals Framework (CCF) by Emery and Flora [18] is a model that identifies, records, and organizes the existing capital into a community to analyze change from a systems perspective by identifying the assets in each capital (stock), the types of capital invested (flow), the interaction among the capitals, and the resulting impacts across capitals. Economic development professionals, prior to engaging in a project to increase economic development in an area, would need to identify and organize the existing community capitals that may be strategically leveraged in the development project. Emory and Flora identify seven (7) capitals that exist in a community:

Financial – Income, security, wealth, credit, investment

Built – Water systems, sewers, utilities, health systems

Political – Inclusion, voice, power

Social – Leadership, groups, bridging and bonding networks, trust, reciprocity

Human – Self-esteem, education, skills, health

Cultural – Language, rituals, traditional crops, dress

Natural – Air, soils, water (quality/quantity), landscape, biodiversity

While the specific capitals outlined relate to economic development within developed communities, the CCF can be used by project managers and evaluators to trace how an investment in each capital might impact the other capitals, both positively and negatively, and allow a better understanding of the strategic nature of funded projects and their impact. A manager begins by collecting and recording the *context*, a detailing of the pre-existing conditions and structures presently available in each of the community capitals. Once detailed, a manager focuses on the *process*, the detailed actions, investments, and interventions that the project seeks to involve. A clear representation of the project, actors or groups involved, actions anticipated, and the project timeframe are recorded, along with the anticipated community capitals expected to be impacted. Finally, the manager details the *outputs and outcomes* by highlighting the results of the actions of the process. These outcomes would be the measures and indicators that each capital should be expecting to observe, both positive and negative.

Building on the work originated by Emory and Flora, a small institution, just like a local community, needs to assess its capital in a variety of areas prior to the design and deployment of a Science DMZ to ensure a successful and long-running implementation. As we have observed throughout this project, the need for a Science DMZ to connect science researchers with other science researchers and their assets is a high-value endeavor. However, many of the key capitals that need to be considered prior to the start of the Science DMZ implementation often go unidentified or unconsidered in the overall scope of the project. To parallel Emory and Flora, we identified the following five (5) capitals that form the Science DMZ Capital Framework (SCF) from the major observed factors of Chapter 5:

1. Network Capital – The detailed assessment of the existing communications network design, function, and assets that may or may not be leveraged in the final implementation of the Science DMZ.
2. Financial Capital – The funding necessary to sustain a Science DMZ buildout, as well as the fortitude to maintain funds in reserve to refresh the Science DMZ hardware based on network management best practices.
3. Political Capital – Critical mass and clout to recommend the shift of the existing network operation practices and procedures to support the inclusion of the Science DMZ in the network, as well as the power to convince the higher authorities that a Science DMZ investment is the best direction to pursue.
4. Science Capital – Researcher demand to keep the Science DMZ operating at a level that will justify the asset investment beyond the external pressure of following the actions of comparator institutions.
5. Human Capital – Staff members with the knowledge, or staff members able to acquire the knowledge, to operate a Science DMZ as well as any other network or system modifications necessary to accommodate the Science DMZ.

Figure 43 illustrates, via a Venn diagram, the intersection of each of the five capitals and how each capital can relate to and influence the others.

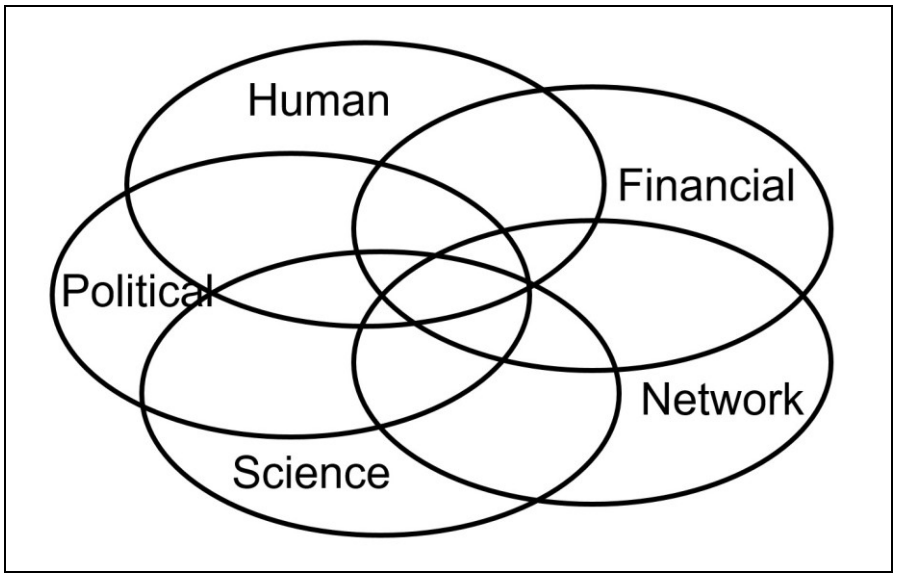

Figure 43: Science DMZ Capital Framework Diagram

These capitals, when reviewed and studied in their entirety prior to embarking on the design and implementation of a Science DMZ, offer the Science DMZ project leadership a synthesized view of the full landscape of the small institution and the key elements that need to be addressed in any proposal to modify their existing network structure. The SCF offers a new viewpoint from which to analyze the changes to the campus environment. The framework encourages project leadership to think systematically about strategies and subsequent projects beyond the initial Science DMZ.

The framework theory proposes that investments made in one or more capitals may influence successful growth in another capital. For instance, building more network capital, as we have seen in a few NSF-funded projects, is likely to increase the network capacity and may lead to more science projects taking place at the small institution. The increase in science projects could lead to new faculty choosing to take positions at the small institution, thus increasing the human capital. The increase in science projects and new faculty working at the small institution would be communicated through the public affairs office of the college and receive public recognition from the community or

engaged alumni. When the public or alumni understand the benefits of this new science research taking place at the small institution, a potential increase in political or financial capital may result.

