NetSage
Award #1540933
Year 3 Annual Report
1 Feb 2017 through 31 Jan 2018

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 resources. In addition, this collected data is also made available for use by the NOC for day-to-day operations and support for end-to-end performance troubleshooting. Highlights of Year 3 included a shift in the portal infrastructure from a bespoke framework to Grafana, receiving supplemental funding for additional Tstat deployments on archives, and collection of SNMP and perfSONAR data from all of the resources.

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 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.

NetSage services 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 3 of the project focused on the next stage of analysis based on our additional data sources, including the science registry. Going back to our guiding list of questions (https://docs.google.com/spreadsheets/d/1cnjrYGoUqWXzm-MNz1ly7Bu7yRpCHJU71jSl2OeItS0/edit?usp=sharing), this included:

- #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?

Year 4 will focus on additional analysis of the data being collected, especially flow data, as well as expanding flow data collection across the full set of IRNC backbones and exchange points, and collecting Tstat data from archives.

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

2. Staffing

At the beginning of Year 3, funded project 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 Manoa, co-PI - visualization oversight
- Alan Whinery, UH Manoa, senior personnel - perfSONAR and coordination with backbones
- Alberto Gonzalez, UH Manoa, graduate research assistant - visualization developer
- Eric Wu, UH Manoa, graduate research assistant - visualization developer

During Year 3, there were several staff changes.

At IU, both interns, Sampathkumar and Singh, finished their MS degrees and are now employed in Industry. In addition, Predrag Radulovic decided to pursue other opportunities and left the project in June and In July, IU hired Brian Tierney as a part time consultant, paid through non-NetSage funds, although he will continue to assist with the project no more than 10 hours a month. During Quarter 3, Uwe Dahlman

left for the commercial sector. In August, Ed Moynihan, IU, joined the team to help support the science registry work. Lisa Ensman joined during the summer of 2017 to help with work on the Science Registry.

At LBNL, Brian Tierney retired in June. In July, Monte Goode and Chris Tracy shifted to other projects. In August, Dr. Anna Giannakou, a recent PhD graduate from INRIA Rennes with expertise in security, systems, cloud computing, and networking, joined the project to support LBNL's network analytics efforts in the coming months. In October, Jon Dugan, LBNL, went on paternity leave. In December, Dipankar Dwivedi, with expertise in machine learning, joined in LBNL's network analytics efforts.

At Hawaii, Erick Wu graduated and left in May 2017. Tyson started at UH as a student on the project during Summer 2017.

With these changes, at the end of Year 3, funded staff included:

- Jennifer Schopf, IU, PI - overall project director
- Ed Balas, IU, system architect - collection and reporting
- 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
- Ed Moynihan, IU, Science Registry Data support
- Lisa Ensman, IU, developer - Science Registry
- Johnathan Stout, IU, developer - Science Registry
- Sean Peisert, UC Davis and LBNL, co-PI - security, privacy, design
- Jon Dugan, LBNL/ESnet, senior personnel - monitoring architecture
- Anna Giannakou, LBNL Post Doc, measurement analysis
- Dipankar Dwivedi, LBNL Post Doc, measurement analysis
- Jason Leigh, UH Mānoa, co-PI - visualization oversight
- Alan Whinery, UH Mānoa, senior personnel – perfSONAR, PIREN coord.
- Alberto Gonzalez, UH Mānoa, graduate research assistant - viz developer
- Tyson Seto-Mook, UH Mānoa, graduate research assistant - viz developer

3. Collaborations, Travel, and Training

NetSage staff participated in various meetings to support ongoing deployment, collaboration, and training. Note that several of these were funded by other sources but relevant to NetSage. Details for Q1-3 meetings are given in those quarterly reports, and these included:

- Schopf attended the CENIC quarterly meeting, March 19-23, University of California San Diego, http://www.cvent.com/events/the-right-connection-cenic-2-0/archived-fc937869fad34c9baccb10ffc904d910.aspx

- Schopf, Tierney, Balas, and Whinery attended the Internet2 Global Summit, April 23-26, Washington DC, https://meetings.internet2.edu/2017-global-summit/
- Schopf attended TNC17, May 29 - June 2, Linz, Austria https://tnc17.geant.org/
- Schopf attended PEARC17, July 9-13, New Orleans, LA https://www.pearc17.pearc.org/
- Schopf and Moynihan attended Earth Science Information Partners (ESIP) Summer Meeting 2017, July 25-28, Bloomington, IN http://www.esipfed.org/meetings/upcoming-meetings/esip-summer-meeting-2017
- Gonzalez attended IEEE Big Data Hawaii, June 25-30, Honolulu, Hawaii, http://www.ieeebigdata.org/2017/
- Schopf attended National Research Platform Workshop, August 7-8, Bozeman, MT http://prp.ucsd.edu/events/the-first-national-research-platform-workshop
- The full team participated in an All Hands Meeting, Chicago, IL, August 10-11
- Schopf spent a day at NSF in late September briefing the program office on the status of NetSage and related projects.
- Schopf attended the Internet2 Technical Exchange, October 15-18, San Francisco, CA https://meetings.internet2.edu/2017-technology-exchange
- Schopf, Balas, Leigh, and Gonzalez attended SuperComputing '17, November 12-17, Denver https://sc17.supercomputing.org/. They participated in an IRNC PI Meeting jointly hosted by the IRNC NOC and NetSage projects. They also briefed NSF program officers on the NetSage project and progress. NetSage was also part of the SCiNet Monitoring display.
- The full team participated in an All Hands Meeting, January 18-19, Honolulu, Hawaii, where planning for Year 4 took place.

4. Project Coordination

Internal project coordination continued with weekly meetings of the majority of the team. We have also implemented a weekly technical call to be able to dive-down into more detailed topics with those NetSage members who are interested.

Two All Hands Meetings were held across the full team, August 10-11, 2017, in Chicago, and January 18-19, 2018, in Honolulu. The August meeting focused on the portal shift from the bespoke framework it was prototyped with to a Grafana-based framework, which was then demonstrated at SC'17. The January meeting focused on planning for Year 4, primarily agreeing to next steps in analytics and visualization using the available flow data and expanding the Tstat archive deployments.

Work with the IRNC-funded backbones continued, and we now have SNMP and perfSONAR data from all of the original circuits for the IRNC backbone projects.

Sampled and Tstat flow data is being collected from NEAAR and TransPAC, and AmPath is contributing sampled data. In Year 4 we expect to expand Tstat and flow data collection to PIREN. We also need to set up SNMP and perfSONAR data for the two new PIREN links that were stood up at the end of Year 3, one from Honolulu to LA and the other from Honolulu to Guam.

