SDN for End-to-end Networked Science at the Exascale (SENSE)

Final Technical Report

Department of Energy Office of Science
University of Maryland Award number DE- SC0016585

Principal Investigator: Tom Lehman

Period of Performance: 08/01/2016- 07/31/2019

Last Update: December 2, 2019

Table of Contents

Executive Summary **3**

1 Introduction **4**

2 Project Motivation **5**

3 Activities and Results **7**

3.1 SENSE Overview **7**

3.2 SENSE Key Functions **10**

3.3 SENSE Architecture Components **11**

3.3.1 SENSE Orchestrator 11

3.3.2 SENSE Resource Manager 12

3.3.3 SENSE Services 13

4 Use Cases **14**

5 Testbed Deployment **16**

6 Products, Documents, Presentations **18**

7 Summary and Conclusions **20**

References **22**

Executive Summary

The Software-defined network for End-to-end Networked Science at Exascale (SENSE) research developed intelligent network services to accelerate scientific discovery in the era of big data driven by Exascale, cloud computing, machine learning and artificial intelligence. The SENSE system allows National Labs and Universities to request and provision end-to-end intelligent network services for their application workflows leveraging Software Defined Network (SDN) capabilities.

The SENSE system includes a model-based architecture, implementation, and deployment which enables end-to-end network service instantiation across administrative domains. An intent based interface allows applications to express their high-level service requirements. Intelligent orchestrator and software defined systems allow for custom tailoring of scalability and realtime responsiveness based on individual application and infrastructure operator requirements. This allows the science applications to manage the network as a first-class schedulable resource in a manner similar to instruments, compute, and storage. The SENSE includes a model-based orchestration system which operates between the SDN layer controlling the individual networks/end-sites, and science workflow agents/middleware. In addition, SENSE has developed model-based "Network Resource Manager" and "End-Site Resource Manager" components which enable advanced features in the areas of multi-resource integration, real time responsiveness, and user interactions.

The SENSE project was motivated by a belief that Domain Science application workflows need real-time, interactive, end-to-end orchestrated SDN services across large, distributed, multi-domain networks. Domain science applications and workflow processes are currently forced to view the network as an opaque infrastructure into which they inject data and hope that it emerges at the destination with an acceptable Quality of Experience. There is little ability for applications to interact with the network to exchange information, negotiate performance parameters, discover expected performance metrics, or receive status/troubleshooting information in real time. The SENSE project was motivated by a vision for a new smart network and smart application ecosystem that will solve these issues and provide a more deterministic and interactive environment for domain science workflows.

The SENSE project developed an architecture and implementation to address this vision. Key contributions of this work include the architecture definition, reference implementation, and deployment as the basis for further innovation of smart network services. The SENSE system with its model-based, realtime, multi-resource cyberinfrastructure awareness provides a data source and framework to leverage artificial intelligence and machine learning technologies to enhance network monitoring, provisioning, optimization, troubleshooting, and workflow services operations.

The SENSE project team is led by Energy Sciences Network (ESnet) and included participation by Lawrence Berkeley National Laboratory (LBL), Fermi National Accelerator Laboratory (Fermilab), Argonne National Laboratory (ANL), California Institute of Technology (Caltech), and University of Maryland/Mid-Atlantic Crossroads (UMD/MAX).

1 Introduction

The Software-defined network for End-to-end Networked Science at Exascale (SENSE) [1] research developed intelligent network services to accelerate scientific discovery in the era of big data driven by Exascale, cloud computing, machine learning and artificial intelligence.

The SENSE system allows National Labs and Universities to request and provision end-to-end intelligent network services for their application workflows leveraging Software Defined Network (SDN) capabilities. For purposes of this project, end-to-end flows focused on data transfers between Data Transfer Node (DTN) clusters at multiple Science DMZ locations. This capability is meant to serve science workflows and cyberinfrastructure resources at these facilities including scientific instruments, high performance computation facilities, and data repositories.

SENSE includes a model-based orchestration system which operates between the SDN layer controlling the individual networks/end-sites, and science workflow agents/middleware. In addition, SENSE has developed a "Network Resource Manager" and "End-Site Resource Manager" which enable advanced features in the areas of multi-resource integration, real time responsiveness, and user interactions.

The SENSE system defines the mechanisms needed to dynamically build end-to-end deterministic and policy-guided Layer 2/3 network services. An intent-based interface allows applications to express service requirement in a high-level domain science specific context, with mechanisms for interactive and full-service lifecycle coordination with workflow automation systems. The SENSE system has two key objectives:

- Facilitate science workflow related provisioning and life-cycle management for a variety of end-to-end network services.
- Provide the intelligence and control to enable more optimal and efficient use of network infrastructures.

The impact of these capabilities includes an ability for science applications to manage the network as a first-class schedulable resource akin to instruments, compute, and storage. This will enable workflow driven optimization of resources spread across a vast geographic footprint such as those used in science domains like high-energy physics and basic energy sciences.

These types of network services will be needed to support the next generation of science workflows that will be driven by Exascale computing resources. Also driving the need for these types of network services is the emerging DOE Superfacility concept. This vision includes the seamless integration of multiple, complementary DOE Office of Science user facilities into a virtual facility to fundamentally transform and accelerate the scientific discovery workflow. The SENSE system provides the mechanisms needed to synchronize and coordinate the connection of multiple distributed compute, storage, and instrument resources with deterministic performance. The SENSE intelligent interactive services also provide methods to assist in the application driven workflow planning/operations which will be needed to realize the Superfacility vision.

The SENSE project included initial work for the integration of SENSE services with several use cases including the Large Hadron Collider File Transfer Services, Exascale for Free

Electron Lasers (ExaFEL), DOE Superfacility, Big Data Express, and distributed compute infrastructures. A common vision for these integrations is the provisioning of SENSE Layer 2 and Layer 3 services based on knowledge of current and planned data transfers. SENSE allows workflow middleware to redirect traffic at granularities ranging from a single flow, specific end-system, or an entire end-site onto the desired SENSE provisioned service. The SENSE Layer 2 services provide deterministic end-to-end resource guarantees, including the network and Data Transfer Node (DTN) elements. The SENSE Layer 3 service provides the mechanisms for directing desired traffic onto specific Layer 3 VPN (L3VPN) for policy and/or quality of service reasons. Additional details regarding these use case integrations are presented in Section 4 Use Cases below.