In order to observe a change in the Science DMZ capitals, an assessment of the present should take place at the onset of a project. This initial point will serve as a zero marker to determine capital growth or decline as we complete the Science DMZ project and as the Science DMZ affects the small institution environment. In Table 1 below, we collect a number of the observed changes in capital in each category over the 18 NSF CC* Science DMZ projects.

Capital	**Change in Capital**
Network	New border router supporting 10 Gbps traffic installed. New core Science DMZ switch supporting 10 Gbps traffic installed. Upgraded WAN connection to 10 Gbps installed. New Data Transfer Node with 10 Gbps installed. New perfSONAR network monitoring installed.
Financial	Capital costs of Science DMZ project at $300,000 received. Equipment maintenance included in annual network operating expenses. Replacement costs for Science DMZ equipment included in plant funds.
Political	Federal grant award to build Science DMZ announced to public. Network upgrade for research will benefit all campus users.
Science	Three to five science drivers are impacted by Science DMZ project. New science drivers for the Science DMZ have emerged.
Human	IT staff gain additional experience in operating Science DMZ equipment. Researchers can access data faster, increasing their productivity. Students have greater involvement in the use of data for research.

Table 1: Science DMZ Capital Framework Influencing Changes

While the change collection exercise illustrated in Table 1 would be most effective when done for only one college, the overall review could be illustrative of the process and could be useful for project leaders to adopt going forward.

The SCF we have identified offers a mechanism for a systemic evaluation process that focuses on the overall impact of the Science DMZ, beyond the specific goal of building the network, to the small institution campus community as a whole. Applying the framework allows the mapping of outcomes by capital and to identify indicators that can measure the degree of change anticipated. The SCF allowed us to study the interaction among capitals that could result in "success leading to success." As we observed in many of the Science DMZ deployments, the Science DMZ aspect is a next step in the development of a robust investment in the small campus network.

Building a Science DMZ, while the focus of the NSF-funded project, is not the long-term goal of this effort. Rather, looking to the future benefits of a Science DMZ installation is key to the use of the Science DMZ Capital Framework. By outlining the Science DMZ Capital Framework (SCF), developers will achieve a greater sense of the connectedness of each capital to the other and realize quickly any deficiencies or absences in areas of the Science DMZ project that may be less technical in nature, as the technical capital is only one of the five Science DMZ capitals to be consulted.

CHAPTER 8

CONCLUSION

In reviewing the 18 National Science Foundation (NSF) CC* project proposals, we observed that they all flow and address the same major points in the design and construction of a Science DMZ. This common pattern is due to the detailed instructions that were part of the NSF proposal solicitation documents (NSF 14-521 and NSF 15-534) as well as the selection criteria of the review panels that sat and awarded these 18 grant projects. Each solicitation criteria set was explicit in wanting to see a network upgrade approaching 10 Gbps, the design and construction of a Science DMZ as defined by ES.NET, some form of analysis using perfSONAR as a common tool throughout the research community, and the use of a separate Data Transfer Node (DTN) on the network in some form to assist with the offloading of data exchange between campus and external research partners. Beyond those core criteria, project proposers were free to identify any other potential benefits that could be achieved through targeted investment of grant funds in the upgrading of the small institution campus network.

Through panel presentations at the last three NSF-sponsored CC* Principal Investigator (PI) meetings in February, 2015, October, 2016, and October, 2017, we identified several other PIs who faced similar successes and challenges as we did in developing a small institution Science DMZ. While we are quickly reminded that small institutions cannot build full Science DMZ configurations in exactly the manner that ES.NET describes on their website, we observed that there are many more factors

involved in the construction of the Science DMZ on the campus of a small institution, and our research has recorded several factors that we have classified and organized to make the future implementation easier to prepare than what existing small institution leaders have had to encounter.

The literature is very sparse on the topic of Science DMZs in a general sense, in spite of its relative popularity nationally as a key tool in the propagation of data-enabled science and research. Most papers that reference Science DMZs in some manner point back to the definitive design paper from ES.NET as we have done. Other than a few additional papers that describe some of the proposed modifications to the base design, particularly focused on security enhancements, [15, 16, 17] not much is written on the extension of the Science DMZ as a research tool. Nothing appears in the literature on the design modifications, challenges, and direction options that small institution Science DMZ designs need to consider. Through anecdotal and research protocol conversations, we have collected the only known group of data on the design and implementation choices of small institution Science DMZs nationally.

Our research proposed collecting network designs and implementation details on the 18 small institution Science DMZs deployed across the US. Through the review of NSF CC* awarded proposals, annual outcomes, final reports, and personal interviews of half of the Principal Investigators (PI), we collected and categorized the design decisions that have been deployed, including summaries on the architecture, management scheme, best practices, policies, security, sustainability costs and other details of each project. We observed that many of the designs follow conceptually the Science DMZ model that ES.NET has proposed. Yet, we also observe design decisions that align with the small

institution's network design choices, their operating conditions, and the major factors involved in the development of a Science DMZ for small institutions as outlined above in Chapter 5. We already recognized that the small institution's factors are not entirely the same as the large institution's factors, and have observed many common factors in the design and deployment of the Science DMZs among the 18 design awards as detailed in Chapter 6.

In spite of the investments made by the NSF in the initial capital outlay towards the design and installation of the Science DMZ on the campuses of small institutions, the sustainability challenges remain. Just as no one perfect Science DMZ network model exists to drop into the small institution network, no one sustainability model exists to ensure that anything built that resembles a Science DMZ will thrive on the network of the small institution. Yet, by following the Science DMZ Capital Framework (SCF) presented in Chapter 7, the major issues identified in this research can be addressed in every Science DMZ design. Examples of those major issues are:

- the financial costs in maintaining the equipment within the Science DMZ network;
- the political costs in continuing the presence of the Science DMZ, even for a small number of users;
- the knowledge costs in making researchers aware of the existence of the Science DMZ, how to use it, and the skills required by technical staff to maintain and support the Science DMZ;
- the demand to advance science research at every institution regardless of size; and

- the design of the network necessary to upgrade the existing resource to support higher speeds.