In Quarter 4 we finally were able to get SNMP and PerfSONAR data for StarLight, so we now have full coverage of these datasets for all of the IRNC exchange points. CENIC and AmLight are also sharing sampled flow data. We are discussing Tstat deployments with both groups, and will begin discussions with StarLight to collect flow data in Year 4.

We continue ongoing discussions and coordination with the IRNC NOC. They are taking the lead at collecting SNMP data, for example. Both teams continue to use the shared archive. In Year 4 we will also begin investigation of alarming and alerts on the NetSage data to support NOC activities.

At the end of Year 3 we received supplemental funding to support additional deployments of Tstat sensors on archives associated with the IRNC partners. This is a focus area for Year 4.

5. Data Collection

NetSage staff are involved in the development and deployment of various pieces of software to support active and passive measurements, monitoring, archiving, analysis, and visualization. This section details those projects.

5.1 System Architecture

The main change to the system architecture in Year 3 was the shift the visualization front end from a java script based approach to a Grafana based approach. This now enables us to use multiple backend databases more easily, expanded the visualization and analysis team to include additional staff members as part of the development team, and gave us an authentication framework to use as we move forward with work with the flow data. While this delayed some analytical work, the benefits far outweighed that concern.

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 time series 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 session of a particular ASN.

In Year 3, the primary focus of TSDS development involved exploring and developing an integration plugin with Grafana. It showed us that using Grafana was feasible and would allow us to create dashboards that combine flow, SNMP, and perfSONAR data in the same visual space with a reduced dependence on custom software. The initial prototype was enhanced, improving stability and adding additional query expressibility, and release on the NetSage github code repository: https://github.com/GlobalNOC/tsds-grafana). Four releases were drafted and deployed to the NetSage systems.

TSDS version 1.5.4 was released, which included several bug fixes for issues identified during the integration process and a new aggregation feature (https://github.com/GlobalNOC/tsds-services/releases/tag/1.5.4). TSDS version 1.6.0 was released in November to support Centos7 deployment.

Work continued through the end of the project year to fix bugs and expand the feature set as new needs were discovered. We expect similar development to continue in Year 4.

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.

Discussions with CENIC were initiated in Quarter 1 in conjunction with the IRNC NOC. A plan for tagging which interfaces to collection on was worked through by CENIC and IU. A technique for controlling the collection of interface data based on interface description tags was put into production. SNMP collections were successfully added in August and incorporated into visual elements. StarLight data was successfully added in November.

In Year 4, we will work with the IRNC NOC to collect SNMP data for the two new PIREN circuits that went live in 2018.

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 uses perfSONAR for its active measurements of bandwidth and throughput, and archives them in the NetSage archive using TSDS.

In Project Year 3, we continued to expand the coverage of the test mesh available at http://data.ctc.transpac.org/maddash-webui/index.cgi?dashboard=IRNC%20Mesh. Additional tests were added to cover new 100G links in AMPATH.

Team members were also involved in the production perfSONAR consortium, which oversees the production development and support of the perfSONAR toolkit. June 7-10 was the annual PS face-to-face meeting with the Developer Team in Ann Arbor, MI to assist in plotting the course of development for the coming year. Team members also made contributions to the 4.0 and 4.0.1 releases of the software suite, including rolling out the software to deployments involved in NetSage.

We identified and resolved performance issues with tests from our New York test point to London. For the New York to London tests, a routing issue was identified and resolved. For the Starlight node, we continue having issues with the node deployment and do not consider this a fully functional site yet. We had to upgrade the port being used from 1GE to 10GE, and we then had to correct the host OS level tunings.

We also worked with the AmLight team on known long standing performance issues related to a few of their aging perfSONAR nodes. Several calls to go over their current state and possible recommendations were shared. Work is ongoing with the Hawaii team to stabilize their test points.

In Year 4, additional testing paths will be added as well as a node in Guam to provide coverage of the new PIREN circuits. In addition, we will develop a version of the map dashboard that uses perfSONAR data instead of SNMP data to visualize this data.

5.5 Tstat on Backbones and Exchange Points

Tstat (http://tstat.polito.it/) is part of the EU Measurement 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, including congestion window size and number of packets retransmitted, which can be used to analyze the health and performance of the link.

In Year 3, we deployed two 100G taps with associated sensors, one into New York for NEAAR and the other into Seattle for TransPAC. This data is now being collected and will be used within the Grafana visuals. In Year 4, we hope to extend the Tstat deployments to other IRNC-funded backbones, as well as CENIC and StarLight.

5.6 NetFlow and sFlow Data Collection from Backbones and Exchange Points

In addition to Tstat's enhanced flow data collection, we also collect normal router-based sampled NetFlow or sFlow data, depending on what the router supports. This allows us to compare the results of Tstat-based header analysis with sampled flow data from routers.

In Year 3, we resolved issues related to AmPath flow data integrity. We expanded coverage to add CENIC sflow data collection. CENIC support required a new feature where by a connector can control which flows are retained by adding an interface description tag to their router configuration. In Year 4, we will continue our work with PIREN for flow data collection as well as starting a conversation about this with the StarLight team.

5.7 Tstat Data Collection from Archives

In addition to collecting Tstat by mirroring traffic on IRNC backbone routers, we are also experimenting with collecting Tstat directly on a number of archives. This will provide additional insight on the overall health and performance of data transfers. This Tstat archive analysis work is being done in partnership with the DOE funded RAMSES project (https://sites.google.com/site/ramsesdoeproject), lead by ANL.

In Year 3, we continued to collect Tstat data from NERSC and ESnet archives, and did an assessment of the overhead of these deployments, which was very low. We also submitted a supplement request to expand the data collection from additional archives across the IRNC projects. In Year 4, we will expand the deployment of Tstat to additional archives, including those associated with UH Mānoa, CENIC, and PRP. This is part of the work that was defined in our supplementary funding request.

5.8 Science Registry

A key to understanding the science use of the IRNC links is having a Science Registry that maps IP addresses to AS Numbers, universities, science project, and science domain. The science registry is a key part of the de-identification pipeline so that we can tag flow data without retaining personally identifiable information (PII) in the form of IP addresses.

At the start of Year 3 we stood up the first beta version of the registry. Data entry began in the late June timeframe using TransPAC top talker data to drive the data entry workflow. Several bug fixes and enhancements were made to the registry software, resulting in four releases. Team members are continuing to refine the data entry process. In Year 4, we will expand this data to include sites for NEAAR, CENIC, and AmLight/AmPath.