The SENSE project team is led by Energy Sciences Network (ESnet) and included participation by Lawrence Berkeley National Laboratory (LBL), Fermi National Accelerator Laboratory (Fermilab), Argonne National Laboratory (ANL), California Institute of Technology (Caltech), and University of Maryland/Mid-Atlantic Crossroads (UMD/MAX).

As the SENSE research project ends, efforts continue to transition this system to production status in support of science workflow operations and cyberinfrastructure related systems.

2 Project Motivation

Network designs are evolving at a rapid pace toward programmatic control, driven in large part by the application of software to networking concepts and technologies, and evolution of the network as a critical subsystem in global scale systems. This is of particular interest to major science collaborations that incorporate large scale distributed computing and storage subsystems. This software-network innovation cycle is important as it includes a vision and promise for greatly improved automated control, configuration and operation of such systems, in contrast to the labor-intensive network deployments of today. However, even the most optimistic projections of software adoption and deployment do not put networks on a path that would make them behave as a truly smart or intelligent system from the application or user perspective, nor one capable of interfacing effectively with facilities supporting highly automated data analysis workflows at sites located across the world.

Today, domain science applications and workflow processes are forced to view the network as an opaque infrastructure into which they inject data and hope that it emerges at the destination with an acceptable Quality of Experience. There is little ability for applications to interact with the network to exchange information, negotiate performance parameters, discover expected performance metrics, or receive status/troubleshooting information in real time. As a result, domain science applications frequently suffer highly variable (from excellent to poor) performance, especially so in highly distributed data intensive environments.

Indeed, the ability for a science application to interact and negotiate with network infrastructure within a science ecosystem, should be a hallmark of truly smart networks and smart applications. The current static, non-interactive network infrastructures do not have a path forward to assist or accelerate domain science application innovations. We therefore envision a new smart network and smart application ecosystem that will solve these issues and enable future innovations across many Research and Education domain science communities.

The SENSE solution was built upon previous Software Defined Networking (SDN) work, which has been a subject of much discussion and research over the past decade. The crux of the SDN concept is software control and programmability of network elements and resources in a manner which enhances network services, management, and resource utilization. Multiple frameworks and systems have been developed which accomplish the basic goal of software-controlled services across a heterogenous mix of network elements. While many of these systems are open source, the reality is that significant deployments in the field have been few and limited. These deployments are typically either small isolated systems, or vertically integrated systems from larger operators who have complete control over all the network resources needed by the higher-level applications being served. A lesson learned is that basic software-defined control functionality does solve many of the issues which must be addressed when trying to enable a more organic deployment of end-to-end SDN services across general cyberinfrastructure. These issues include handling of multiple administrative/control domains, resource state hiding/visibility, scalability, and realtime responsiveness, all of which need to be tailorable for specific deployments and application requirements.

In summary, current SDN and orchestration technologies have the following issues which inhibit development of an integrated, interactive smart network and smart application ecosystem:

- Current SDN Technologies - The SDN Controller Northbound API solutions are narrowly focused designs which are typically driven by an underlying southbound API feature set. The opposite approach should be utilized, where the user-facing API should be developed based on user/application requirements, with the southbound API and feature sets correspondingly interfaced and adapted. With such a layer-based orchestration architecture, the SDN Controller Northbound API should be constructed with the orchestration layer as its user.

- Current Orchestration Technologies - The current orchestration architecture and associated implementation projects have not defined the requirements and features needed for the API between the Orchestrator and underlying SDN Controllers. This would be the SDN Controller Northbound Interface, which would also be the Orchestrator Southbound Interface. These requirements and the orchestrator functionality should be driven by the user/application requirements and therefore reflected in the Orchestrator Northbound Interface.

- Combined SDN and Orchestration Technologies - The current solutions are focused on traditional service provisioning, customer onboarding, and operations/maintenance. While new technologies such as Network Functions Virtualization (NFV) and automated provisioning are being employed in service of this paradigm, the service and use model is not much changed from a customer perspective, outside of the ability to initiate automated functions. These systems are not currently on a path to the degree of realtimeness and interaction needed for the smart network and smart application ecosystem envisioned by the Research and Education (R&E) community. For example, the information exchange between the Orchestrator and SDN Controllers is not designed with the degree of flexibility to i) include/exclude real-time states, ii) adjust the degree of resource/topology sharing/hiding as required by local policy and/or user requirements, and iii) tailor

operations to optimize scalability or real-time responsiveness. In addition, the orchestrator functions are not designed to take advantage of enhanced interactions with underlying SDN controllers in the form of interfaces which are focused on interactive based and focused on realtime responsiveness and intelligent services.

- End-to-End Solutions - Current SDN and Orchestration solutions are not end-to-end in the context of application workflows. Domain science application workflows need solutions which manage all resources along the end-to-end path. Note that this includes the networking stack inside the end systems, as well as the devices along the network path.

Current SDN and orchestration systems do not provide the degree of interaction, realtimeness, and intelligent network services needed for the next generation of domain science workflows. An integrated smart-network, smart-application ecosystem is needed. Such an ecosystem would enable application workflows to ask questions, iterate on solutions, receive recommendations, and access full life-cycle status and troubleshooting information. Future SDN enabled infrastructure needs mechanisms to provide topology and state information in realtime, based on fine-grained policy, scalability, and service objectives. End system resource management needs to be integrated into the orchestration of end-to-end network resources. In the longer-term end-to-end system operation needs to be monitored at several layers with sufficient granularity, to provide a foundation for future system optimization through such mechanisms as reinforcement learning. Domain science application workflows need real-time, interactive, end-to-end orchestrated SDN services across large, distributed, multi-domain networks.

The SENSE project was motivated by a belief that Domain Science application workflows need real-time, interactive, end-to-end orchestrated SDN services across large, distributed, multi-domain networks.

3 Activities and Results

3.1 SENSE Overview

The SENSE project team is led by Energy Sciences Network (ESnet) and included participation by Lawrence Berkeley National Laboratory (LBL), Fermi National Accelerator Laboratory (Fermilab), Argonne National Laboratory (ANL), California Institute of Technology (Caltech), and University of Maryland/Mid-Atlantic Crossroads (UMD/MAX).

There are three main components of the SENSE System with lead organizations noted for each:

- SENSE Orchestrator – Integrates resource descriptions from the Resource Managers and includes computation services to process model based descriptions to respond to user requests and coordinate provisioning actions. UMD/MAX led the development of this component.
- SENSE Network Resource Manager (SENSE Network-RM) - This component covers network resources such as ESnet and dynamically constructs model based descriptions for consumption by the SENSE Orchestrator. ESnet led the development of this component.

