Technical Report 2023-10-01
Internetworking and Media Communications Research Laboratories
Department of Computer Science, Kent State University
http://medianet.kent.edu/technicalreports.html

A SCI.DMZ AND DATA TRANSFER NETWORK NODE FOR KENT STATE UNIVERSITY

Javed I. Khan[1] and Philip Thomas[2]
e-mail: Javed@ kent.edu | plthomas@ kent.edu

[1] Internetworking and Media Communications Research Laboratories
Department of Computer Science

[2] Information Systems Division
Kent State University
233 MSB, Kent, OH 44242
December 2023

Abstract: This document articulates the key design of a wide-area-network (WAN) and a science DMZ that we have built. It has been built and is sharable by the 8-campuses of Kent State University (KSU) spread across northeastern Ohio. This science DMZ (sDMZ) houses a node in the national Data Transfer Network (DTN)- capable of supporting high-volume high-data-rate (HV-HDR) transfers. The shared sDMZ connects to OARnet's optical exchange and is designed to provide an immense (upto) 100 Gbps unimpeded transfer rate capacity. We have also engineered a special virtual DMZ perimeter built over a highly responsive managed-delay WAN. This allows researchers from all other allied regional campuses access to the facility with uniform and consistent user experience.

1.Introduction

1.1. Sci.DMZ Network Architecture

The first objective of this grant was to create a separate network to avoid campus traffic and DTN traffic interference. Our partner ISP in this grant, OARNET, upgraded our Campus network to a 200 Gb/s pipe. This connection using DWDM, dense wavelength-division multiplexing, allowing OARNET to virtually supply the Kent State University user community with a pipe that could run a 100Gb/s link and a separate 100Gb/s link dedicated to the DTN.

As part of our Internet 2 provision, we could use the full 100Gb/s to transfer files to other Internet to participating organization as part of our Internet 2 membership. Through University funding we added I1 connection at 5Gb/s to handle software updates and non Internet2 traffic to amazon and other locations. OARNET installed the DWDM fiber optic device in our campus termination point and our campus network services team extended sing mode fiber directly to the Arista router we installed in our Computer Science data center where the DTN servers were going to be installed.

