Enabling Performance Intelligence for Application Adaptation in the Future Internet*

Prasad Calyam, Mukundan Sridharan, Yingxiao Xu,
Kunpeng Zhu, Alex Berryman, Rohit Patali and Aishwarya Venkataraman
OARnet/Ohio Supercomputer Center, The Ohio State University
1224 Kinnear Road, Columbus, Ohio, USA.
{pcalyam, yxu, kzhu}@osc.edu, {sridhara, berryman, rpatali, venkatar}@oar.net

Abstract

Today's Internet which provides communication channels with best-effort end-to-end performance is rapidly evolving into an autonomic global computing platform. Achieving autonomicity in the Future Internet will require a performance architecture that: (a) allows users to request and own 'slices' of geographically-distributed host and network resources, (b) measures and monitors end-to-end host and network status, (c) enables analysis of the measurements within expert systems, and (d) provides performance intelligence in a timely manner for application adaptations to improve performance and scalability.

We describe the requirements and design of one such "Future Internet Performance Architecture" (FIPA), and present our reference implementation of FIPA called 'OnTimeMeasure'. OnTimeMeasure comprises of several measurement-related services that can interact with each other and with existing measurement frameworks to enable performance intelligence. We also explain our OnTimeMeasure deployment in the Global Environment for Network Innovations (GENI) infrastructure - a collaborative research initiative to build a sliceable Future Internet. Further, we present an application-adaptation case study in GENI that uses OnTimeMeasure-enabled performance intelligence in the context of dynamic resource allocation within thin-client based virtual desktop clouds. We show how a virtual desktop cloud provider in the Future Internet can use the performance intelligence to increase cloud scalability, while simultaneously delivering satisfactory user quality-of-experience.

*This material is based upon work supported by the National Science Foundation under award numbers CNS-1050225 and CNS-0940805, VMware, and Dell. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, VMware or Dell.

1 Introduction

1.1 Future Internet Overview

Today, the Internet primarily acts as a communication channel between end-systems without any performance guarantees in the quality of end-to-end connectivity provided. This is popularly called as the 'best-effort' service model. The main difficulty in providing quality guarantees arises from the fact that no single service provider or entity controls all the network links and hosts in an end-to-end path. The current Internet is formed by federations of individual networks which peer each others traffic for mutual benefit to extend their network boundaries. Yet, they do not share each other's performance measurements as it may expose bottlenecks that are undesirable when competing for business from the same customers. Although this federated nature has been a barrier to performance transparency and quality guarantees, it is the architectural cornerstone that has enabled the current Internet to scale successfully across the world.