5.9 Additional Data Sources in Year 4

In Year 4, the project will investigate the inclusion of information from actual data transfers as an additional data source to be added to the testpoint. Currently, some of the Pacific Research Platform (PRP) are using Fiona nodes to understand larger-scale gridFTP data transfer behaviors, which are then displayed in a mesh, similar to that used with standard perfSONAR measurements. We will explore expanding the current set of perfSONAR tests (using the pScheduler extension) as well as possible use of the FIONA software to begin gathering data from actual file transfers

In our original project plan, we had proposed to do monitoring for SDN starting in Year 4. However, with the current state of SDN, and lack of production support for a

standard implementation, this will not make sense. The primary SDN use cases are simple Layer2 circuit setup, which is already captured in the current tools.

6.Visualization Tools for NetSage

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, as was a visualization 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." These two core pieces use only SNMP and perfSONAR data, and the current version is available at http://portal.netsage.global

Year 3 began with creating a number of prototype visualizations for flow data, such as Sankey diagrams and Chord diagrams. The team also transitioned to using Grafana as the new visualization framework in the hope that this will help with the longer term maintainability of NetSage. As with any generic framework, there are challenges that need to be overcome to customize it to create the types of charts that were present in the previous version of NetSage. An early prototype was demonstrated at SC'17. The visualization team also prepared guidelines to help the full NetSage team to become more aware of how to construct good visualizations when using Grafana.

In Year 4, work will focus on innovative displays of flow behaviors. A soft release of flow dashboards will be made for the IRNC project staff at the Internet2 Global Summit in May, after which we will take their feedback and do a more complete release of this data. We expect a full release including editing rights for the PIs by mid-2018. We will also adapt the current map visualization to accept other data, beginning with bandwidth data from perfSONAR. And we will continue the adaption of Sankey graph techniques for the Grafana framework and as a novel way of showing how data flows over the IRNC networks.

7. Data Analysis

In Year 3, the largest shift for the data analysis work was in our shift to using Grafana. This tool simplifies the types of queries that can be made on the databases, and blurs the line between the analysis of the data and its visualization. Some initial prototypes were presented to the IRNC PIs at the Internet2 Global Summit meeting.

7.A. Flow Analysis

We have begun to focus on different types of high-level analysis that can be done using the collected flow data. A key feature we want to leverage is the ability for people to experiment without those visuals becoming immediately public as at times we can create misleading or inaccurate visuals. We plan an initial release of the flow data visualizations (behind an authentication wall) for Year 4 Quarter 1. We

will receive feedback from the IRNC PIs at the May Internet2 Global Summit meeting, and then go forward with an extended release of these dashboards shortly after that.

7.B. Archive Tstat Analysis

In Year 3, we continued to analyze Tstat data from the NERSC and ESnet archives. Analysis initially focused on the top 10 links that demonstrated the highest retransmission percentages per data node and then expanded to include all links that experience packet loss and with flow sizes larger than 500 MB. We are currently developing a lightweight method for packet loss prediction. Our method originally deployed a support vector algorithm and incorporated a combination of path-related properties (such as round trip time) along with host related properties (such as tcp congestion window). Most recently, we shifted to a random forest technique, which is showing higher accuracy. We are currently evaluating the capabilities of our predictor using different combination of tstat fields. The final set of fields that we include are: throughput, size of the file being transferred, source and destination IP addresses, round trip time, duration of the flow, and TCP congestion window.

7.C. Other Analysis

In addition, in Year 4 we will be working with the IRNC NOC to define a series of predictions and conditions to include alarm and alert capability. We will work with several smoothing algorithms to see if they can assist with some of the visualization approaches. And we will also begin work comparing the sampled sFlow or NetFlow collected with the Tstat data also collected for the same resource.

8. Data Privacy and Security

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

The current security documents have been accepted by the IRNC PIs. We are working on additional documentation and planning for an expanded portal, including identification of data that the PIs may prefer to be behind an authentication wall. As a reminder, NetSage does not collect PII.

For the dashboards, with the inclusion of flow data, we are integrating Shibboleth into Grafana to enable 3rd party authentication to the dashboard that may collect sensitive information. This extension will then be incorporated back into the open source Grafana code base.

9. Response to EAB Review

We thank the review panel for their thoughtful comments. Our response to the reviewer suggestions are included below the italicized recommendations language from the feedback document.

1. *The archive created by the project is a great resource. Can it be leveraged to do more, such as provide alerts and notifications to participating networks?*

The original project plan included investigation into alerts and notifications to start in Year 4, which begins February 1, 2018. This is currently part of the updated WBS included in the annual report.

2. *The panel encourages the team to think how can the current methodology migrate to future networks where virtualization will be increasingly used.*

The NetSage project is unclear as to the review panel's use of the concept of a virtual network. If the panel means use of networks to support virtual resources, such as cloud access, the current IRNC links are not used for cloud access as that is done locally not via international backbones. Furthermore, testing against endpoints (such as perfSONAR requires) can not function in a cloud environment, as explained during the review.

If the panel means tracking the use of VLANs, we are collecting SNMP data for VLANS already part of the NetSage infrastructure. It is unclear about what modifications are being requested. It is possible that some VLANs could be part of the science registry and we could analyze that data, but this was not part of the original science registry plan.

3. *The panel encourages the team to investigate new data sources that can be added to the project. There are many other data plane measurement projects in the community including the ARK project at CAIDA (KC Claffy), the ANT project at ISI (John Heidemann), the RIPE ATLAS project in Europe, and the Internet Atlas at Wisconsin (Paul Barford). The panel feels that project should carefully investigate such community projects to determine whether there are opportunities for data, tool or methodology sharing.*

The NetSage project reviews available networking data sources on an ongoing basis for consideration of inclusion in the project. For example, in addition to the original scope, we now include Tstat data for archives, and are working toward the inclusion of FIONA node GridFTP testing data in Year 4. That said, the data sources the panel is recommending are targeted mostly for commercial broadband services. The probes involved in these projects are built with Raspberry PI hardware, which cannot test at more than 1G. They also require AC power and ethernet plugs, which are not available at the IRNC exchange points or network endpoints.

Alternatively, it is possible that the review panel meant to suggest that we incorporate existing probe data from these projects into our visualizations. For

example, the RIPE data archive has a public API, and collects/publishes loss data on commercial broadband networks where it is deployed. However, in that the probe data is collected for commercial broadband (non R&E) networks, and we only display data associated with the IRNC-funded resources, this is clearly out of scope for the current project. A project to do this work would require substantially more resources in terms of staffing and equipment than the current project.

4. *Control plane data sources form projects such as Route Views and BGPMon should also help detect routing anomalies and track/verify IP prefixes advertised by the science ASes.*

Control plane data sources are of interest to many projects when routes change. However, the IRNC-funded network infrastructure monitored by NetSage does not have routing changes. For example, the NEAAR 100G trans-Atlantic link has a route from New York to London that is fixed. Any changes are due to circuit cuts and temporary route-arounds, are extremely rare, and are reversed as soon as possible.