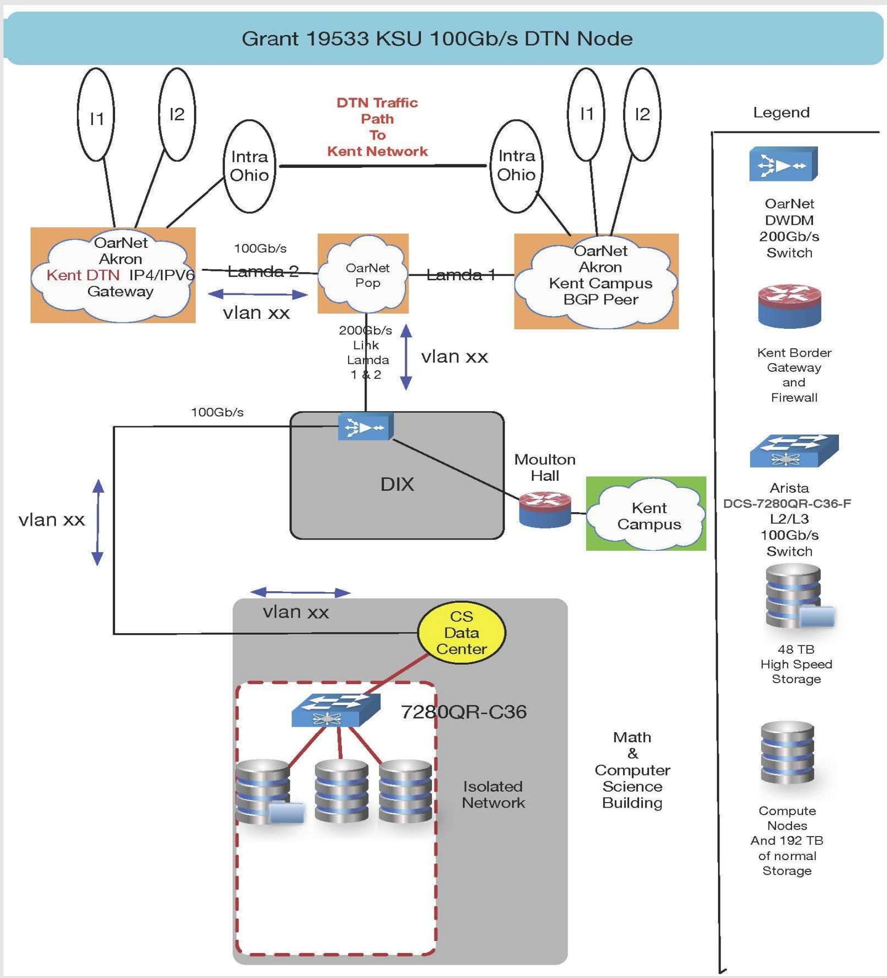

Fig-1a Kent State Regional Campus and sci.DMZ Network

2.Sci.DMZ- Network Configuration

The sci.DMZ network has been built up as a completely separate enclave though it is physically located in the Kent Campus in the Computer Science Data Center located in Mathematical Sciences Building. Fig-1 provides the general design of the network.

2.1. Front End Connectivity

A DTN node needs a 40-100 Gbps data connectivity to support movement of very large data files to and from other DTN nodes around the world.

OARNET is kent's ISP. It maintains a DWDM switch in Kent to provide Kent connectivity through customer border router located at DIX location. The first task was to provide the sci.DMZ DTN with an independent 100 Gbps optical connectivity of its own to our regional REN. In the new configuration OARNET upgraded its equipment to provide us with two 100 Gbps lambdas.

A 100 Gbps lambda-1 wave now connects the DTN's front end Arista switch/router to OARNET's router at Akron PoP. A second 100 Gbps lambda-2 wave for of rest of Kent campus's connecting at Kent's main BGP border gateway/router at Moulton. This redesign increased Kent older connectivity also.

2.2. Inter Campus Connectivity

The main campus located at Kent acts as the hub-campus. Besides its AKRON PoP, OARNET also maintains a redundant route to hub-campus traffic through its Cleveland PoP to this hub.

OARNET- through its 'Ohio-Intranet' also provides Kent the connectivity to it regional campuses. The last mile's of those networks are provisioned by third-party links to the nearest OARNET PoP. Kent pays point-to-point link to nearest OARNET hub. These paths have in general limited subscribed bandwidth with variable quality. However, once the traffic is in OARNET it is physically on Ohio-Intranet. OARNET eBGP-external and private AS numbers are used to bring all regional campus traffic back to Kent campus.

From the hub-campus, OARNET provides the single outgoing connectivity to outging world for traffic for all campuses. OARNET provided connectivity to Internet 2 sites (such as to all REN research and education network sites)- referred as I2 as well as to commodity internet (referred as I1).

2.3. Layer-3 Configuration

We again worked with our ISP partner to configure the eBGP for IPv4 and IPv6 routes and to set up a local DNS, a separate Domain Name Service, server to support the IPv6 and IPv4 address resolution. The current campus DNS did not support IPv6. We had to directly talk to our ISP provider to provide the DNS service for the DTN node. To keep things separate from the campus network our ISP provided us a class C IPv4 space different from our campus public class B network. Kent State University already had b IPv6 space that wasn't being used, so we carved a portion of that spaced out for OARNET to use to support the DTN network. OARNET Akron PoP now advertises Kent DTN for both IPV6/IPV4 as DTN gateway. It is also a BGP peer for rest Kent campus KSU network.

2.4. Faculty Access to sci.DMZ and DTN

The sci.DMZ is a separate network zone from any campus. Thus, all faculty researchers who use Global (for large files transfer when both nodes have DTN) or SCP data transfer (for smaller files) are given account. They can remote access the server from known IP.

