1. INTRODUCTION

Advancements in network connectivity have improved scientists’ ability to collaborate and have helped accelerate the rate of scientific discovery. Modern science relies on the global Internet to create large and physically distributed scientific work environments. As a result, scientists have become more dependent on networking, and network problems have an increasingly significant impact on science. The pursuit of modern network-based science is, therefore, dependent upon cyberinfrastructure designed to support the efficient diagnosis of network issues along end-to-end paths that traverse multiple physical networks.

Understanding network performance poses a number of challenges for distributed collaborators. Scientific communities (i.e., virtual organizations) usually span many physical organizations. The distributed nature of these virtual organizations makes it difficult for campus network engineers or end-users to effectively find or fix end-to-end performance problems. Campus networks are usually interconnected via regional GigaPoPs or Regional Optical Networks (RONs) and national backbones. A typical physical path from one scientist to a remote instrument or collaborator requires the crossing of at least five administrative domains (i.e., campus, regional, backbone, regional, and campus). Each of those domains usually maintains explicit control over who can access diagnostic information and infrastructure and what performance data is shared with peer domains, if any. Often, considerable time and effort must be expended to convince the physical organizations to begin testing. Once testing has begun, it can still be a difficult task to isolate the issue, and even more time can be spent evaluating different solutions. As a result, it is increasingly difficult to perform tuning and troubleshooting in distributed environments due to the need to access information from many different sources.

Network performance data is important for more than just addressing network pathology. Computer and network researchers have long examined the use of network information to improve distributed computing applications. Application adaptation based on network performance, or network proximity, has been somewhat successful thus far, but the scientific community is only just beginning to reap the benefits. Some Grid computing platforms take rudimentary network information as input to their scheduling algorithms. Currently, measurements for this purpose are made from the “edge”, but it is clear that reporting on the state of the network with the collusion of the network elements, through common cyberinfrastructure, would be desirable.

Over the past 20 years, numerous projects have developed tools that can measure, monitor, share, and/display network performance data. Most such tools have only attempted to tackle part of the problem, such as data gathering or analysis and visualization. While the proliferation of data-gathering tools (active and passive) has provided many new insights for network researchers, the lack of widespread, uniform deployments has significantly curtailed their potential for end-to-end problem diagnosis and application feedback. Similarly, analysis and visualization tools have been hampered by the relative lack of standards for sharing performance data, limiting the breadth of data (both in type and deployment) available to be consumed. A few projects, such as AMP [1], MonALISA [2], IEPM-BW [3], and Surveyor [4], have attempted to bundle data generation, data sharing, and analysis with visualization. However, such systems have failed to achieve critical mass in deployment due to lack of open source code, their packaging approach, the unitary solution imposed on heterogeneous organizations (virtual and physical), the limited nature of the problem addressed, and/or the lack of community involvement in their development. The absence of effective tools and the presence of poor mechanisms for sharing performance data have made it difficult to make high-performance networking widely available, as it tends to take a team of network experts from multiple physical organizations to solve the all-too-frequent performance problems.

1.1 Problem Statement

Today, numerous scientific disciplines have developed globally distributed work environments that depend on the network. Some examples include:
• The international High Energy and Nuclear Physics (HENP) community is nearing completion of the Large Hadron Collider (LHC) facility located at CERN in Switzerland. The LHC will allow thousands of scientists, engineers, and university professors access to petabytes of particle physics data. Data will be stored at multiple national repositories and processed at dozens to hundreds of sites around the globe.

• Radio astronomy’s electronic Very Long Baseline Interferometry (e-VLBI) community is using high-performance networks to move radio telescope data from multiple sites to a central correlator. The composite image, with very fine resolving power, is then used to examine astronomical objects, track spacecraft, or measure the movements of the earth’s tectonic plates.

• The University Corporation for Atmospheric Research (UCAR) Unidata project provides raw atmospheric and oceanic science data and analysis tools that can be used by scientists and educators to enable a better understanding of earth-science. Data is shipped to multiple institutions in “real-time” or “near real-time” enhancing the learning process for students.

• The National Ecological Observatory Network (NEON) project is constructing a national scale infrastructure to support studies on major environmental challenges at regional and continental scales. Distributed sensor networks will capture in situ measurements of soil and water biogeochemistry and send the results to an archival storage system. Supercomputers and interactive display systems will assist in analyzing current and historical data, while also generating multiple timescale forecasts.

The rapid detection and correction of network performance problems and infrastructure faults represent unresolved problems facing all these scientific communities. It is difficult to overstate the amount of effort that large-scale distributed projects such as these spend on network debugging and performance tuning. That this effort is duplicated across projects represents a significant waste of time and energy. Moreover, many scientists are forced to become part-time network engineers, which is an inefficient use of their time and effort.

Network pathology alone does not drive the need for network performance information; even when a precisely pinpointed problem is found in the network, there is often nothing that can be done. In these cases, application communities must be able to adapt their use of the network based on feedback about its current state. Job distribution in environments like the Open Science Grid can make use of metrics of network performance to determine the most advantageous location to launch jobs [5]. This will become even more prevalent in future applications.

There are two major problems this proposal is designed to address:

1. A widespread demand for network performance information by:
   • Researchers who depend on the network for their work need feedback about how well the network is working.
   • Advanced network-based applications (such as those emerging in Grid computing research efforts) that could use network performance information to tailor their execution.
   • Network operations staff who need end-to-end information from other network domains to diagnose network problems.

2. A lack of flexibility in extracting performance data from the network resulting from:
   • Network technology that changes relatively quickly and, as a consequence, the useful diagnostic data is a moving target.
   • New types of analysis that are continually being developed. Sometimes, these are done using completely new tools that generate information on completely new metrics. Often, they represent new ways of correlating existing diagnostic data types.
   • Lack of effective techniques to share performance data across domains

This desired functionality of gathering and exchanging network performance metrics has been partially addressed by many projects and tools. Listed below are a number of requirements for a network performance system that are not available in toto in any extant software.
• Ease of extensibility to new data sources – new tools are produced frequently and the system must not be biased toward any particular tool, or type of tool.
• Secure, policy-based access – the ability to provide protection for who can access what and under what conditions at a local level is critical for a single system framework to be accepted.
• Community-driven – the architecture of the system must be completely open (and open source) to encourage the reintegration of extensions and the piecemeal upgrading of modular components.
• A single, monolithic, global measurement system is unrealistic. Federation and neutral information exchange standards-based formats and a modular mix-and-match architecture are necessary. This preserves local autonomy while promoting global utility.