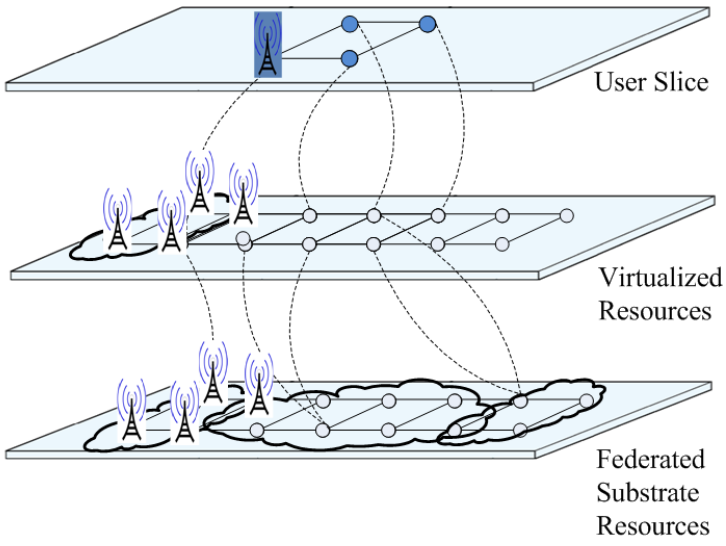


Figure 1: A User Slice in the Future Internet

Given the demanding nature of emerging multimedia-and-data rich applications that require quality guarantees, slowly but steadily a new Internet – referred to as the Future Internet (FI), from here on – is evolving. Key for the success of the FI is to retain the fundamental strength of the current Internet, i.e., the federated nature of networks, but also provide end-users with better control and performance transparency over the core Internet resources. This evolving FI is being driven primarily by improvements in virtualization [1] technology at end-hosts and core network switches/routers. The Global Environment for Network Innovation (GENI) [2] project is the early realization

of such an FI vision. GENI is funded through the U.S. National Science Foundation and is a 'sliceable' Internet infrastructure being co-developed by academia and industry. A GENI user as shown in Figure 1 can request a slice of virtualized resources from both end-host substrates –wired and wireless– and core network for experimentation of an FI application. The goal of GENI is to foster innovations in FI architectures, protocols and applications through experimentation in both controlled and real-world environments with actual users [3].

1.2 Autonomic Future Internet Applications

As the Internet evolves, the current general purpose network will need to tightly couple with applications to form a global computing platform. This platform will have customizations to deliver quality guarantees for different applications, each in their own respective user slices. In addition, every application in this environment needs to be aware of the status of host and network components that support its execution, and should have the ability to quickly adapt to any changes that limit performance and scalability within user slices. The idea of having networks being tightly coupled with applications can already be seen in the current Internet trends in the context of content delivery networks (e.g., Akamai) and private point-to-point links connecting datacenters of major application service providers (e.g., Amazon). In such tightly-coupled platforms, applications and their resources will need to be managed in a distributed manner, the applications need to rapidly-scale with user demand, and the rate-of-failure as well as time-to-repair should be as small that it does not impact user quality-of-experience (QoE). Thus, applications within user slices need the FI to have automated adaptations with autonomicity attributes such as being self-configuring, self-managing, self-monitoring, self-healing and self-optimizing.

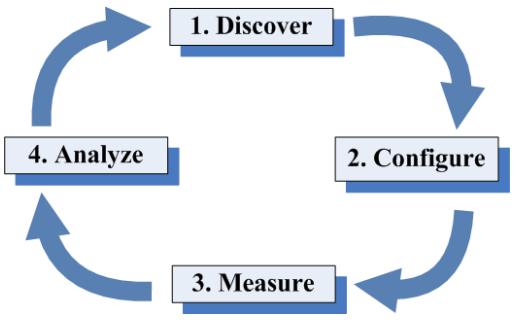


Figure 2: Autonomic FI application life-cycle

A typical autonomic FI application life-cycle is likely to go through four phases in a cyclical manner as shown in Figure 2:

1. *Discover:* In this phase, the application would identify the host and network resources needed, discover them and reserve them in a geographically-distributed user slice.

2. *Configure:* Next, the application would configure the reserved application resources meet the quality guarantees expected by the user; it would also configure other instrumentation and measurement resources during this phase at both the host and network.

3. *Measure:* Next, the application would start monitoring the performance of the underlying host and network resources, as well as its own behavior.

4. *Analyze:* Lastly, the application would feed the measurements into expert-systems to derive *performance intelligence*, which can inturn be used to timely trigger reconfiguration of resources or healing of application behavior to improve performance and scalability within a user slice.

Although these phases are functionally separate, they are likely to be automated to operate in a continuous manner during an application's lifetime within a user slice. Nevertheless, some of the tasks in the FI application life-cycle phases will still require human involvement at the application service provider level, particularly for adaptations that may be critical and could affect large numbers of user slices.

