NetSage
Award #1540933
Year 2 Annual Report
1 Feb 2016 through 31 Jan 2017

PIs: Jennifer Schopf (IU), Sean Peisert (UCD, LBNL), Jason Leigh (UHM)

Summary
The goal of the NetSage project is to collect data from the IRNC-funded backbone and exchange points to better understand the use of the networks. In addition, this collected data is also made available for use by the NOC for day-to-day operations. Highlights of Year 2 include gathering SNMP data from all backbones, gathering perfSONAR data from all IRNC backbones and exchange points, initial collection of sflow and Tstat data, initial development of analysis tools, and extended visualization of available data, along with the deployment of http://portal.netsage.global.

1. NetSage Overview
NetSage is building and deploying advanced measurement services that will benefit science and engineering communities, focusing on:

- Better understanding of current traffic patterns across IRNC links, and the ability to better understand growth trends for capacity-planning purposes;
- Better understanding of the main sources and sinks of large, elephant flows, to know where to focus attention on outreach and training; and
- Better understanding of where packet loss is occurring, whether or not the loss is caused by congestion or other issues, and the impact of this on end-to-end performance.

When fully operational, NetSage services will provide a combination of passive measurements (including SNMP data, flow data, and deep packet header inspection), as well as active measurements (mainly perfSONAR) for longitudinal network performance data visualization. Year 2 of the project focused on additional measurement information for the IRNC-funded backbones and exchange points, updates to the common archive, extending the visualization tools, and some work with analytics.

Year 3 will focus on the next stage of analysis based on our additional data sources, including the science registry. Going back to our guiding list of questions (http://bit.ly/2mrpudp), this will include:

- #3.5 Which links are experiencing packet loss;
- #5 What are the top sites that use the IRNC links?
- #6 What are the top science projects that use the IRNC links?
- #12 What is the nature of Elephant flows that use the links?
- #15 What is the max, min, and average duration of elephant flows?
- #19 How can we best identify a list of top talkers for each link?
- #20 How many flows experiencing issues also have small buffer sizes?

This report details the staffing, collaboration, tool development, deployment, and planning for the project.

2. Staffing

At the end of Year 1, funded staff included:

- Jennifer Schopf, IU, PI - overall project director
- Andrew Lee, IU, senior network engineer - measurement project management, IRNC project coordination
- Ed Balas, IU, system architect - collection and reporting
- Mitch McCracken, IU, software developer - PerfSonar, Data processing pipeline
- Dan Doyle, IU, TSDS development and support
- Sean Peisert, UC Davis, co-PI - security, privacy, performance experimental design
- Brian Tierney, UC Davis, staff scientist - monitoring architecture, performance experimental design, privacy
- Jonathan Ganz, UC Davis, graduate research assistant - configuration of and experimentation with performance monitoring tools
- Jason Leigh, UH Mānoa, co-PI - visualization oversight
- Alan Whinery, UH Mānoa, senior personnel - perfSONAR and coordination with backbones
- Alberto Gonzalez, UH Mānoa, graduate research assistant - visualization developer
- Eric Wu, UH Mānoa, graduate research assistant - visualization developer

In mid-April, Mitch McCracken was replaced by Michael Johnson on this project. In addition, over the course of the year, the IU development team expanded to employ (part time) CJ Kloote, for general development, Johnathan Stout, for work on the Science Registry, Uwe Dahlmann, for work analyzing the 100Gbps sensor, and Sangho Kim, for sensor development. For any given month, the time commitment for Dahlman, Doyle, Johnson, Kim, Kloote, and Stout (referred to in the budget as the IU development team) is equal to one full FTE - different development staff are used dependent on the task at hand.

IU hired on 5 summer interns, in part funded by the NetSage project: Abhishek Singh (MS at IU), Abhinandan Sampathkumar (MS at IU), Ayush Kohli (BS at Southern Illinois University), Tina Yu (BS at UIUC), and Sydney Lyon (BA at IU). They focused

on initial prototypes of analysis tools using flow data from the ACE and TransPAC projects. Singh and Sampathkumar continued as hourly employees after the summer.

Andrew Lee,IU, has shifted to having additional TransPAC4 responsibilities and is no longer funded by NetSage. Predrag Radulovic, IU, began attending meetings to weigh in on the network analysis side of the project and to assist as needed. He oversees the interns for the IU group, and has a deep background in the field due to his prior role with the GLORIAD InSight project.

Jonathan Ganz, UC Davis, left the project on July 31, 2016. Monte Goode and Chris Tracy, both LBNL, work on the project at intervals, part time. In January, Jon Dugan, LBNL, started with the project to take over the liaison role that prior to this Brian Tierney had filled. Tierney will be retiring in June, 2017.

At the end of Year 2, funded staff included:

- Jennifer Schopf, IU, PI - overall project director
- Predrag Radulovic, IU, analysis
- Ed Balas, IU, system architect - collection and reporting
- Uwe Dahlmann, IU, system engineer - 100Gbps sensor
- Dan Doyle, IU, developer - collection and reporting
- Michael Johnson, IU, developer - collection and reporting
- Sangho Kim, IU, Intern - sensor development
- CJ Kloote, IU, developer - collection and reporting
- Johnathan Stout, IU, developer - Science Registry
- Abhi Sampthkumar, IU, Intern - Network analysis
- Abhi Singh, IU, Intern - Network analysis
- Sean Peisert, UC Davis and LBNL, co-PI - security, privacy, performance experimental design
- Brian Tierney, UC Davis and LBNL/ESnet, staff scientist - monitoring architecture, performance experimental design, privacy
- Jon Dugan, LBNL/ESnet, senior personnel - monitoring architecture
- Monte Goode, LBNL/ESnet, senior personnel - monitoring architecture
- Christopher Tracy, LBNL/ESnet, senior personnel - capacity planning
- Jason Leigh, UH Mānoa, co-PI - visualization oversight
- Alan Whinery, UH Mānoa, senior personnel - perfSONAR and coordination with backbones
- Alberto Gonzalez, UH Mānoa, graduate research assistant - visualization developer
- Eric Wu, UH Mānoa, graduate research assistant - visualization developer

3. Collaborations, Travel, and Training

NetSage staff participated in various meetings to support ongoing deployment, collaboration, and training. During Year 2, these included:

- NetSage All Hands meeting, February 4-5, University of Hawaii Manoa, Hawaii. The full team attended this meeting that focused on Year 2 planning.
- Internet2 Global Summit (http://meetings.internet2.edu/2016-global-summit/), May 15-18, Chicago, IL. Schopf, Lee.
- TNC16 (https://tnc16.geant.org/), June 12-16, Prague, Czech Republic. Tierney, Schopf, and Lee. Internet2 Technical Exchange: https://meetings.internet2.edu/2016-technology-exchange/, SEPT 25-28, 2016,Miami, FL. Schopf. Paid by other funding: Tierney, Balas, Doyle, Johnson.
- NSF Campus Cyberinfrastructure PI and Cybersecurity Innovation for Cyberinfrastructure PI Workshop (http://www.thequilt.net/public-event/2016-nsf-campus-cyberinfrastructure-program-pi-workshop/), Oct. 19-21 2016, Philadelphia, PA. Tierney.
- IEEE Visualization (http://ieeevis.org/year/2016/info/vis-welcome/welcome), Oct 23-28, 2016, Baltimore, MD. Gonzalez. IEEE Viz is the primary conference for learning about the latest advances in data visualization. Gonzalez made valuable connections and learned about the latest advances in the field to be able to incorporate them into NetSage.
- SC'16, http://sc16.supercomputing.org/, Salt Lake City Nov 14-17, 2016. Schopf, Balas, Gonzalez, Leigh. Paid by other funding Radulovic, Tierney, Peisert. In addition to meetings with IRNC PIs, the team briefed NSF on their activities over the last year.
- January perfSONAR planning meeting, LBNL, Jan 24-25, 2017. Schopf, Doyle, Johnson. Attending via other funding were Tierney and Chevalier (IU). NetSage is becoming a bigger user of perfSONAR, and this meeting allowed us to make sure our requirements and upcoming needs were taken into account as part of their planning and development cycle.
- January All Hands Meeting, LBNL, Jan 25-26, 2017. Full team attended. This meeting focused on Year 3 planning, which will include an extension of the analysis, Additional flow data collection, and visualization of the new information.

4. Project Coordination

4.1 Internal Coordination

Internal project coordination continued with weekly meetings of the majority of the team. Toward the end of Year 2, some of these meetings focused on specific technical topics, and so in Year 3 we plan to implement a separate Development call with a rotating focus (analysis, tools, viz).