2. The perfSONAR Framework

The perfSONAR consortium (http://www.perfsonar.net), initiated in partnership by Internet2 [6], the European Union-funded GÉANT2 project1 [7], and DoE Energy Sciences Network (ESnet) [8], has been working, in recent years, to provide an open set of protocols and a reference implementation to address cross-domain sharing of network measurements and metrics. This project is focused on building open source, flexible, modular, and extensible performance middleware to greatly simplify the process of gathering and sharing network performance information across multiple administrative domains to simplify the process of debugging network issues and to provide performance data to network-sensitive applications without redundant measurements. Early prototypes of the perfSONAR framework are deployed in Europe and in North and South America and serve as a central component of the network monitoring functionality for the global network dedicated to serving the LHC project. The rapid adoption of the early prototype of the perfSONAR system by various National Research and Education Networks (NRENs) and the LHC project is a testament to the need that it fills.

In developing a framework to facilitate expanded deployment of services to support scientific collaboration, perfSONAR aims to significantly enhance global research, which increasingly requires interaction between geographically-distributed sites, multiple teleconferencing tools, and rapid exchange of large data files. By using perfSONAR, complex collaborations in a broad range of cutting-edge research communities between multiple scientists will be able to function in a manner similar to that found at a single site. They will do this by using standard authentication credentials to enhance self-diagnosis of problems without the need for coordinated efforts between network domains. With these efforts, core perfSONAR services in the U.S. will be integrated, hardened, and deployed, serving as core infrastructure for current use and future community extension of the perfSONAR framework. In the process, these measurement services will address critical challenges to research collaboration, and will benefit U.S. scientists by providing them with an avenue for easy access to network performance information. This effort will also have an impact on the EU-funded perfSONAR efforts, in which the PI team is already an active participant.

The perfSONAR project is a global effort designed to unify today’s measurement and monitoring infrastructures into a single, open cyberinfrastructure framework, breaking down barriers that exist between sources of performance information. This framework specifically defines a federated network measurement infrastructure that can be used by a large number of research scientists and scientific collaborations. The perfSONAR project is still at a very early phase, with much work to be done yet, but it is beginning to have a significant impact already2 and generate interoperable, independent development of additional services, which is a testament to the level of community support.

The perfSONAR project originated with the desire to unify the efforts of Internet2’s End-to-End Performance Initiative Performance Environment System (E2E piPES) project [9] and the EU’s GÉANT2 Joint Research Activity on monitoring and measurement (GN2-JRA1) [10], ultimately involving parallel

1 GÉANT2 is a four year project begun in 2004 that is funded by the European Union’s Sixth Research & Development Framework Programme with a contribution of 93 million Euros and matching contribution from the European National Research and Educational Networks (NRENs). It includes funding the network of the same name as well as research and support activities.
2 For deployment information, please see http://www.perfsonar.net/deploy.html.
efforts in the U.S. DoE’s ESnet [11]. The desired result of this project, a common performance middleware, is referred to as the perfSONAR framework. This project seeks to enable consistent access to network performance data across a large fraction of the world’s advanced research and education networks. The effort is in its second year of funding in the EU and receives a small amount of seed money from Internet2 and ESnet in the United States. Some components of this framework exist in prototype form, but, much remains to be done.

The University of Delaware, Internet2, and ESnet, have been participating in this effort since its inception, and Georgia Institute of Technology (GaTech) and the Stanford Linear Accelerator Center (SLAC) have recently joined as well, but the primary locus of development has been among DANTE [12] and the European NRENs participating in the GÉANT2 project, due to the significant support of the EU. We propose a collaborative effort between the University of Delaware, Internet2, GaTech, and SLAC to extend and harden the existing perfSONAR implementation for the U.S. and international scientific community. This will facilitate additional U.S. acceptance and deployment of newly-developed performance middleware tools from the perfSONAR project.

The perfSONAR system seeks to enable federated authentication and rich policy expression in terms of authorization and resource utilization, thereby allowing network operations staff and end-users alike to access and generate network performance data, subject to the locally-controlled access policies of the parties involved. The system will also provide open protocols and Application Programming Interfaces (APIs) to data gathering tools (e.g. SNMP [13], BWCTL/Ipert [14, 15], OWAMP [16]) and to analysis and visualization tools (e.g. CNM [17], MonALISA, perfSONAR UI [18], ICE [19], Visual perfSONAR [20]). This will allow network researchers to innovate in the data gathering arena and in the analysis and visualization arena without being shackled by the lack of a common performance middleware. The goal is to spark a proliferation of both classes of tools (in effect, a competition among data gathering tool developers and a competition among analysis and visualization tool developers) by achieving a critical mass of deployed performance middleware that immediately provides access to a wide range of data gathering tools (for new analysis and visualization software) and to a wide range for analysis and visualization software (for new data gathering tools).

The perfSONAR project has adopted the Open Grid Forum (formerly Global Grid Forum) Network Measurement Working Group (OGF NM-WG) schema [21] for the data representation model and protocol. This is an extensible XML data representation designed to allow new data formats to be represented in ways such that intermediate parts of the framework do not need to understand the details of new types of network performance data – but those intermediate parts of the framework can still be used to share that data widely. This allows for maximum code reuse and aids extensibility. The perfSONAR development team has leadership positions in the OGF NM-WG, ensuring that any lessons learned from the development and deployment of these reference implementations are transitioned back into the work of the OGF NM-WG.