Following this framework and approaching the development of the Science DMZ on a small institution campus will have a much greater opportunity of entering into a successful deployment than discovering the needs of the environment partway into its deployment.

Finally, there is a challenge that has emerged among the higher education research community that this research squarely addresses. As the 2017 NSF CC* PI Workshop was progressing on October 3 in Albuquerque, New Mexico, one of the PI participants approached the microphone during a question and answer period after one of the session presentations and made the bold statement, "If an institution is not connected [with a Science DMZ], then they are not participating in the advancement of science."

In one interpretation of this statement, we might consider the speaker to be making a very arrogant statement relative to his or her own institution. This person's institution is connected with large network paths and with at least one Science DMZ serving a perfSONAR box, at least one DTN, and several high-data-generating research devices with scores of scientists performing cutting-edge, basic research. There may even be a Nobel Prize winner on the faculty of that institution. To believe that only well-networked research equipment institutions are the only locations capable of advancing science is not just arrogant, it is ignorant of the basic research that is created on whiteboards, notebooks, and in laboratories that do not generate vast amounts of digital data all over the country.

However, the second interpretation, which I like to think is the intended reason for this statement, is designed to be a challenge for the research community as a whole. While scientific discovery is often credited to one individual, or a small team of individuals working in harmony, the overall scientific discovery and its development and deployment is as likely a result of many other individuals and teams working to validate and verify the discovery independently. When the verification and confirmation that a discovery is valid involves recreating the experiments and calculations from papers and other knowledge transfer mechanisms, the verification process is delayed. In today's data-intensive science fields, tools exist that can duplicate the data and the environmental conditions of the digital experiment just as it was originally done by the primary researcher. Moving that digital collection manually via physical hard drive or transferring that digital collection over low bandwidth links that require more time than moving the data physically will result in delayed discovery verification and non-participation by those researchers at institutions where the high bandwidth network links and data transfer tools are not available.

Therefore, the challenge that this statement above should evoke is one of partnership, where large institutions have an obligation to partner with small institutions to bring them to a state of connectedness to allow the small institutions to participate. And, in parallel, small institutions need to identify the best mechanisms available to increase their capacity for connectivity and bring their researchers and students to a place where they are participatory in the advancement of science as the speaker above is attempting to spark. This research into the major factors associated with the development of a Science DMZ at a small institution should be the primary tool that small and large

institutions consult and utilize as they move forward to increase research production at all institutions of learning.

CHAPTER 9

FUTURE WORK

In the context of our detailed examination of the design of Science DMZs for small institutions, we identified a series of follow-on research that warrants a deeper understanding of problems related to the Science DMZ concept beyond those specific to small institutions. While these six areas are not exhaustive, they are worthy of understanding how they may impact the Science DMZ concept at larger institutions as well as what considerations they may offer for smaller institutions.

9.1 Non-Research Institution Science DMZ Access

While the Science DMZ was designed to support the movement of data between research institutions to expand scientific discovery, there are several organizations that are not research institutions that have very hard problems that require scientific tools to be applied to finding solutions. Commercial organizations collect large amounts of data that require processing and examination to learn what the data may be communicating. These organizations often call upon research institutions to assist in data modeling and processing using unique resources such as high performance computing (HPC) systems and large amounts of secure storage, both local and in the cloud. The difficulty that these organizations have is moving the datasets to be researched between the organization and the research institution. While the research institution may have a Science DMZ and data transfer node (DTN) in place, the commercial organization has no such entities available

to them. Even the commercial Internet service provider (ISP) for the commercial organization likely would prevent large, sustained data transfers that could nearly saturate the commercial organization's data pathway or the upstream data pathway for the provider. The nationwide research and education network known as Internet2 is reluctant to allow commercial entities to directly connect to this research network. However, there needs to be some form of network configuration that would allow a commercial entity with large datasets to be able to transfer that research dataset over a non-commercial network link in a fashion that is not inhibited by throttling, connection restrictions, and network delay. Many commercial organization, when considering how they share their research data with partners, turn to the practice of packing physical disks in boxes and shipping those disks between physical data centers. This practice was the norm for the research community prior to the design of the Science DMZ, and unfortunately continues in many locations today. Consequently, a solid design that doesn't violate the network's purpose while supporting the operational requirements needs to be developed to encourage more interaction between the commercial organization and the scientific research community.

9.2 Virtual Circuits

One optional variant to the ES.NET Science DMZ model is the deployment of virtual circuits to create Science DMZ pathways from the gateway router of the LAN. Virtual circuits, or virtual local area networks (VLAN), are a logical means of defining a network connection between two entities (devices, networks, locations). Rather than creating a physical network connection between the border router and the Science DMZ

network, a VLAN could be put into place allowing traffic destined for the Science DMZ network to pass unrestricted. VLANs require the use of packet tagging, which even at the lowest logical levels of the network protocol stack, could lead to increased throughput over the physical network connection. A Science DMZ network configuration should be designed and studied to be compared with the ES.NET configuration to determine any network operational differences.

9.3 100 Gigabit Ethernet

With many of the large institutions having migrated their network WAN connection and campus Science DMZ to support 100 Gigabit Ethernet services, there is a strong influence that other Science DMZ network operators should consider upgrading as well. Since the overall possible round-trip network data rate is dependent on the network link with the lowest network bandwidth rate, small institutions may be pressured to upgrade their network connections to utilize the 100 Gbps rate of the Internet2 backbone. Every institution, small or large, must strongly consider the financial investment required for advanced network connectivity in relation to the science data to be transferred, its frequency, and the overall campus return on investment. The Science DMZ Capital Framework will be a useful tool in performing such an investigation.

9.4 Software-Defined Networking

With the emerging stability of the software-defined network (SDN) standards and devices, there are some key benefits that could propel small institutions to build SDN-based Science DMZs. One benefit is the relatively lower cost for hardware than tradition

network switches. With SDN's design approach of unintelligent hardware with robust software controllers, the cost for the installation of a Science DMZ might be lower than the traditional ES.NET model. Another benefit might be found in the operation of the SDN such that the WAN bandwidth could be diverted via a SDN flow at times on the network when traffic is light, such as late in the overnight, so as to not impede on the flows for the remainder of the campus network. In this manner, the small institution would not be upgrading the campus WAN for dedicated science, but for the whole campus, and using the available WAN connection via SDN flows at times that are most advantageous for the campus network administrator to spare.

