# Sampling Approaches for Multi-Domain Internet Performance Measurement Infrastructures

***Final Scientific/Technical Report***
***for DOE Award # DE-SC0001331 to Ohio State University Research Foundation***
***Period of Performance: August 1st 2009 – December 31st 2012***

**Prasad Calyam, Ph.D. (Principal Investigator)**
**Calyam.1@osu.edu**
**Ohio Supercomputer Center/OARnet, The Ohio State University**

*This material is based upon work supported by the Department of Energy under Award Number: DE-SC0001331. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.*

This progress report provides a concise description of the project overview, workplan status, accomplishments (i.e., major activities, results and findings), outreach and collaborations, and cost status. For specific details regarding the research experiments, results and findings, and for additional project-related information, please contact the PI – Prasad Calyam at calyam.1@osu.edu. The following website can also be referred to obtain latest project information - http://www.oar.net/initiatives/research/projects/multidomain_sampling.

## I. Executive Summary:

In this project, we worked towards helping the high-performance networking communities that are supporting DOE scientists to overcome the "network awareness" gap (illustrated in Figure 1). Specifically, we developed multi-domain network status sampling techniques and tools to measure/analyze multi-layer protocol performance on the Internet. The project outcomes include enhanced scheduling algorithms, measurement federation policies, and tools to sample multi-domain and multi-layer network status with measurements obtained through frameworks such as perfSONAR. We also validated our algorithms and policies with measurement analysis tools for network weather forecasting, anomaly detection, and fault-diagnosis.

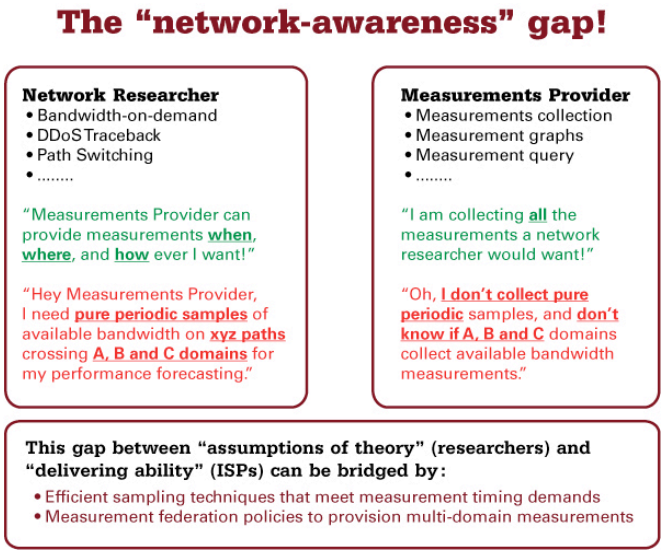

Figure 1: "Network Awareness" gap illustration

The deployment of the project outcomes has been performed on testbeds to support networking for DOE science. For e.g., we deployed our tools in the DOE's E-Center for network performance monitoring of Tier-1 to Tier-2 Large Hadron Collider sites consuming data feeds from CERN (Tier-0), as well as within the ESnet network monitoring dashboard framework based on Nagios.

The project team members are:

- Prasad Calyam, Ph.D. (Principal Investigator)
- Weiping Mandrawa (Software Engineer)
- Lakshmi Kumaraswamy (Student Research Assistant)
- Pu Jialu (Student Research Assistant)
- Thomas Bitterman (Software Engineering Consultant)

The project collaborators include: ESnet, Lawrence Berkeley National Laboratory, FermiLab, Bucknell University, University of Delaware, and Internet2.

## II. Workplan Schedule

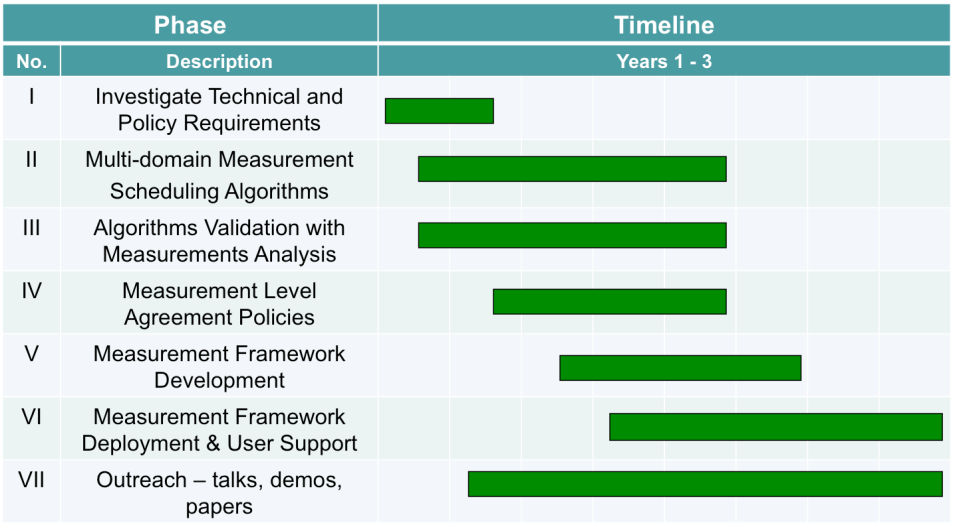


Figure 2: Project Milestones Gantt chart

Figure 2 shows the project milestones and workplan in the form of a Gantt chart. We progressed in the project in a timely manner to accomplish the goals and objectives that were established for the project period.