Building a large community invested in deploying the perfSONAR measurement framework is as important as developing the actual software. The perfSONAR framework is being developed in an open community by consensus, with all community-developed code given a BSD-style open-source license. Strategically, this was seen as the best way to encourage widespread development and deployment of the framework, as every organization knows that it can walk away with the entirety of the work at any time (and therefore is more willing to contribute to the effort). When individual institutions have a sense of ownership of the resulting software infrastructure, continued involvement is more likely. The model preserves the ability for any member of the community to produce new innovative tools. (Other models often preserve uniformity at the cost of innovation.) This strategic approach has been validated by the rapid growth of the perfSONAR community across the world.

The perfSONAR framework defines seven basic types of services as follows:

- **Measurement Point Service:** Creates and/or publishes monitoring information related to active and passive measurements.
- **Measurement Archive Service:** Stores and publishes monitoring information retrieved from Measurement Point Services.
• Lookup Service: Registers all participating services and their capabilities.
• Authentication Service: Manages domain-level access to services via tokens.
• Transformation Service: Offers custom data manipulation of existing archived measurements.
• Resource Protector Service: Manages details regarding policies on system resource consumption.
• Topology Service: Offers topological information about networks.

These services support a variety of tools that measure, monitor, analyze, and report performance problems in networks. Some services (e.g. Measurement Point [MP], Measurement Archive [MA]) come in many varieties (e.g. SNMP MP, RRD MA), while others may be standard (e.g. the Authentication Service). Collectively, this extensible, modular set of services is the performance middleware upon which new tools and analysis services can be built. This methodology has the following benefits:

• New tools can be developed and widely deployed. The extensible perfSONAR framework provides a common set of services that allow users to find and use measurement and analysis tools. Normally, new tools face a steep barrier to wide-scale deployment; users must learn how to operate the tools and how to interpret the results. Network operators who don’t use these tools are reluctant to accept trouble reports based on the tool results. perfSONAR provides a standard (XML-based) interchange format that normalizes data output and allows users to focus on solving problems instead of becoming tool experts.
• One common troubleshooting technique is a divide-and-conquer approach where sections of the end-to-end path are tested in sequence until a suspect section is identified. The major difficulty with this approach is finding the possible test points along any arbitrary path. The perfSONAR Lookup and Topology Services can help address this problem. Users will be able to find and use both active and passive measurement points, without requiring assistance from an administrator.
• Trouble reports can easily be generated and sent to the appropriate administrator for resolution. All too often, users submit trouble reports that either lack useful data or the data is presented in a format unfamiliar to the network operator. Delays occur while the network operators request clarification of reported problems. perfSONAR addresses this issue by using a standard web services interface to access data from local and remote measurement archives. In addition, individual tools will convert their output into an OGF NM-WG XML schema. This allows the storage and retrieval of data in a tool-, user-, or domain-dependant format.

In summary, the perfSONAR framework will ultimately:
• Be easily extensible to new types of measurement and analysis tools. Tool developers will be able to focus on how the tool operates, not on how to deploy the tool widely enough for it to become useful.
• Scale to global size. Multi-national science communities will be able to focus on science, not just run networks.
• Contain flexible policy for sharing across administrative domains. Local domains will retain control over the policies that define how specific resources can be accessed and what data will be shared with whom according to uniform, organization-agreed-upon policies, not ad hoc decisions based on personal relationships.
• Use common data exchange formats. OGF NM-WG schemata will be used to ensure that different tools and archives can exchange data effectively.
• Use common discovery mechanisms. Users will be able to find measurements points and archival data to assist in troubleshooting an application’s behavior.
• Contain an evolvable set of currently discoverable services and information.
• Use generic service components for tracking resource utilization and user authorization to acquire those resources for performance testing.
• Piggyback on existing federated trust authentication infrastructures.
3. RELATION TO EXISTING PROJECTS

Network researchers and end-users have always had a need for network measurement, monitoring, and analysis tools. This need dates back to the early days of the Internet when the “ping” command was written. While tools like “ping” and “traceroute” command have become, and will continue to be, the front-line tools used by most network engineers, current security pressures to limit or block the packets used by these tools shows that additional tools and access models are needed. The new tools and how they are used should answer both basic connectivity type questions and more complex “what’s wrong with this path” type questions.

Previous efforts to develop these new tools can be sorted into three major classes: (1) individual tools that can be installed on Internet hosts, (2) measurement tools deployed in a single test domain, and (3) measurement tools that monitor more than just the network.

The first class of projects deals with the development of individual measurement tools. These refer to two types of measurement tools: passive and active. Passive tools (e.g., tcpdump [22], OCx-mon [23], Netflow [24], and Web100 [25]) make use of application-generated data packets that flow through the network. A copy of the data packet is passed to the tool for some type of processing. Each tool typically generates its own unique output format. As with active tools (described below), the output of the tool is usually handed over to a tool expert for interpretation and analysis.

Active tools (e.g., Iperf, Thrulay [26], NDT [27], NPAD [28]) use a client/server model to generate test packets. Typically these tools require remote login access to both the client and server hosts, or the server may run a daemon process that accepts connections under some local policy control. Each individual tool usually presents data using a unique display format, forcing users to spend time and energy learning exactly how each tool works, how to interpret the results, and the failure modes. Incomplete results and developer assumptions mean that both false positive (the tool reports a problem that does not exist) and false negative (the tool reports no problem even when one exists) reports can be generated. Finally, most tools do not save enough background information to make the data useful when reporting problems.

The major advantage of these tools is that individuals and small communities can download and install the tools on an as needed basis. This ensures that tools are used in a wide variety of operating environments, much larger than the developers could have achieved. The major disadvantage is that these tools may not be already installed when problems are reported. Another deterrent is that remote administrators may be reluctant to install them when they are not sure how the tool works or what its reporting capabilities are. Finally, tools that require manual action at both ends usually require some kind of coordination between remote administrators and an implicit trust relationship, reducing ease of use.