The weekly meetings were complimented by full team face-to-face meetings held every 6 months in rotating locations. In February 2016, we held our third All Hands Meeting at University Hawaii Manoa. This set out a plan for additional data collection, next steps for the visualization tools. We emphasized completion of the perfSONAR test mesh, the flow data collection pipeline, and initial portal design.

In May 2016, we held an All Hands Meeting at IU the week before the Internet2 Global Summit, and overlapping one evening with the PerfSONAR quarterly review meeting. The main topics were coordination with the research AMIS projects, finishing the policy documents for review by the IRNC PIs, coordination with the IRNC NOC for performance trouble shooting, especially with the IRNC perfSONAR mesh, and understanding next steps for data collection, analysis, and visualization.

In January, the Fifth All Hands Meeting in Berkeley focused on Year 3 plans, primarily collecting flow data from all the backbones, additional analysis tools, and how to incorporate ELK analysis and prototype visualizations into the NetSage portal.

4.2 IRNC Project Coordination

We continue to coordinate with the AMI research projects. Lee attended several AMIS project calls. A representative was sent to the May 2016 NetSage All-Hands meeting. Staff attended the AMIS project meeting (held in conjunction with TechEx), however it was decided that the instrument they are deploying was not yet ready for larger-scale deployment or to include that data in the NetSage archive at this time.

The other IRNC Research Monitoring project, InSight, is currently only working in a domestic space. We worked to establish a relationship with the new project leadership, but a vendor representative was sent to eh May meeting who could not speak to the broader project. Follow-up from this meeting has been challenging.

Work with backbones continued, and we now have SNMP and PerfSONAR data from all projects. In May, this included shifting from data collection for TransPAC3 in Los Angeles to working with TransPAC4 in Seattle as they brought their 100G circuit to ASia online. In September, a replacement for ACE was announced: the Networks for European, American, and African Research (NEAAR) project. This project began to bring up a new 100G link between the US and Europe in December 2016, with it going live in January. We are currently coordinating to include perfSONAR, SNMP data, and flow/Tstat data starting in February 2017.

Work with the exchange points is lagging slightly. CENIC and AmLight joined the IRNC perfSONAR mesh in October, with StarLight finally deploying a node in January 2017. CENIC and AmLight have also agreed in principle to share flow data. The IRNC NOC continues to coordinate with them for SNMP data. Progress with StarLight for flow data or SNMP data is less optimistic.

We continue ongoing discussion with the NOC. They are taking the lead at collecting SNMP data from the exchange points, for example. Both teams continue to use the shared archive.

5. Software Development

Over the course of the project, NetSage staff will be involved in the development and deployment of various pieces of software to support active and passive measurements, monitoring, archiving, analysis, and visualization. This sections details those projects.

5.1 System Architecture

During Quarter 1, we conducted a review of loss flow data and the flow size frequency distributions to guide system design and tuning for the IRNC archive and associated systems. This resulted in the decision to discard flows under 500 MBytes. This decision also assisted in some agreement for the privacy concerns, as email is included in this discarded category.

Initial testing of the flow processing pipeline was successfully performed using data from the TransPAC Los Angeles facility, resulting in flow data stored in TSDS. This testing included both IP to Geo-location lookups and the de-identification of IP address information before sending flows to the central IRNC archive. No significant issues in the pipeline design or TSDS storage model were identified during this testing.

During Quarter 2 and 3, we developed a netflow importer for the packet processing pipeline to support cases where we do not have access to packet capture data but are able to get netflow export from routers. We are now collecting netflow data from TransPAC and ACE, and we are working on a deployment with the AMLight team for the links in Miami.

In Quarter 4, we have refined the processing pipeline to include step where we store de-identified individual flows for short term analysis and generation of aggregates. At the end of year 3 we have a functioning flow processing pipeline that can take input data from sflow, netflow or Tstat, process it to add organization and geolocation metadata, de-identify according to defined policy and then store it in a central archive in both canonical / raw and indexed / queryable form.

5.2 Time Series Data System (TSDS)

The Time Series Data System (TSDS) (http://globalnoc.iu.edu/software/measurement/tsds.html) is a software suite that provides well structured and high performance storage and retrieval of timeseries data, including interface throughput rates, flow data, CPU utilization, and number of peers on a router. Along with the raw data, the TSDS suite is capable of tracking and reporting based on metadata, for example viewing interface throughput from the viewpoint of a VLAN or BGP peer sessions from a particular ASN.

In Project Year 2, there were three TSDS releases to improve stability and query language capability along with the integration of perfSONAR loss, latency, and

bandwidth data. Focus in quarter 3 and 4 for TDSD were on stability of the underlying Mongo database and revisions to the aggregation of individual flows.

5.3 Simple Network Management Protocol (SNMP)

The Simple Network Management Protocol (SNMP) is an application–layer protocol defined in RFC1157 for collecting and organizing information about managed devices on IP networks. SNMP is commonly used by routers and switches to monitor networks for conditions that warrant administrative attention. This data is commonly collected and openly archived by most R&E networks.

In Project Year 3, we completed the collection of SNMP based interface statistics with storage in TSDS. Work has begun to integrate exchange point usage data in collaboration with the IRNC NOC, which is leading the collection effort.

5.4 perfSONAR

perfSONAR (http://www.perfsonar.net/) is a network measurement toolkit designed to provide federated coverage of paths, and help to establish end-to-end usage expectations. The NetSage project will be using perfSONAR for its active measurements of bandwidth and throughput, and archiving them in the NetSage archive using TSDS.

In Project Year 3, the IRNC PerfSONAR components were upgraded to version 3.5 and a maddash test mesh was configured for the IRNC projects. Several issues were identified and addressed with partners as we work to establish a normal performance baseline. Tests with CENIC were added to the IRNC mesh and a node was deployed at starlight with tests added to the mesh.
http://data.ctc.transpac.org/maddash-webui/index.cgi?dashboard=IRNC%20Mesh

The addition of a node at StarLight in January, 2017, now means that we have perfSONAR data for all of the IRNC backbones and exchange points.

5.5 Tstat

Tstat (http://tstat.polito.it/) is part of the EUMeasurement Plane (mplane) FP7 project developed by Munafó and Mellia at Politecnico di Torino. Tstat can be used to analyze either real-time or captured packet traces, and rebuilds each TCP connection by looking at the TCP header in the forward and reverse direction. Tstat reports a number of useful TCP flow statistics, such as congestion window size and number of packets retransmitted, which can be used to analyze the health and performance of the link.

We developed a component to parse Tstat logs and import into the flow processing pipeline. The Los Angeles trial deployment for TransPAC3 of this is described above in Section 5.1. After additional testing, Tstat was deployed for the ACE links in WIX as well. In Quarter 3, we began evaluating these data sources, and identified several anomalies. Some were attributed to data normalization issues with how Netflow and Tstat data is represented within the packet processing pipeline. These were

resolved in Quarter 4. A second set dealt with functional correctness, for example, we identified instances where UDP flows were reported with impossibly large data transfers for the number of reported packets. Work is ongoing to evaluate the correctness of the data.

5.6 Flow Data Collection from Backbones

In addition to Tstat's enhanced flow data collection, we will also collect normal router-based Netflow or sflow data, depending on what the router supports. This will allow us to compare the results of Tstat-based header analysis with sampled flow data from routers, and will provide a backup source of flow data in case there are problems with the Tstat collectors.

Work was begun to import sflow data from TransPAC, ACE and AMLight netflow data into the archive in Year 2.

TransPAC3 data was being collected in LA by the project, as was ACE data from WIX. NetSage staff worked with those teams to include de-identified versions in TSDS for future use with the NetSage analysis tools. When the TransPAC4 link between Seattle and Tokyo went live in May, we began collecting flow data from Seattle as well.

Over the summer, work with AMLight flow data began, with several different technical approaches being considered. Currently, data is being collected, however the timing data is incomplete due to a problem with the data collection by the Brocade router. A software workaround will be implemented in Year 3 Q1.

5.7 Tstat Data Collection from DTNs

In addition to collecting Tstat by mirroring traffic on IRNC backbone routers, we are also experimenting with collecting Tstat directly on a number of Data Transfer Nodes (DTNs) directly. As not all DTN traffic crosses IRNC links, this will provide additional insight on the overall health and performance of DTN transfers. This Tstat DTN analysis work is being done in partnership with the DOE funded RAMSES project (https://sites.google.com/site/ramsesdoeproject), lead by ANL.