Logically remote access traffic from all campuses are treated equally and routed to sci.DMZ are treated as external traffic and routed via OATNET. This applies also for scientists who are working from hub-campus, despite its geo-proximity.

We have also deployed a perSONAR network. This enables continuous monitoring of regional network and ensure regional campus network latency and throughputs are adequate for consistent user experience. See details in [xx]

2.5. Security

A cauterized model of security measures is now being used for the sci.DMZ/DTN. This is very new for Kent. The DTNs principal data transfer path to/and from OARNET has no firewall- which would have slowed down the traffic making it impossible to meet the data transfer rate needed for large science data.

All communications, blocked at source/endpoint/servers. Only selected and scrutinized Globus and SCP ports are enabled rest are closed. DTN servers have only limited and known software. Data transfer services are performed by scientists (or proxy engineers) only by Globus or SCP with validated accounts and they can connect to other external node/end-points. For management access otherwise only the verified IP and port tuples can connect to the DTN servers.

The sci.DMZ is a separate zone from any campus. Thus, its policy does not impact security policies for the main campus which are more geared for enterprise specific practice.

3.Data Transfer Network & Kent DTN Node

Fig-2 diagram shows the hardware that makes up the DTN Transfer node set up. The Kent DTN Node is comprised of three servers and one SAS storage array. It uses an airport model as proposed in mind.

Hardware Schedule			
Node	Server Model	Purpose	Notes
1	DELL R740XD	High Speed Storage	48TB raw Storage
2	DELL R640	Application Box	Globus, Ansible, DNS,Perfsonar Manager, Perfsonar Data Store
3	DELL R640	PerfSonar 100Gb/s test box	Bare Metal test Box
4	Dell ME4012	Storage Hanger	144TB raw Storage

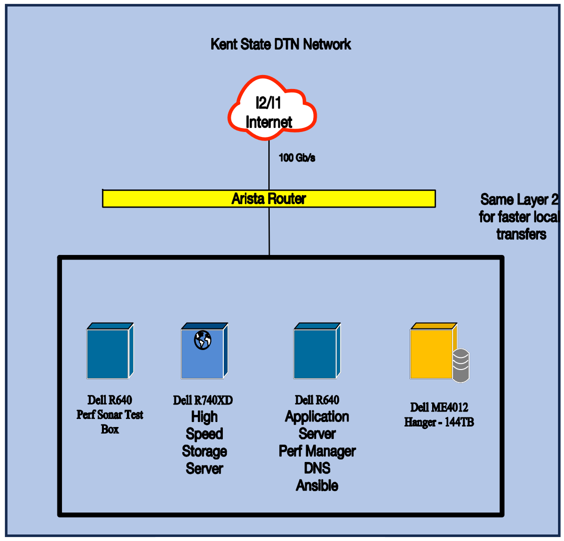

Fig 2 Kent DTN

3.1. Choice of Router

A new Arista 7280QR-32 was selected as sci.DMZ networks' main transfer router. To support 100 Gbps transfer rate the frontend router must have low latency. Also, at such high speed a very important factor is the depth of the buffers- in case of traffic buildup. Arista 7280QR-32 has less than four microsecond latency and has eight (8) GB buffer to avoid packet loss during any traffic delays at such highspeed.

3.2. Design of Storage

It has 48TB of highspeed raw NWMe storage to use as the runway for data transfers. High speed NVMe storage is expensive, and we devised the concept of storing data that was not being transferred to be installed on a cheaper but larger 144 TB Dell ME4012 SAS of raw data. This is a hanger type area where outgoing transfer workloads would wait until transfer was needed. Conversely for incoming transfer workload data would be moved off the high-speed storage to the slower storage after transfer was completed from another university or agency.

For applications to perform the transfer, A DELL R640 server was added. Further another standby DELL R640 was added as a bare metal test box to the framework to enable the perfSONAR telemetry closet to the runway.