5. *The passive DTN data collection presents a great opportunity for the project. While it acknowledges that it is hard, the panel encourages the PIs to continue to pursue this direction. Instrumenting the LSST nodes for example would be a good start and will provide valuable information from a major IRNC application.*

In Year 3, the NetSage project requested each of the backbone and exchange point projects to share a list of their most used archives and their contact points at those locations to assist with additional Tstat deployments on archives. The response rate on this has been very low, but efforts will continue in this area.

6. *The panel recommends that the project continue to pursue more integration with the PET Team.*

The NetSage team remains fully committed to interactions and support for the PET.

7. *The project should investigate how to share data with the community. Data related to security research can be collected, classified and shared through programs such as IMPACT (www.impactcybertrust.org)*

The NetSage PIs agree that the current datasets could have significant value to the research community if shared. However, this was explicitly excluded from the project scope in the proposal due to likely privacy concerns by the network and exchange point owners. As such, it is also excluded from the current privacy agreements with the other IRNC PIs. We are open to having conversations with the PIs to adapt this policy, however we have been informed already by some partners that due to concerns related to the privacy laws in other countries, it is unlikely that this sort of data sharing will be acceptable.

9. Full Plans for Year 4

Year 4 plans include work in additional data collection, analysis and visualization, as well as ongoing upkeep and maintenance of the current software products in use. For each activity we cross reference the section of this report where it is discussed and the WBS number for Year 4 (see Section 10).

9.A Data Collection

In Year 4 we will try to get all four data types (SNMP, perfSONAR, flow data, Tstat data) from all of the IRNC funded backbones and exchange points. This will include:

- Set up all data for two new PRIEN links (Guam-Honolulu; Honolulu-LA) (SNMP – Section 5.3, WBS 1.5.2.5 and 1.5.2.6; perfSONAR – Section 5.4, WBS 1.4.2.10 and 1.4.2.11; Flow data – Sections 5.5 and 5.6, WBS 1.7.12.6 and 1.7.12.7)
- Add Tstat/Flow for PIREN (Section 5.5, WBS 1.7.12)
- Add Tstat for Miami (Section 5.5, WBS 1.7.11)
- Add Tstat/Flow with StarLight (Sections 5.5 and 5.6, WBS 1.7.16)
- Add Tstat for CENIC/PRP (Section 5.5, WBS 1.7.14)

In addition, we will have staff members work with the other IRNC funded projects to expand the data included in the Science Registry. In the current prototype, we only have data on projects and institutions from TransPAC, so we will extend this NEAAR, AmLight, and CENIC to expand the usability of the analysis (Section 5.8, WBS 2.4.3).

There will be a focus on deploying Tstat testpoints on archives across the project (Section 5.7). This will start with the data archives hosted by the University of Hawai'is Institute for Astronomy (WBS 1.8). We will also work with both CENIC and PRP to instrument their archives (WBS 1.7.14). In addition, we will continue to work with the other IRNC-funded projects to identify highly used archives that can be instrumented.

Finally, we will investigate the feasibility of including file transfer time data similar to what is being collected by the FIONA nodes used in the PRP monitoring system (Section 5.9, WBS 1.15)

9.B Analysis and Visualization

We are now grouping analysis and visualization together, as with Grafana the analysis is primarily done through queries that result in different visuals of the data.

At the end of Year 3, we now have access to flow data from NEAAR, TransPAC, AmLight, and CENIC. These data sets (and any additional flow data collected during the year) will be the corner stone to the analysis and visualization work in Year 4. We will have a set of dashboards for top ten talker data with a small release with limited functionality before the Internet2 Global Summit, being held on May 6-10 (Section 6, WBS 2.13). The feedback from the IRNC PIs at this meeting will be used

to develop a more extensive release, which will include dashboards that the PIs can adapt the queries more readily later in the year (WBS 2.24).

Part of our original plan included working with the IRNC NOC to define some alarms and alerts using data collected by NetSage (. The most challenging part of this work will be in determining predictions of behaviors, and then thresholds to alert on when those predictions change. We will meet with the IRNC NOC in Quarter 1, and start work on this shortly after (Section 7.C, WBS 2.24).

There will be some solely visualization work in Year 4 as well (Section 6). This will include adapting the current portal's map widget to view perfSONAR data instead of SNMP bandwidth data (WBS 3.12). We are also investigating the use of Sankey graphs to see how data is flowing between institutions (WBS 3.20).

Effort will also be spent on some more in-depth analysis projects. For example, Giannakou and Dwivedi are conducting research and analysis to predict retransmit behavior on archives, and to better understand what factors play a role in being able to predict these behaviors. Research includes experimentation with a variety of features and machine learning techniques (Section 7.B, WBS 3.16).

We also plan to look into understanding any differences we see in the collected sampled flow data from routers as opposed to the unsampled Tstat data (Section 7.C, WBS 2.27), and the effect of applying smoothing techniques to raw data on prediction accuracy (WBS 2.22). The results of theses analyses will also be integrated into the NetSage portal, and we will investigate additional methods for displaying analysis and prediction results.

9.C Ongoing support

Effort will also be spent in Year 4 to maintain support for all of the current deployments, both in terms of hardware and software (WBS 1.12.1). Additional releases of the TSDS archive are expected. The IRNC perfSONAR MadDash will continue to be supported jointly between NetSage and the IRNC NOC PET team (Section 5.4, WBS 1.4). And we will incorporate some of our Grafana extensions back into that open source project (Section 5.4, WBS 1.4.3). For example, we have are adding in expanded authentication support that includes Shibboleth (Section 8, WBS 3.11.3).

10. WBS For Year 3-4

The table below includes the work completed in Year 3 and the planned work for Year 4. WBS numbers are included for the items that were included preciously (WBS Year 3) and added for the new items (WBS Year 4). Following this report, we will remove references to all items completed prior to Year 4.