Highlights from Year 2 in this area include:

- In Quarter 1 we installed Tstat on the NERSC production DTNs, and the ESNet test DTNs, and started working on analysis of this results.
- In Quarter 2 we built a data analysis pipeline consisting of the open source tools RabbitMQ, Logstash, and Elastic. We also built a data analysis GUI using Kibana.
- In Quarter 3 we started collecting GridFTP log files as well. We built prototype dashboards in both Kibana and Grafana to compare the strengths and weaknesses of each.
- In Quarter 4 we continued to refine the tools used to collect and analyze Tstat data from the ESnet and NERSC DTNs. We identified and fixed a bug in our

method for computing the packet retransmit rate from the Tstat logs, and refactored how we store the data in ElasticSearch.

At the end of year 2 we are now confident that we have proper infrastructure in place, and enough historical data, to begin serious analysis of that data. LBNL has hired a new post-doc, who will start in July 2017, to work on this. This will be a focus of Project Year 3.

5.8 Science Registry

In Quarter 4 we began the design and initial implement of a science resource registry. This system is used to map IP addresses to specific organizations, research disciplines, research project or type of component. This data allow us to tag flows with metadata before we de-identify them. The user interface design, shown in Figure 1, and the database design were completed and presented at the Berkeley All Hands Meeting in January. In Year 3, we will finish the development and then deploy this as part of our processing pipeline. This work is a joint effort with the Australian network, AARNET.

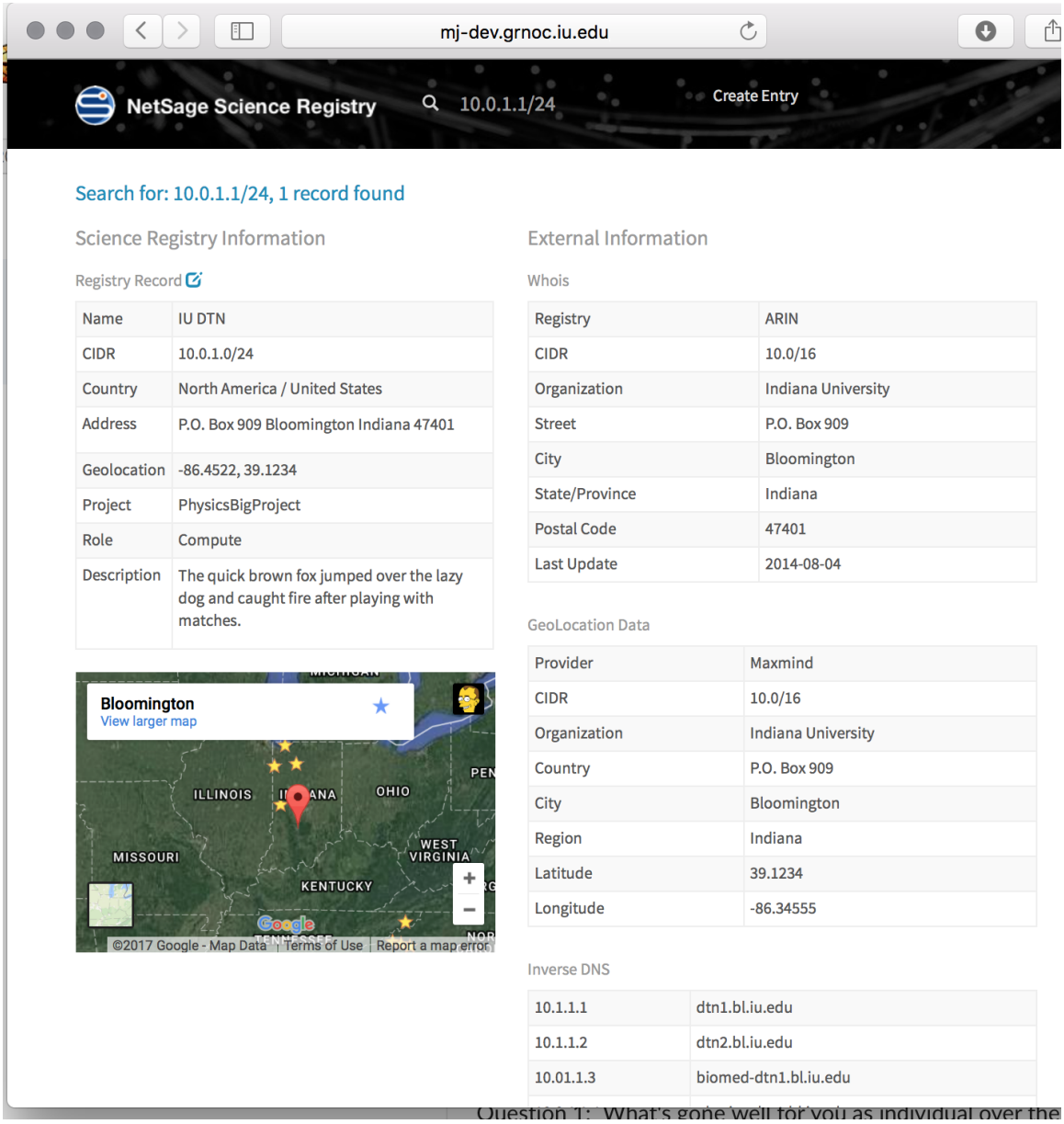

Figure 1: Sample user interface for the Science Registry..

6.Visualization Tools for NetSage

In Year 1, the visualization team defined several NetSage use case queries to determine which were feasible in the short term and in the long term. We used this list to define and prioritize needed data sources and privacy concerns, as well as the visualization approach. A list of these questions can be found at: http://bit.ly/2mrpudp

In Year 2, the first prototype of NetSage that visualizes the data for the first prioritized query (What is the Max, Min, Avg, bandwidth between links?) was completed. The visualization depicted a topology map and additional charts that

show the distribution of bandwidth use and data transferred. Furthermore a capability was developed to enable visualizations to be easily shared with network administrators as URLs to help accelerate troubleshooting.

In the second quarter of Year 2, the visualizations were further enhanced to answer the query "what is the duration and are there any period patterns or peak periods in bandwidth use and network loss across the IRNC links." This was accomplished with a heatmap visualization, shown in Figure 2, which enables network administrators to see patterns quickly at a glance.

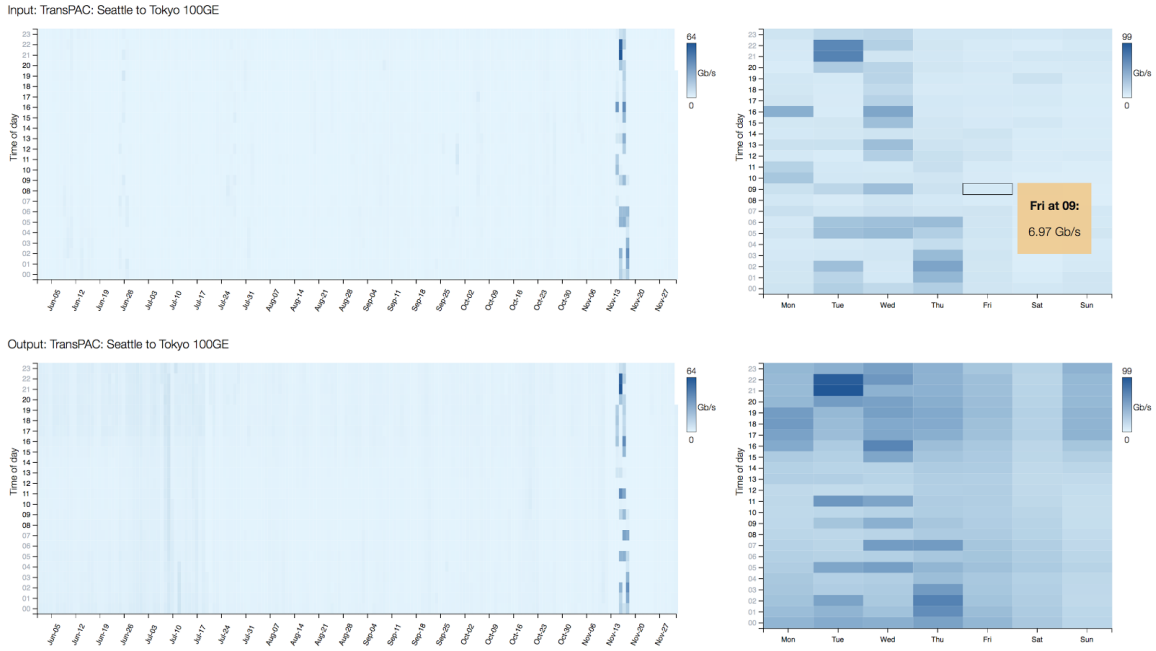

Figure 2: NetSage Heat Map to illustrate period patterns of network bandwidth utilization.

TSDS PerfSonar queries were also integrated to enable the visualization of loss in the networks. Lastly the visualization codes were refactored to incorporate the newest version of D3, which introduced substantial changes in the visualization framework.