## III. Project Accomplishments

In this section, we discuss our project accomplishments categorized by the milestones during the period of performance. More specifically, we describe the major project activities, significance of the outcomes, results and findings. Note that Milestones V and VI are closely linked to Milestones I – IV, and so we combine the related accomplishments in our descriptions below.

### Milestone-1: Technical and Policy Requirements Gathering

### A. Major Activities

We conducted a survey of earlier sampling research papers in wired and wireless networks to identify the technical challenges in measurement sampling to meet analysis objectives. In

addition, we conducted a survey of perfSONAR web-service schemas for understanding the existing measurement policies being standardized in Open Grid Forum's NMWG for exchange of active and passive network measurements between ISP domains.

### B. Significance of Outcomes

An understanding of the technical and policy requirements for provisioning measurements in multi-domain measurement federations was crucial for planning the work activities in the subsequent phases of the project, and to successfully meet the project objectives. The perfSONAR measurement framework has been widely deployed in the DOE networking communities, and hence our survey provided us insight into the commonly used schemas to query performance measurements that indicate bottlenecks affecting large-scale file transfers.

### C. Results and Finding

The salient findings of the technical and policy requirements were as follows:

Technical Requirements:

- Intra-domain and Inter-domain measurement probes access
- Measurement conflicts avoidance
- Measurement request/response protocols
- Measurement sampling frequency guarantees
- Measurement orchestration flexibility (e.g., centralized and distributed)
- Data fusion of multi-metric/layer/timescale measurements
- Expert-systems for "network-aware" applications

Policy Requirements:

- Measurement Level Agreements
  - Share topologies, allowed duration of a measurement, permissible bandwidth consumption for measurements, …
- Semantic Priorities
  - Some measurement requests have higher priority than others
- Authentication, Authorization, Accounting
  - Determine access control and privileges for users or other federation members submitting measurement requests
- Measurement Platform
  - Operating system, Hardware sampling resolution, TCP flavor for bandwidth measurement tests, fixed or auto buffers, …

## Milestone-2: Multi-domain Measurement Scheduling Algorithms

### A. Major Activities

We developed a distributed meta-scheduler as a perfSONAR extension to enable perfSONAR to override current tool-specific resource brokers with programmable measurement orchestration to: (a) meet monitoring objectives accurately and timely using strategies such as adaptive sampling, and (b) leverage concurrent execution when possible to increase number of measurement requests that can be handled network-wide - to enforce measurement-level agreements within enterprise federations.

Figure 3 shows our proposed architecture with the Authentication and Authorization layer and the Resource Protection Layer that are important to be included for enterprise network monitoring. The Authentication and Authorization layer provides a user, access to measurement resources of

multiple domains within an enterprise, based on the modes of access determined in the enterprise security and measurement-level policies. Enforcement of the enterprise policies on the measurement resources is done in perfSONAR in the Resource Protection Layer. In the open architecture of perfSONAR, there have not been any efforts in the perfSONAR development community to develop services for the Resource Protection Layer that are needed to realize enterprise-specific deployments.

We integrated customized perfSONAR instances with our meta-scheduler services within an exemplar DOE enterprise viz., the DOE E-Center developed by a team comprised of Fermilab, BNL, ORNL and SLAC. The E-Center is the enterprise implementation of perfSONAR within the DOE community to orchestrate and manage DOE enterprise user requests for network performance measurements accessible via perfSONAR web services. Figure 3 shows an example use case of our perfSONAR extension as a Resource Protection Service in E-Center. The E-Center administrator maintains two databases, one for AAA and one for policy. The policy-inference and meta-scheduler services run in a distributed manner at some or all of the measurement points controlled by E-Center. When a measurement request arrives, it is checked whether there is an associated authentication token for the DOE enterprise user (e.g., DOE scientist, ESnet network engineer). If a token exists, it is checked against the AAA and policy databases at E-Center to infer the access privileges and priority of the measurement request. If there is no authentication token, the request is treated as one arising from a general Internet user. Based on the policy inference, our meta-scheduler is invoked to modify the measurement schedules at the affected measurement points to successfully handle the measurement request. As illustrated in Figure 4, the DOE enterprise users get higher priority in initiating measurements compared to the casual Internet user on the resource protected measurement points, as opposed to their random priority in openly accessible measurement points.