A second class of projects deploy network-only measurement or monitoring tools inside some test domain. In addition to individual tools, these larger research projects (AMP, Surveyor, IEPM PingER [29]) have been established to generate large-scale measurement and monitoring programs. These projects tend to use both active and passive test tools to perform their measurement and monitoring functions. A typical deployment strategy is to purchase a physical PC or develop a software installation package that can be installed on a remote PC. Donor institutions are then approached and asked to provide a physical home (power and network connectivity) for this test box. Remote management of these remote boxes by the research project is another common attribute of these projects.

The main advantage of these projects, and the systems they develop, is that they deploy boxes at a number of remote locations. Testing usually is done on a regular schedule, providing a long-term baseline from which to observe changes. The major disadvantage is that the host institution does not typically review the test results and feels little ownership of the measurement infrastructure. This leads to the condition where the research team may have a better understanding of a site’s network connectivity, but they are not involved in the troubleshooting process. Moreover, systems developed in such projects have lacked community involvement in the design and are ultimately limited by the ability of the research team to support such an infrastructure. Experience has shown that such projects fade away as funding shifts and people move on to new projects and that subsequent projects rarely build on what has come before.
A third class of projects also exists. This group (MonALISA, NWS [30], NetLogger [31]) monitors more than the network; they try to capture additional information (CPU utilization, Disk I/O rates, memory swap rates) to generate a complete end-to-end picture of a specific application. The main advantage of these projects is that they provide a comprehensive view of how an application or high-level service is operating. The main disadvantage is that they are usually complete systems that are hard for non-project members to extend. These projects are the class of projects most similar to the perfSONAR framework. Many members of the perfSONAR development team have been directly involved with creating or using these tools and the lessons learned during those developments have been incorporated into the design and development of perfSONAR.

NetLogger provides tools for instrumenting applications, but does not focus on providing data about the network. To be useful in this wide context, all network applications would need to be instrumented using the NetLogger toolkit. Additionally, there are no provisions for simply providing network diagnostic information that is generated by the network devices themselves.

NWS works by setting up sensors across the network. It uses the collected data to attempt to “forecast” future network conditions based on the current data and historical values. This is interesting and relevant work that could be made even better by having more sources of data made available to it.

MonALISA is very similar to perfSONAR in that it deploys a set of dynamic services that work in a coordinated fashion to provide relevant information. It is a well developed set of tools that provides much of the same functionality that a perfSONAR framework would provide. It is uses Sun’s JINI for discovery and is Java-only technology. MonALISA is produced by a closed development effort that requires any potential developer to agree to a license that signs over intellectual rights to Caltech for any related effort and limits the right to deploy the software to a finite period of time. The server-side code is not available for redistribution or development by outside groups, so new diagnostic tools must rely upon the MonALISA project to incorporate the data produced by those new tools. MonALISA also lacks the community involvement, consensus-building, and investment we believe is needed to achieve widespread ownership of the performance framework and, therefore, widespread deployment.

In contrast, the perfSONAR framework defines a series of mid-level services that can be used to create a measurement infrastructure. It does not mandate the use of a specific tool or set of tools to be used to generate or capture measurement data. Rather, it defines a standard XML-based storage schema that allows multiple tools to generate and consume data. This increases the usability and reach of these tools, making them more applicable to a larger number of potential users. perfSONAR does not require real-time intervention to deploy tools or to coordinate the data collection process. Instead, a standard authentication mechanism allows interested parties (e.g., end-users, scientists, dedicated support staff, network operators) to pre-register themselves in a domain-managed database. Once registered, these users can present their credentials to a remote domain’s federated authorization service and, if approved (according to locally-determined policies, and home institution defined identity), they can begin troubleshooting a problem.

perfSONAR also provides Lookup and Topology Services that allow users to find measurement points along a path of interest. A prototype example is the traceroute mapping tool [32] developed by ESnet. The user simply inputs the output from a traceroute command into a web page form and clicks a submit button. The web service contacts the perfSONAR Lookup Service to determine if monitoring data exists for the specific router interfaces. If so, the data is automatically extracted from a measurement archive (MA) and displayed in the user’s browser.

We believe the perfSONAR framework provides the necessary services and infrastructure components to support numerous scientific communities. Domain scientists will be able to use these services to determine what, if anything, is impacting their specific application. Network researchers will be able to develop, test, and widely deploy new tools and higher-level services that can be used by domain scientists and network operators.
4. IMPROVEMENT and SUPPORT PLAN

4.1 CURRENT STATUS

As it currently stands, perfSONAR is an established consortium, a set of protocol standards for interoperability between components of a measurement infrastructure, and a partially complete reference implementation (release v2.0 due in February, 2007). The perfSONAR architecture has support from a reasonably large segment of the research community. Such support takes the form of active contributions to the development of core perfSONAR services, the development of analysis and visualization software dependent on the perfSONAR middleware, and the installation of the current system.

perfSONAR’s current shortcomings are primarily in two areas: first, development of a few of the not immediately critical, yet important, infrastructure components has yet to be done, such as the development of the authentication/authorization components needed to hook perfSONAR into existing authentication frameworks. Second, all the existing services need additional work to be turned into production quality software. These shortcomings are especially problematic in making the software easy to deploy.

4.1.1 DEPLOYMENT STATUS

Prototype perfSONAR services have been widely deployed on many research networks in the national and international community. ESnet, GÉANT2, RNP (Brazil) have been partners in the development of the perfSONAR architecture. Specifically, the EU has invested significant resources toward the development and deployment of perfSONAR via the GN2-JRA1 project. Because of this centralized investment in the infrastructure, deployments across Europe have flourished. For example, as of the time of this writing there are 13 EU-related national research network deployments of network utilization measurement archives. (A list of current deployments can be seen at http://wiki.perfsonar.net/jra1-wiki/index.php/RRD_type_MA_Service_Installations.) There are far fewer U.S.-based deployments: two as of this writing. This does not represent a lack of interest on the part of the U.S. research community; however, the current status of the implementation is a barrier to adoption in some cases. It takes several hours (more than a day with significant hand-holding for most system administrators) to install and configure even the simplest services. Additionally, the current services have significant performance issues. The EU participants have been given incentives via the GN2-JRA1 grant to compensate them for this installation effort, which has lowered the barrier of entry for many of the EU sites. With commensurate incentives, we believe the scale of US deployments can be significantly advanced through the development of product-quality services released through package management systems well-known to U.S.-based system administrators. In other words, the Java-based approach of the reference-implementation with its many dependencies is proving to be a significant barrier to adoption for U.S.-based administrators. We propose to implement the production-quality services developed by this proposal in Perl and use package management solutions such as CPAN (www.cpan.org) to ease the installation hurdle.