Item	Y3 WBS	Y3 WBS	Notes
Data Collection	1	1	
Prep work for data from backbones	1.1	1.1	Completed
Meet with backbone PIs to understand current practices	1.1.1	1.1.1	Completed
Coordinate with Viz team to understand data needs for viz questions	1.1.2	1.1.2	Completed
Establish what data is to be collected in Year 2	1.1.3	1.1.3	Completed
Prep work for data from exchange points (XP)	1.2	1.2	Completed
Meet with XP PIs to understand current practices	1.2.1	1.2.1	Completed
Coordinate with Viz team to understand data needs for viz questions	1.2.2	1.2.2	Completed
Establish what data is to be collected in Year 2	1.2.3	1.2.3	Completed
Signoff on CENIC for data collection	1.2.4	1.2.4	Completed
Signoff on Miami for data collection	1.2.5	1.2.5	Completed
Signoff on StarLight for data collection	1.2.6	1.2.6	Completed
Collector set up	1.3	1.3	Completed
PerfSonar Related Tasks	1.4	1.4	Ongoing
Define and deploy PS test mesh for backbones	1.4.2	1.4.2	Ongoing
Input TransPAC LA data	1.4.2.1	1.4.2.1	Completed
Input TransPAC Seattle data	1.4.2.2	1.4.2.2	Completed
Input Ampath data	1.4.2.3	1.4.2.3	Completed
Input PIREN data	1.4.2.4	1.4.2.4	Completed
Make everything not orange	1.4.2.5	1.4.2.5	Completed
PS node in ManLan for NEAAR circuit	1.4.2.6	1.4.2.6	Completed
define and deploy new perfSONAR nodes for Exchange Points	1.4.3	1.4.3	Completed
deploy new server to Starlight and add host to mesh	1.4.3.1	1.4.3.1	Completed
Add cenic hosts to PS mesh	1.4.3.2	1.4.3.2	Completed
Set up perfSONAR MA at IU for data collection	1.4.4	1.4.4	Completed
Add PS node in Tokyo		1.4.2.7	Completed
Add PS node in London		1.4.2.8	Completed
Verify PS node in Miami		1.4.2.9	Completed
Add in node for Honolulu-LA link		1.4.2.10	Year 4

Add in node for Honolulu-Guam link		1.4.2.11	Year 4
Ongoing support for IRNC PS mesh		1.4.3	Ongoing
SNMP related tasks	1.5	1.5	Ongoing
Evaluate and tune SNMP to TSDS integration	1.5.1	1.5.1	Completed
SNMP data from Backbones	1.5.2	1.5.2	Ongoing
Input Miami SNMP	1.5.2.1	1.5.2.1	Completed
Input ACE WIX SNMP	1.5.2.2	1.5.2.2	Completed
Input Hawaii SNMP Data	1.5.2.3	1.5.2.3	Completed
Input Transpac 100G link	1.5.2.4	1.5.2.4	Completed
Input PIREN-LA SNMP data		1.5.2.5	Year 4
Input PIREN GUAM SNMP data		1.5.2.6	Year 4
SNMP data from Exchange points	1.5.3	1.5.3	Completed
Input Starlight SNMP data	1.5.3.1	1.5.3.1	Completed
Input MIami XP SNMP data	1.5.3.2	1.5.3.2	Completed
Input CENIC XP SNMP data	1.5.3.3	1.5.3.3	Completed
General FlowData Initial Handling	1.6	1.6	Completed
TSDS FLow: add sparse storage model to TSDS for fow data with per flow metadata tagging	1.6.1	1.6.1	Completed
TSDS Flow: non aggregated, first Proof of concept	1.6.2	1.6.2	Completed
TSDS Flow: storage model stress testing and refinement	1.6.3	1.6.3	Completed
TSDS Flow: design refinement and scaling	1.6.4	1.6.4	Completed
Pipeline: design modular de-identification rabbitmq based processing pipeline	1.6.5	1.6.5	Completed
Pipeline: reference flow de-identification processor	1.6.6	1.6.6	Completed
TSDS Flow: flow stitching for both histograms and flows over a date line	1.6.7	1.6.7	Completed
TSDS Flow: flow stitching for parallel flow stitching, ex: gridftp - use HINTES heuristics	1.6.8	1.6.8	Not needed
TSDS Flow: flow stitching phase 2- parallel flows	1.6.9	1.6.9	Not needed
Make TSTAT easier to install as a product (also monitoring tstat process)	1.6.10	1.6.10	Completed
Create default aggregates for all metadata that we have in flows as of August 2017		1.6.11	Completed
Tstat/Flow deployment	1.7	1.7	
tstat logs into pipeline	1.7.1	1.7.1	Completed
Write up a document for De-Identification pipeline including data delete on source, etc	1.7.2	1.7.2	Completed
Bro and Tstat analysis	1.7.3	1.7.3	Completed
evaluate tstat for scalability (TCP retransmit)	1.7.4	1.7.4	Completed
Develop initial config for Bro	1.7.5	1.7.5	OBE

Develop initial config for tstat	1.7.6	1.7.6	Completed
configure tstat at TransPac LA	1.7.7.1	1.7.7.1	Completed
Input TP Seattle Sampled Flow Data into TSDS	1.7.7.2	1.7.7.2	Completed
Input TP Seattle UNsampled (TSTAT) Flow Data into TSDS	1.7.8	1.7.8	Completed
Input ACE/WIX Flow Data-Sampled	1.7.9	1.7.9	Completed
Input ACE/WIX Flow Data-UNSAMPLED using TSTAT	1.7.10	1.7.10	Completed
Input Ampath Flow Data	1.7.11	1.7.11	Ongoing
Talk to Julio and Jeronimo about Tstat vs sFlow vs AMIS data being available (SFLOW feed over ipsec tunnel)	1.7.11.1	1.7.11.1	Completed
Incorporate sampled flow data into TSDS	1.7.11.4	1.7.11.4	Completed
Evaluate AMIS data vs TSTAT	1.7.11.2	1.7.11.2	Completed- AMIS not fully functional
Purchase and deploy equipment if needed to support TSTAT at Ampath	1.7.11.3	1.7.11.3	Year 4
Incorporate unsampled flow data (tstat) from Ampath	1.7.11.5	1.7.11.5	Year 4
Input PIREN Flow/tstat Data	1.7.12	1.7.12	Ongoing
Talk to Lassner/David Wilde about Tstat and sFlow data being available	1.7.12.1	1.7.12.1	Started in Year 3, waiting on PIREN response
Purchase/Deploy flow equipment if needed for Tstat at PIREN	1.7.12.2	1.7.12.2	Waiting on 1.7.12.1
Incorporate sampled flow data from PIREN into TSDS	1.7.12.3	1.7.12.3	Waiting on 1.7.12.1
Deploy PIREN tstat data collection	1.7.12.4	1.7.12.4	Waiting on 1.7.12.1
Incorporate tstat data from PIREN into TSDS	1.7.12.5	1.7.12.5	Waiting on 1.7.12.1
Talk to someone about Guam flow/Tstat data		1.7.12.6	Waiting on 1.7.12.1
Talk to someone about HNL-LA Tstat/flow data		1.7.12.7	Waiting on 1.7.12.1
Input NEAAR Flow/tsta Dtata	1.7.13	1.7.13	Completed
Talk to NEAAR team about Tstat and sFlow data being available	1.7.13.1	1.7.13.1	Completed
Purchase/Deploy flow equipment if needed	1.7.13.2	1.7.13.2	Completed
Incorporate sampled flow data from NEAAR into TSDS	1.7.13.3	1.7.13.3	Completed
Deloy NEAAR tstat data	1.7.13.4	1.7.13.4	Completed
Incorporate tstat data into TSDS	1.7.13.5	1.7.13.5	Completed
Input CENIC Flow Data (Year 3)	1.7.14	1.7.14	sflow completed, TSTAT in Year 4
Talk to CENIC team about Tstat and sFlow data being available	1.7.14.1	1.7.14.1	Completed