The DYNES experiment [48, 49] was originally envisioned as a precursor to the deployment of SDN and the Science DMZ. While the project concluded with some small insights into the potential delivery of faster data flows as dynamically controlled by the end nodes, hence the origin of the DYNES name, the model was not adopted nationally.

SDN still suffers from a definition crisis where the community is torn between those defining SDN as a solution that utilizes the OpenFlow protocol to move data flows through the network to maintain traffic performance, while others view SDN as a software solution to provide automated network redefinition to make logical data paths through the network to maintain traffic performance using VLANS. [27, 30] Today's SDN remains a network-controlled environment, though the promise of end-client control via software application continues to be on the SDN roadmap. When implemented, the ability for the client to request network flows and control network resources will make Science DMZs very powerful tools that could automate the movement of data to support

advanced scientific discovery. This model should be designed and examined against the ES.NET model for potential benefits and issues that would impact small institutions.

9.5 Regional Science DMZs

One concept that has been raised within the small institution community is the creation of a Regional Science DMZ. This concept has two prevailing definitions that have not been fully defined and are presently used interchangeably. One primary definition and design of a Regional Science DMZ is a network communications tool that serves a region of institutions with low capacity network links. Those institutions with low-capacity links slowly transfer data to and take data from DTNs on the Regional Science DMZ physically located nearby the collection of small institutions, maybe even being hosted by one and shared by all. Researchers from those small institutions would continue to use their existing low-bandwidth network links to the Regional Science DMZ, but the Regional Science DMZ could be used to transmit and receive data from research partners and later moved to the small institution at more tolerable transmission times, such as early in the morning when network traffic is at its lowest rate of contention. By regionalizing the Science DMZ service, small institutions would only be responsible for moving their data to and from the regional Science DMZ, while the Science DMZ services would then connect to the other parties involved in exchanging data with the small institution.

A second definition and design of a Regional Science DMZ concept might be the banding together of institutions, large and small, that all have a local Science DMZ network that, if connected together and sharing resources in a coordinated manner, could

define a new tool that could be greater than the sum of its parts. Science DMZs across the region, with high-capacity network links, can connect resources together (similar to an in-facility environment) to address a problem that might have regional significance (like climate issues, pollution, epidemics, and other problems).

While the concept of a Regional Science DMZ seems like a good method to promote more small institutions to participate in data exchange for research, the model doesn't address the challenges that the small institution maintains with getting data to and from the Regional Science DMZ in either definition, which is often the key barrier to small institutions participating in data science in the first place. An example of an EPSCoR state embarking on a Regional Science DMZ concept is the OneOklahoma Friction Free Network. [35] With a vision for the incorporation of many of the future factors included in this chapter, more examples of how 100Gb networks and SDN can be included in the networks of a regional collection of small institutions would be valuable for locations across the country that could benefit from working in harmony with other institutions in close proximity.

9.6 Commodity Science DMZs

Since this research uncovered a set of typical patterns of Science DMZ design and deployment based on starting small institution campus networks, we conjecture that a commodity model of typical components and configuration could be generated to offer small institutions a simple cookbook kit to deploy a Science DMZ that starts at a place that is close to their needs. While hardware components vary in price daily via a number of factors, a point-in-time retail cost could be offered alongside the basic Science DMZ

design to support the future adoption of the Science DMZ deployment by small institutions. Using the Science DMZ Capital Framework, technologists looking to include a Science DMZ as a data transfer solution on the network would be able to determine the exact needs based on the existing Science DMZ capital and pick from the commodity kit to enhance the existing network for the small institution.

9.7 Three-Years-Hence Review

While this research examined Science DMZ deployments that began in 2014 and 2015, with some deployments still not completed, a review of the current state of operation of these Science DMZ deployments three years in the future, in 2020, would be an interesting follow-on study. Questions to consider would include:

- Did the design serve the needs of data-intensive researchers?
- Did the network hardware survive operation over the additional three years?
- Has the campus properly sustained the Science DMZ environment?
- Were there modifications made to the Science DMZ design that was deployed? Why? How do those changes appear? What problems did the modifications attempt to address?
- Are the same scientists and information technology leaders and supporters in place at the small institution?
- How much additional science research was completed as a result of using the Science DMZ?

While these questions are not designed to be a complete set, they are representative of the questions that we might ask the Science DMZ designers and operators of the future knowing what we know about the past and present.

9.8 Science DMZ Security

All Science DMZ deployments, not only the small institution locations, have as an open question the need to develop more and better security mechanisms to protect scientific and other protected data. While the "deny all" approach and disconnectedness of existing Science DMZs are valid at protecting access to devices on the Science DMZ today, significant human interaction is required to maintain the protected environment. Human configuration errors, purposeful or accidental, can expose small institutions to significant harm. As large institutions look towards developing better security methods and tools to support the Science DMZ model, small institutions should monitor their efforts to determine the feasibility of additional security approaches in the future.

LIST OF REFERENCES

[1] About CERN. home.cern/about. Accessed 19 Nov. 2017.

[2] Allcock, W., et al. GridFTP: Protocol Extensions to FTP for the Grid. Apr. 2003. www.ggf.org/documents/GWD-R/GFD-R.020.pdf.

[3] Atkins, D., et al. "Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure." *2nd IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID'02)*, 2003, doi:10.1109/ccgrid.2002.1017106.

[4] Borman, D., Braden, B., Jacobson, V., and Scheffenegger, R. TCP Extensions for High Performance. IETF RFC 7323. Sept. 2014. www.ietf.org/rfc/rfc7323.txt?number=7323.

[5] Calyam, Prasad, et al. "Wide-Area Overlay Networking to Manage Science DMZ Accelerated Flows." *2014 International Conference on Computing, Networking and Communications (ICNC)*, 2014, doi:10.1109/iccnc.2014.6785344.