4.1.2 IMPLEMENTATION STATUS

The currently implemented perfSONAR services can be classified in two groups: first, services that support the perfSONAR infrastructure and, second, services that actually provide information about network performance.

The currently implemented infrastructure-level services include the Lookup Service and, to a lesser degree, the Topology Service. The currently deployed Lookup Service is not distributed, so the only services a client can discover are services that register with that specific Lookup Service. The PIs of this proposal have developed a prototype version of a hierarchical, distributed Lookup Service [34]. This prototype needs to be further developed into a production quality piece of software. This will allow the global discovery of perfSONAR resources that register with any Lookup Service participating in the global perfSONAR community. Additionally, the currently deployed Lookup Services has significant scaling problems with respect to the number entries in its internal database. These issues are not fundamental to the design, but, rather, they represent problems with the current implementation. There is currently effort in GÉANT2 to prototype a Topology Service that will offer information regarding the
physical and logical layout of a network. This service currently only provides Layer 3 (IP) information regarding the network, however as circuit switched network paths become more prevalent in the near future, it will become important to both monitor dynamic network state. The topology service is seen as being complementary to higher-level measurements such as using traceroute.

The information service with the largest current deployment footprint is the RRD MA service written in Java. This implementation is used to make network utilization data available across a fairly wide number of research networks. It works fairly well for that purpose. This service is currently somewhat difficult to install and configure, given the typical knowledge base of network operation center (NOC) personnel (most are not particularly versed in Java system dependencies). Overlay style services that expose the data that network administrators already collect using existing systems will be critical because it will allow more incremental adoption of perfSONAR.

### 4.2 Improvement Overview

The goal of this proposal is to support the development and hardening of critical perfSONAR services. Production implementations of specific framework services must be developed and widely deployed to ensure that the perfSONAR framework will become more than a short-lived research project. Prototypes of many services already exist, but these prototypes must be enhanced to ensure that they work reliably over a wide range of conditions. Prototype services are now being tested in several major national (Abilene, ESnet) and international (GÉANT2, RNP, various European NRENs) research networks. The test results will be evaluated and revised to ensure that production-level services are routinely available to a large number of scientific communities and to provide the cornerstone of further community development.

We intend to make the base infrastructure available with some fairly simple (but useful) tools. We expect the network research community will assume responsibility for the creation of better diagnostic and analysis tools using this framework. There is already evidence of this, as noted previously. In the greater perfSONAR project, several international research groups have already begun developing tools that rely on data being stored in a perfSONAR measurement archive service.

Our specific focus is to create production quality versions of specific perfSONAR services that are critical to current scientific user communities and to future network researchers. Without this kind of sustained production-level support, we fear that few, if any, of these communities will ever benefit from the development and deployment of the perfSONAR services. These goals will only be realizable if the core services are stable enough for perfSONAR-aware applications to be developed. We will then build atop these core services by developing example clients (both interactive and report based) that will enable access to such services.

The PIs and several international participants have already contributed their time and energy to the perfSONAR project with the goal of making various measurement components interoperable through a common performance middleware. The early work laid the foundation for a larger development project by defining how the various perfSONAR services will operate. However, it is increasingly clear that we can no longer rely solely on the goodwill of those participants to move the perfSONAR framework from the concept phase to a working infrastructure. A permanent infrastructure that will benefit multiple scientific communities and one which receives full community investment is the next logical step. Funding this proposal will translate these concepts into working infrastructure components.

### 4.3 Specifics

The first phase of this effort will be to harden and deploy the common components that make it possible to federate³ perfSONAR services. Of prime importance is the Lookup Service. This is the most critical component to adoption and must be a highly available production service. In conjunction with this proposal we intend to deploy two servers on the Internet2 network to perform the function of root Lookup Service servers. The Lookup Service is based on open XML standards and implementations, to allow for easy expansion, and thus is completely non-proprietary. We have recently developed a prototype of this service based on open-source XML tools. The protocol that we have designed is based on XML tooling

---

³ Services from multiple administrative domains seamlessly interoperate to form a single logical service infrastructure.
and is, therefore, completely adaptable to storing data of any sort. Even if the perfSONAR infrastructure evolves, we will be in a position to adapt the registrations stored in the Lookup Service. Indeed, this infrastructure could be beneficial to any distributed system wishing to bootstrap lookup information.

The local Lookup Service currently holds a complete view of perfSONAR services. An additional requirement will be to make topologically-related queries for services. For example, this is necessary to find measurement points near a given network path. We propose to create an Information Service that is capable of combining queries that have topological and service-oriented components. This could be implemented as a transformation service that interacts with both a Lookup Service and a Topology Service. For performance reasons, it may be more realistic to implement this additional service interface within the same services and have it include both Lookup and Topology interfaces.

Another commonly required component is the Authentication Service. The perfSONAR architecture does not intend to build a new authentication infrastructure; instead, we are working with European partners to leverage the existing work from Shibboleth [35] and eduGAIN [36] to provide identity management using existing SAML [37] standards. However, both systems were written with a specific application in mind: browser interaction with a web-server. The perfSONAR project has been working closely with those efforts to develop ways to integrate those identity management systems with perfSONAR. However, this is a fairly complex development and perfSONAR will require components and protocols to bridge these two models. This will be a particularly difficult deployment problem and will require a strong product-quality effort for the first deployments. A key part of this deployment will be to ensure that this Authentication Service works in a federated way with similar efforts being supported by the EU.