- SENSE End-Site/DTN Resource Manager (SENSE EndSite/DTN-RM) – This component covers the End Site resources including the Data Transfer Nodes (DTNs) and associated networking. In addition to creating model-based descriptions, the SENSE EndSite/DTN-RM includes capabilities to instantiate per flow priority as part of SENSE end-to-end provisioning. Caltech led the development of this component.

SENSE includes a model-based orchestration system which operates between the SDN layer controlling the individual networks/end-sites, and science workflow agents/middleware. In addition, SENSE has developed a "Network Resource Manager" and "End-Site Resource Manager" which enable advanced features in the areas of multi-resource integration, real time responsiveness, and user interactions. Figure 1 presents an overview of the SENSE system architecture.

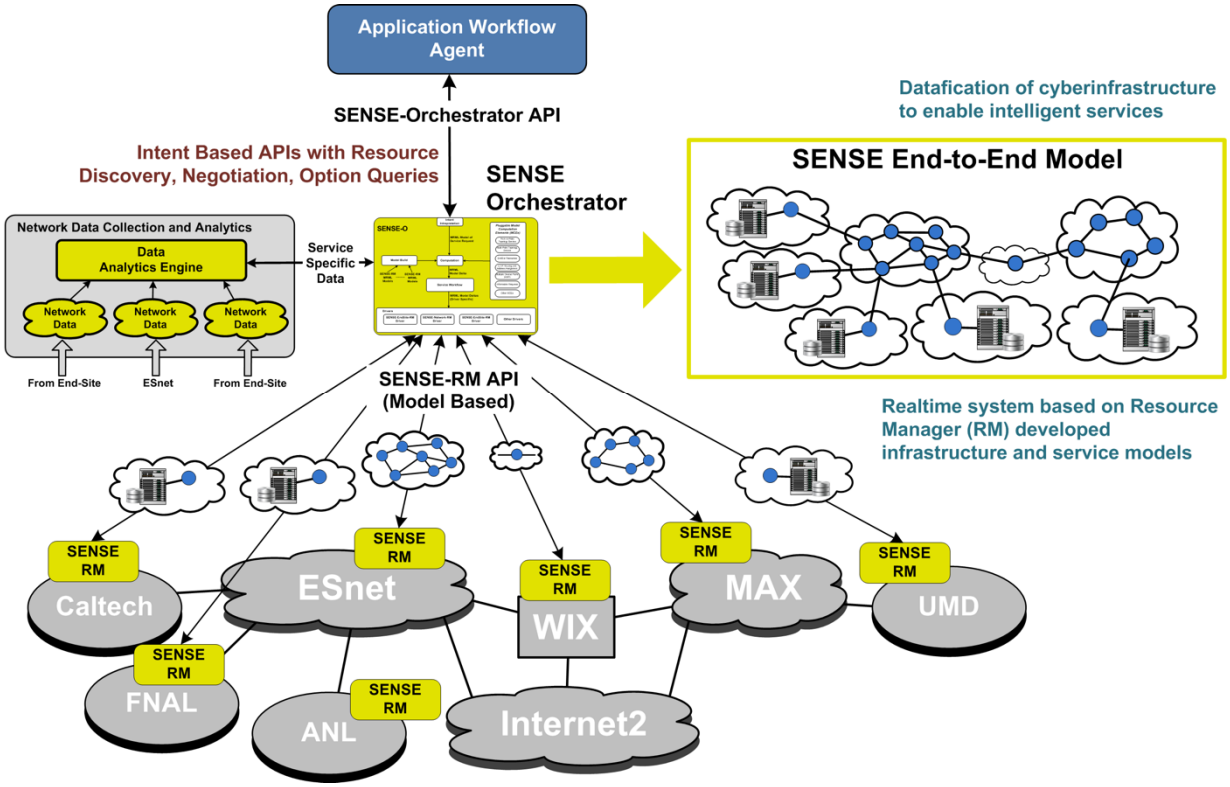

Figure 1 SENSE Architecture

The SENSE system defines the mechanisms needed to dynamically build end-to-end deterministic and policy-guided Layer 2/3 network services. An intent-based interface allows applications to express service requirement in a high-level domain science specific context, with mechanisms for interactive and full-service lifecycle coordination with workflow automation systems. The SENSE system has two key objectives:

- Facilitate science workflow related provisioning and life-cycle management for a variety of end-to-end network services.
- Provide the intelligence and control to enable more optimal and efficient use of network infrastructures.

The impact of these capabilities includes an ability for science applications to manage the network as a first-class schedulable resource akin to instruments, compute, and storage. This will enable workflow driven optimization of resources spread across a vast geographic footprint such as those used in science domains like high-energy physics and basic energy sciences.

As summarized in the introduction project motivation sections, there are several key features which are missing from current solutions as they relate to domain science research and associated cyberinfrastructure. The SENSE system enables a new application to networked system interaction paradigm which includes the following capabilities in response to the solution objective and problem statement:

- Intent Based: The ability for an application to submit a service request in the form of a high-level statement of desired results or outcomes, as opposed to a specific set of network centric inputs. The format of an Intent based interface will be customized based on application specific requirements. In some situations, an intent may be expressed based on a highly abstracted network view with performance metric annotations. In other situations, the intent form will be expressed in the context of application specific resources, end points, and references. The SENSE system is designed to apply a DevOps model to the interface construction, which is enabled by a rich semantic web model-based infrastructure description which allows for variable levels of abstractions and infrastructure/services relationship tracking.

- Interactive: The ability for an application workflow agent to engage in a bi-directional exchange aka "conversation" with the network as part of workflow planning. This conversation can include discovery of available services, asking "what is possible" or "what do you recommend" types of questions, engaging in iterative negotiations prior to actual service requests, or full-service life-cycle status and troubleshooting queries. This can be extended to processes that drive adjustments or remedial actions to maintain system performance and/or task progress, and the balance among competing demands on the available resources as needed.