Purchase/Deploy flow equipment if needed for CENIC	1.7.14.2	1.7.14.2	Waiting on 1.7.14
Incorporate sampled flow data for CENIC into TSDS	1.7.14.3	1.7.14.3	Completed
Deploy CENIC tstat data	1.7.14.4	1.7.14.4	Waiting on 1.7.14
Incorporate tstat data from CENIC into TSDS	1.7.14.5	1.7.14.5	Waiting on 1.7.14
Input AMPATH Ex Pt Flow Data (Year 3)	1.7.15	1.7.15	Completed
Input StarLight Flow Data (Year 4)	1.7.16	1.7.16	Year 4
Talk to StarLight team about Tstat and sFlow data being available	1.7.16.1	1.7.14.1	Waiting on 1.7.16
Purchase/Deploy flow equipment if needed for StarLight	1.7.16.2	1.7.14.2	Waiting on 1.7.16
Incorporate sampled flow data for StarLight into TSDS	1.7.16.3	1.7.14.3	Waiting on 1.7.16
Deploy StarLight tstat data	1.7.16.4	1.7.14.4	Waiting on 1.7.16
Incorporate tstat data from StarLight into TSDS	1.7.16.5	1.7.14.5	Waiting on 1.7.16
Evaluation of 100G Tstat collection	1.7.17	1.7.17	Completed
Instrumentation of Data Archives	1.8	1.8	Ongoing
Get data to kevin about CPU and I/O overhead of tools (tstat vs others)		1.8.1	Completed
Generate an RPM and/or better documentation on how to install tools on archives that will forwards tstat data to IU		1.8.2	Needs final documentation
Deploy Tstat on Hawaiian astronomy archives		1.8.3	Year 4
Deploy Tstat on CENIC/PRP archives		1.8.4	Year 4
Use top talkers list to identify likely DTNs		1.8.5	Year 4
Instrument NCAR Archive		1.8.6	Year 4
JMS to follow up with ESIP NASA guy for NASA DTN instrumentation		1.8.7	Year 4
Other possible DTNs from IRNC partners?		1.8.8	Year 4
Use of Globus DTN software from Princeton		1.13	Cant use - Log files only available to sys admins
Contact Raj about permission levels needed		1.13.1	Cant use - Log files only available to sys admins
Input SDN Data (Year 4)	1.9	1.9	OBE
Evaluate Argus vs tstat	1.1O	1.1O	OBE
Examine options for extended data services	1.11	1.11	Completed
Evaluate ELK and JetStream for larger faster data stores	1.11.1	1.11.1	Completed
JetStream account setup	1.11.2	1.11.2	Completed
Port data to JetStream for ELK work	1.11.3	1.11.3	Completed
Additional software framework Upkeep	1.12	1.12	
TSDS maintenance	1.12.1	1.12.1	Ongoing

Update TSDS for PS 4.0 archiving	1.12.2	1.12.2	Completed
TSDS-Elk integration	1.12.3	1.12.3	Completed
Deploy new version of TSDS		1.12.4	Completed
Add a keep alive notification for Tstat sensors (modify package)		1.12.5	Year 4
Migrate everything into new Git repo		1.14	Completed
Data transfer information (ie fiona) as additional data source		1.15	Year 4
Investigate approaches to including data transfer info		1.15.1	Year 4
Decision about inclusion		1.15.2	Year 4
Find guinea pigs for data transfer inclusion		1.15.3	Year 4
Analysis	**2**	**2**	
Topology publication Service	2.1	2.1	Completed
Capacity Planning tools	2.2	2.2	Not needed
Design capacity planning tools	2.2.1	2.2.1	Not needed
Implement and refine capacity planning tools- report generation hopefully starting for Q2	2.2.2	2.2.2	Not needed
Top Talkers scripts	2.3	2.3	Completed
Recreate AS to Science Project Data base (Science Registry)	2.4	2.4	In progress
Develop basic data base framework	2.4.1	2.4.1	Completed
Make sure framework has easily updated front end	2.4.2	2.4.2	Completed
Input data from Gloriad database	NOT	NOT	Data contaminated - not going to happen
input data	2.4.3	2.4.3	Ongoing
Get TransPAC to add data to science registry		2.4.3.1	Started Year 3, ongoing
Get NEAAR to add data to science registry		2.4.3.2	Started Year 3, ongoing
Get Ampath to add data to science registry		2.4.3.3	Year 4
Get PIREN to add data to science registry		2.4.3.4	After flow data collection
Get CENIC to add data to science registry		2.4.3.5	Year 4
Get StarLight to add data to science registry		2.4.3.6	After flow data collection
Extensions to basic science registry framework		2.4.4	Year 4
Add “short name” to flow tagging for organization names		2.4.4.1	Planned Y4Q1
Science registry metadata exporter		2.4.4.2	Year 4
Flow tagging based on SR metadata		2.4.4.3	Year 4
Read-only public mode for SR		2.4.4.4	Year 4
Form to submit changes to SR		2.4.4.5	Year 4
More science disciplines and ability to edit list		2.4.4.6	Year 4
More roles and ability to edit list		2.4.4.7	Year 4
Notes field for SR		2.4.4.8	Year 4

