CHAMELEON USER MEETING 2019

ELASTIC DATA TRANSFER INFRASTRUCTURE FOR A DYNAMIC SCIENCE DMZ

BY:

- **Joaquin Chung**, Zhengchun Liu, Tekin Bicer, and Rajkumar Kettimuthu

Background

- **The Science DMZ Model:** provides a network design pattern for research institutions to support high-speed wide-area data transfers
  – Dedicated data transfer nodes (DTNs) with network capabilities that match that of the wide area network (WAN) and that run high-performance data transfer tools such as GridFTP
  – Performance monitoring hosts with network-monitoring software such as perfSONAR to conduct both active and passive network measurements
  – Security policies and tools that can be applied specifically to the science-only traffic

Motivation

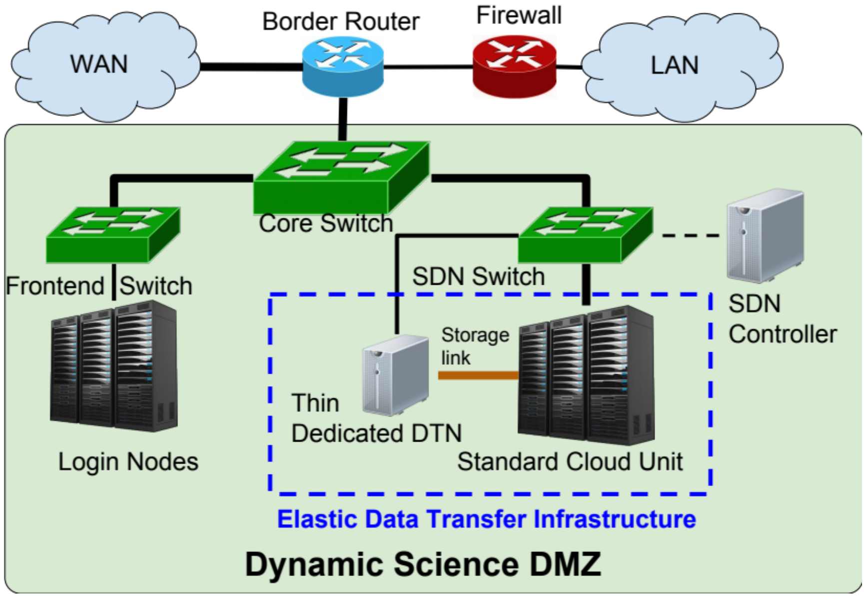
- We analyzed of ~40 billion GridFTP command logs totaling 3.3 exabytes and 4.8 million transfers logs collected by the Globus transfer service from 2014/01/01 to 2018/01/01
- Findings:
  - On average, DTNs are completely idle (i.e., there is no transfers) for 94.3% of the time
  - 80% of the endpoints were active less than 6% of the online time

These grossly underutilized DTN resources can be put into good use by pooling them with the compute resources and using cloud technologies to acquire them on demand

DTN resources can be put into good use by pooling them with the compute resources and using cloud technologies

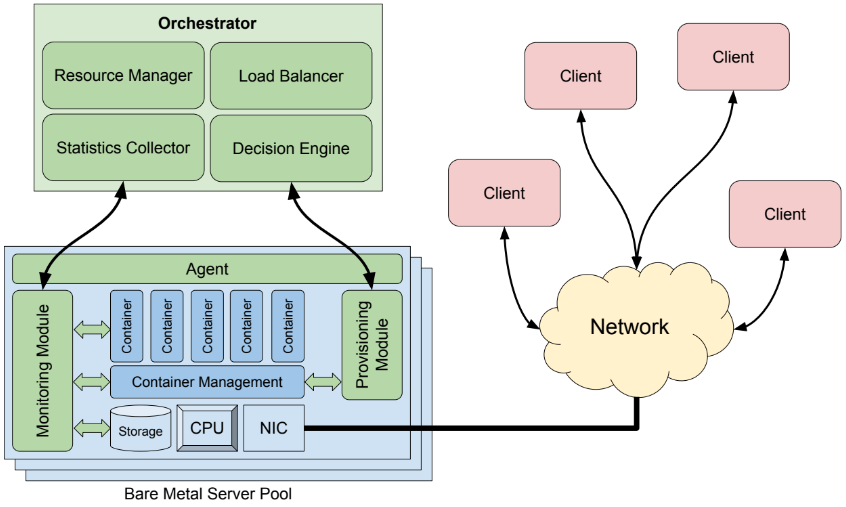
Proposed Architecture

- A dynamic model that securely extends the Science DMZ all the way to virtualized compute resources
- Uses an elastic data transfer infrastructure that expands and shrinks dynamically to conserve resources
- Reserves minimal resources that will work as a dedicated DTN (thin dedicated DTN)



Elastic Data Transfer Infrastructure (DTI)

- **Orchestrator:** maintain a global view of resource utilization across the infrastructure and make decisions about how to allocate resources efficiently
- **Agents:**
  - Collect usage information and report it back to the orchestrator
  - Execute configuration changes on the behalf of the orchestrator