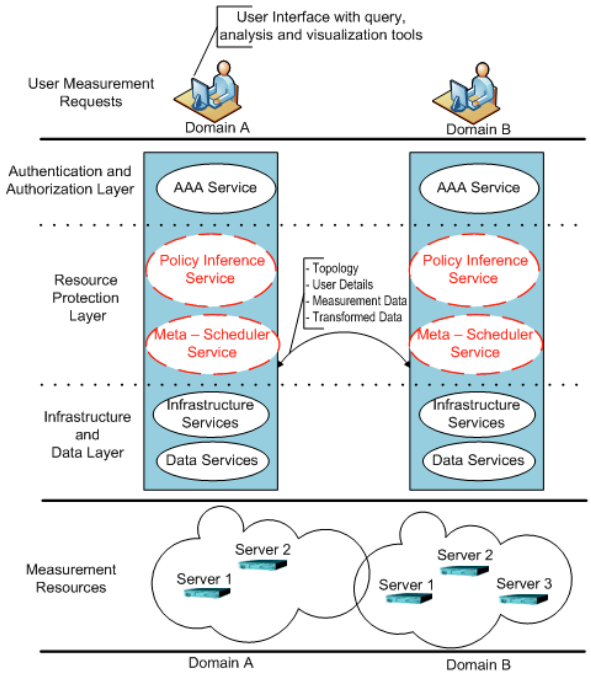


Figure 3: Resource Protection Layer formed through our perfSONAR extension services

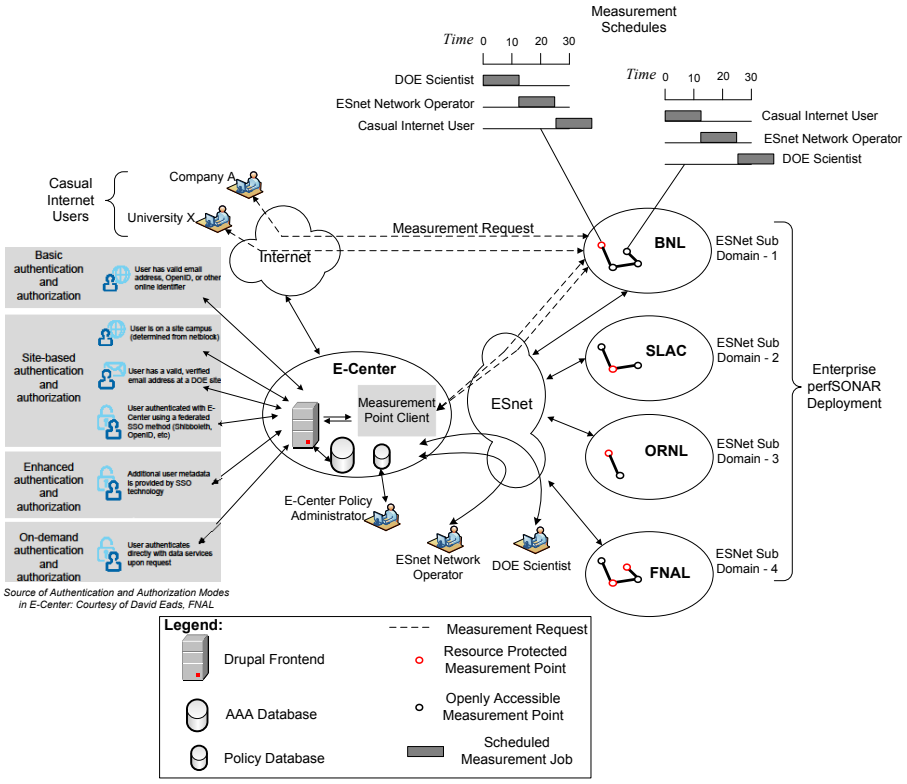


Figure 4: Example use case of our perfSONAR extension for Resource Protection

## B. Significance of Outcomes

Measurement schedulers should handle diverse sampling requirements of users to assist in their measurement analysis objectives. In addition, efficient scheduling algorithms will allow more consumers of perfSONAR measurements (e.g., network operators, researchers) to sample network paths. They also can better support on-demand measurement sampling with quick measurement response times to rapidly troubleshoot network bottlenecks using customized measurements. Our activities are geared towards understanding what characteristics of scheduling algorithms are better suited for handling the large-scale network on-going and on-demand measurement needs in the DOE networking communities.

## C. Results and Findings

We have studied the effect of conflict-free measurement scheduling on periodic measurement tasks. In addition, we have studied effects of scheduling measurement tasks with mixtures of sampling pattern requirements in the context of full-mesh, tree and hybrid topologies for increasing number of measurement servers, measurement tools and concurrent-execution bounds (to limit measurement traffic on a network path). We evaluated an offline Earliest Deadline First (EDF) based scheduling algorithm and compared its performance with an offline Heuristic Bin Packing algorithm in the context of scheduling active measurement tasks in large-scale network measurement infrastructures. We have identified measurement scheduling instances where there is measurement tasks starvation, soft deadline misses and schedule infeasibility for tasks with hard deadlines. Through our E-Center integration, we were able to validate that our perfSONAR web-service extensions will enable consumers of perfSONAR measurements (e.g., network operators, researchers) to directly control the sampling parameters on the network paths of interest, and thus provider greater flexibility for network performance monitoring affecting large-scale file transfers and other latency-sensitive end-applications.