[6] "Campus Cyberinfrastructure – Data, Networking, and Innovation Program (CC*DNI)." *Campus Cyberinfrastructure - Data, Networking, and Innovation Program (CC*DNI) (nsf15534)*, www.nsf.gov/pubs/2015/nsf15534/nsf15534.htm.

[7] "Campus Cyberinfrastructure – Infrastructure, Innovation and Engineering Program (CC*IIE)." *Campus Cyberinfrastructure - Infrastructure, Innovation and Engineering Program (CC*IIE) (nsf14521)*, www.nsf.gov/pubs/2014/nsf14521/nsf14521.htm.

[8] "Campus Cyberinfrastructure - Network Infrastructure and Engineering Program (CC-NIE)." *Campus Cyberinfrastructure - Network Infrastructure and Engineering Program (CC-NIE) (nsf12541)*, www.nsf.gov/pubs/2012/nsf12541/nsf12541.htm.

[9] "Campus Cyberinfrastructure - Network Infrastructure and Engineering Program (CC-NIE)." *Campus Cyberinfrastructure - Network Infrastructure and Engineering Program (CC-NIE) (nsf13530)*, www.nsf.gov/pubs/2013/nsf13530/nsf13530.htm.

[10] “Campus Cyberinfrastructure Program (CC*).” *Campus Cyberinfrastructure Program (CC*) (nsf16567)*, www.nsf.gov/pubs/2016/nsf16567/nsf16567.htm.

[11] “Campus Cyberinfrastructure Program (CC*).” *Campus Cyberinfrastructure Program (CC*) (nsf18508)*, www.nsf.gov/pubs/2017/nsf18508/nsf18508.htm.

[12] Claise, B., et al. Cisco Systems NetFlow Services Export Version 9. Request for Comments (Standard) 3954. The Internet Society. Oct. 2004.

[13] Data Transfer Nodes. fasterdata.es.net/science-dmz/DTN/. Accessed 30 Nov. 2017.

[14] Dart, Eli. "Optimizing Data Management at the Advanced Light Source with a Science DMZ." GlobusWorld Conference. 17 Apr. 2013.

[15] Dart, Eli, et al. “The Science DMZ: A Network Design Pattern for Data-Intensive Science.” *Scientific Programming*, vol. 22, no. 2, 2014, pp. 173–185., doi:10.1155/2014/701405.

[16] Dart, Eli and Metzger, J. "The Science DMZ." LHCOPN/LHCONE Meeting, Internet2. Washington, DC. 13 Jun. 2011.

[17] Dart, Eli and Sinatra, M. “The Science DMZ.” Winter Join Techs, Clemson, SC. 2011.

[18] Emery, Mary, and Cornelia Flora. “Spiraling-Up: Mapping Community Transformation with Community Capitals Framework.” *Community Development: Journal of the Community Development Society*, vol. 37, no. 1, 2006, pp. 19–35., doi:10.1080/15575330609490152.

[19] Events Website of The Quilt. www.thequilt.net/the-quilt-events/. Accessed 13 Aug. 2017.

[20] FIPS PUB 199: standards for security categorization of federal information and information systems. Computer Security Division, Information Technology Laboratory, NIST. 2004, csrc.nist.gov/publications/fips/fips199/FIPS-PUB-199-final.pdf. Accessed 14 Jan. 2017.

[21] Foster, Ian. “Globus Online: Accelerating and Democratizing Science through Cloud-Based Services.” *IEEE Internet Computing*, vol. 15, no. 3, 2011, pp. 70–73., doi:10.1109/mic.2011.64.

[22] Globus Website. www.globus.org. Accessed 29 Nov. 2016.

[23] Hanemann A., et al. PerfSONAR: A Service Oriented Architecture for Multi-domain Network Monitoring. In: Benatallah B., Casati F., Traverso P. (eds) Service-

Oriented Computing - ICSOC 2005. ICSOC 2005. *Lecture Notes in Computer Science*, vol. 3826. Springer, Berlin, Heidelberg. 2005.

[24] Hethmon, P. and Elz, R. Feature negotiation mechanism for the File Transfer Protocol. IETF RFC2389. Aug. 1998. www.ietf.org/rfc/rfc0959.txt?number=2389.

[25] Home Page. www.internet2.edu. Accessed 30 Nov. 2017.

[26] Hwang, T. "NSF GENI cloud enabled architecture for distributed scientific computing," Proceedings of the 2017 IEEE Aerospace Conference, Big Sky, MT, 2017, pp. 1-8, doi: 10.1109/AERO.2017.7943855

[27] Ibarra, Julio, et al. "Benefits Brought by the Use of OpenFlow/SDN on the AmLight Intercontinental Research and Education Network." *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 2015, doi:10.1109/inm.2015.7140415.

[28] Johnston, W.E., et al. "Addressing the Problem of Data Mobility for Data-Intensive Science." *Proceedings of the Eighth International Conference on Engineering Computational Technology*, doi:10.4203/ccp.100.2.

[29] Kissel, Ezra, et al. "Efficient Wide Area Data Transfer Protocols for 100 Gbps Networks and Beyond." *Proceedings of the Third International Workshop on Network-Aware Data Management*, 2013, doi:10.1145/2534695.2534699.

[30] Large Scale Networking (LSN) Workshop on Operationalizing SDN. ccit.clemson.edu/research/lsn-sdn-workshop/. Accessed 29 Nov. 2017.

[31] Lustre Filesystem Website. www.lustre.org. Accessed 29 Nov. 2016.

[32] Magri, Dino Rafael Cristofoleti, et al. "Science DMZ: Support for e-Science in Brazil." *2014 IEEE 10th International Conference on e-Science*, vol. 2, 2014, doi:10.1109/escience.2014.53.

[33] Mandrichenko, Igor. GridFTP Protocol Improvements. 11 Jul. 2003. www.ggf.org/documents/GWD-I-E/GFD-E.021.pdf.