Each server is directly connected to the Arista Layer two and Layer Three switch on its own 100Gb/s port. All devices are on the same layer. They are arranged in two broadcast domains allowing for faster transfer between the boxes. For all nodes, Arista is the only layer three hope out to the Internet. This flat design gives the lowest latency and lowest hop count. It enables the fast transfer speed. The design is only limited by the read and write speeds of the drives.

3.3. Equipment Configurations

The Data Transfer Network Node currently has 40 Physical Cores / 80 Logical CPU Cores; 768GB RAM; 217.92GB SSD Disk Space; 144TB Disk Space. Below are detail specifications of the launchpad.

Node1 (Data transfer Endpoint) is 1 x Dell R740XD
• 2 x Xeon Gold-5222 • 4/8 cores at 3.8GHz each • 256GB RAM • Mellanox ConnectX-5 EX Dual Port 40/100 Gbps ethernet • 12 x Intel P4510 4TB NVMe SSD (48 TB Raw Disk Space) • 11 x Samsung PM9A3 15.36TB NVMe SSD (168.96 TB Raw Disk Space) • RAID 5 (Linux MD Software Array)
Node2 (VM Hosting - VMware ESXi) is 1 x Dell R640 with
• 2 x Xeon Gold-6254 (18/36 cores at 3.1GHz each) • 256GB RAM • Mellanox ConnectX-5 EX Dual Port 40/100 Gbps ethernet • 1 x 480GB SATA SSD (480 GB Raw Disk Space)
Node3 (SonarPerf / Network Monitoring) is 1 x Dell R640
• 2 x Xeon Gold-6254 (18/36 cores at 3.1GHz each) • 256GB RAM • Mellanox ConnectX-5 EX Dual Port 40/100 Gbps ethernet • 1 x 480GB SATA SSD (480 GB Raw Disk Space)
Node4 (SAS Storage Array) is 1 x Dell ME4012 with
• 12 x 12TB NLSAS HDD (144 TB Raw Disk Space) • 2 x RAID 6 (Hardware Array)

3.4. Charge Model for sci.DMZ Connectivity

The award paid the first two years of connectivity. Now it has been absorbed into universities information services revenue model as anticipated. There are three key components to the DTN science connectivity charge for a high data rate science connectivity namely 1) connection, 2) bandwidth, and 3) thyperload. The first part is the cost for the high capacity physical link and hardware infrastructure. The purchased hardware needs to be not only supporting peak need but also future ready. Installed physical capacity may not be used or even lighted. The second component is the sustained bandwidth charge- it is normally, close to the typical bandwidth charge model. It is set at most prevalent bandwidth that science transfer applications together may use almost daily basis from its science workloads (typically one tenth of a TB). It is very similar to internet bandwidth charge. The third component is the charge for the ability to case-by-case use the link for hyper data transfer (> 1 TB). Hyper data transfer is infrequent and is needed only few times is a month/year. As a sample, currently university is paying $1500/month for physical link for DTN. It also includes 3 Gbps bandwidth for DTN to Internet-1 transfer. While in future there might an extra charge currently OARNET also provides the DTN the capacity to support upto100 Gbps to Internet-2. For example, OARNET provided 45 Gbps I1/I2 bandwidth to AWS/AZURE sites when we moved ~170 TB data.

4.Sample Data Transfer Experience

4.1. RESEARCH DATA TRANSFER USE-CASE

The goal of this project is to map neural degeneration through the genome in a way to detect patterns for Parkinsons and Alzheimer's. If this were a success, in theory, a simple blood test could be used for early neural degeneration detection.

The data set was computer at AZURE by the research team requested the DTN to move the result after processing to Kent. 19.3TB of results process at Azure cloud. 62498 items transferred in 5.6 hours (20,338 seconds).