URL field for SR		2.4.4.9	Year 4
Admin section functionality for SR		2.4.4.10	Year 4
Need to get routing table dumps	NOT	2.18	IRNC Noc isn't collecting this as a time series
Largest transfer per month analysis	2.6	2.6	Completed
Summer Student work 2016	XXX	2.19	Completed
Tstat Analysis scripts (non flow, retransmits etc)- walk through Kibana experiment	2.7	2.7	Year 4
Work on caching analysis results in TSDS (storage of derived metrics)	2.9	2.9	Completed
Traffic characterization based on HNTES (of Elephant flows) - sort of part of Ed's work too	2.1O	2.1O	Completed
Data cleaning		2.2	Ongoing
Flow analysis capacity based on ESnet tool (Year 3)	2.11	2.11	Not needed
Design detailed flow analysis capacity based on esnet tool	2.11.1	2.11.1	Not needed
Implement detailed flow analysis capacity	2.11.2	2.11.2	Not needed
Indepth analysis for packet loss	2.12	2.12	Year 4
Analysis for top X		2.13	Planned Y4Q1
by organization	2.13	2.13.1	Planned Y4Q1
By country		2.13.2	Planned Y4Q1
By Protocol		2.13.3	Planned Y4Q1
For each link		2.13.4	Planned Y4Q1
Americas Greatest Networks - most reliable		2.13.5	Year 4
Soft release for IRNC PI meeting at I2 (May 2018)		2.13.6	Year 4
Incorporate feedback from May 2018		2.13.7	Year 4
Full release mid-2018		2.13.8	Year 4
Analysis for top science projects	2.14	2.14	Year 4 -waiting on registry data
Analysis for elephant flows - min, max and duration	2.15	2.15	Planned Y4Q1
Analysis of buffer size issues	2.16	2.16	Year 4
PIREN analysis for astronomy data		2.20.	Waiting on Tstat on UH archives
ENNETIX contacts and eval		2.21	OBE
Evaluate moving average for additional smoothing in graphs		2.22	Year 4
Evaluate Elastic X-Pack for data analysis and anomaly detection		2.23	Year 4
Out year analysis projects	2.17	2.17	
Alarms and alerts for NOC	2.17.1	2.24	Year 4 & 5
Meet with NOC for initial questions		2.24.1	
Identify properties needed to alert on		2.24.2	

Basic prediction		2.24.3	
Hook into ticketing or email system		2.24.4	
Feedback from NOC		2.24.5	
Traffic analysis in SDN and multi-tenant networks	2.17.2	2.17.2	OBE
Use of BGP metrics with analysis of flow systems	2.17.3	2.17.3	OBE - paths don't change on monitored links
Evaluation of PS tests and sampling	2.17.4	2.25	Year 4 & 5
Incorporate BGP information	2.17.5	2.17.5	OBE - paths don't change on monitored links
Develop SDN monitoring prototype	2.17.6	2.17.6	OBE
Compare sampled and un-sampled flow data		2.27	Year 4
Compare active and passive measurement data		2.26	Year 5
Flow data dashboards with variable queries		2.28	Year 4
Visualization Tasks	**3**	**3**	
create a google sheet with all network details needed to generate vis	3.1	3.1	Completed
Integrate TSDS database queries into prototype.	3.2	3.2	Completed
Default Summary View viz	3.3	3.3	Completed
Test prototype against real available data (waiting on CORS)	3.4	3.4	Completed
Develop viz prototypes based (flow data)	3.5	3.5	Completed
Develop hierarchical visualization column interface & widgets	3.5.1	3.5.1	Completed
Create basics of visualization	3.5.2	3.5.2	Completed
Research how to save user preferences	3.5.3	3.5.3	Completed
Create initial map view with all the nodes loaded	3.5.4	3.5.4	Completed
Aesthetic improvements	3.5.5	3.5.5	Completed
Develop Visualization of Heatmaps		3.5.7	Completed
Develop Bandwidth data visualization		3.5.8	Completed
Develop Losses data visualization		3.5.8	Completed
Develop Latency data visualization		3.5.10	Completed
Bind visualization (interface) to TSDS database (waiting on CORS)	3.5.6	3.5.6	Completed
Define for visualization team	3.6	3.6	Completed
Define what an elephant flow is	3.6.1	3.6.1	Completed
concatenate TransPac flow data, and generate histogram of that data	3.6.1.1	3.6.1.1	Completed
Define what "loss" is	3.6.2	3.6.2	Completed
Define what "top talkers" means	3.6.3	3.6.3	Completed

Show prototype to IRNC community & gather feedback (perhaps via Youtube video)	3.7	3.7	Completed
Develop video to show to gather feedback	3.7.1	3.7.1	Completed
Develop second version of video (after top talkers is in)	3.7.2	3.7.2	Completed
Jason Zurawski or Eli Dart			Completed
Kevin Thompson	3.7.3	3.7.3	Completed
Shawn McGee	3.7.4	3.7.4	OBE
Chris Rob (NOC PET)		3.7.5	Completed
NOC rep (Luke? Jent?)		3.7.6	Completed
Backbone owners		3.7.7	Completed
Exchange Point Owners		3.7.8	Completed
CIO/NW Planner		3.7.9	OBE
Incorporate feedback and revise visualization	3.8	3.8	Completed
Grafana base portal setup		3.9	Completed
Basic sw install		3.9.1	Completed
Accounts for development		3.9.1.1	Completed
Public vs private portal settings		3.9.1.2	Completed
Build basic frame other pieces will fit into		3.9.1.3	Completed
Define basic check in/testing process		3.9.1.4	Completed
Investigate UI auto testing frameworks		3.9.2	Completed
Use testing list and create screen shots for verification		3.9.2.1	Completed
Style guide definition		3.9.3	Completed
Get grafana base portal to match Aug 2017 portal		3.9.4	Completed
Left most summary data box		3.9.4.1	Completed
Map widget		3.9.4.2	Completed
bandwidth distribution		3.9.4.3	Completed
Link naming		3.9.4.4	Completed
Traffic volume		3.9.4.5	Completed
Total data		3.9.4.6	Completed
A-Z bandwidth and Z-A bandwidth		3.9.4.7	Completed
Line charts at bottom of portal (with zoom in)		3.9.4.8	Completed
Testing pass Oct 20		3.9.4.9	Completed
Release November 1		3.9.4.10	Completed
Heat maps to grafana		3.10.	Completed
Second release of flow data		3.11	Year 4
Navigation through views – not a pull down menu (maybe a table?)		3.11.1	Year 4
Grafana - CJ to set up shib access for grafana, set all to read only		3.11.2	Planned Y4Q1
Incorporate Shib extensions back to Grafana		3.11.3	Year 4
[Additional features for second release will be added based on feedback]			