## Milestone-3: Algorithms Validation with Measurements Analysis

### A. Major Activities

We evaluated a network performance "plateau-detector" algorithm that was used in earlier large-scale measurement deployments such as the NLANR AMP and SLAC IEPM-BW. The evaluation involved analyzing anomaly detection performance for both synthetic as well as DOE lab site perfSONAR measurement traces. Based on our analysis experiences, we developed perfSONAR extensions that will allow users to analyze and visualize uncorrelated and correlated performance anomalies. Our extensions that we packaged as "OnTimeDetect" tool works in real-time or offline modes for bottleneck detection and remediation as well as adaptations in network resource provisioning. Figure 5 shows the workflow between the OnTimeDetect software and a perfSONAR deployment. The OnTimeDetect GUIs include dialog-based applications such as the one shown in Figure 6, Twitter notifications of anomaly events, and the anomaly annotated Google Charts. We integrated OnTimDetect within the DOE E-Center in their Anomaly Detection Service as shown in an exemplar screenshot in Figure 7.

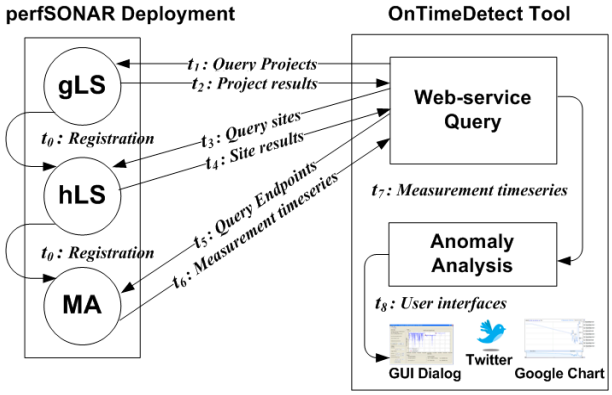


Figure 5: Workflow between OnTimeDetect tool and a perfSONAR deployment

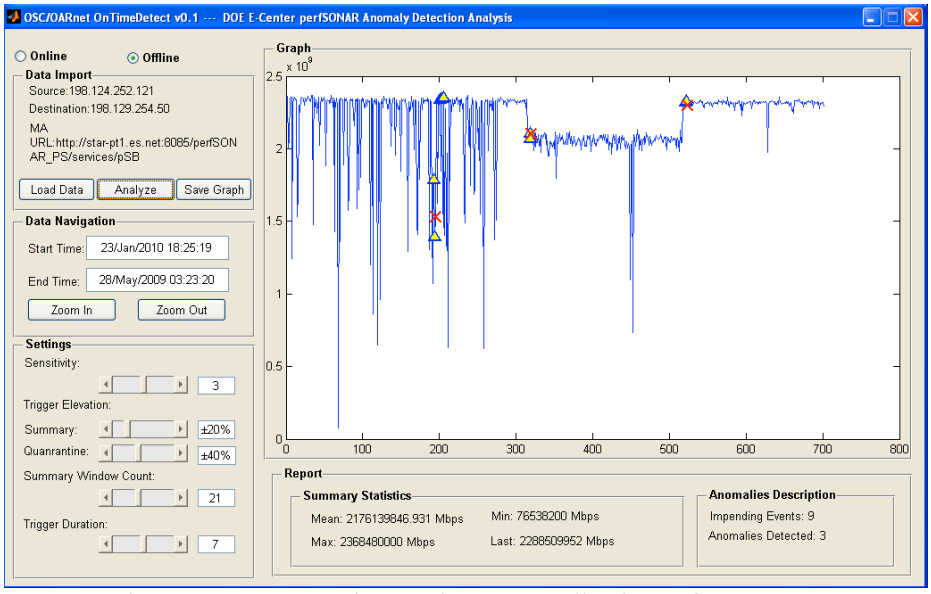


Figure 6: OnTimeDetect GUI tool to analyze anomalies in perfSONAR measurements


Figure 7: OnTimeDetect integration in E-Center user interface with annotated anomaly graphs

In addition to addressing anomaly detection monitoring objectives, we also evaluated a "dynamic winning-predictor selection" algorithm on ESnet perfSONAR measurement traces and analyzed forecasting performance. This algorithm has been widely used in other measurement frameworks such as the Network Weather Service for addressing the monitoring objective of predicting network performance to help in network control and management frameworks.

## B. Significance of Outcomes

Figure 8 shows how our sampling and analysis algorithms and tools fit with the different abstraction layers, network management frameworks and expected measurement outputs. At the highest level, applications require network awareness, and information on predictable or guaranteed network performance. They need measurements to indicate which optimal paths can be used to transfer large files or stream high-definition video streams. Our results bridge the "network awareness" gap illustrated in Figure 1 and make measurement frameworks such as perfSONAR more accessible and functional for application requirements of performance intelligence.

With increased programmability and extensibility enabled by our results, DOE community users and network operators can test from their end-hosts in greater numbers by leveraging concurrent execution principles in OnTimeSample, and gain better visibility into network performance through OnTimeDetect and its GUI extensions. Moreover, they can quickly run on-demand measurements without disrupting any on-going measurements, thus they can quickly get assessments of network weather and identify ideal network paths or circuit resources for meeting data movement timeliness.