The Authentication Service really is only necessary for measurement services in the context of authorization. When a service needs to determine if it should allocate resources for a request, there are two basic questions that need to be answered: first, is there enough of the given resource, and, second, is this particular identity allowed to consume the available resource? (Notice that this is more complicated than the normal authentication/authorization model needed for determining if users are allowed to access the resources on a typical web server, which most federated authentication/authorization frameworks were initially designed to protect.) The consumable resource aspect of the problem is addressed by the Resource Protector Service in the perfSONAR architecture. It keeps track of resource allocation and utilization in both absolute and chronological (scheduling) aspects. The Resource Protector Service has the job of accounting for current and future resource utilization and then makes requests to the Authentication Service to determine rights. This is a more dynamic authorization model than most systems need to deal with, again, requiring more effort during deployment. Making the Resource Protector an independent service provides users with the ability to aggregate the usage of common resources. Multiple services might make use of a single common resource (say a common backbone link). If this utilization accounting was only done on a per-service basis, two different measurement points might run a test that consumed excessive network throughput at the same time. The RP allows services to consult an entity specifically deployed to take into account (and protect) this limited, shared resource.

Therefore, in conjunction with the Lookup Service and the Authentication Service, we will integrate and deploy a Resource Protector Service for each of the limited resources on the new Internet2 Network [38]. This service controls the use of other services by applying policy based on current resource load and rights and identity of the requesting entity. (As with most modern Authentication/Authorization frameworks, this identity need not be fully disclosed to the resource but will be auditable using the authentication framework.) The Resource Protector will be used to implement the policy currently applied in the Internet2 Network Observatory [39]. This service allows users to perform tests to points in the “core” of the network, which is critical for isolating problem areas. The hosts and links that support these measurements are limited resources and, without protection, could easily be overloaded, reducing the accuracy of their measurements and rendering them useless.

The next major phase of this effort will be to provide specific, useful example diagnostic services. The infrastructure is only useful if it provides information to the users. As stated before, we expect the network research community to build new innovative diagnostic tools using this infrastructure. However,
this cannot happen unless there are high-quality examples showing how to use the framework. Additionally, it will not be possible to convince system administrators to deploy the software unless it provides added value. The specific diagnostic services we intend to provide as part of this proposal will be based on a set of commonly used network tools (BWCTL, OWAMP, Traceroute, ping, SNMP data collection, Layer-2 link status). We intend to make the active tools available as on-demand services as well as providing data archives of any regularly scheduled measurements of these types. The interesting part of this task is not the actual tools being presented, but in allowing federated access to those tools and standard mechanisms to combine their outputs for diagnostic purposes.

To provide more information about the how the many physical networks are interconnected, a Topology Service that will provide information for Layer 2 and Layer 3 information will be created. This will provide information on a per domain basis and hence will rely on the use of a federated Lookup Service to provide complete end-to-end information.

In order to demonstrate the flexibility and usefulness of the perfSONAR framework, a set of core analysis examples will be constructed. These tools will demonstrate the ability to quickly aggregate and visualize network topologies and performance data using best practice techniques. Examples will include interactive graphs of network nodes with performance charts and static reports of the network (a weathermap). It will demonstrate the integral use of all perfSONAR services; using the Authentication Service to query Topology Services and Measurement Archives for real network information. All such services will be discovered through a query of the Lookup Service. On-demand diagnostics and alarm services based on statistical analysis can then be run to further investigate into potential network issues.

The final phase of this effort will be to promote the perfSONAR infrastructure to the research community. With available hardened examples and widespread involvement by the major research networks, it will be possible to first interest and then train researchers and their students to use the infrastructure. This last step will be needed to ultimately fulfill the promise by allowing new interesting tools and, specifically, new analysis to take place.

In summary, the specific objectives and tasks of this proposal are:

4.3.1 Objectives

- Production quality versions of: a Distributed Lookup Service, the Resource Protector, the Authentication Service and the Information Service
- Widespread deployment of these services. Currently, NOC staff from Internet2, ESnet, and GÉANT2 are deploying prototype implementations of these services. As most end-to-end problems occur near the ends of a network path, it is particularly important to bring regional network operators, GigaPoP operators, and campus network operators into the fold.
- Integration of these tools and services into the daily operations of the NOC for all three networks listed above. The above services should provide a wealth of information to the NOC staff; however, the usefulness of this data depends on the degree to which it is used by the NOC staff to identify, locate, and solve problems on a daily basis.
- Expand the uptake of these services to other geographical locations.

4.3.2 Tasks

- Integrate and deploy two ‘root’ Lookup Service instances in very well connected, geographically-distributed locations.
- Integrate and deploy an Authentication Service federated with GÉANT2 (EU) authentication services allowing intercontinental authenticated and authorized monitoring applications.
- Integrate and deploy a Resource Protector service that can protect the Lookup Service from abuse as well as implementing the current protected measurement functionality of the Internet2 Network Observatory
- Deploy diagnostic Measurement Point and Measurement Archive Services to support SNMP, OWAMP, BWCTL, Traceroute, and ping diagnostic tools on three campuses and the Internet2 backbone network.
• Provide documentation and self-funded workshops to promote the use of perfSONAR in the NOC and research communities.

5. INTELLECTUAL MERIT

The perfSONAR services represent the base components needed by any large-scale network measurement infrastructure. Properly implemented, these services should become nearly invisible to the general network user community. In addition, a new generation of researchers and students will use these services in as-yet unimagined ways to create new tools and higher-level services that can help us understand why an application failed to operate as expected. That said, there are still many challenges that must be overcome before the perfSONAR services reach this level of ubiquity. Three of the fundamental challenges that we will resolve are:

1. Extensibility – The perfSONAR services are infrastructure components that will be used by other tools and services to make network management easier. We will work actively to generate API’s that will meet the needs of the network measurement and application development communities.