[34] Monga, Inder, et al. "Software-Defined Networking for Big-Data Science - Architectural Models from Campus to the WAN." *2012 SC Companion: High Performance Computing, Networking Storage and Analysis*, 2012, doi:10.1109/sc.companion.2012.341.

[35] Neeman, Henry, et al. "The OneOklahoma Friction Free Network: Towards a Multi-Institutional Science DMZ in an EPSCoR State." *Proceedings of the 2014 Annual*

Conference on Extreme Science and Engineering Discovery Environment, 2014, doi:10.1145/2616498.2616542.

[36] PEARC17 Home Page. www.pearc17.pearc.org. Accessed 30 Nov. 2017.

[37] Postel, J. and J. Reynolds. File Transfer Protocol (FTP). Request for Comments (Standard) 959. The Internet Society. Oct. 1985.

[38] Postel, John. Internet Protocol. Request for Comments (Standard) 791. The Internet Society. Sep. 1981.

[39] Postel, John. Transmission Control Protocol. Request for Comments (Standard) 793. The Internet Society. Sep. 1981.

[40] Quilt Regional Network Mesh. quiltmesh.onenet.net/maddash-webui/. Accessed 26 Oct. 2017.

[41] Schmuck, Frank and Roger Haskin. "GPFS: A Shared-Disk File System for Large Computing Clusters." *Proceedings of the Conference on File and Storage Technologies (FAST'02),*28–30 January 2002, Monterey, CA, pp. 231–244.

[42] "Science DMZ Network Architecture." *Wikipedia*, Wikimedia Foundation, en.wikipedia.org/wiki/Science_DMZ_Network_Architecture. Accessed 4 Oct. 2016.

[43] Seetharam, Sripriya, et al. "ADON: Application-Driven Overlay Network-as-a-Service for Data-Intensive Science." *2014 IEEE 3rd International Conference on Cloud Networking (CloudNet)*, 2014, doi:10.1109/cloudnet.2014.6969014.

[44] Shiers, Jamie. "The Worldwide LHC Computing Grid (Worldwide LCG)." *Computer Physics Communications*, vol. 177, no. 1, 2007, pp. 219–223., doi:10.1016/j.cpc.2007.02.021.

[45] Sulakhe, Dinanath, et al. "High-Performance Data Management for Genome Sequencing Centers Using Globus Online: A Case Study." *2012 IEEE 8th International Conference on E-Science*, 2012, pp. 1–6., doi:10.1109/escience.2012.6404443.

[46] Tanenbaum, Andrew S. *Computer Networks (4th Ed.)*. Prentice Hall, 2003.

[47] Zurawski, J., et al. "perfSONAR: On-board Diagnostics for Big Data." *1st Workshop on Big Data and Science: Infrastructure and Services.* Co-located with IEEE International Conference on Big Data. 2013.

[48] Zurawski, Jason, et al. “The DYNES Instrument: A Description and Overview.” *Journal of Physics: Conference Series*, vol. 396, no. 4, 2012, doi:10.1088/1742-6596/396/4/042065.

[49] Zurawski, Jason, et al. “Scientific Data Movement Enabled by the DYNES Instrument.” *Proceedings of the First International Workshop on Network-Aware Data Management*, 2011, doi:10.1145/2110217.2110224.

APPENDIX A

GLOSSARY OF TERMS

Access Control List: ACL; An access control list is a set of permissions attached to an object, typically a switch or router, that specifies the allowed functions that are permitted by a particular user or device. ACLs provide an additional layer should an attacker wish to compromise the security of a system.

Autonomous System Number: ASN; The autonomous system number (ASN) is the globally unique value that is assigned to an autonomous system (AS), or sometimes called a routing domain. On the Internet, an autonomous system is the unit of router policy that represents either a single network or a group of networks that is controlled by a common network administrator (or group of administrators) on behalf of a single administrative entity (such as a university, a business enterprise, or a business division).

Networks within an autonomous system communicate routing information to each other using an Interior Gateway Protocol (IGP). An autonomous system shares routing information with other autonomous systems using the Border Gateway Protocol (BGP).

Campus Cyberinfrastructure: CC*; Campus Cyberinfrastructure is the name of the National Science Foundation program that invests in coordinated campus-level networking improvements. The program was established in 2012.

Data Transfer Node: DTN; A data transfer node is a computer system that is purposely-built and dedicated for the purpose of wide area data transfers between systems. DTNs are typically PC-based Linux servers built with high-quality components and having access to high-speed, high-capacity local storage, running software tools designed for high-speed data transfer to remote systems, and connecting to high-speed network links.

Demilitarized Zone: DMZ; In computing, a DMZ or demilitarized zone (sometimes called a perimeter network) is a logical or physical subnetwork that houses the external-facing services of an organization to an untrusted network, usually a larger network such as the Internet. The DMZ adds an additional layer of security to an organization's local area network (LAN); an external computer on the network can access only what is available in the DMZ, while the rest of the organization's network is protected behind a firewall. The DMZ functions as a small, isolated network positioned between the Internet and the private network.

The name DMZ is derived from the term "demilitarized zone", an area between nation states in which military operation is disallowed.

Data, Networking, and Innovation: DNI; The National Science Foundation program sub-name of the Campus Cyberinfrastructure program that invests in coordinated campus-level networking improvements. This NSF program name was in operation in 2015 only.

European Council for Nuclear Research: CERN; The name CERN is derived from the acronym for the French, *Conseil Européen pour la Recherche Nucléaire*, or European Council for Nuclear Research, a provisional body founded in 1952 with the mandate of establishing a world-class fundamental physics research organization in Europe. At that time, pure physics research concentrated on understanding the inside of the atom, hence the word "nuclear".

File Transfer Protocol: FTP; The File Transfer Protocol (FTP) is the standard network protocol/application used for the transfer of computer files and data between two computers, or hosts, on a computer network. The computer that initiates the transfer is called the client, while the participating computer in the exchange is called the server.

Gigabits per Second: Gbps; A unit of measure to compare download speeds on network links. Each gigabit is approximately equal to 1 billion bits, the smallest unit of digital data.