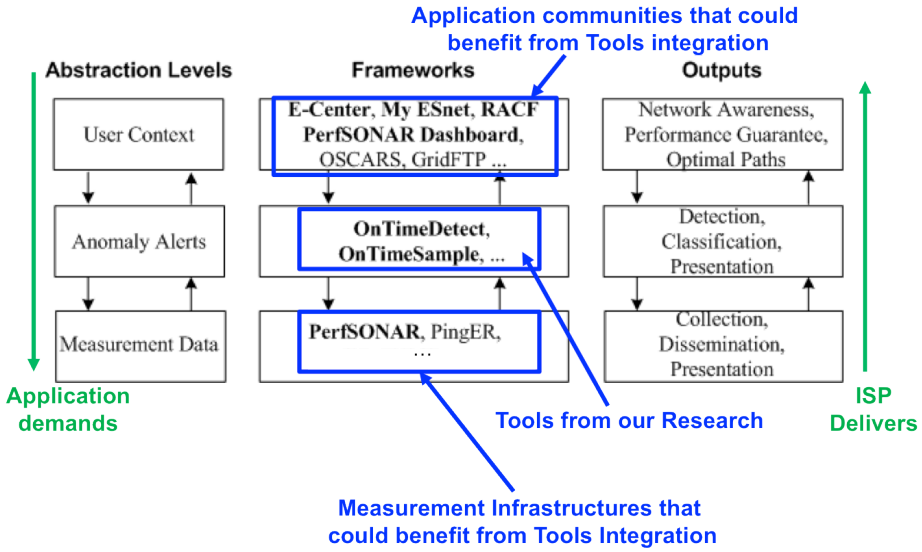


Figure 8: Our research outcomes and tool relevance to next-generation network management

## C. Results and Findings

We characterized the nature of network performance plateaus that affect network norm and anomaly event threshold levels. The threshold levels are calculated based on the statistical properties of historic and current measurement samples. Figure 9 shows an example ESnet perfSONAR BWCTL tool measurement trace with plateau anomaly event that was used in our study. Using our characterizations, we developed a dynamically adaptive scheme to be used in the plateau-detector algorithm to configure threshold levels for avoiding the triggering of false alarms.

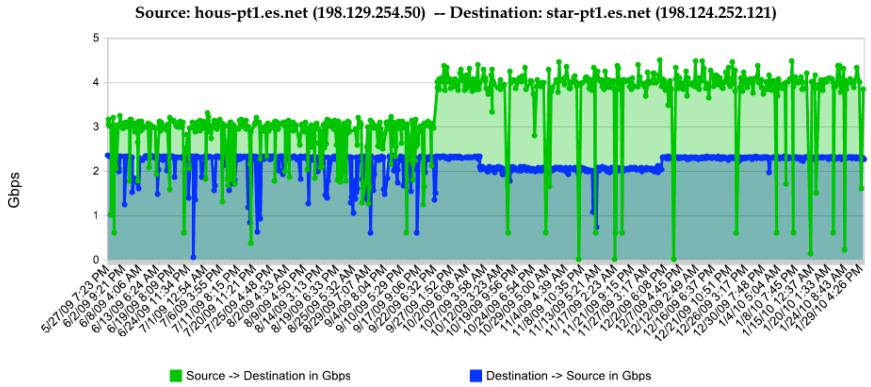


Figure 9: ESnet perfSONAR BWCTL tool measurement trace

We found that consumers of perfSONAR web-services for measurement data in the DOE community do not have access to automated techniques and intuitive tools to analyze anomalies in real-time and offline manner. Hence our evaluations and OnTimeDetect tool provided the DOE community with anomaly detectors that produce minimum false alarms and detect bottleneck events rapidly.

In our anomaly detection study, we validated our network-topology aware scheme for correlated anomaly detection using traceroute information and one-way delay measurements collected over 3 months involving 216 paths between the 17 DOE national lab network locations, published via perfSONAR web services. We showed how our adaptive plateau detection algorithm can be used to generate uncorrelated anomaly events with high accuracy and much less noise, than using

traditional static threshold based schemes, which can be extremely dense and noisy in terms of false alarms, even over short periods of data analysis. Using the critical hop/link based spatial filtering, we were able to diagnose bottleneck at edges caused by several of the correlated anomaly events. Further, we showed that the burstiness information has to be supplemented with the number of common events information in prioritization of critical paths for troubleshooting, and the paths with relatively higher burstiness and common events need to have higher priority during troubleshooting.

In our network weather forecasting study, we determined the statistical distribution of the types of forecasting schemes that produce low root-mean square error (i.e., accurate forecasts). We observed that only a handful of schemes such as the last value, sliding window average, and running mean are most often selected to generate forecasts. Other schemes are selected less often and their selection can be traced to interesting network status variations. These observations have motivated us to further investigate which schemes are better suited for the different network performance patterns observed in the measurement traces. They also suggest the potential for tuning sampling frequency and limiting for e.g., "last value" scheme based predictions that indicate over-sampling i.e., increased levels of measurement traffic that could have been allocated for actual application traffic.

We found out that consumers of perfSONAR web-services for measurement data in the DOE community do not have access to automated techniques and intuitive tools to: (a) forecast network performance, and (b) tune spatio-temporal sampling on network paths to obtain network-wide performance forecasts. Hence our evaluations in network weather forecasting studies provided the DOE community with network status mean predictors over different sampling intervals that are vital for network control and management relating to large-scale data transfers.