2. Federated Lookup Service – For each network domain, there is an inherent tradeoff between local autonomy and the ease with which components can communicate and share information about themselves. A critical goal of the perfSONAR system is to allow network domains to control local measurement and management infrastructure while allowing that same infrastructure to be used by peer networks, connector networks, and end-users. Clearly, it is desirable to control what or how much of a resource can be shared and part of that is controlling how much information is advertised about local infrastructure. Alongside that goal is the system requirement of scalable information sharing. While both topics have been addressed in formal terms, building a system that realizes both goals in working software is non-trivial.

3. Policy – Related to the above goal is the need to express measurement policy with a rich vocabulary. Given that there may be measurement constraints based on availability of recent measurements, other ongoing measurements, and the credentials of the individual user, it should be obvious that we need a language and infrastructure that can express and control these policy goals. Policy definitions must likewise have federation in mind and many of the challenges described for the Federated Lookup Service apply.

6. BROADER IMPACT, EDUCATION, AND OUTREACH

To understand the impact for users, scientists, support staff, and network researchers, it is informative to look at how these groups go about solving problems today. [40, 41] As a sample case study, consider what might happen when a scientist wants to join a video conference. Upon joining the call, the user finds the audio stream is very poor and the video stream is non-existent. The scientist contacts his local network administrator to report a problem. The administrator suspects packet loss is causing this problem, so she decides to use the Iperf tool to examine the end-to-end path. Her approach will be the ‘divide-and-conquer’ method where tests will be run from one end of the path to various intermediate points along the path. This will help her isolate where the loss is occurring and who should be tasked with fixing this problem. To run these Iperf tests, she must first contact her peer administrators at the various network domains in the end-to-end path. Several hours of phone calls, telephone tag, and interactively running the Iperf tool finally show that the loss is occurring on one specific link in this end-to-end path. A repair is scheduled and the scientist is informed that the video will be available after the repair is completed. Thus, several days of effort by multiple network administrators are required to find and fix this specific problem.

Reducing the hours and number of participants required to find and report these types of problems is a major goal of the perfSONAR project. Not only is the time to repair reduced, but the frustration level of the scientist is also reduced. Knowing where a problem is, and when it is likely to be repaired, can help the scientist develop contingency plans.

In addition to making manual trouble reporting and resolution easier, the perfSONAR infrastructure can assist network operators with the day-to-day monitoring tasks that can automatically find and report problems. Waiting for trouble reports from users before noticing a problem is another serious issue with today’s complex network. While a typical operations staff quickly notices and fixes a major outage or a
problem with a well-used application, a niche application may fail for weeks or months before a trouble report makes it into the NOC. In the mean time, the end-user is frustrated and unable to effectively use the network to perform a specific task.

At a higher level, the impact of this work on society will mirror that of the Internet itself; by making networks easier to operate and use, scientists and end-users alike will have an improved experience. This translates to scientific and commercial benefits that are well understood at this point.

6.1 Impact on Networking Research

Network research scientists and engineers strive to build better tools to capture and analyze network behavior. A major impediment to the widespread deployment of tools that require active intervention at two sites is the need for remote login access to one or both of the test boxes at one or both of the end sites. A partial solution to this problem is to run one end as a daemon process and interactively use the other end during the test period. While this is a common solution, it still requires that the remote administrator approve the long-term use of this tool and that the tool is sufficiently robust to restart itself if and when it fails. The perfSONAR Resource Protector Service addresses this issue and makes it easier to deploy tools based on the client-server model that can be used effectively by a large number of people.

A second problem facing the network research community is the lack of a common data exchange format. A typical active measurement test tool will generate new data for every test and then the tool will perform some basic analysis and present a summary to the tool operator. In some cases, this tool operator would like to pass the raw data over to some network expert for further analysis. In a perfSONAR network, this is done in a straightforward manner because the data is represented in a well-defined format. Specifically, the data is converted into a generic format using the OGF NM-WG schema. The Measurement Archive Service will save the data for later use.

A third problem facing this community is the need to have tools both generate data and analyze that data to investigate problems. This forces a developer who wants to analyze a specific set of data, to first create a tool that generates that data. The perfSONAR framework will allow multiple tools to generate or consume data from a Measurement Archive Service. In addition, the perfSONAR Transformation Service may also be invoked to convert the data from one format to another, meaning that the analysis developer can simply write a transformation program instead of developing a data generation program.

A perfSONAR-enhanced network will promote strong growth in several fields of network research. Researchers in data generation tools can spend their time developing new tools and algorithms that create specific data flows. Researchers in data analysis tools can focus on their craft, and possibly obtain a large cache of input based on the Transformation and Measurement Archive Services.

6.2 Impact on Broader Scientific Community

Scientists and their support staff rely on high-performance networks for a multitude of communication services. Everything from email to interactive communication (Instant Messenger, video, remote instrument control) to bulk transfer of large amounts of scientific data moves over today’s networks. As networks become an increasingly critical piece of scientists’ work environments, finding and fixing performance problems takes on a whole new meaning.

Currently, network administrators spend an extraordinary amount of time coordinating debugging efforts due to multi-domain issues, even for relatively trivial problems. Looking to the future of more switched networks, this looks to be even increased in the future. The perfSONAR project will enable remote network administrators to debug performance issues much more efficiently by reducing the need for coordination. Specifically, coordination will be a one-time authorization configuration using a federated model. This will lead to more responsive problem resolution, which will, in turn, lead to more effective utilization of the network by the scientific community at-large.

6.3 Impact on Education

The perfSONAR project will have two major impacts on educators and students. A perfSONAR-enabled infrastructure will make it easier for educators to bring the network into the classroom. Daily use of the network for routine teaching activities depends on a network infrastructure that can meet the needs
of the teaching application. A perfSONAR-enabled network will allow educators and support staff to
easily verify and demonstrate that the network is operating properly.

A second benefit stems from the fact that perfSONAR services are infrastructure components. Educators will be able to incorporate these services into the learning process. New avenues of research into tool design will emerge where some tools specialize in generating specific traffic flows to probe the network for very specific problems. Other tools will emerge that analyze data stored in perfSONAR archives. These specialized tools can be easily and widely deployed because they rely on deployed infrastructure, not on the goodwill of end users. Funding the perfSONAR project will ensure that the base infrastructure will become available to a wide range of educators and students.