Infrastructure, Innovation and Engineering: IIE; The National Science Foundation program sub-name of the Campus Cyberinfrastructure program that invests in coordinated campus-level networking improvements. This NSF program name was in operation in 2014 only.

Internet Protocol: IP; Internet Protocol (IP) is an Internet standard (RFC793) that defines how data is transmitted from one computer, or host, to another on a network. IP, in some cases, works closely with other protocols, such as Transmission Control Protocol (TCP) to define the basic rules of communication on the Internet.

Internet Service Provider: ISP; An Internet service provider, or ISP, is a company that provides individuals, businesses, and organizations with access to the worldwide Internet through the use of equipment and telecommunications cabling located in a specific geographical area. ISPs often provide other related services such as email services, website construction, and virtual hosting.

Local Area Network: LAN; A local area network (LAN) is a group of computers and associated devices, such as printers and other shared peripherals, that share a common communications link, both wired or wireless, usually within a close geographic area, such as a floor, building, or campus.

Megabits per Second: Mbps; A unit of measure to compare download speeds on network links. Each megabit is approximately equal to 1 million bits, the smallest unit of digital data.

Multi-mode Fiber: MMF; In optical fiber technology, multi-mode fiber is optical fiber that is designed to carry multiple light rays, or modes, concurrently, each at a slightly different reflection angle within the optical fiber core. The size of the glass core is usually between 50 and 62.5 microns in diameter, making multi-mode fiber best used for relatively short distances between devices on the network because the modes tend to disperse over longer lengths (called modal dispersion).

National Science Foundation: NSF; The National Science Foundation (NSF) was created by Congress in 1950 to serve as an independent federal agency "to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense..." NSF supports basic research and people to create knowledge that transforms the future. The annual budget of the NSF is $7.5 billion (FY 2017), and serves to fund nearly 24 percent of all federally-supported basic research performed by U.S. colleges and universities.

Network Address Translation: NAT; Network Address Translation, or NAT, is the virtualization of Internet Protocol (IP) addresses on a local area network (LAN). By using NAT, organization can minimize the number of real IP addresses required by mapping non-routable IP addresses through a NAT router.

Network Infrastructure and Engineering: NIE; The National Science Foundation program sub-name of the Campus Cyberinfrastructure program that invests in coordinated campus-level networking improvements. This NSF program name was in operation in 2012 and 2013 only.

No-Cost Extension: NCE; The National Science Foundation grant award category that allows awarded grants to file for an automatic one year extension to complete proposed grant work without any additional budget to carry out the extended work. NCEs are granted for projects that may have started late, had significant delays in acquiring resources, or simply needed more time to complete the grant project as defined in the project scope.

Point-of-Presence: POP; A point-of-presence (POP) is an access point from one location to the rest of the worldwide Internet. An Internet service provider (ISP) has at least one point-of-presence on the Internet, with each point having its own unique Internet Protocol (IP) address.

Power over Ethernet: PoE; Power over Ethernet (PoE) is a technology for wired Ethernet local area networks (LAN) that allows the electrical direct current (DC) necessary for the operation of each network device to be carried by the data cables rather than by power cords. In doing so, the number of wires that must be strung in order to install the network is minimized, as the same wire that carries network

communication can also carry the power necessary for the operation of the network device.

Science DMZ: A Science DMZ is a network, or a portion of a network, that is built as close to the exterior edge of the campus network as possible, which is usually attached to the border router or gateway of the campus. The Science DMZ is designed such that the equipment, configuration, and security policies are optimized for high-performance scientific applications rather than for general-purpose business computing systems.

Sponsored Educational Group Participant: SEGP; A Sponsored Educational Group Participant (SEGP) is a classification of organization that has gained access to the services of Internet2, the national research and education network. SEGP members gain access to Internet2 through a primary member organization, such as a university within the state.

Single-mode Fiber: SMF; In optical fiber technology, single-mode fiber is optical fiber that is designed to carry only one light ray, or mode, within the optical glass fiber core, which is usually between 8 and 10 microns in diameter. Single-mode fiber is best used for long distances between devices on the network because the tight size of the core minimizes the amount of reflections that take place within the core, minimizing light dispersion over distance.

Software-Defined Network: SDN; Software-defined networking (SDN) is an umbrella term that represents several kinds of networking technology focused on making the network as flexible as the virtualized server and storage infrastructure of the modern data center. The goal of SDN is to allow network engineers and administrators to respond quickly to changing business requirements. In an SDN, a network administrator can shape traffic from a centralized control console without having to touch individual switches, and can deliver services to wherever they are needed in the network.

Storage Area Network: SAN; A storage area network (SAN) is a dedicated high-speed network (or subnetwork) that interconnects and presents shared pools of storage devices to multiple servers, allowing each server to access shared storage as if it were a drive directly attached to the server. A SAN moves storage resources off the common user network and reorganizes them into an independent, high-performance network.

Terabytes: TB; A unit of measure of digital data storage. Each Terabyte (TB) is equivalent to 1 trillion bytes.

Transmission Control Protocol: TCP; Transmission Control Protocol (TCP) is an Internet standard (RFC793) that defines how to establish and maintain a network conversation via which application programs can exchange data. TCP works in close conjunction with the Internet Protocol (IP), which defines how computers

send packets of data to each other. Together, TCP and IP are the basic rules defining communication on the Internet.

Virtual Local Area Network: VLAN; A virtual local area network (VLAN) abstracts the concept of the LAN, allowing computers and other devices to appear connected in their own separate network traffic area through the definition of specific ports on specific switches. By default, systems on one VLAN don't see the traffic associated with systems on other VLANs on the same physical network. VLANs allow network administrators to partition their networks to match the functional and security requirements of their systems without having to run new cables or make major changes in their current network infrastructure. IEEE 802.1Q is the standard that defines VLANs and their use.

Wide Area Network: WAN; A wide area network (WAN) is a geographically distributed private telecommunications network that interconnects multiple local area networks (LANs). Typically, a router or other multifunction device is used to connect a LAN to a WAN. Multiple WAN connections form the backbone of the worldwide Internet.