1.3 Future Internet Performance Intelligence

Given that enabling performance intelligence in a timely manner is central to support the autonomic nature of FI applications, measurements that indicate 'host-and-network status awareness' in a user slice need to be carefully instrumented, orchestrated, archived and analyzed. Thus, we can envisage a performance architecture for the FI viz., "FIPA" comprising of several services that interoperate with each other and with existing measurement frameworks such as Cricket [4] and perfSONAR [5]. Together, they will satisfy specific performance intelligence use cases of FI applications both at the user as well as application service provider levels. FIPA should also enable FI applications to use performance intelligence information at the hop, link, path and slice levels for determining actionable instances for adaptation. Further, FIPA should facilitate validation of the adaptation effectiveness with additional performance intelligence in the next life-cycle iteration, and so on. Existing works that share our vision of performance intelligence for management of the FI include [6] - [8].

1.4 Paper Organization

The remainder of this paper is organized as follows: In Section II, we detail the design motivations and functionalities of our envisioned FIPA services. In Section III, we present our reference implementation of FIPA called 'OnTimeMeasure' [9] in the GENI infrastructure. We describe the various OnTimeMeasure software modules that can be deployed in both centralized and distributed manners. In addition, we describe how the OnTimeMeasure software modules interoperate with other GENI components at the substrate plane, application plane, and measurement plane. Further, we describe how primitives in OnTimeMeasure can be customized to monitor application-specific metrics. In Section IV, we present an application-adaptation case study in a GENI slice. The application-adaption uses OnTimeMeasure-enabled performance intelligence in

the context of dynamic resource allocation within thin-client based virtual desktop clouds (VDCs). We show how a virtual desktop cloud provider (CSP) in the FI can use the performance intelligence to increase cloud scalability by supporting a higher number of virtual desktops per datacenter, while simultaneously delivering satisfactory user QoE. Section V concludes the paper.

2 Future Internet Performance Architecture

2.1 Design Motivations

The core design motivations of the FIPA can be categorized into two broad perspectives. The first perspective relates to satisfying the use-cases in the FIPA design from the user side as well as from the application service provider side. The second perspective relates to design principles that make the FIPA extensible, fault-tolerant, standards-compliant and secure. In the following, we describe the design motivations for FIPA based on these two perspectives. We remark that the below descriptions are a combination of the authors' original ideas, as well as are based on feedback from group discussions in the GENI community forums [10].

The primary use cases from the user side include: (i) Have I got the system and network resources I asked (i.e., purchased) in my slice? (ii) Is my slice environment functioning to my expectation over my slice lifetime? (iii) Can a problem occurence in my slice environment that impacts my QoE be identified and notified to my application to adapt and heal? (iv) Can problem information also be shared with me and my application service provider if my application cannot automatically heal itself?

Correspondingly, the primary use cases from the application service provider side include: (i) Can I check if the user got the system and network resources he/she asked (i.e., purchased) in user's slice? (ii) Can I monitor all the detailed active (e.g., Ping, Traceroute, Iperf) and passive (e.g., TCPdump, Netflow, Router-interface statistics) measurements at end-to-end hop, link, path and slice levels across multiple federated ISP domains? (iii) Can I analyze all the measurements to offline provision adequate resources to deliver satisfactory user QoE, and online i.e., real-time identify anomalous events impacting user QoE?

To make the FIPA extensible, fault-tolerant, standards-compliant and secure, the FIPA can be designed based on principles of service-oriented-architecture. In that vein, the FIPA for extensibility will need to be comprised of a set of specialized services, each with a defined functionality and web service interface specification. This should allow the services to interoperate with each other through communication layers by exchanging *control* messages in order to request, generate, archive, discover, analyze and present *data* from various measurement sources. The architecture is inherently fault-tolerant because of the distributed and modular nature of the design. Entire services or individual modules can be replicated to avoid failure or for load-balancing purposes. In terms of standards-compliance, the potential choices could include SOAP/XML-RPC based messaging services for information exchange between the various services, Resource Description Framework formats adopted in GENI for resource discovery and setup [11], as well as request/response schemas of measurements being standardized