- Realtime: This term has many different meanings and time scales depending on the context of the situation. For the purposes of this project, the problem space is large scale multi-domain, orchestrated SDN services. Each of the full lifecycle activities of resource discovery, provisioning, service status, troubleshooting, and feedback response loops may have their own requirements as it relates to realtime operations. In these contexts, realtime typically means a time scale of seconds to minutes. For example, provisioning an end-to-end path which consists of two Department of Energy (DOE) Laboratories High Performance Computing (HPC) facilities connected across a single wide area network, may have a response time in the 10s of seconds. A more complex end-to-end path with ten or more separate administrative domains, may have a response time of several minutes. A key design objective of the SENSE design is to provide the mechanism where a tradeoff between realtimeness and scalability can be made at runtime by dynamic configuration. The SENSE model-based interface between the orchestration and SDN layer is designed to allow this tradeoff via controls that dynamically vary the real-time states which are monitored as part of the topology distribution, in addition

to a more focused interaction for specific real-time information discovery and negotiations.

- End-to-End: The SENSE notion of end-to-end orchestrated SDN includes the multi-domain wide area, regional, and end-site networks as well as the network stack inside the end systems. The inclusion of the end-system networking stack is very important from deterministic and automated service provisioning, monitoring, and troubleshooting perspective. The practical application of this approach is to manage the networking stack all the way to the network socket of the host operating system, virtual machine, or container where the application process is interacting with the network. This is designed to provide a foundation for applications such as science workflow management systems that coordinate the use of computational, storage and network resources.

- Full-Service Lifecycle Interactions: In order to optimize performance and adjust to changing conditions, applications need mechanisms to discover status and states during the service provisioning as well as during the service operational phase. This includes resource discovery, provisioning, service status, troubleshooting, and feedback response loops activities. A continuous conversation between application and network for the full-service duration which allows for application situational awareness is part of the SENSE vision.

3.2 SENSE Key Functions

There are four main functions of the SENSE system:

- SENSE Orchestrator North Bound Interface - A highly customizable interface for application workflow agents to query regarding possible actions, recommendation, and/or request specific service instantiation. While a standard northbound interface has been defined, this interface is designed to be easily and rapidly customized for individual user requirements. The SENSE system has much data and intelligence regarding underlying networked system. This information can be customized for user consumption in a highly detailed or abstract manner.

- SENSE Orchestration - This includes the integration of resource model-based descriptions from the underlying network infrastructures, the computation services to process resource models to respond to user requests, and the coordination of provisioning actions.

- SENSE Orchestrator South Bound Interface - Provides for a continuous exchange of topology descriptions which include an ability for the resource owners to tailor the level of abstraction and realtime states in accordance with local policies and service objectifies. This is one of the key innovations of the SENSE system and is based on semantic web-based graph models which provides for high degree of service flexibility and infrastructure owner-controlled customizations.

- SDN Layer - The SENSE architecture does rely on an underlying SDN layer, however it does not require a particular SDN controller or system implementation. The SENSE architecture accepts that there will be a variety of deployed SDN solutions which will cover specific network and administrative regions. The SENSE system

provides the mechanisms and infrastructure to leverage these systems and provide guidance for how they can be fully integrated into the SENSE system. This requires existing SDN systems to implement the SENSE Orchestrator Southbound Interface as their controller Northbound Interface. Existing systems may accomplish this via native implementation of the SENSE API or via thin layer on top of their existing API which provides the proper interface. This technique of adopting underlying SDN systems for SENSE system integration has been used successfully as part of the SENSE system deployment on ESnet and other R&E infrastructures. Systems based on OpenDaylight (ODL), Network Services Interface (NSI), On-Demand Secure Circuits and Advance Reservation System (OSCARS), and Open Network Operating System (ONOS) have all been integrated into SENSE orchestrator operations. The SENSE development and testing activities have demonstrated that valuable orchestrated services can be provided using these existing SDN systems as they are with no internal modifications. In addition, more advanced SENSE services can be enabled by making some changes to these systems in the areas of topology description, abstraction, real-time states inclusion, and computation to support negotiations. SENSE has developed "Network Resource Manager" and "End-Site Resource Manager" components which enable advanced features in the areas of multi-resource integration, real time responsiveness, and user interactions.

3.3 SENSE Architecture Components

Within the SENSE orchestration architecture, there are two distinct functional roles, Orchestrator and Resource Manager (RM). The interaction of Orchestrator(s) and RM(s) follows a hierarchical workflow structure whereby the Orchestrator accepts requests from users or user applications and determines the appropriate RMs to contact and coordinate the end-to-end service request. The RMs are (administrative or technology) domain specific and are responsible for committing and managing local resources.

As illustrated in Figure 1, this hierarchical structure of Orchestrator and RM components separates application-facing service control functions from infrastructure-facing resource control functions.

3.3.1 SENSE Orchestrator

The SENSE Orchestrator is expected to be closely associated with a domain science collaboration/application (e.g. LHC/CMS [10], ExaFEL [11]) and processes “high-level” context sensitive intents to determine what resources are needed and coordinate the requests of “lower-level” (or sub) intents to the corresponding RMs. As such, the SENSE Orchestrator performs the following functions:

- Receives model-based resource descriptions from multiple RMs
- Receives and responds to the user’s “high-level” intent requests (which is defined within the context of the user’s domain science collaboration/application).
- Renders the user’s “high-level” descriptive intent request into “low-level” prescriptive requests for required resources
- Performs multi-constraint resource computation (based on AuthN (authentication) / AuthZ (authorization), resource availability, etc.) to determine the appropriate and necessary resources needed and corresponding RMs to contact. The AuthN functions are utilized to create the trusted relationship between the SENSE Orchestrator and the individual SENSE-RMs. The SENSE orchestration layer

AuthZ/AuthN functions are based on industry and community standards such as OAuth2, OpenID Connect and InCommon.
- Coordinate requests and replies from RMs and feedback the results to the user accordingly.
- Support queries by the user for status and state.
- Provide resource notifications to users as necessary.

The SENSE Orchestrator can take on different functionality, customized to the domain science needs and resources available to it (e.g. experiment, compute, storage, network, etc.). It provides a SENSE Orchestrator North Bound Interface (SENSE Orchestrator NBI) to accommodate such needs via service-oriented intent based interactions. This north bound interface is discussed further including details on intent design, negotiation mechanism, and workflow operations in Section 3.