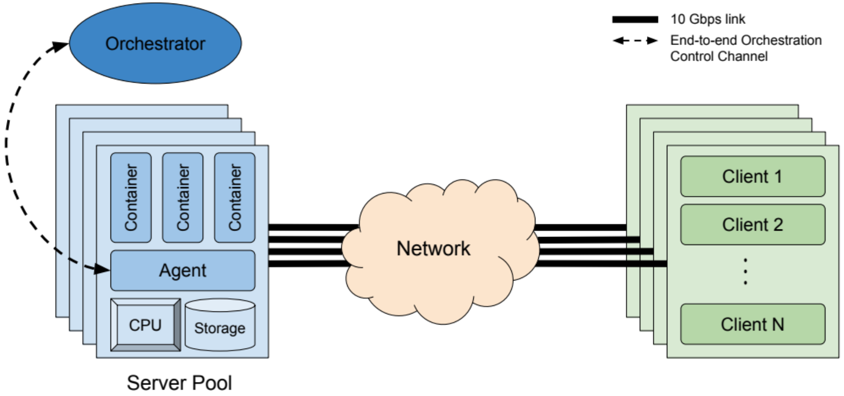


Elastic DTI Design

1) Virtualized infrastructure (how to provision/remove)
   - Use existing compute resources as DTNs
   - Make idle DTN resources available to the campus
2) Provisioning and removal schemes (when to provision/remove)
   - Thresholds

Chameleon Prototype

- **Physical infrastructure:** 8 bare metal nodes with 48 cores and 128 GB of RAM running Ubuntu 16.04.5 LTS
- **Software stack:**
  - Docker version 18.09.0 to manage containers
  - Globus GridFTP server version 12.12 as DTN software
  - globus-url-copy version 9.29 to manage data transfers



Experiments

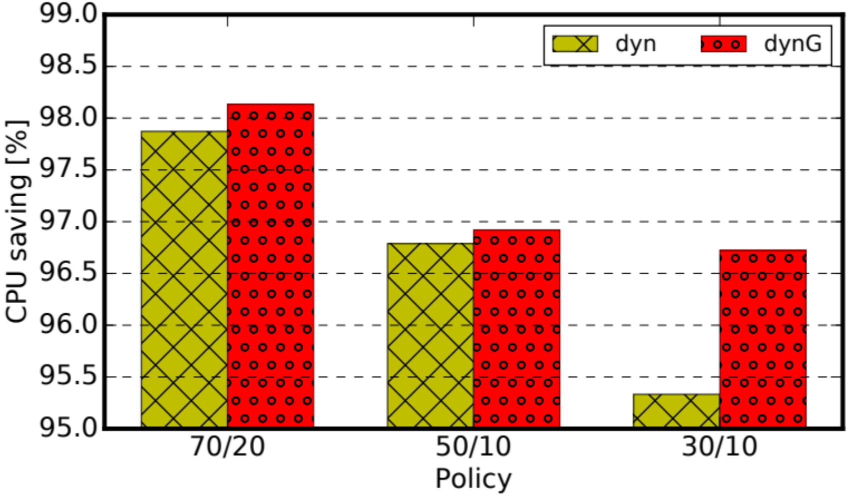
Objective: to compare a static DTN vs. the elastic DTI

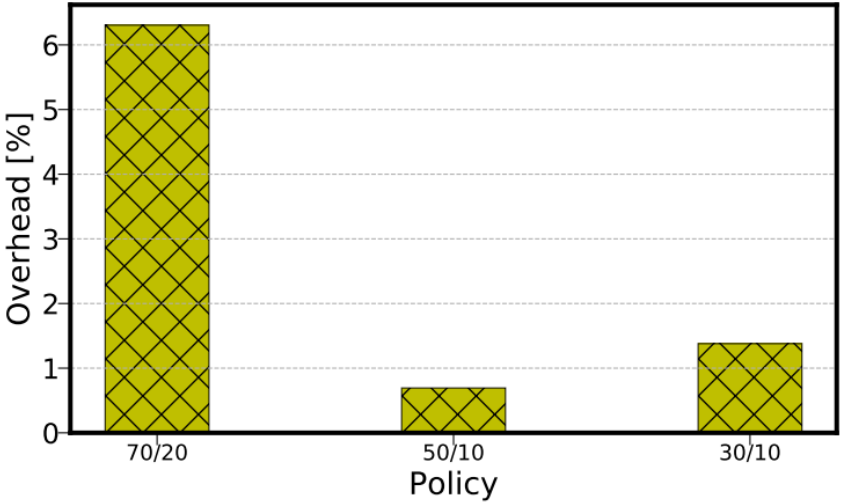
Trace: 30 min. Poisson distribution with lambda 3 sec. and dataset size representative of our previous study

Baseline: 4 standard DTNs with 12 cores and 96 GB of RAM each

Name	Scheme	High (%)	Low(%)
dyn-30/10	*dyn*	30	10
dyn-50/10	*dyn*	50	10
dyn-70/20	*dyn*	70	20
dynG-30/10	*dynG*	30	10
dynG-50/10	*dynG*	50	10
dynG-70/20	*dynG*	70	20

Preliminary Results





Analysis - Pros

- Bare metal nodes are very powerful, a single node provides 48 cores, which is 4 times more cores than the average DTN
- State of the art network: each node is connected to the network with a 10 Gbps NIC, top of rack switches has a 40 Gbps uplink, and each site (TACC and UChicago) has a 100 Gbps fabric
- Network switches are compatible with OpenFlow, a key enabler of software-defined networking (SDN)
- Both sites are connected through a 100 Gbps link

Analysis – Cons

What we wanted	What we deployed
Parallel file system for server “site”	Each server had a replica of the dataset We could have deployed a parallel file system on software
A network load balancer on the server “site”	A service running on the orchestrator that will tell clients where to send their transfer requests We tried HA proxy, but it did not performed well
Muti-site experiment	Run client and server at one site *by the time we started, sharedwan1 wan not available*

Future Resource Requirements/Capabilities

- Test more software-defined networking/storage capabilities
- Creating custom topologies as in GENI (Global Environment for Network Innovations), but using slices of Chameleon hardware switches
- Test P4 switches or any other kind of programmable dataplanes
  - Perhaps using existing Chameleon FPGA resources for networking research.

THANKS! … QUESTIONS?

- e-mail: chungmiranda@anl.gov

Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.