In the third quarter, the user-interface was improved to introduce a workflow-like design whereby a user can simply follow a checklist to produce a desired visualization, as shown in Figure 3.

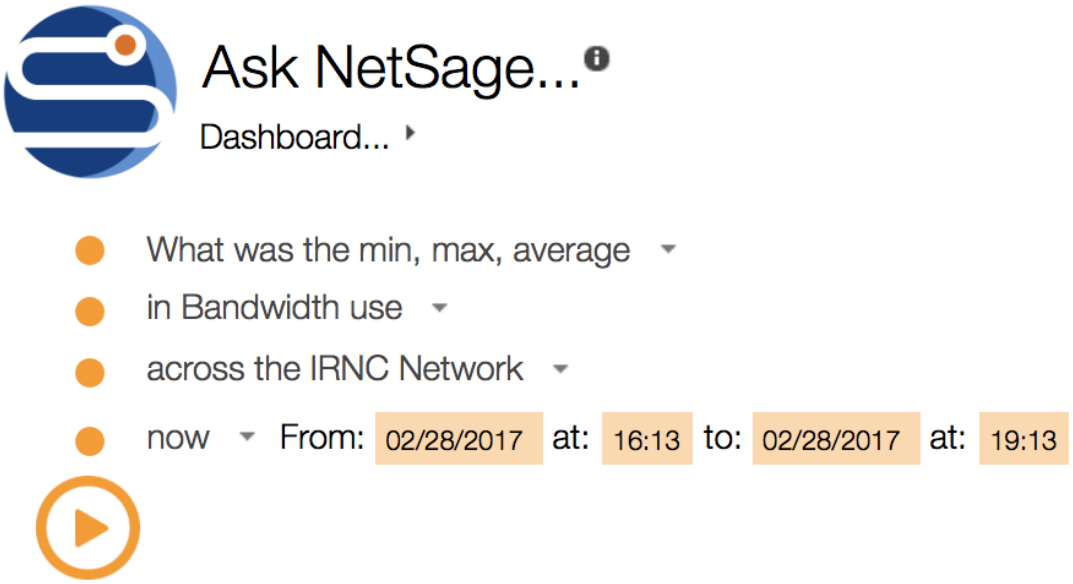

Figure 3: Revised NetSage querying interface providing a workflow-like design.

Furthermore a dashboard has been implemented as the default launch page for NetSage that shows the current state of the IRNC networks over the most recent three-hour period. Lastly a youtube video explaining the current state of the user interface has been produced to begin to collect user feedback. The video can be accessed at: https://youtu.be/Bip-5t-KSLU.

In the fourth quarter the NetSage visualization service moved to a more permanent hosting location on IU's server rather than a developmental Amazon server, in anticipation of supporting more production use of the portal. Improvements were also made to the dashboard by introducing a bandwidth distribution chart to enable quick comparison of bandwidth use across all the IRNC links. Furthermore the prior histogram view was expanded to permit more detailed view of how bandwidth use per link is distributed in a given time interval, as shown in Figure 4.

The latest development version of NetSage can be found here:
http://dev.netsage.global/netsage/dashboard.html

At the January All Hands Meeting, the visualization team met with the developers of my.es.net to begin to share visualization capabilities. With Dugan joining the project, this will also be a focus of Year 3.

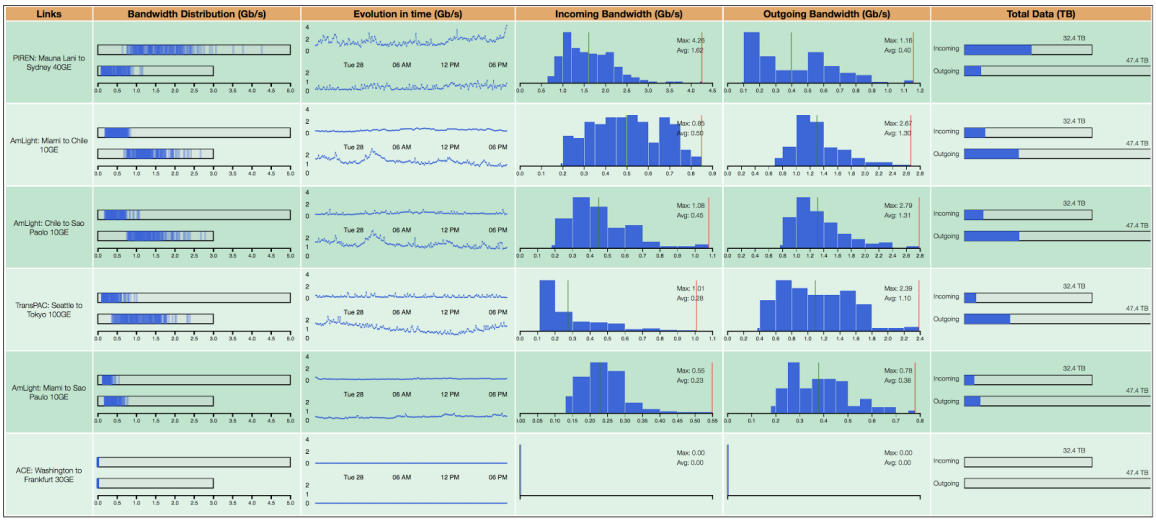

Figure 4: New NetSage Dashboard charts incorporating bandwidth distribution information.

6. Deployment, Data Collection, and Analysis

In Year 1, the project focused on the three IRNC backbone projects, and the collection of SNMP, perfSONAR, and Flow data from them, and by the end of the first project year we had SNMP data from all three backbones and were working toward inclusion of perfSonar data sets. In Year 2, we were successful in gathering SNMP data from all of the backbones as well as perfSONAR data from all the backbones and exchange points. We are also collecting sFlow data from ACE, and TransPAC, as well as Tstat data from ACE.

We have been able to move forward with development of scripts to analyze Top Talker and Largest Data Transfer in a time frame, using data collected from the TransPAC 10G LA-Tokyo link. This work is jointly funded with TransPAC.

In the Summer of 2016, International Networks at Indiana University hired five interns to work on netflow analysis. This included three undergraduate students, from three different universities, participating in the Summer Research Opportunities in Computing (SROC) at IU's School of Informatics and Computing (SoIC), and two graduate students from IU. The main project they worked on jointly involved analysis of netflow data from the IRNC's TransPAC project, including data collected in 2016. Students tried out various analytics tools and finally ElasticSearch, Logstash, and Kibana (ELK) toolset was selected due to its great scalability in large datasets ("big data analytics"). Analysis included searches for largest and longest flows, top talkers and flow profiles for elephant flows. One of the undergraduate students also worked on pattern analysis of BGP routing table archive from the R&E community.

The two MS summer students continued to work with the project to produce prototypes of possible data analysis tools based on TransPAC and ACE data in house. These tools include:

- Processing netflow records that do not contain Layer3 information (source and destination ASnum)
- Generating flow profiles for significant (large) flows from netflow data
- Automation of quarterly Top10 Talker traffic reporting for IRNC projects based on netflow data

In addition, we conducted a scaled trial use of ELK for analysis of individual de-identified flows. For this effort we fed a copy of flow data from our pipeline into a ELK cluster we deployed on the Jetstream infrastructure. The trial was successful, pointing to a possible long term value in using JetStream for certain analytical use cases with clustered systems like ELK. A demonstration dashboard was created and demoed at SC16. This work lead to the refinement of the flow processing pipeline to include ongoing integration of ELK into our processes and tool suite.

NetSage team member Chris Tracy was temporarily pulled into another project, and so the work on capacity planning techniques was pushed back to Year 3. Currently, all of the IRNC backbones are greatly undersubscribed, so this will not impact IRNC planning.

In Year 3, with the addition of a postdoc at LBNL for the purpose of data analysis, we will be expanding our efforts in network data analysis to better characterize patterns in flow data, including baselines, trends, and anomalies.

7. Data Privacy and Security

Basic security measures are being maintained, and there were no security incidents to report for this project year.

At the February All Hands meeting we decided to take an alternative approach. We drafted the needed security documents in house (led by Sean Peisert), and had them reviewed by the other IRNC PIs and outside experts. Specifically, we produced three documents: a Network Data Privacy Policy, to be posted on the project website, and a Project Network Data Collection Memorandum of Cooperation, and a Network Data Retention Policy, all to be agreed to by our project partners.

All drafts have been updated from their initial versions based on feedback from the IRNC PIs. The updated drafts are on the NetSage website at http://www.netsage.global/home/netsage-privacy-policy, and have been shared with the IRNC PIs.