Using the guidance provided by our analysis with actual perfSONAR measurement data sets, we found that coupling outputs with additional information sources (e.g., router logs, maintenance activity logs) within frameworks such as NICE [23] can more effectively determine the "ground truth". Our findings provide insights to better understand how network operators can effectively and easily handle several network-wide anomaly event occurrences with varying degrees of severity showing up as 'red' alerts on their monitoring dashboards.

## Milestone-4: Measurement Level Agreement Policies

### A. Major Activities

We now briefly describe our work on the policy-inference service that co-operates with the meta-scheduler service as shown previously in Figure 3. Our policy-inference service can be operated by an enterprise policy administrator who has access to enterprise security and measurement-level policies. The policies are input into a database such that ontology trees can be constructed to capture the semantic relationships between different enterprise entities, and also to allow easy additions, modifications and deletions of policy rules. When a new user measurement request arrives in an enterprise, an inference engine in the policy-inference service communicates with Topology layer services and services in the Authentication and Authorization layer, and processes the ontology trees correspondingly to determine the access privileges and relative priority of the new measurement request considering all of the already scheduled measurement requests.

Figures 10 and 11 illustrate how for example, policy and user ontologies, respectively can be constructed within an enterprise comprising of multiple measurement domains. Enterprises typically are part of measurement federations comprising of multiple domains within the enterprise (e.g., SLAC, BNL) as well as other external domains (e.g., Internet2, LHCOPN).

Given the fact that Internet performance monitoring is inherently multi-domain in nature (since end-user applications traverse end-to-end paths through multiple ISP domains), measurement resources for a network path measurement could involve both 'intra-domain' resources and 'inter-domain' resources. Based on such resource types, resource policies can be specified in the policy ontology shown in Figure 10. Also, all of the domains within a federation could have common policies to co-operate with each other, and such policies could be specified in the policy ontology. Enterprise users can have IDs/names, and roles as an internal domain user or a federation (external-domain) user. User ontology example shown in Figure 11 can be used to capture such user IDs/names, roles and other detailed user preferences (i.e., sampling patterns, semantic monitoring objectives) that are part of measurement requests. Further, user authentication can be handled using 'tokens' that can be included in the ontology. The user tokens are sent along with the user measurement request details to the Resource Protection service, which then communicates with services in the Authentication and Authorization layer to determine access privileges and relative priority before scheduling the request on the measurement resources.

We developed the user and resource policy ontologies using the open-source protégé-OWL. The protégé-OWL editor supports Web Ontology Language (OWL) and supports a knowledge-based framework. We also used Semantic Web Rule Language (SWRL) to manage the rule-base in the inference engine while enforcing the enterprise security and measurement-level policies. SWRL was used because of its easy portability and extensibility. For example, SWRL can be used to specify a rule such as - if the resource policy is set as intra-domain, then measurement requests of intra-domain users get higher semantic priority versus the federation (external-domain) users.