3.3.2 SENSE Resource Manager

The SENSE Resource Manager (SENSE-RM) is tied to a domain with physical resources, e.g. a WAN (with network resources), a site (with Science DMZ resources), etc., and is responsible to facilitate the management of domain-specific resources. The SENSE-RM is responsible for the following functions:

- Provides (appropriately scoped and abstracted) model-based resource descriptions
- Receives and responds to the "low-level" intent requests from the Orchestrator
- Performs multi-constraint resource computation (based on resource availability, authentication/authorization, administrative policy etc.) to determine the local resources appropriate and necessary to service the request
- Coordinates resource allocations/commitments, provisioning and de-provisioning with local controllers (e.g. NMS, etc.) as necessary
- Supports queries by the Orchestrator for status and state
- Provides resource notifications to the SENSE Orchestrator as necessary

The SENSE-RMs are specific to an administrative domain. However, within a single administrative domain, multiple instances of RMs may be deployed based on the distinct technology regions (e.g. Data Transfer Nodes (DTNs), optical packet/transport, Layer2 OpenFlow, etc.). Conversely, a SENSE-RM may model multiple technology domains as a single resource description. For example, a network may have distinct switches and routers which provide layer 2 and layer 3 services correspondingly. However, the domain may instantiate a single RM which provides a unified resource description characterizing both sets of resources. The current SENSE implementation includes the following types of SENSE-RMs: Network Resource Manager (Network-RM) and Data Transfer Node Resource Manager (DTN-RN). There are multiple types of Network-RM which are tailored toward interoperation with specific types of underlying SDN systems. The DTN-RM is currently evolving into a Site Resource Manager (Site-RM) where it will manage some of the local site networking in addition to the end hosts networking stacks. In some situations, a Site-RM may be responsible for a complex set of end-system and network resources, to the point where it is effectively performs orchestration services across multiple site systems. This in line with the overall SENSE architecture where the resource manager identifies a set of resources for which it is responsible. The higher-level orchestrator does not need to know that there may be more levels of resource managers or controllers within a site.

3.3.3 SENSE Services

The SENSE system has been developed to operate in "Development Operations (DevOps)" mode, where custom services can be rapidly developed in response to individual application requirements. The general system philosophy is that while not "every" service imaginable can be implemented, almost "any" service can be. That is, the system design is such that resource states and capabilities are sufficiently available to allow the construction of many different services. The user requirements will be utilized to form the basis of the actual services. For each of these services, the user can interact with SENSE in the following modes:

- Immediate Provision: If SENSE finds a resource path which satisfies the application request, provisioning starts immediately (after routine confirmations from both sides).
- What is Possible?: In this mode, SENSE simply conducts a "Resource Computation" and provides the results back to the requestor. No provisioning action is taken without further explicit requests from the user.
- Negotiation: One or more rounds of Resource Computation requests with subsequent provisioning request by the application user if desired.

In the context of SENSE services, the "network" includes the switching and routing elements AND the network stacks of the end systems, such as Data Transfer Nodes inside Science DMZ facilities. The data plane capabilities associated with these services are:

- Layer 2 point-to-point with Quality of Service (QoS)
- Layer 2 multi-point with QoS
- Layer 3 Flow QoS
- Layer 3 Virtual Private Network (Create and Attach)

From the user application perspective, the SENSE Orchestrator provides application services via a programmable northbound interface, called the SENSE Orchestrator NBI. While the SENSE Orchestrator can support modular intelligent computation and almost arbitrary orchestrated services, exposed through the SENSE Orchestrator, the SENSE Orchestrator NBI is a select intent based API with an emphasis on end-to-end network connection discovery and computation, featured with intelligent bandwidth and schedule negotiation and workflow assistance

Through ontology-based resource modeling and intelligent orchestration, SENSE provides a powerful solution for end-to-end QoS across many network domains. Bandwidth QoS only represents one aspect of Quality of Experience for data transfer application users. Many users also want deterministic or predictable time schedules for data transfers. To add some more specificity to these ideas of smart network services, we present four examples of SENSE service capabilities that represent the intent sophistication used in describing different levels of Quality of Experience.

- Immediate QoS Provisioning - Our supported QoS classes include guaranteedCapped (no burst over the capped limit), softCapped (allowing for bursting over the cap when extra bandwidth is available) and bestEffort. For users who are not sure how much bandwidth to ask but want to first query for the maximum available, the intent can include a " maximum-bandwidth" query statement as shown in the earlier example in Section 3.2. Once the intent negotiation is concluded, the service will be

reserved across all domains, and then immediately provisioned once it is committed.

- Time-Block-Maximum Bandwidth (TBMB) - Through the example intent snippet below, we introduce the "bandwidth scheduling" intent. In this example, the user asks for the same 10G connection without requiring immediate provisioning. Instead the intent is to schedule a bandwidth service to start and end at specific times in the future, and to provide the maximum possible bandwidth that is continuously available during that block of time.

- Bandwidth Sliding Window (BSW) - SENSE also implements an interesting feature for end-to-end bandwidth scheduling based on the "sliding window" concept. As an example, a user intent is to schedule a service lasting for 4 hours that can be scheduled flexibly within the next 2 days. This particular intent is called a "bandwidth-sliding-window".

- Time-Bandwidth Product (TBP) - Another interesting SENSE service intent is based on the concept of the "Time-Bandwidth Product" (TBP). For 8 hours of transfer at a bandwidth of 10Gbps, the TBP represents a data volume of 36000 gigabytes or 36 terabytes Allowing users to query and negotiate bandwidth and schedule based on a given TBP will provide a better quality of experience in many cases where bulk data transfer is the goal, as TBP is a good estimate of the total amount of data to transfer. In the example of intent "queries" given below, the user tries to find a schedule to transfer an estimated 10,000 gigabytes (10 terabytes) of data within a 2 day time window after October 1st 2018 8:00ET. The user wants to check for the fastest possible transfer speed using a "use-highest-bandwidth = true" option. Alternatively, the user can ask for the least bandwidth (or widest schedule) using a "use-lowest-bandwidth = true" option, or a bandwidth-bounded schedule using both "bandwidth-mbps >=" and "bandwidth-mbps <=" options. The latter will return a feasible schedule that satisfies both the time-bandwidth-product and the bandwidth upper and lower bounds.

Through implementation of these sophisticated capabilities, SENSE represents an innovative end-to-end SDN service paradigm. In this paradigm, the network control plane is an intelligent system that integrates and orchestrates arbitrary end-to-end services through ontology based real-time resource modeling and modular model computation. Users or client agents can then ask the intelligent system complex "What is possible" questions and operate their services in an intent based, interactive and negotiable workflow. In the SENSE project, we are building a reference implementation that is specific to the big science models, controlling primarily data transfer and network resources.