APPENDIX B

NATIONAL SCIENCE FOUNDATION AWARDS

Table B1 is a listing of the National Science Foundation (NSF) CC* awards made in 2014 and 2015 in the area of Campus Design for the Small Institution. Table B2 is the same listing of award numbers and the award amounts assigned to each project. Further details about these awards can be obtained via the NSF website at http://www.nsf.gov/awards.

NSF ACI Award Number	Award Title	Award Institution	City, State
1440617	CC*IIE Campus Design - Internet2 Infrastructure	Northwest Indian College	Bellingham, WA
1440648	CC*IIE Campus Design: Network Upgrade and Science DMZ to Enable High-Performance Data Transfer	Wabash College	Crawfordsville, IN
1440661	CC*IIE Campus Design: Saint Anselm Science DMZ	Saint Anselm College	Manchester, NH
1440686	CC*IIE Campus Design: Building The 10Gbps Network for Big Data in the Sciences	Saint Olaf College	Northfield, MN
1440689	CC*IIE Campus Design: Network Infrastructure for Improved Science Discovery and Education	Earlham College	Richmond, IN
1440704	CC*IIE Campus Design: Building a Next-Generation Research Network for Vassar College	Vassar College	Poughkeepsie, NY
1440729	CC*IIE Campus Design: Upgrading the Juniata Collaborative Science Infrastructure	Juniata College	Huntingdon, PA

1440786	CC*IIE Campus Design: Advancing Research at Carleton College via 10Gbps Upgrade to the Northern Lights GigaPOP and Implementation of a Science DMZ	Carleton College	Northfield, MN
1541307	CC*DNI Campus Design: Building a State-Of-The-Art Research Network at Franklin and Marshall College	Franklin and Marshall College	Lancaster, PA
1541342	CC*DNI CAMPUS DESIGN: Supporting Scientific Research Using Technology at Malone and Other Small Institutions	Malone University	Canton, OH
1541344	CC*DNI Campus Design: Trinity College Next Generation Science Network and DMZ	Trinity College	Hartford, CT
1541348	CC*DNI Campus Design: Internet2 Infrastructure to Enable Research in Big Data Science and Engineering at Tennessee State University	Tennessee State University	Nashville, TN
1541352	CC*DNI Campus Design: Northern's Network Expansion for Large Science and Engineering Data Flows	Northern New Mexico College	Espanola, NM
1541373	CC*DNI Campus Design: Networking Upgrades for Colorado State University Pueblo	Colorado State University-Pueblo	Pueblo, CO
1541376	CC*DNI Campus Design: Midtown WAN Redesign for GIS and CIS Science Research	Harrisburg Area Community College	Harrisburg, PA
1541394	CC*DNI Campus Design: Enhanced Data Delivery at Fort Hays State University	Fort Hays State University	Hays, KS
1541428	CC*DNI Campus Design: Western New Mexico University's Small Campus Cyberinfrastructure Grant	Western New Mexico University	Silver City, NM
1541456	CC*DNI Campus Design: Developing a Scientific Research Network to Support Data-Driven Research at the University of Arkansas at Pine Bluff	University of Arkansas at Pine Bluff	Pine Bluff, AR

Table B1: Summary of NSF CC* Awards 2014-2015

NSF ACI Award Number	Award Institution	City, State	Award Amount
1440617	Northwest Indian College	Bellingham, WA	$349,896
1440648	Wabash College	Crawfordsville, IN	$347,107
1440661	Saint Anselm College	Manchester, NH	$300,000
1440686	Saint Olaf College	Northfield, MN	$327,640
1440689	Earlham College	Richmond, IN	$347,228
1440704	Vassar College	Poughkeepsie, NY	$332,724
1440729	Juniata College	Huntingdon, PA	$349,924
1440786	Carleton College	Northfield, MN	$349,897
1541307	Franklin and Marshall College	Lancaster, PA	$350,000
1541342	Malone University	Canton, OH	$319,748
1541344	Trinity College	Hartford, CT	$340,657
1541348	Tennessee State University	Nashville, TN	$349,144
1541352	Northern New Mexico College	Espanola, NM	$350,000
1541373	Colorado State University-Pueblo	Pueblo, CO	$306,663
1541376	Harrisburg Area Community College	Harrisburg, PA	$208,946
1541394	Fort Hays State University	Hays, KS	$350,000
1541428	Western New Mexico University	Silver City, NM	$349,751
1541456	University of Arkansas at Pine Bluff	Pine Bluff, AR	$281,488

Table B2: Summary of NSF CC* Award Amounts 2014-2015

APPENDIX C

INSTITUTIONAL REVIEW BOARD APPROVALS

University *of* New Hampshire

Research Integrity Services, Service Building
51 College Road, Durham, NH 03824-3585
Fax: 603-862-3564

20-Dec-2016

Valcourt, Scott A
310 Nesmith Hall
Computer Science Department
Durham, NH 03824

IRB #: 6598
Study: The Identification of Major Factors in the Deployment of a Science DMZ at Small Institutions
Approval Date: 20-Dec-2016

The Institutional Review Board for the Protection of Human Subjects in Research (IRB) has reviewed and approved the protocol for your study as Exempt as described in Title 45, Code of Federal Regulations (CFR), Part 46, Subsection 101(b). Approval is granted to conduct your study as described in your protocol.

Researchers who conduct studies involving human subjects have responsibilities as outlined in the document, *Responsibilities of Directors of Research Studies Involving Human Subjects.* This document is available at http://unh.edu/research/irb-application-resources. Please read this document carefully before commencing your work involving human subjects.

Upon completion of your study, please complete the enclosed Exempt Study Final Report form and return it to this office along with a report of your findings.

If you have questions or concerns about your study or this approval, please feel free to contact me at 603-862-2003 or Julie.simpson@unh.edu. Please refer to the IRB # above in all correspondence related to this study. The IRB wishes you success with your research.

For the IRB,

Julie F. Simpson
Director

cc: File
Bartos, Radim