Fig-3 captured the data transfer. It reached a max of 9.6 Gbs. This is amazing for a internet1 connection and had an average of 4.24 Gbps. Having such capability allows our researchers to be able to do mass compute jobs at amazon and Supercomputers centers and move masses amount of data in reasonable time frames, avoiding huge costs of sending drives off and on site.

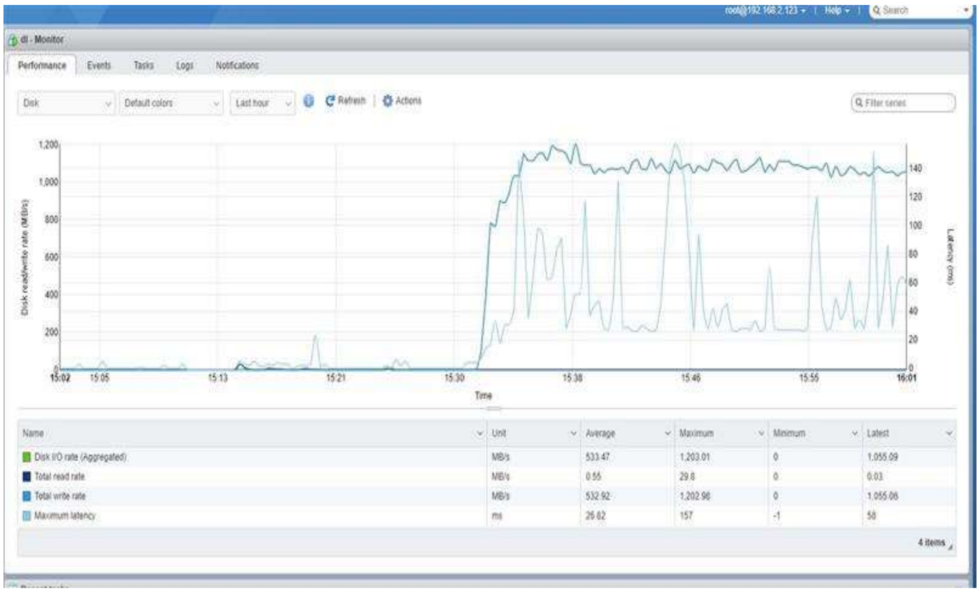

Fig-3 This diagram illustrates a data transfer of 19.3 TB over I1 from Azure Storage to the DTN for Genomics Workload