Year 3 Planning

In Year 3 we will take advantage of the collected data, and expand it to include additional Tstat and flow data records for the rest of the backbones, as well as some of the exchange points. These data sources, as well as the fully functional Science

Registry, will enable us to look more in depth at understanding the use of the IRNC resources. Not only will we be able to investigate factors around what types of science are using which links, but we will be able to directly address issues around the largest flows. With the loss data collected by both Tstat and perfSONAR, we will be able to begin to analyze packet loss as well as the role of buffer size for flows that are not performing well.

With John Dugan of my.es.net joining the team, we will investigate ways to take advantage of their visualization tools in addition to our own. We will also work on expanding the current set up to include data from the extended archives used by our ELK set up, instead of only TSDS.

Lastly, we will be exploring ways to generalize this framework for additional deployments. The Advanced Networks for the Atlantic (ANA) consortium, the Global Network Architecture (GNA) for research networking group, and the University of Hawaii have all expressed an interest in possibly deploying NetSage to work with their network infrastructure. This work has the potential to give us a way to continue support for the project past its current funding.

8. WBS

The Table below shows an updated Work Breakdown Schedule (WBS) for Year 3. Updated numbering for the new items is shown. In general, work is on track from the original project plan. Extended tasks are additional detail, fleshing out the preliminary plans.

WBS Year 2	WBS Year 3	Data Collection	% done	FTE Days	Start	End	Notes
1.1	1.1	Prep work for data from backbones	100 %	200d	2/1/16	11/4/16	
1.1.1	1.1.1	Meet with backbone PIs to understand current practices	100 %	200d	2/1/16	11/4/16	
1.1.2	1.1.2	Coordinate with Viz team to understand data needs for viz questions	100 %		ongoing		
1.1.3	1.1.3	Establish what data is to be collected in Year 2	100 %	60d	6/1/16	8/23/16	
1.2	1.2	Prep work for data from exchange points (XP)	98%	400d	10/1/15	4/12/17	
1.2.1	1.2.1	Meet with XP PIs to understand current practices	100 %	400d	10/1/15	4/12/17	
1.2.2	1.2.2	Coordinate with Viz team to understand data needs for viz questions	100 %		ongoing		
1.2.3	1.2.3	Establish what data is to be collected in Year 2	100 %	60d	6/1/16	8/23/16	
1.2.4	1.2.4	Signoff on CENIC for data collection	100 %	5d	8/1/16	8/5/16	
1.2.5	1.2.5	Signoff on Miami for data collection	100 %	1d	8/8/16	8/8/16	
1.2.6	1.2.6	Signoff on StarLight for data collection	25%	10d	10/3/16	10/14/16	
1.3	1.3	Collector set up	100 %				
	1.3.1	Purchase equipment (joint with NOC)	100 %				
	1.3.2	Initial build	100 %				
	1.3.3	Open source TSDS	100 %				
1.4	1.4	PerfSonar Related Tasks	0.9	364d	11/16/15	4/6/17	
1.4.1	1.4.1	Evaluate and refine PerfSONAR to TSDS integration - needs the TSDS collector turned on for the current Maddash	100 %	15d	11/16/15	12/4/15	
1.4.2	1.4.2	Define and deploy new PS test mesh for backbones	20%	30d	2/25/17	4/6/17	
1.4.2.1	1.4.2.1	Input TransPAC LA data	100 %				

1.4.2.2	1.4.2.2	Input TransPAC Seattle data	100%				
1.4.2.3	1.4.2.3	Input Ampath data	100%				
1.4.2.4	1.4.2.4	Input PIREN data	100%				
	1.4.2.5	Make everything not orange	100%				
	1.4.2.6	PS node in ManLan for NEAAR circuit	20%	30d	2/25/17	4/6/17	
1.4.3	1.4.3	define and deploy new perfSONAR nodes for Exchange Points	100%	191d	5/12/16	2/2/17	
1.4.3.1	1.4.3.1	deploy new server to Starlight and add host to mesh	100%	100d	5/12/16	9/28/16	
1.4.3.2	1.4.3.2	Add cenic hosts to PS mesh	100%	100d	9/18/16	2/2/17	
1.4.4	1.4.4	Set up perfSONAR MA at IU for data collection	100%	8d	11/30/15	12/9/15	
1.5	1.5	SNMP related tasks	0.32	523d	11/12/15	11/13/17	
1.5.1	1.5.1	Evaluate and tune SMP to TSDS integration	100%	1d	11/30/16	11/30/16	
1.5.2	1.5.2	SNMP data from Backbones	100%	87d	11/12/15	3/11/16	
1.5.2.1	1.5.2.1	Input Miami SNMP	100%				
1.5.2.2	1.5.2.2	Input ACE WIX SNMP	100%				
1.5.2.3	1.5.2.3	Input Hawaii SNMP Data	100%	20d	11/12/15	12/9/15	
1.5.2.4	1.5.2.4	Input Transpac 100G link	100%	5d	3/7/16	3/11/16	
1.5.3	1.5.3	SNMP data from Exchange points	4%	217d	1/15/17	11/13/17	
1.5.3.1	1.5.3.1	Input Starlight SNMP data	0%	20d	10/17/17	11/13/17	
1.5.3.2	1.5.3.2	Input MIami XP SNMP data	5%	30d	1/15/17	2/23/17	
1.5.3.3	1.5.3.3	Input CENIC XP SNMP data	5%	30d	3/17/17	4/27/17	
1.6	1.6	General FlowData Initial Handling	0.36	363d	3/14/16	8/2/17	
1.6.1	1.6.1	TSDS FLow: add sparse storage model to TSDS for flow data with per flow metadata tagging	100%	40d	3/14/16	5/6/16	
1.6.2	1.6.2	TSDS Flow: non aggregated, first Proof of concept	100%	11d	5/9/16	5/23/16	
1.6.3	1.6.3	TSDS Flow: storage model stress testing and refinement	80%	75d	5/24/16	9/5/16	
1.6.4	1.6.4	TSDS Flow: design refinement and scaling	80%	30d	8/1/16	9/9/16	
1.6.5	1.6.5	Pipeline: design modular de-identification rabbitmq based processing pipeline	100%	25d	3/14/16	4/15/16	

1.6.6	1.6.6	Pipeline: reference flow de-identification processor	100 %	7d	5/24/16	6/1/16	
1.6.7	1.6.7	TSDS Flow: flow stitching for both histograms and flows over a date line	100 %	16d	5/24/16	6/14/16	
1.6.8	1.6.8	TSDS Flow: flow stitching for parallel flow stitching, ex: gridftp - use HINTES heuristics	0%	284d	7/1/16	8/2/17	
1.6.9	1.6.9	TSDS Flow: flow stitching phase 2 (YEAR 3)		100d	3/3/17	7/20/17	
NEW	1.6.10	Make Tstat easier to install as a product (also monitoring tstat process)	100 %	45d	8/15/16	10/14/16	
1.7	1.7	Tstat/Flow deployment	42%	484d	11/30/15	10/5/17	
1.7.1	1.7.1	tstat logs into pipeline	100 %	15d	2/15/16	3/4/16	
1.7.2	1.7.2	Write up a document for De-Identification pipeline including data delete on source, etc	95%	35d	6/20/16	8/5/16	
1.7.3	1.7.3	Bro and Tstat analysis	100 %				
1.7.4	1.7.4	evaluate tstat for scalability (TCP retransmit)	100 %				
1.7.5	1.7.5	Develop initial config for Bro	100 %				
1.7.6	1.7.6	Develop initial config for tstat	100 %				
1.7.7	1.7.7.1	configure tstat at TransPac LA	100 %	15d	11/30/15	12/18/15	
1.7.15	1.7.7.2	Input TP Seattle Sampled Flow Data into TSDS	100 %	15d	8/8/16	8/26/16	
1.7.8	1.7.8	Input TP Seattle UNsampled (TSTAT) Flow Data into TSDS	50%	45d	12/15/16	2/15/17	
1.7.9	1.7.9	Input ACE/WIX Flow Data-Sampled	100 %	15d	6/2/16	6/22/16	
1.7.10	1.7.10	Input ACE/WIX Flow Data-UNSAMPLED using TSTAT	100 %	15d	6/2/16	6/22/16	
1.7.11	1.7.11	Input Ampath Flow Data	0.66	217d	6/15/16	4/13/17	
1.7.11.1	1.7.11.1	Talk to Julio and Jeronimo about Tstat vs sFlow vs AMIS data being available (SFLOW feed over ipsec tunnel)	100 %	45d	6/15/16	8/16/16	
1.7.11.4	1.7.11.4	Incorporate sampled flow data into TSDS	75%	75d	11/16/16	2/28/17	
1.7.11.2	1.7.11.2	Evaluate AMIS data vs TSTAT	100 %	40d	9/1/16	10/26/16	
1.7.11.3	1.7.11.3	Purchase and deploy equipment if needed to support TSTAT	10%	30d	3/3/17	4/13/17	
1.7.11.5	1.7.11.5	incorporate unsampled flow data (tstat) from Ampath		30d	3/3/17	4/13/17	
1.7.12	1.7.12	Input PIREN Flow/tstat Data	0.03	93d	1/17/17	5/25/17	