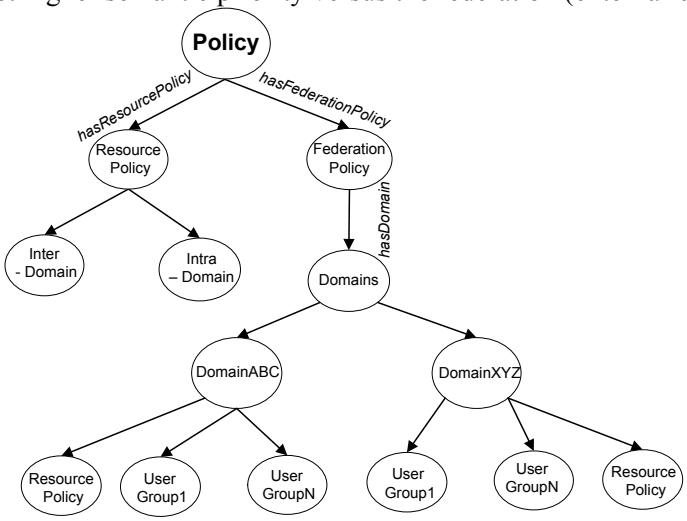


Figure 10: Policy Ontology

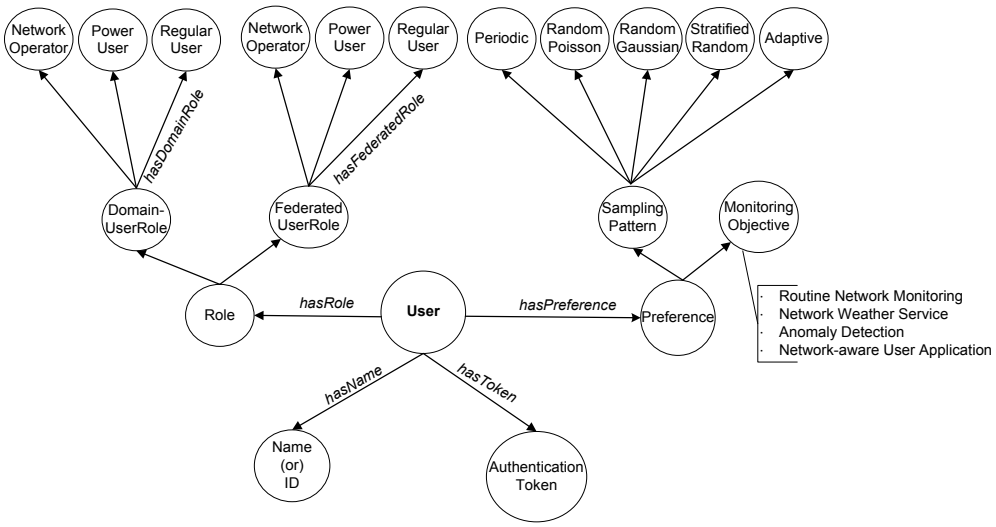


Figure 11: User Ontology

Figure 12 shows the ontology-based semantic meta-scheduler service modules we have developed. The priority calculator receives measurement requests (i.e., tasks to be scheduled) from end-users and uses the inference gained from ontology trees and a runtime solver to dynamically calculate the priority of each task. To capture the relative importance of the tasks, "initial state" and a "run state" priority calculations are used. The initial state priority calculation is based on policy and user ontologies. The "user role based priority", "resource policy based priority" and "sampling preference based priority" account for "initial state priority". Run state priority is set by the runtime solver based on oversampling penalty value obtained from the predictor scheme that satisfies the monitoring objective of the user. Final priority is calculated as the weighted difference between initial state priority and run state priority. Measurement requests are ordered based on decreasing initial state priority which is known a priori. The semantic scheduler module receives these ordered measurement requests and generates schedule table which is sent to measurement points in perfSONAR, which in turn initiate tools to sample performance data. The collected performance data is compared with the predicted data computed, and their corresponding mean square error (MSE) value is calculated. In the "run state" oversampled tasks are identified by the runtime oversampling detector and are penalized by reducing their relative priority. The schedule is altered based on the newly computed priority and the schedule output is sent to measurement points and the process repeats as detailed above.

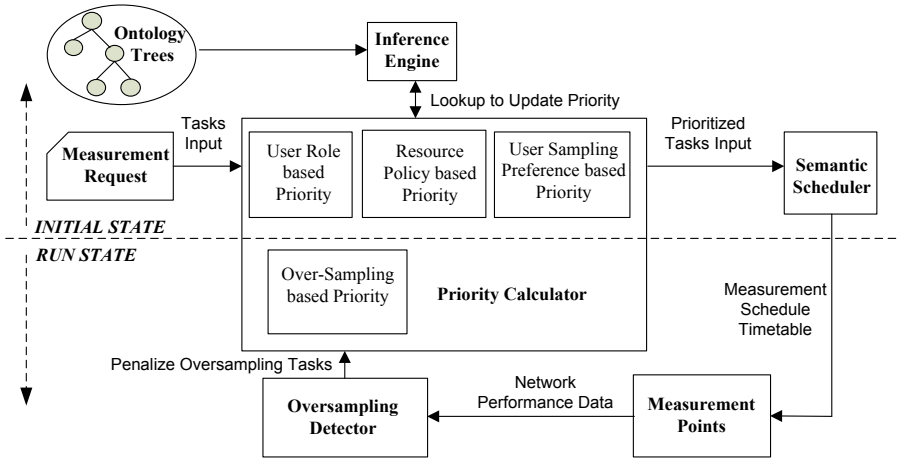


Figure 12: Ontology-based Semantic Meta-scheduler to cater Measurement Level Agreements

## B. Significance of Outcomes

Given the rising trend in recent times among ISPs to deploy openly accessible measurement frameworks, a need has arisen to develop measurement schedulers that handle semantic priorities. ISPs are using these frameworks to create "measurement federations" that facilitate measurements across multiple domains for reaping the mutual benefits of performing end-to-end path measurements. The most widely adopted is the perfSONAR framework that has been adopted by over a 100 user communities that include major regional, national, and international ISPs in academia and universities. Given that measurement resources (i.e., tool servers, network bandwidth) are limited, it might not be possible for a measurement scheduler to accommodate all user requests, i.e., generate completely feasible schedules under high measurement request loads. Consequently, measurement requests that could not be scheduled might adversely affect monitoring accuracy needed in critical resource adaptation decisions. Moreover, lack of semantic priorities might block intra-domain measurement requests that are more important than inter-domain measurement requests from an ISP's perspective. In such cases, there is a need to prioritize measurement requests by using semantic priorities based on user and resource policies. Semantic priorities of measurement requests can indicate cases of urgency to a measurement scheduler, which can then generate measurement schedules in a manner that supersedes typical scheduling priorities, i.e., period, laxity, execution time. Our work in this project activity is motivated by the fact that none of the existing measurement scheduling algorithms have the ability to handle semantic priorities that are important considerations in measurement level agreements in federated network monitoring.