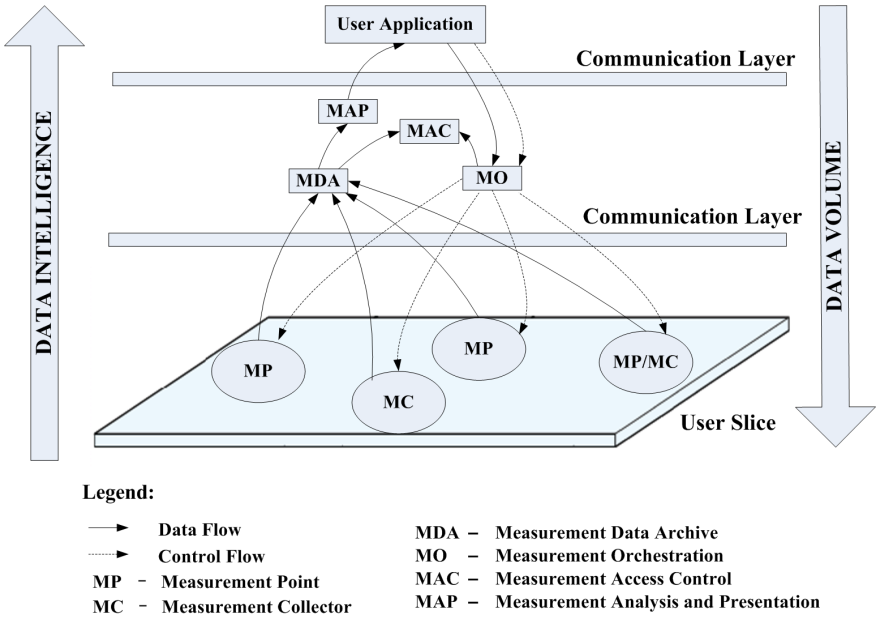


Figure 3: FI Performance Architecture

in the Open Grid Forum [5]. Further, the security issues of FIPA can be encompassed by existing advances in federated identification such as OpenID and Shibboleth [10].

2.2 FIPA Services

Figure 3 shows the FIPA services interacting with each other via communication layers in such a manner that the measurement data volume collected at the slice resources is 'comprehensive' i.e., large in terms of metrics to quantify performance of the end-to-end hop, link, path and slice levels. However, as the performance intelligence is processed and delivered to the application, the measurement data volumes are 'abstract' i.e., small in terms of metrics to guide actionable instances for adaptation. In the following, we briefly describe the composition and functions of each of our envisioned FIPA services:

Measurement Point (MP) Service: An MP refers to the instrumentation that taps into "measurement data sources" such as network hops and/or system nodes (i.e., passive measurements), or probes end-to-end network paths (i.e., active measurements) to capture and format measurement data. Several MPs are installed within the user slice such that they execute the tap/probe control utilities. MP services provision wired network measurements such as availability, TCP/UDP throughput, available bandwidth, delay, jitter and loss. In addition, MP services provision wireless network measurements such as utilization, signal strength, and signal-to-interference/noise ratio.

Measurement Collection (MC) Service: MC is comprised of programmable components that collect, and locally store measurement data using database design principles

that allow efficient storage and rapid query capabilities. They can co-reside with MPs on the same host resources or can be installed on separate hosts within an user slice. MC services support measurement data query with pull/push or pub/sub mechanisms using protocols such as SNMP, SCP, or HTTP.

Measurement Orchestration (MO) Service: This service relies on an interpreter to gather measurement requests from authorized users via command-line or graphical user interfaces. It manages various MPs/MCs via schedulers to perform on-going sampling (e.g., periodic, random, stratified random, adaptive) and on-demand sampling of both active and passive measurements. The MO service is aware (through discovery/lookup interfaces) of slice topology, control actors, control actions, data sources and data types in order to cater measurement requests with any specific timing demands. The MO service further enforces resource policy constraints such as how much bandwidth to use, and which measurements have higher priority when concurrent measurements need to be orchestrated under resource contention scenarios.