1.7.12.1	1.7.12.1	Talk to Lassner/David Wilde about Tstat and sFlow data being available	50%	1d	1/17/17	1/17/17	
1.7.12.2	1.7.12.2	Purchase/Deploy flow equipment if needed	10%	30d	3/3/17	4/13/17	
1.7.12.3	1.7.12.3	Incorporate sampled flow data into TSDS		30d	2/17/17	3/30/17	
	1.7.12.4	Deploy PIREN tstat data		30d	3/30/17	5/10/17	
	1.7.12.5	Incorporate tstat data into TSDS		30d	4/15/17	5/25/17	
	1.7.13	Input NEAR Flow/tsta Dtata	3%	72d	2/1/17	5/11/17	
	1.7.13.1	Talk to NEAAR team about Tstat and sFlow data being available	50%	1d	2/1/17	2/1/17	
	1.7.13.2	Purchase/Deploy flow equipment if needed	10%	30d	2/15/17	3/28/17	
	1.7.13.3	Incorporate sampled flow data into TSDS		30d	3/1/17	4/11/17	
	1.7.13.4	Deploy NEAAR tstat data		30d	3/15/17	4/25/17	
	1.7.13.5	Incorporate tstat data into TSDS		30d	4/1/17	5/11/17	
1.7.13	1.7.14	Input CENIC Flow Data (Year 3)	0.06	83d	1/17/17	5/11/17	
	1.7.14.1	Talk to CENIC team about Tstat and sFlow data being available	50%	1d	1/17/17	1/17/17	
	1.7.14.2	Purchase/Deploy flow equipment if needed	10%	30d	3/15/17	4/25/17	
	1.7.14.3	Incorporate sampled flow data into TSDS		30d	4/1/17	5/11/17	
	1.7.14.4	Deploy CENIC tstat data		1d	4/15/17	4/15/17	
	1.7.14.5	Incorporate tstat data into TSDS		1d	5/1/17	5/1/17	
1.7.14	1.7.15	Input AMPATH Ex Pt Flow Data (Year 3)		45d	2/1/17	4/4/17	
1.7.15	1.7.16	Input StarLight Flow Data (Year 4)		60d	7/15/17	10/5/17	
	1.7.17	Evaluation of 100G Tstat collection	100%	30d	10/1/16	11/10/16	
1.8	1.8	Instrumentation of DTNs	23%	220d	7/1/16	5/4/17	
1.8.1	1.8.1	Find the DTNs to instrument using top talkers list	10%	20d	8/8/16	9/2/16	
1.8.2	1.8.2	Add in nersc data as a test data set for internal use (?)	60%	45d	7/1/16	9/1/16	
1.8.2	1.8.2	Add in ESNet data as a test data set for internal use (?)	60%	45d	2/2/17	4/5/17	
1.8.3	1.8.3	NCAR DTN as endpoint - Other NCAR Person		45d	8/8/16	10/7/16	
1.8.4	1.8.4	Add tstat to DTN #1		45d	3/3/17	5/4/17	
1.8.5	1.8.5	Add tstat to DTN #2		45d	3/3/17	5/4/17	

1.9	1.9	Input SDN Data (Year 4)		30d	8/15/19	9/25/19	
1.1O	1.1O	Evaluate Argus vs tstat	10%	200d	9/1/16	6/7/17	
NEW	1.11	Examine options for extended data services	100%	345d	10/1/16	1/25/18	
NEW	1.11.1	Evaluate ELK and JetStream for larger faster data stores	100%	30d	10/1/16	11/10/16	
NEW	1.11.2	JetStream account setup	100%	10d	10/1/16	10/13/16	
NEW	1.11.3	Port data to JetStream for ELK work	100%	30d	12/16/17	1/25/18	
	1.12	Additional software framework Upkeep		82d	3/1/17	6/22/17	
	1.12.1	TSDS maintenance		30d	5/1/17	6/9/17	
	1.12.2	Update TSDS for PS 4.0 archiving		30d	3/1/17	4/11/17	
	1.12.3	TSDS-Elk integration		60d	4/1/17	6/22/17	
2	2	Analysis and Exp. Year 2	6%	862d	10/15/15	2/1/19	
2.1	2.1	Topology publication Service	100%	862d	10/15/15	2/1/19	
2.2	2.2	Capacity Planning tools	29%	324d	10/15/15	1/10/17	
2.2.1	2.2.1	Design capacity planning tools	30%	268d	10/15/15	10/24/16	
2.2.2	2.2.2	Implement and refine capacity planning tools-report generation hopefully starting for Q2		7d	1/2/17	1/10/17	
2.3	2.3	Top Talkers scripts	30%	100d	9/12/16	1/27/17	
2.4	2.4	Recreate AS to Science Project Data base (Science Registry)	39%	113d	1/1/17	6/6/17	
2.4.1	2.4.1	Develop basic data base framework	75%	46d	1/1/17	3/3/17	
2.4.2	2.4.2	Make sure framework has easily updated front end		90d	1/15/17	5/18/17	
2.4.3	NOT	Input data from gloriad database	100%	90d	2/1/17	6/6/17	
2.4.4	2.4.3	Update all data		90d	2/1/17	6/6/17	
2.5	NOT	Need to get routing table dumps (Discuss at F2F)	100%	90d	12/12/16	4/14/17	
2.6	2.6	Largest transfer per month analysis	20%	67d	10/10/16	1/10/17	
XXX	XXX	Summer Student work 2016	100%	106d	3/7/16	8/1/16	
2.7	2.7	Tstat Analysis scripts (non flow, retransmits etc)- walk through Kibana experiment	100%	180d	5/12/16	1/18/17	

2.8	Moved to Vuz	Work with Viz guys to adapt TSDS framework for analysis - start with Hans top 10 scripts	100%	90d	6/8/16	10/11/16	
2.9	2.9	Work on caching analysis results in TSDS (storage of derived metrics)	25%	100d	10/10/16	2/24/17	
2.1O	2.1O	Traffic characterization based on HNTES (of Elephant flows) - sort of part of Ed's work too	100%	7d	6/21/16	6/29/16	
		Data cleaning					
2.11	2.11	Flow analysis capability based on ESnet tool (Year 3)		126d	5/1/17	10/23/17	
2.11.1		Design detailed flow analysis capability based on esnet tool		60d	5/1/17	7/21/17	
2.11.2		Implement detailed flow analysis capability		60d	8/1/17	10/23/17	
	2.12	Analysis for packet loss		30d	4/1/17	5/11/17	
	2.13	Analysis for top sites		30d	5/1/17	6/9/17	
	2.14	Analysis for top science projects		30d	6/1/17	7/12/17	
	2.15	Analysis for elephant flows - min, max and duration		30d	7/1/17	8/10/17	
	2.16	Analysis of buffer size issues		30d	8/1/17	9/11/17	
	2.17	Out year analysis projects		522d	1/1/19	12/30/20	
2.12	2.17.1	Troubleshooting tools for NOC		522d	1/1/19	12/30/20	
2.13	2.17.2	Traffic analysis in SDN and multi-tenant networks		522d	1/1/19	12/30/20	
2.14	2.17.3	Use of BGP metrics with analysis of flow systems		1d	1/1/19	1/1/19	
2.15	2.17.4	Evaluation of PS tests and sampling		1d	1/1/19	1/1/19	
2.16	2.17.5	Incorporate BGP information		1d	1/1/19	1/1/19	
2.17	2.17.6	Develop SDN monitoring prototype		1d	1/1/19	1/1/19	
3	3	Visualization Tasks					
3.1	3.1	create a google sheet with all network details needed to generate vis	100%	30d	1/1/16	2/11/16	
3.2	3.2	Integrate TSDS database queries into prototype.	100%	30d	1/1/16	2/11/16	
3.3	3.3	Default Summary View Visualization	100%	184d	2/7/16	10/19/16	
3.4	3.4	Test prototype against real available data (waiting on CORS)	100%	90d	2/12/16	6/16/16	
3.5	3.5	Develop viz prototypes based (flow data)	100%	194d	2/12/16	11/9/16	