## C. Results and Findings

Our novel ontology-based semantic priority scheduling scheme handled the resource contention by dynamically prioritizing measurement requests based on user roles, user sampling preferences, and resource policies. It was able to efficiently satisfy network monitoring objectives such as network weather forecasting, anomaly detection and fault-diagnosis across multi-domain measurement federations. To the best of our knowledge, our semantic scheduling scheme is the first to apply the ontology concept and an inference engine rule base to offline prioritize measurement requests and generate schedules. In addition, our semantic scheduling scheme also has the ability to online detect and mitigate oversampling in measurement requests to further improve the measurement schedulability under high measurement loads. Our performance evaluations that used real-world measurement request parameters and multi-domain considerations demonstrated that the semantic scheduling scheme outperforms existing measurement scheduling algorithms such as round-robin and other heuristic based algorithms that are based on typical scheduling priorities, e.g., period, laxity and execution time. More specifically, we showed how the semantic scheduling algorithm can improve the satisfaction ratio among users and also how the semantic scheduling algorithm reduces the average stretch to ensure fairness in handling measurement requests. Thus, our semantic scheduling scheme and evaluation results foster the deployment and management of large-scale multi-domain measurement infrastructures used for meeting monitoring objectives in support of next-generation applications and networks.

# IV. Project Outreach

## A. Peer-reviewed Papers

- P. Calyam, L. Kumarasamy, C. -G. Lee, F. Ozguner, "Ontology-based Semantic Priority Scheduling for Multi-domain Active Measurements", *Springer Journal of Network and Systems Management (JNSM)*, 2014.

- P. Calyam, M. Dhanapalan, M. Sridharan, A. Krishnamurthy, R. Ramnath, "Topology-Aware Correlated Network Anomaly Event Detection and Diagnosis", *Springer Journal of Network and Systems Management (JNSM)*, 2013.
- P. Calyam, S. Kulkarni, A. Berryman, K. Zhu, M. Sridharan, R. Ramnath, G. Springer, "OnTimeSecure: Secure Middleware for Federated Network Performance Monitoring", *IEEE Conf. on Network and Service Management (CNSM) (Short Paper)*, 2013.
- P. Calyam, L. Kumarasamy, F. Ozguner, "Semantic Scheduling of Active Measurements for meeting Network Monitoring Objectives", *IEEE Conference on Network and Service Management (CNSM) (Short Paper)*, 2010.
- P. Calyam, J. Pu, W. Mandrawa, A. Krishnamurthy, "OnTimeDetect: Dynamic Network Anomaly Notification in perfSONAR Deployments", *IEEE Symposium on Modeling, Analysis & Simulation of Computer & Telecommn. Systems (MASCOTS)*, 2010.

### B. Talks and Demos

- "Multi-domain Internet Performance Measurement Algorithms and Tools: Relevance to ESnet", Talk at ESnet/LBNL, Berkeley, 2011.
- "Anomaly Detection Integration and Deployment in DOE Monitoring Infrastructures", Internet2/ESCC Joint Techs, Fairbanks, Alaska, 2011.
- "Experiences from developing analysis techniques and GUI tools for perfSONAR users", perfSONAR Workshop, Arlington, VA, 2010.
- "Multi-domain Internet Performance Sampling and Analysis Tools", Internet2/ESCC Joint Techs, Columbus, OH, 2010.
- "OnTimeDetect Tool for network anomaly notification in perfSONAR deployments", Presentation/Demo at Internet2 Spring Member Meeting "Network Tools Tutorial" session, Arlington, VA, April 2010.
- "Sampling and Analysis Tools for E-Center to support Multi-domain Internet Performance Measurement", Winter ESnet Site Coordinators Committee Meeting, Salt Lake City, Utah, February 2010.
- "Multi-domain Internet Performance Measurement: Sampling and Analysis", Presentation/Demo at ESnet/Internet2 Joint Techs Conference, Salt Lake City, UT, February 2010.
- "Sampling Approaches for Multi-Domain Internet Performance Measurement", Presentation at Kickoff Meeting for DOE/SC/ASCR Network Research Projects, Fermilab, Batavia, Illinois, September 2009.

### C. News Articles

- "Using perfSONAR to Find Network Anomalies", *ESnet Network Matters Blog*, 2011.
- "Monitoring Advanced Network Health Status", *OSC Research Report*, 2010.
- "Research seeks to improve service for users of next-generation networks", OSC Press Release, October 2009.

## V. Conclusion

Our project's overall goal was to bridge the "network awareness" gap in the high-performance networking community that is supporting DOE science. Towards this end, we developed multi-domain network status sampling techniques (e.g., conflict-free scheduling algorithms, multi-domain measurement policies) and open-source software tools (e.g., OnTimeDetect, OnTimePredict, OnTimeSample) to measure/analyze multi-layer protocol performance on the Internet. We believe that we have made substantial contributions in the project. By virtue of our efforts, we expect network operators as well as DOE scientists to obtain better network

performance transparency and the ability to rapidly troubleshoot bottlenecks that affect large-scale file transfers and other latency-sensitive applications. Our project outcomes were geared towards becoming integrated into the existing DOE-led efforts such as the perfSONAR measurement framework deployments at ESnet sites, and the E-Center web-portal for site-to-site and hop-by-hop performance monitoring on ESnet and other peer network paths.