Further developments of this implementation will continue to provide more sophisticated intents adapted to complex situations encountered in actual field operations, and possible optimized responses using machine learning.

4 Use Cases

The SENSE project is now in a phase where use case integration is a key focus area. The main use cases currently under test are as described below.

Data Transfer Node Priority Flow: Science DMZ located Data Transfer Nodes (DTNs) [6] are a common method for moving data to/from compute facilities in the R&E community. For this use case, SENSE services are utilized to enable a "DTN Priority Flow Service". Since SENSE services are provisioned across the switching and routing elements and the network stacks of the end systems, this allows the creation of QoS-enabled paths that can be utilized for specific flows such that deterministic performance can be achieved regardless of the background traffic. The concept of operation is that these "SENSE enabled DTNs" can either be placed adjacent to current production DTNs as standalone transfer nodes, or SENSE software can be installed directly on the production DTNs. In either case, standard DTN operations and flows across the best effort routed IP paths continue as normal. When a SENSE flow is established between DTNs, this flow will receive priority access to network and host level resources. The best effort flows will continue, but maybe at a reduced rate. This SENSE capability includes a Layer 2 point-to-point, Layer 2 multipoint service, and. Layer 3 QoS and VPN based services. The workflow agent for this use case utilizes the "Time-Block-Maximum Bandwidth", the "Bandwidth-Sliding-Window", and the "Time-Bandwidth-Product (TBP)" SENSE services to instantiate Layer 2 paths with QoS. This workflow also utilizes the "What is Possible?" and "Negotiation" features to demonstrate those feature sets. A description of SENSE services, as well as additional information regarding testing for this use case is available here [12].

LHC/CMS Use Cases: The SENSE project is also working on use cases that integrate with Large Hadron Collider/ Compact Muon Solenoid (LHC/CMS) data movement and analysis workflows. The Worldwide LHC Computing Grid (WLCG) relies on the File Transfer Service (FTS) as the data movement middleware for moving sets of files from one site to another. Currently, we are focusing on multiple implementations with SENSE Services:

- Rucio - the next-generation of Distributed Data Management system benefiting from recent advances in cloud and "Big Data" computing to address high-energy physics experiment scaling requirements. Rucio was originally developed to meet the requirements of the high-energy physics experiment ATLAS and now is extended to support not only the LHC experiments but also other diverse scientific communities. Rucio uses File Transfer Service (FTS) to globally distribute the majority of the LHC data across the WLCG infrastructure. We are implementing a plugin for FTS Service to be able to request network resources depending on the transfer queues and sizes between participating sites and do bulk transfer of files reliable from one site to another.

- Another ongoing implementation is a new compact event form called the "nanoAOD" [4] that enables the rapid widespread distribution, ingest and real-time processing through a set of "PhysicsTools" of entire datasets of one to a few terabytes, that can be subsequently further analyzed on user's desktops and laptops.

The associated CMS analysis workflows and Distributed Data Management implementations currently under development, are planned to be accelerated and scaled up in terms of the number of simultaneous workflows supported, through the use of SENSE's interactive bandwidth allocation and management services, together with the RM services at a number of CMS sites, and high throughput data transfer applications such as FTS, XRootD and Caltech's open source Fast Data Transfer (FDT).

Exascale for Free Electron Lasers (ExaFEL): The objective of this use case is to stream nano crystallography diffraction data from SLAC National Accelerator Laboratory (SLAC) to National Energy Research Scientific Computing Center (NERSC) over the network in order to perform analysis on Cori, a Cray XC40, which has a peak performance of about 30 petaflops with 2,388 Intel Xeon "Haswell" processor nodes, 9,688 Intel Xeon Phi "Knight's Landing" nodes, and a 1.8 PB Cray Data Warp Burst Buffer. The feedback is provided afterwards to the beamlines in the form a 3D electron structure visualization. The workflow not only uses SENSE components to stream the data from the LCLS online cache at SLAC to NERSC compute nodes over network but also orchestrates the SFX analysis processes to give real-time or near-real-time feedback to the experiment. For this use case, the ExaFEL application workflow agent utilizes the "Time-Block-Maximum Bandwidth" SENSE Service to provision the network path. This includes establishment of Layer 2 paths with QoS with time domain scheduling. Additional information regarding testing for this use case is available here [2].

BigData Express: BigData Express provides schedulable, predictable, and high-performance data transfer service for DOE's large-scale science computing facilities (LCF, NERSC, and US-LHC computing facilities, among others) and their collaborators. This project seeks to orchestrate the system, storage, and network resources involved in high-performance data transfers. From a network services perspective, BigData Express focuses on controlling local network resources supporting the end systems. For wide-area service, BigData Express system utilize SENSE services to provision paths across ESnet. The BigData Express workflow agent utilizes the "Time-Block-Maximum Bandwidth" and the "Bandwidth-Sliding-Window" SENSE services to instantiate Layer 2 paths with QoS. This application also utilizes the "What is Possible?" and "Negotiation" features sets to co-schedule across multiple end-sites and network resources. Additional information regarding testing for this use case is available here [3].

AutoGOLE and SENSE Integration: The Research and Education community has standardized network provisioning through the Network Service Interface (NSI) developed within the Open Grid Forum (OGF). The AutoGOLE is a worldwide community of networks that actively supports NSI. Beyond the horizon of network standardization, there are plans to develop multi-domain and multi-resource provisioning and scheduling to include the end site resources. This typically includes end site based Data Transfer Nodes (DTNs) as dedicated high-throughput servers. The SENSE system is being utilized to enable multi-resource service provisioning and coordinated provisioning of network and end site resource including the DTNs. The AutoGOLE/NSI and SENSE systems both utilize a model-based resource description system based on the Network Markup Language (NML)[14]. This common resource description language provides a good starting point for the integration of AutoGOLE and SENSE provisioned services. This work is in the prototype phase and is expected to continue as part of the Global Networking Advancement Group (GNA-G) of the Global Network Architecture collaboration [13]

5 Testbed Deployment

The SENSE solution architecture aims to address the problem of real-time interactive end-to-end SDN orchestration, which consists of a complex of realistic problem features including distributed resource management, real-time modeling, multi-domain data integration, end-to-end orchestration, and intelligent service interface and interaction. The methodology for SENSE solution architecture design is more empirical than quantitative.