3.5.1	3.5.1	Develop hierarchical visualization column interface & widgets	100 %	176d	2/12/16	10/14/16	
3.5.2	3.5.2	Create basics of visualization	100 %	30d	2/12/16	3/24/16	
3.5.3	3.5.3	Research how to save user preferences, visualization configuration...(cookies, login users?)	100 %	15d	2/12/16	3/3/16	
3.5.4	3.5.4	Create initial map view with all the nodes loaded	100 %	26d	9/11/16	10/14/16	
3.5.5	3.5.5	Aesthetic improvements	100 %	20d	5/6/16	6/2/16	
		Develop Visualization of Heatmaps	100 %	94d	7/1/16	11/9/16	
		Develop Bandwidth data visualization	100 %	53d	7/1/16	9/13/16	
		Develop Losses data visualization	100 %	15d	10/11/16	10/31/16	
		Develop Latency data visualization	100 %	15d	10/11/16	10/31/16	
		Stable prototype for SC	100 %	22d	10/11/16	11/9/16	
3.5.6	3.5.6	Bind visualization (interface) to TSDS database (waiting on CORS)	100 %	91d	2/12/16	6/17/16	
3.6	3.6	Define for visualization team https://docs.google.com/a/lbl.gov/document/d/1F4nodKtHQHE3ZKAGmrxz-QxYmQ7FVs0TyCptb6FeodU/edit?usp=sharing	100 %	7d	2/12/16	2/22/16	
3.6.1	3.6.1	Define what an elephant flow is	100 %	5d	2/12/16	2/18/16	
3.6.1.1	3.6.1.1	concatenate TransPac flow data, and generate histogram of that data	100 %	5d	2/12/16	2/18/16	
3.6.2	3.6.2	Define what "loss" is	100 %	7d	2/12/16	2/22/16	
3.6.3	3.6.3	Define what "top talkers" means	100 %	7d	2/12/16	2/22/16	
3.7	3.7	Show prototype to IRNC community & gather feedback (perhaps via Youtube video)	24%	310d	6/16/16	8/23/17	
	3.7.1	Develop video to show to gather feedback	100 %	120d	6/16/16	11/30/16	
	3.7.2	Develop second version of video		30d	4/1/17	5/11/17	
	3.7.3	Kevin Thompson	50%	60d	6/1/17	8/23/17	
	3.7.4	Shawn McGee		60d	6/1/17	8/23/17	
		Find "Shawn McGee" for NOAA/NCAR		60d	6/1/17	8/23/17	
		Chris Rob (NOC PET)		60d	6/1/17	8/23/17	
		NOC rep (Luke? Jent?)		60d	6/1/17	8/23/17	

		Backbone owners		60d	6/1/17	8/23/17	
		Exchange Point Owners		60d	6/1/17	8/23/17	
		CIO/NW Planner (think someone submitting a cc* proposal)		60d	6/1/17	8/23/17	
3.8	3.8	Incorporate feedback and revise visualization		30d	8/24/17	10/4/17	
	3.9	Vizualize top 10 talkers	48%	106d	11/15/16	4/11/17	
	3.9.1	Get examples of current approaches and discuss	100%	10d	11/15/16	11/28/16	
		Visualization flow data top talkers perhaps using myesnet		11d	3/28/17	4/11/17	
	3.1	Visualize largest transfer per month					
	3.11	visualize most used science projects					
	3.12	Viz for packet loss		30d	5/1/17	6/9/17	
	3.13	Viz for top sites		30d	6/1/17	7/12/17	
	3.14	Viz for top science projects		30d	7/1/17	8/10/17	
	3.15	Viz for elephant flows - min, max and duration		30d	8/1/17	9/11/17	
	3.16	Viz of buffer size issues		30d	9/1/17	10/12/17	
	3.17	Test mergepart of myesnet graph inside netsage		6d	3/21/17	3/28/17	
	3.18	Bugs and Fixes		11d	3/8/17	3/22/17	
	3.18.1	Fix the UI when query goes to a new tab, so that the query parameters are reflected in the UI		6d	3/8/17	3/15/17	
	3.18.2	Fix the evolution in time chart so you can just hover your mouse on a specific time and it just highlights the closest data point and values.		11d	3/8/17	3/22/17	
	3.17	Outyear Viz projects	40%	522d	2/2/17	2/2/19	
3.9	NOT	Develop viz prototype for Tstat data (Year 3)	100%	1d	2/2/17	2/2/17	OBE
3.1O	NOT	Refine/refactor Year 3 prototype	100%	1d	2/2/17	2/2/17	OBE
3.11	NOT	Evaluate third set of prioritized queries (Year 3)	100%	1d	2/2/17	2/2/17	OBE
3.12	NOT	Release and evaluate prototype (Year 3)	100%	1d	2/2/17	2/2/17	OBE
3.13	3.17.1	Dev viz prototype using longitudinal data analysis (Year 4)		1d	2/2/18	2/2/18	
3.14	3.17.2	Refine/refactor Year 4 prototype		1d	2/2/18	2/2/18	

4.15	3.17.3	Evaluate next set of prioritized queries (Year 4)		1d	2/2/18	2/2/18	
3.16	3.17.4	Release and evaluate prototype (Year 4)		1d	2/2/18	2/2/18	
3.17	3.17.5	Refine previous prototype, bug fixing (Year 5)		1d	2/2/19	2/2/19	
3.18	3.17.6	Release and evaluate final (Year 5)		1d	2/1/19	2/1/19	
3.19	3.19	Kabana view prototype	100%	30d	7/25/16	9/2/16	
4	4	Project Coordination	1%	301d	12/9/15	2/1/17	
4.1	4.1	Project management and coordination					
4.1.1	4.1.1	Weekly project meetings			ongoing		
4.2	4.2	Coordinate with NOC	100%	301d	12/9/15	2/1/17	OBE
4.3	4.3	AMIS coordination	100%	143d	10/10/16	4/26/17	
4.3.1	4.3.1	Help with TSDS integration	100%	90d	10/10/16	2/10/17	OBE
4.3.2	4.3.2	Integrate other data with TSDS (Year 3)	100%	60d	2/2/17	4/26/17	OBE
	4.3.3	Assist in close out if requested					
4.4	4.4	InSight / Gloriad coordination	100%	173d	10/10/16	6/7/17	
4.4.1	4.4.1	Evaluate integration of data sources	100%	60d	10/10/16	12/30/16	OBE
4.4.2	4.4.2	Integrate data if feasible (Year 3)	100%	90d	2/2/17	6/7/17	OBE
	4.4.3	Assist in close out if requested					
4.5	4.5	NetSage Data Privacy Policy	0.99	320d	12/10/15	3/1/17	
4.5.1	4.5.1	Draft web page message	100%		12/10/15		
4.5.2	4.5.2	Draft partner policy	100%				
4.5.3	4.5.3	Get document feedback	100%	83d	2/8/16	6/1/16	
4.5.3.1	4.5.3.1	Get feedback from IRNC PIs	100%	20d	5/5/16	6/1/16	
4.5.3.2	4.5.3.2	Get feedback from Kim Milford	100%	70d	2/8/16	5/13/16	OBE
4.5.3.3	4.5.3.3	Get feedback from Erin from CAIDA	100%				
4.5.4	4.5.4	Final draft of web page message posted	100%	1d	1/1/17	1/1/17	
4.5.5	4.5.5	Final draft of partner policy	100%	1d	1/1/17	1/1/17	
4.5.6	4.5.6	Partner policy agreed to by		1d	3/1/17	3/1/17	

4.5.6.1	4.5.6.1	TransPAC	100%	1d	3/1/17	3/1/17	
4.5.6.2	4.5.6.2	ACE	100%	1d	3/1/17	3/1/17	
4.5.6.3	4.5.6.3	PIREN		1d	3/1/17	3/1/17	
4.5.6.4	4.5.6.4	AMPATH		1d	3/1/17	3/1/17	
4.5.6.5	4.5.6.5	Miami XP		1d	3/1/17	3/1/17	
4.5.6.6	4.5.6.6	StarLight		1d	3/1/17	3/1/17	
4.5.6.7	4.5.6.7	CENIC	100%	1d	3/1/17	3/1/17	
4.6	4.6	Year 1 reporting	100%	1d	2/1/16	2/1/16	
		Y1 annual report (with Q3)	100%	1d	2/1/16	2/1/16	
4.7	4.7	Year 2 reporting	100%	263d	2/1/16	2/1/17	
4.7.1	4.7.1	Updated budgets for Year 2	100%	1d	2/1/16	2/1/16	
4.7.2	4.7.2	Y2Q1 report	100%	1d	5/1/16	5/1/16	
4.7.3	4.7.3	Y2Q2 report	100%	1d	8/1/16	8/1/16	
4.7.4	4.7.4	Y2Q3 report	100%	1d	11/1/16	11/1/16	
4.7.5	4.7.5	Y2 annual report (with Q4)	100%	1d	2/1/17	2/1/17	
4.7	4.7	Year 3 reporting	0%	291d	2/1/17	3/14/18	
4.7.1	4.7.1	Updated budgets for Year 3	0%	30d	2/1/17	3/14/17	
4.7.2	4.7.2	Y3Q1 report	0%	30d	5/1/17	6/9/17	
4.7.3	4.7.3	Y3Q2 report	0%	30d	8/1/17	9/11/17	
4.7.4	4.7.4	Y3Q3 report	0%	30d	11/1/17	12/12/17	
4.7.5	4.7.5	Y3 annual report (with Q4)	0%	30d	2/1/18	3/14/18	
4.11	4.11	Domestic Travel Year 2	100%	39d	2/12/16	4/6/16	
4.11.1	4.11.1	All Hands Meeting - February U Hawaii Manoa	100%				
4.11.2	4.11.2	AMIS PI Meeting (February El Paso)	100%	1d	2/12/16	2/12/16	
4.11.3	4.11.3	May 12,13 2016 All Hands Meeting (Bloomington)	100%	14d	3/20/16	4/6/16	
4.11.4	4.11.4	Global Summit (May Chicago)	100%				
4.11.5	4.11.5	SuperComputing Nov 2016	100%				