Current map with PS data instead of SNMP data		3.12	Prototype completed Jan 2018, release mid 2018
Recreate my.es.net view of top talkers into portal		3.13	OBE
Bugs and Fixes	3.18	3.15	Ongoing
Updates for bandwidth dashboard		3.22	
Update SNMP Map on Bandwidth Dashboard		3.22.1	
Add PIREN LA link to SNMP map		3.22.1.1	Dependent on data gathering
Add PIREN Guam Link to SNMP Map		3.22.1.2	Dependent on data gathering
integrate existing worldview code into grafana		3.22.1.3	Completed
create example map for SNMP data		3.22.1.4	Completed
create example map from perfsonar data		3.22.1.5	Completed
create example map from ELK data		3.22.1.6	Completed
add worldview plugin to opensource git repo		3.22.1.7	Completed
Add exchange point info		3.22.1.8	Year 4
work to get map added to mainline grafana widgets		3.22.1.9	Year 4
logarithmic scale		3.22.1.10	Year 4
opacity based legends		3.22.1.11	Year 4
apply config changes without reload		3.22.1.12	Year 4
lines based on different functions		3.22.1.13	Year 4
invert legend		3.22.1.14	Year 4
additions to hover text box		3.22.1.15	Year 4
mapping improvements between datasource and text displayed		3.22.1.16	Year 4
dynamic wireframes, hide not present elements		3.22.1.17	Year 4
dynamically scale legend from dataset		3.22.1.18	Year 4
wireframe editor in grafana instead of outside		3.22.1.19	Year 4
get map added to grafana project		3.22.1.20	Year 4
investigate flows per country map (IU Communications style)		3.22.1.21	Year 4
Check all A-Z and Z-A mappings		3.22.2	Year 4
Bottom Graph Updates		3.22.3	Year 4
Make bottom graphs have same color for same network in each		3.22.3.1	Year 4
Make bottom graph hover listing sort according to largers		3.22.3.2	Year 4
Viz for Max sending vs retransmits		3.16	Dependent on research
Outyear Viz projects	3.17	3.17	
Develop viz prototype for Tstat data (Year 3)	NOT	NOT	OBE
Refine/refactor Year 3 prototype	NOT	NOT	OBE
Evaluate third set of prioritized queries (Year 3)	NOT	NOT	OBE
Release and evaluate prototype (Year 3)	NOT	NOT	OBE

Dev viz prototype using longitudinal data analysis (Year 4)	3.17.1	3.17.1	OBE
Refine/refactor Year 4 prototype	3.17.2	3.17.2	OBE
Evaluate next set of prioritized queries (Year 4)	3.17.3	3.17.3	OBE
Release and evaluate prototype (Year 4)	3.17.4	3.17.4	OBE
Refine previous prototype, bug fixing (Year 5)	3.17.5	3.17.5	OBE
Release and evaluate final (Year 5)	3.17.6	3.17.6	OBE
Kabana view prototype	3.19	3.19	OBE
Sankey Grafana Integration		3.20.	Began in Year 3
Add Sankey prototype (with mock data)		3.20.4	Year 4
refactor Sankey prototype to handle actual data		3.20.5	Year 4
integrate a specific query of flow data with Sankey		3.20.6	Year 4
generalize data processing for Sankey		3.20.7	Year 4
Review over the questions that guide Viz		3.21	Year 4
Which are still valid?		3.21.1	Year 4
Match questions with visualizations that we have		3.21.2	Year 4
Gather additional questions		3.21.3	Year 4
Design portal in order to answer the questions.		3.21.4	Year 4
Project Coordination	4	4	
Project management and coordination	4.1	4.1	Ongoing
Weekly project meetings	4.1.1	4.1.1	Ongoing
Refresh NetSage website home page		4.1.2	Ongoing
REU funding for testers		4.1.3	Year 4
Coordinate with NOC	4.2	4.2	Ongoing
AMIS coordination	4.3	4.3	Completed
InSight / Gloriad coordination	4.4	4.4	Completed
NetSage Data Privacy Policy	4.5	4.5	Completed
Draft web page message	4.5.1	4.5.1	Completed
Draft partner policy	4.5.2	4.5.2	Completed
Get document feedback	4.5.3	4.5.3	Completed
Get feedback from IRNC PIs	4.5.3.1	4.5.3.1	Completed
Get feedback from Kim Milford	4.5.3.2	4.5.3.2	Completed
Get feedback from Erin from CAIDA	4.5.3.3	4.5.3.3	Completed
FInal draft of web page message posted	4.5.4	4.5.4	Completed
Final draft of partner policy	4.5.5	4.5.5	Completed
Partner policy agreed to by	4.5.6	4.5.6	Completed
TransPAC	4.5.6.1	4.5.6.1	Completed
ACE	4.5.6.2	4.5.6.2	Completed

PIREN	4.5.6.3	4.5.6.3	Completed
AMPATH	4.5.6.4	4.5.6.4	Completed
Miami XP	4.5.6.5	4.5.6.5	Completed
StarLight	4.5.6.6	4.5.6.6	Completed
CENIC	4.5.6.7	4.5.6.7	Completed
Year 3 reporting	4.7	4.7	Completed
Updated budgets for Year 3	4.7.1	4.7.1	Completed
Y3Q1 report	4.7.2	4.7.2	Completed
Y3Q2 report	4.7.3	4.7.3	Completed
Y3Q3 report	4.7.4	4.7.4	Completed
Y3 annual report (with Q4)	4.7.5	4.7.5	Completed
Year 4 reporting		4.8	Ongoing
44Q1 report			Year 4
Y4Q2 report			Year 4
Y4Q3 report			Year 4
Y4 annual report (with Q4)			Year 4
Travel Year 3	4.13	4.13	Completed
AHM January 2017 Berkeley	4.13.1	4.13.1	Completed
AHM July Chicago	4.13.2	4.13.2	Completed
Invite people for feedback on video/usability		4.13.2.1	Completed
I2 global summit April 2017	4.13.4	4.13.4	Completed
TNC May 2017	4.13.5	4.13.5	Completed
TechEx Oct 2017	4.13.6	4.13.6	Completed
What demo for tech ex? Science registry?		4.13.6.1	Completed
Side meetings for tech ex?		4.13.6.2	Completed
SC Nov 2017	4.13.7	4.13.7	Completed
AHM 18-19 January 2018 Hawaii	4.13.3	4.13.3	Completed
Hawaii guys to figure out a location		4.13.3.1	Completed
Plan meeting		4.13.3.2	Completed
Papers Year 3	4.14	4.14	Completed
Write NetSage paper for IEEE Big Data congress	4.14.1	4.14.1	Completed
Write NetSage paper for Terena '17	4.14.2	4.14.2	Completed
Year 4 travel plans		4.15	Year 4
CENIC March 2018		4.15.1	Year 4
Internet2 PI meeting May 2018		4.15.2	Year 4
Brian Tierney - NetSage for APAN March 2018		4.15.3	Year 4
PEARC meeting - June 2018		4.15.4	Year 4
Paper submission to PEARC		4.15.4.1	Year 4
SC '18		4.15.5	Year 4
July 2018 AHM		4.15.6	Year 4
October I2 Tech Ex		4.15.7	Year 4

Jan 2019 AHM		4.15.8	Year 4
General project infrastructure maintenance		4.20	Year 4
Refresh NetSage web page		4.20.1	Year 4
REU funding for testers		4.20.2	Year 4
Contact email for project for outsiders		4.20.3	Year 4