Reference implementations and testbed experiments are our primary means to validating the design. In addition, a real-world, at-scale SENSE testbed deployment helps us evaluate its technical applicability for a wide spectrum of use cases, scenarios and application workflows.

The SENSE architecture, models, and protocols define methods such that new implementations can include the most advanced levels of smart interactive networked services. However, a key part of the SENSE vision is to allow adaptation to and deployment on existing facility deployments interconnected by production networks. The same features that allow for adjustment and optimization of realtimeness vs. scalability, also allow existing SDN deployments to adapt their use of the SENSE functions in a manner which is compatible with their underlying network infrastructure. This allows early deployment of SENSE services for testing and use case development, and also provides guidance for future upgrades of network automation systems. This approach allowed the deployment of SENSE service which operates on top of ESnet, DOE laboratory, and university production and testbed infrastructures. The result is a SENSE testbed which allows for real world testing and the ability to provide services to use cases which include connections to their production resources.

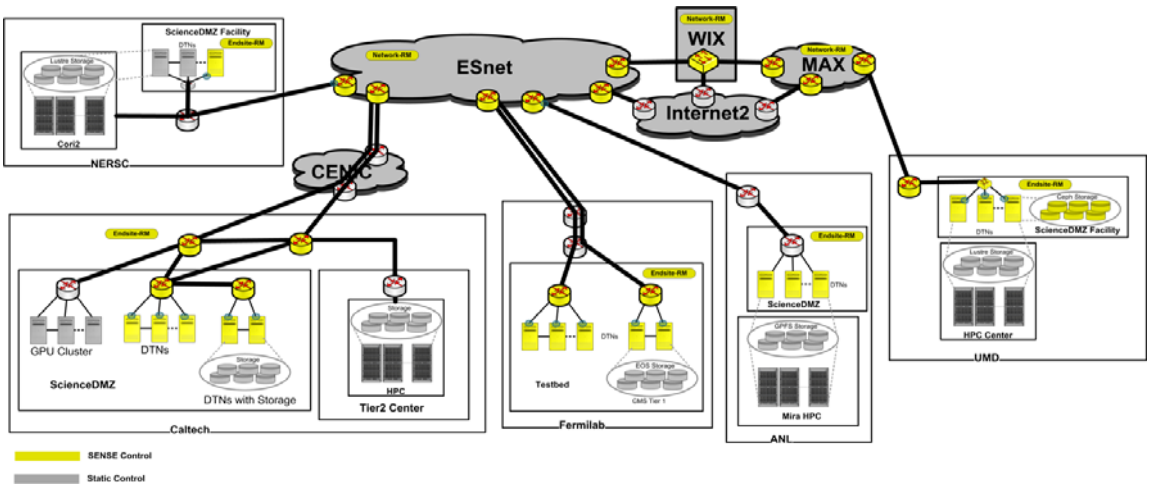

Figure 2 SENSE Testbed

The SENSE testbed deployment had to deal with multiple existing deployed SDN systems. The SENSE system provides the mechanisms and infrastructure to leverage these systems and provide guidance for how they can be fully integrated into the SENSE system. This requires existing SDN systems to implement the SENSE Orchestrator Southbound Interface as their controller Northbound Interface. Existing systems may accomplish this via native implementation of the SENSE API or via thin layer on top of their existing API which provides the proper interface.

This technique of adopting underlying SDN systems for SENSE system integration has been used as part of the SENSE system deployment on ESnet and other R&E infrastructures. Systems based on OpenDaylight (ODL], Network Services Interface (NSI), On-Demand Secure Circuits and Advance Reservation System (OSCARS), and Open Network Operating System (ONOS) have all be integrated into SENSE Orchestrator operations. The SENSE development and testing activities have demonstrated that valuable orchestrated SDN

services can be provided using these existing SDN systems as is, with no internal modifications. The practical implication of this approach so far is that existing SDN system capabilities may limit the degree of realtimeness or interactivity that SENSE can provide to the user's application workflows.

However, we have also identified the needed changes to these systems in the areas of topology description, abstraction, realtime states inclusion, and computations to support negotiation that will allow for full provision of the more advanced SENSE services. There are also opportunities for native implementation of a SENSE based SDN system which further enhances the ability to increase the realtimeness and interactivity of the orchestrated services. This typically involves a tighter coupling between the SENSE defined resource model generation, real-time states tracking, and resource control mechanisms. This native implementation approach was utilized for the SDN layer at end sites, and higher performance was observed in the subsequent testing activities.

A SENSE testbed has been deployed which includes a mix of development and production resources. This testbed is being utilized to develop and test the SENSE software, as well as test with domain science use cases. As shown in Figure 2, this testbed is deployed at several DOE laboratory and university facility sites. For the allocation of network resources, the SENSE system interacts with production provisioning systems of ESnet and other networks. For the end-system resources, a mix of production and prototype DTNs are deployed. For the production DTNs limited access is provided, resulting in tailoring the set of SENSE based dynamic configurations to match local site policies. This approach to use a mix of production and research resources enables experience with various real-world site deployments and considerations.

The initial experimentation on the SENSE testbed were focused on validating the SENSE solution architecture design, evaluating soundness of intent based interactive service workflow for real-world use cases, and obtaining metrics on the key performance factors for realtimeness and scalability.

6 Products, Documents, Presentations

SENSE project activities included work on the following items.

<u>Products</u>

1) SENSE Project Web Site
Description: SENSE Project Web Site which provides overview of the project and activities.
URL: http://sense.es.net

2) SENSE Orchestrator Software Repository
Description: SENSE Orchestrator Software repository and associated WIKI. All software includes a MIT License, GENI Public License, or BSD License.
URLs: https://github.com/esnet/SENSE-Orchestrator/wiki
http://github.com/esnet/StackV.community

3) SENSE Network Resource Manager for NSI based networks
Description: SENSE Network Resource Manager software repository and associated information for NSI based networks.
URLs: https://github.com/esnet/sense-rm

4) SENSE Network Resource Manager for OSCARS based networks
Description: SENSE Network Resource Manager software repository and associated information for OSCARS based networks
URLs: https://github.com/esnet/sense-nrm-oscars

5) SENSE EndSite/DTN Resource Manager
Description: SENSE EndSite/DTN Resource Manager Software repository and associated Web Page. All software includes an Apache License, Version 2.0.
URLs: https://sdn-sense.github.io
https://github.com/sdn-sense

6) SENSE Orchestrator (NorthBound) API
Description: Provides a definition of the SENSE Orchestrator Northbound API
URL: https://app.swaggerhub.com/apis/xi-yang/SENSE-O-Intent-API/0.9.0

Documents
2018 IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS)
The Innovating the Network for Data-Intensive Science (INDIS) workshop
November 11, 2018
Dallas, Texas

"SENSE: SDN for End-to-end Networked Science at the Exascale" Paper

Inder Monga, Chin Guok, John MacAuley, Alex Sim), Harvey Newman, Justas Balcas, Phil DeMar, Linda Winkler, Tom Lehman, Xi Yang, "SDN for End-to-End Networked Science at the Exascale (SENSE)", 2018 IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS) Workshop, November 15, 2018, co-located with SC18, Dallas, Texas.