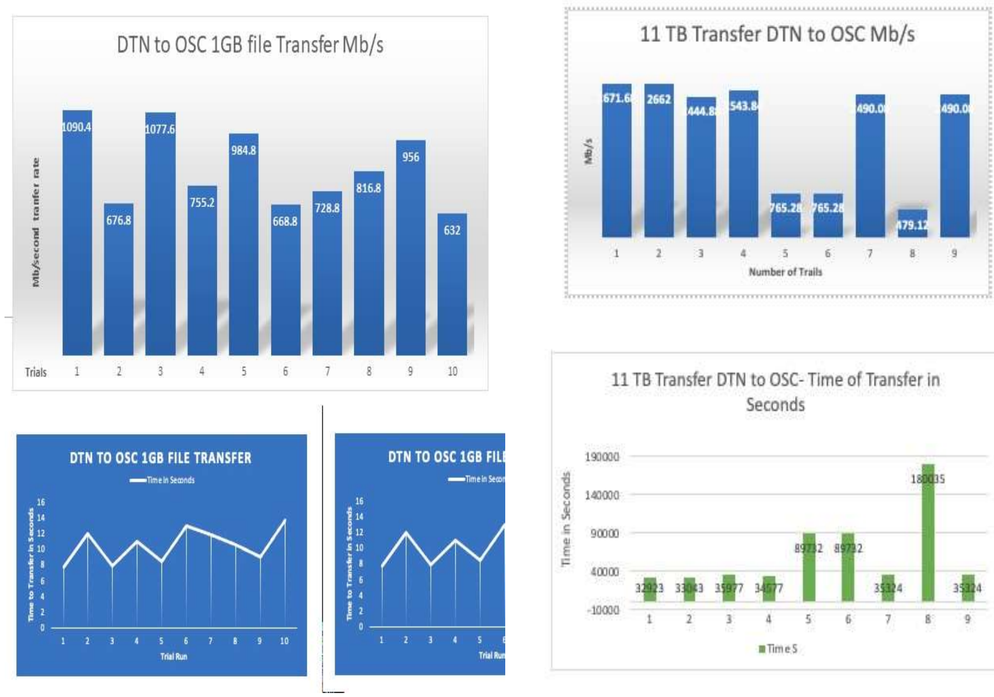

Fig-4 Transfer rate and time for moving 1TB (left diagrams) and 11 TB (right two diagrams) files between DTNs of Kent and OSC

4.2. OSC TRANSFER USECASE

Ohio Supercomputer Center (OSC) will be a major destination for our faculty researchers for science computing. To provide a picture of the speed we performed several tests to transfer large files between the DTNs at Kent and OSC. OSC has a 45Gb/s Globus Endpoint to support large Data Transfers.

Fig-4 shows the observed speeds and total transfer times for 10 runs. To get a realistic picture we performed 10 random runs at various days and various hours.

The 1 TB file consistently transferred in ~8-14 seconds, with a mean = 10.48s, and a standard deviation of 2.02s. The data transfer speed has the mean = 8387.2 Mb/s and standard deviation of 165 Mb/s.

The 11 TB file took about ~32,000-35000 seconds (8 hours). The transfer speed observed mean 812.43 Mb/s and standard deviation of 644.28 Mb/s. This is a noticeably slower performance – but still high. OSC was having work done at their end during some of the transfers. Thus, it game us insight even at a relatively bad day what real science use case may expect.

5.Conclusions

Sci.DMZ represents an important development in research networking. Science research is increasingly dependent on collaboration and access to big data and large-scale computing infrastructure beyond a single institutional boundary. Unfortunately, it is increasingly felt that the classical monolithic campus network architectures are not very suitable for either. This novel deployment of Science DMZ along with the 100 Gbps node for DTN is an experiment with a new campus computing architecture that can support growth of collaboration and support large scale data transfers. This experimental DTN's large data transfer service is a new instrument available to all institutional faculty in Kent's 8-campuses, it is also available to faculty members in surrounding lesser provisioned institutions.

The work is being supported by the NSF Award# 1925678 by National Science Foundation, Cyber Infrastructure Program.

6.References

1. Campus Locations of Kent State University, https://www.kent.edu/campus-locations, last access 9/1/23

2. Castillo, Pickett & Poulini, Implementation of a Science DMZ at San Diego State University to Facilitate High-Performance Data Transfer for Scientific Applications". National Science Foundation. September 10, 2012. Retrieved 2017-12-07.

3. Baseline Expectations for Trust in Federation v2, Web: https://incommon.org/federation/baseline-expectations-for-trust-in-federation/ Last Retrieved November 1, 2023.
4. S. J. Barton, Patricia Book & Shirley Burton, Managing Campus Systems/Communications, Model of Kent State, in Leading America's Branch Campuses, The ACE Series on Higher Education, Samuel Schuman (Editor), R&L Education, Jun 16, 2009].
5. Allen, B., Bresnahan, J., Childers, L., Foster, I., Kandaswamy, G., Kettimuthu, R, Kordas, J., Link, M., Martin, S., Pickett, K., Tuecke, S., "Software as a service for data scientists," Communications of the ACM, vol. 55, no. 2, pp. 81,88, February 2012
6. Javed I. Khan & Philip Thomas, PERFSONAR TELEMETRY IN KENT'S MUTI-CAMPUS SYSTEM, Technical Report, October 2023. [https://www.medianet.cs.kent.edu/techreports/TR2023-10-02-multicampusperfsonar-KT.pdf]