4.12	4.12	International Travel Year 2	100 %				
	4.13	Travel Year 3					
	4.13.1	AHM January 2017 Berkeley	100 %				
	4.13.2	AHM July Chicago			7/11/17		
	4.13.3	AHM January 2018 Hawaii			1/15/18		
	4.13.4	I2 global summit April 2017			4/1/17		
	4.13.5	TNC May 2017			5/1/17		
	4.13.6	TechEx Oct 2017			10/1/17		
	4.13.7	SC Nov 2017			11/1/17		
	4.14	Papers Year 3	32%	88d	1/20/17	5/23/17	
	4.14.1	Write NetSage paper for ieee Big Data congress	90%	33d	1/20/17	3/7/17	
	4.14.2	Write NetSage paper for Terena '17		60d	3/1/17	5/23/17	

9. Financial Reporting Details for Year 2

Item	Univ	Feb-16	Mar-16	Apr-16	May-16	Jun-16	Jul-16	Aug-16	Sep-16	Oct-16	Nov-16	Dec-16	Jan-17	**Total**
STAFF COSTS (INCLUDING BENEFITS, F&A)														
Schopf, Jennifer-PI, 15%	IU	1,672	1,672	1,672	1,672	3,134	4,431	4,431	4,431	4,431	4,431	4,431	4,431	**40,842**
Lee, Andrew - Senior Architect, 25%	IU	1,721	1,721	1,721	1,721	3,808	0	0	0	0				**10,693**
Predrag Radulovic - Anlysis 20%							4,567	4,567	4,567	3,045	3,045	3,045	3,045	**25,883**
Balas, Ed - Senior SW Developer, 10%	IU	1,312	1,312	1,312	1,312	1,312	1,346	1,346	1,346	1,346	1,346	1,346	1,346	**15,985**
McCracken, Mitch - Software Developer, 100%	IU	11,880												**11,880**
IU Dev Team	IU		13,616	13,616	13,616	13,616	13,616	13,616	13,616	13,616	13,616	13,616	13,616	**149,774**
Sampathkumar, Abhi	IU					1,147	3,424	2,293	1,985	1,901	1,901	2,590	3,326	**18,567**
Singh, Abhi	IU					1,148	3,445	2,165	1,997	1,901	1,901	2,590	3,326	**18,473**
Interns	IU				15,216	15,216	7,925							**38,358**
Sean Peisert-UC Davis Lead and LBNL Lead, 5-15%	UCD	1,891	1,891	1,891	2,693	2,782	1,457	6,726	2,953	0	1,665	0	9,630	**33,580**
Brian Tierney-Staff Scientist, 5-15%	UCD	3,041	3,041	3,041	8,250	9,402	3,003	12,309	7,521	7,984	7,984	2,302		**67,878**
Jonathan Ganz-GSR, 45% AY, 100% Sumr	UCD	4,532	4,532	4,532	3,890	3,888	6,625							**27,999**
Sean Peisert-LBNL Lead	LBNL											1,892		**1,892**
Brian Tierney-Staff Scientist, 5-15%	LBNL												9,799	**9,799**
Chris Tracy, Engineer, 5%	LBNL											4,077		**4,077**
Monte Goode - software developer, 25%	LBNL						23,580	4,606					1,375	**29,561**
Jon Dugan - engineer, 5%	LBNL												1,982	**1,982**
Alan Whinery (2 months)	**UH**	4,559	4,559	4,559	4,786	4,786	4,786							**28,035**
Alberto Gonzalez (11 month grad student) [Step 13 salary]	**UH**	3,831	3,831	3,831	3,831	3,831	3,831		3,831	3,831	3,831	3,831	3,831	**42,140**
Eric Wu (11 month grad student) [Step 11 salary]	**UH**	3,542	3,542	3,542	3,542	3,542	3,542		3,542	3,542	3,542	3,542	3,542	**38,959**
Leigh, Jason (2 months)	**UH**				14,783	14,783								**29,566**
TOTAL STAFFING		**37,982**	**39,717**	**39,717**	**75,313**	**82,395**	**85,579**	**52,059**	**45,790**	**41,597**	**43,262**	**43,262**	**59,250**	**645,924**

Travel and Other Costs (including overhead)														
Travel - Schopf SC Austin, TX Nov 2015	IU		2,070											**2,070**
Travel - Lee SC Nov Austin, TX Nov 2015	IU		3,590											**3,590**
Travel - Tierney Hawaii AHM Feb 2016	LBNL		2,329											**2,329**
Travel - McCracken Hawaii AHM Feb 2016	IU		1,740											**1,740**
Travel-Lee Hawaii AHM Feb 2016	IU	1,155	774	1,848										**3,777**
Travel - Schopf Hawaii AHM Feb 2016	IU			147										**147**
Travel - Balas Hawaii AHM Feb 2016	IU	622	743											**1,365**
Travel - Schopf I2 May 2016	IU		990											**990**
Dell poweredge r630 server	IU	5,313												**5,313**
CDW - Power strips	IU		121											**121**
Travel - Leigh Indiana	UH				4,638									**4,638**
Travel - Gonzalez Indiana	UH				2,370									**2,370**
Travel - Lee - Terena, Prague June 2016	IU		1,085					5,189						**6,274**
Travel - Tierney - TNC Prague June 2016	UCD				5,066									**5,066**
Dell - 1100 watt power supply	IU			990										**990**
CDW- samsung pro iTB SSD	IU					1,341								**1,341**
Travel - May meeting parking costs	IU					74								**74**
Travel - Schopf- I2 May 2016	IU								1,965	681				**2,646**
Travel - lee - I2 chicago May 2016	IU						323							**323**
Monitor for SC and Demos	IU					2,639		1,214						**3,853**
Intern set up	IU					4,013			1,094					**5,107**
480G drive - amazon	IU						467	1,094		129				**1,690**
Travel - Schopf NSF may	IU						791							**791**
Travel - Tierney - NSF CiCi Oct 2016	UCD										3,668			**3,668**
Travel - Gonzalez - IEEE Viz Oct'16	UH									3,570				**3,570**
Travel - Schopf - SC'17 SLC - Nov. 2016	IU										1,463	2,962		**4,425**
Travel - Balas SC'17 SLC Nov 2016	IU										1,699	994	396	**3,089**
Travel - Gonazlez SC'17 SLC Nov 2016	UH										2,026			**2,026**
Travel - Balas AHM Jan 2017	IU												1,242	**1,242**
Travel - Schopf I2 May 2017	IU												1,089	**1,089**
Misc sw	IU									129				**129**
Napatech gear	IU											549		**549**
Dell server	IU												4,248	**4,248**
TOTAL TRAVEL		**7,090**	**13,442**	**2,985**	**12,074**	**8,067**	**1,581**	**7,497**	**3,059**	**4,509**	**8,856**	**4,505**	**6,975**	**80,640**

PARTICIPANT SUPPORT														
Meeting support for Hawaii AHM				367										**367**
Meeting support for IU AHM						990								**990**
TOTAL PARTICIPANT SUPPORT		**0**	**0**	**367**	**0**	**990**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1,357**
EQUIP OVER $5K														
Testbed Setup		10,380												**10,380**
Server for 100G card eval								11,524						**11,524**
Napatech Card												8,958		**8,958**
Napatech Card												12,949		**12,949**
TOTAL EQUIP		**10,380**	**0**	**0**	**0**	**0**	**0**	**11,524**	**0**	**0**	**0**	**21,907**	**0**	**43,810**
													TOTAL	**771,732**