INDIS Workshop at SC18 - SENSE Paper URLs:
- http://cmsweb1.es.net/assets/Uploads/76kkEi9wqPrVdrGK4bSFdg.pdf
- https://ieeexplore.ieee.org/document/8648795

Presentations
1) 2018 IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS)
The Innovating the Network for Data-Intensive Science (INDIS) workshop
November 11, 2018, Dallas, Texas

"SENSE: SDN for End-to-end Networked Science at the Exascale" Presentation
- https://scinet.supercomputing.org/workshop/sites/default/files/INDIS18-Monga-SENSE.pdf

2) MAXedge Services using the SENSE System
- 2019 MAX Participants Meeting
  April 11, 2019
  College Park, MD
- https://www.maxgigapop.net/wp-content/uploads/MAXedge-Services.pdf (pp. 43-46)

3) SENSE Demonstration Videos

- Point to Point QoS
  https://www.youtube.com/watch?v=YOAo2A0eXZQ
  https://tinyurl.com/sense-pt-to-pt-sc18

- Multi Point service
  https://www.youtube.com/watch?v=D7hCfJZIU_M
  https://tinyurl.com/sense-multipoint-sc18

- Intent API Intelligent Services
  https://www.youtube.com/watch?v=w5jSVon1ZdU
  https://tinyurl.com/sense-api-sc18

- Layer-2 & Layer-3 VPN Services
  https://drive.google.com/file/d/1j_0YV0Jug3ynENqcb_2iicenHFHqWdM3
  https://tinyurl.com/sense-3vpn-sc19

7 Summary and Conclusions

The SENSE system architecture and implementation utilizes model-driven datafication of cyberinfrastructure to enable intelligent network services. Science applications utilizing Intent-based APIs with automated resources discovery and negotiation enable a significantly different mode of operation as compared to current network usage modes. With dropping costs of 100GE capable devices, powerful end systems are increasingly being placed at edge locations where high-bandwidth connections directly to regional and national networks will be the norm. The Science DMZ is an example of a high-performance end-system edge deployment. As a result, the expectation is that we are entering a cycle where network capacity will be easily overwhelmed by these advanced end-site and edge facilities. This indicates a need for methods to manage network resources and access in a more intelligent manner, which includes providing the application agents with sufficient information so that they can plan and optimize their operations. The SENSE project vision and implementation is focused on these issues to be prepared for the day where unmanaged network utilization and extreme over provisioning is no longer the preferred operational approach.

The SENSE architecture and service plan creates many avenues for investigation and provides a platform to address interesting research questions. These issues revolve around the focus on interaction, negotiation, the degree of realtime state management and consideration at many levels of the decision and control operation process.

As the SENSE architecture and implementation evolves through multi-institution testbed deployment, the focus of the project is to continue integration with domain science use cases and transition the SENSE services to production status for both the network and application operations. As the SENSE research project ends, efforts continue to transition this system to production status in support of science workflow operations and cyberinfrastructure related systems.

As the SENSE research project ends, efforts continue to transition this system to production status in support of science workflow operations and cyberinfrastructure related systems.

Future plans include exploring some of these issues as noted below as part of ongoing development and testing of the SENSE system:

- What are the tradeoffs between scaling and real-time state collection and performance?
- How to make the realtime vs scalability features dynamic and configurable so adjustments can be based on conditions and application needs?
- Which information/states should be routinely exchanged between Resource Manager(s) and Orchestrator? Which information should be accessible on demand? What are the best methods to make this dynamic/configurable to adjust based on different Resource Manager capabilities and policies?
- What is the right level of abstraction for Application Agent to SENSE system interactions? Is it necessary to provide variable levels from highly abstract to very detailed and resource specific?
- What is the best method for realizing multi-domain, multi-resource authentication and authorization? What is the proper granularity for this? User? Project? Domain? Individual network and end system resource elements? Flows?

References

1. SDN for End-to-end Networked Science at the Exascale (SENSE), http://sense.es.net
2. ExaFEL use case testing, http://www.slac.stanford.edu/~wilko/psdm/netdemo/rates.html
3. Big Data Express, http://bigdataexpress.fnal.gov
4. nanoAOD, https://github.com/cms-nanoAOD
5. Fast Data Transfer (FDT), https://github.com/fast-data-transfer/fdt
6. Data Transfer Node (DTN), https://fasterdata.es.net/science-dmz/DTN/
7. StackV SENSE Open Source Orchestration Software Suite, http://sense.es.net/software
8. SENSE Orchestrator North Bound API, https://app.swaggerhub.com/apis/xi-yang/SENSE-O-Intent-API/0.9.0
9. Data Transfer Node Priority Flow testing, https://tinyurl.com/sense-demo
10. Large Hadron Collider/ Compact Muon Solenoid (LHC/CMS), https://home.cern/about/experiments/cms
11. Exascale for Free Electron Lasers (ExaFEL), https://www.exascaleproject.org/project/exafel-data-analytics-exascale-free-electron-lasers/
12. Inder Monga, Chin Guok, John MacAuley, Alex Sim), Harvey Newman, Justas Balcas, Phil DeMar, Linda Winkler, Tom Lehman, Xi Yang, "SDN for End-to-End Networked Science at the Exascale (SENSE)", 2018 IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS) Workshop, November 15, 2018, co-located with SC18, Dallas, Texas. http://cmsweb1.es.net/assets/Uploads/76kkEi9wqPrVdrGK4bSFdg.pdf
13. Global Network Architecture (GNA) international collaboration of national research and education (R&E) networks, https://gna-re.net
14. Multi-Resource Markup Language (MRML), https://github.com/esnet/nml-mrml