6.4 Educational Plan

The educational outreach of this project will have three main components: early prototype components of the perfSONAR system have been featured in courses at the University of Delaware. In Computer Science, the Distributed Computing course (CISC879) has covered the efficient exchange of network performance information for distributed computing extensively, using perfSONAR as a case study. This material covers the OGF NM-WG schema and architecture upon which the perfSONAR framework is largely based. perfSONAR features components necessary for any distributed system and serves as an excellent pedagogical vehicle. In this course, we cover issues such as web service construction and data modeling/normalization. The project in the course included construction of a perfSONAR component plugin.

Members of the perfSONAR team are already engaged with resolving performance issues on production networks. By continuing to work with operations staff, local administrators, and faculty, the team will promote the tools and the advantages of self-service over waiting for third parties to run tests. For example, at GaTech, a campus-wide network monitoring infrastructure is already deployed. The infrastructure will be perfSONAR-enabled, allowing faculty and central support staff to view performance and schedule tests within and beyond campus. The perfSONAR vision will play an important role in the support of GaTech's international campuses, such as in France and China.

Experience and feedback will be used as input to documentation. This will provide the building blocks to educate other sites and content for workshops. In particular, the team will conduct tutorials and workshops at the Joint Tech's Workshops and at regional conferences aimed at support staff in general, not necessarily the research network community. The team will also assess interest in conducting workshops for specific disciplines such as at the Computing in High Energy Physics conference.

7. PROJECT MANAGEMENT PLAN

This is a distributed collaboration that will include frequent communication between the Principal Investigators (PIs) and senior personnel as well as an experienced project management configuration that is essentially already in place for the perfSONAR effort. PI Martin Swany (University of Delaware) and Co-PI Eric Boyd, Director of Internet2's Performance Architecture & Technologies team, will coordinate and manage this project. Boyd will serve as the primary technical manager. In his work with Internet2's E2E piPEs project and in building the perfSONAR partnership, Boyd has experience administering a group of employees and volunteers from multiple institutions and multiple countries in the development and creation of a distributed client/server application. He also has experience working in and directing commercial software development teams. Boyd will handle managerial oversight aspects regarding progress of the project, setting project milestones, priorities, and policies. PI Swany, co-PI Les Cottrell (SLAC), and co-PI Warren Matthews (GaTech) will be engaged in the technical details of this effort.

Five levels of communication will be maintained and available to participants:

1. Susan Evett, a program manager at Internet2, will maintain a permanent project website on the Internet2 server with links to email archives, call transcripts, presentations, and papers. Project milestones, priorities, and policies will be maintained on this site.

2. As a group, we will agree on an instant messenger application (e.g. Yahoo or IRC) for semi-synchronous, non-permanent written interactions and Evett will keep the participants contact information up-to-date on the aforementioned web page.
3. Boyd will maintain an email discussion list at Internet2 for asynchronous archived group discussions.

4. Internet2 will host a weekly conference call meeting (in addition to the current biweekly, international perfSONAR calls in which Boyd, Cottrell, Matthews, and Swany currently participate) for synchronous verbal interactions to review project status and set directions for resource allocations and milestones.

5. We will hold face-to-face meetings every six months (perhaps in conjunction with another conference or workshop) for synchronous face-to-face verbal interactions to review our progress more formally. Each team will be expected to give a short presentation on progress to date, changes in plan, and give 3-month and 6-month objectives. We will archive presentations made during such meetings on the aforementioned web page.

8. **DELIVERABLES, SCHEDULE and MILESTONES**

**Year 1:**
- Expand prototype of the Federated Lookup Service
- Prototype basic credential service (Design and develop authorization infrastructure that works with Shibboleth and eduGAIN authentication services)
- Design and develop some diagnostic tools (essentially integrate existing tools into the perfSONAR framework). Examples include BWCTL, OWAMP, Traceroute, SNMP
- Begin integrating with NMI Build and Test services [42]
- Prototype Topology Service based on US backbone data
- Prototype analysis clients and alerting services

**Year 2:**
- Merge Credential Service and Information Service in new Internet2 Network deployment
- Convert Internet2 Network Observatory Tools to perfSONAR
- Deploy Federated Lookup Services at: Internet2 backbone, ESnet backbone. Coordinate federated Lookup Service deployment with EU-funded GÉANT2 JRA-1 project. (DANTE will deploy on GÉANT2 backbone.)
- Deploy Authentication/Authorization infrastructure including Shibboleth instance. Federate this authentication server with eduGAIN, the EU authentication service, to enable transcontinental monitoring.
- Design and develop additional diagnostic tools (L2 status monitors – specifically write tools that can get statistics from TL1 devices)
- Provide analysis clients to NOCs for feedback

**Year 3:**
- Deploy additional diagnostic tools (based on feedback from users)
- Deploy analysis clients.
- Outreach year – workshops, training, support infrastructure, push university deployments.

9. **PAST NSF PROJECTS**

PI Swany is currently funded by grant CNS-0509170, entitled “An Integrated Approach to Improving Communication Performance in Clusters” currently in its second year. This effort has been very successful, with papers in Supercomputing 2005 [43-SC05] and ParCo 2005[44-Parco]. The project focuses on improving network performance for parallel computing with empirical optimization techniques. Co-PI Boyd was partially funded under grant SCI-0314723, entitled “Targeted Assistance and Instrumentation for Internet2 Applications” led by Matt Zekauskas. Jeff Boote was instrumental in creating the OWAMP and BWCTL tools; Boyd directed the E2E piPEs project, which was the overall framework in which OWAMP and BWCTL fit. In addition, Boyd, Evett, and Zekauskas participated in the Performance Measurement Architecture workshop funded by this grant, in which participants developed the idea of having a measurement framework, and highlighted work that needed to be done, much of which is being addressed by the perfSONAR framework.