Measurement Data Archive (MDA) Service: This service allows MPs and MCs to publish a stable archive of measurement data and meta-data, for purposes such as: sharing performance data, analyzing anomaly events, enabling measurement correlations for ground-truth verification (about cause of anomaly event), and making it possible to provide long term access of measurement data for authorized users. It also allows cataloging measurement data indexes for future search and retrieval purposes.

Measurement Analysis and Presentation (MAP) Service: This service is comprised of programmable components that analyze and visualize measurement data in MDA for authorized users. It supports various configurable dashboard tools, and other data charting toolkits. It uses the measurement data to analyze slice-level metrics; for example, a user will be able to check conformance of resource allocation to the resources assigned (i.e., purchased) with slices, and for on-going monitoring of the user slice environments. This service enables applications to detect impending or perceived problems for timely adaptation.

Measurement Access Control (MAC) Service: At the heart of every FIMA service are the access control and data privacy mechanisms that are needed to authenticate users or service entities in order to ensure intended use of measurement resources and data collections. This service leverages federated user and policy databases to allow other FIMA services to interoperate securely with each other.

3 OnTimeMeasure: FIPA Reference Implementation

3.1 Software Modules

Figure 4 shows the four main software modules in OnTimeMeasure [9] in a GENI slice: (i) Node Beacon, (ii) Root Beacon, (iii) OnTime Beacon, (iv) OnTime Control. The Node Beacon module supports the MP and MC services. The Root Beacon supports the MO, MAP, and MDA services. GENI user installs the Node and Root Beacons i.e., an OnTimeMeasure measurement instance in the user's slice based on the MAC services provided by GENI wired and wireless substrate providers (e.g., ProtoGENI [19], PlanetLab [20]). To control the measurement services hosted by the Node and Root

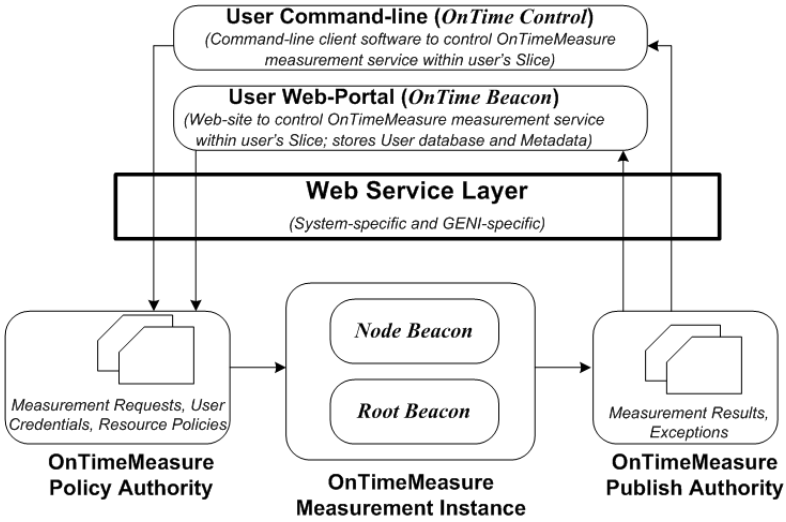


Figure 4: OnTimeMeasure Software Modules

Beacons within a GENI user slice, a GENI user can interactively use either the web-interface client viz., OnTime Beacon, or invoke automated scripts through command-line interface client viz., OnTime Control. The OnTime Beacon also has "Administrator" privileges, where a substrate provider or measurement service provider or even GENI operations can host the user web-portal, manage measurement service user credentials, discover and manage measurement service instances in user slices, and view meta-data about user-initiated slice measurements.

A "Policy Authority" is part of the OnTime Beacon web-portal that determines the list of measurement capabilities that a user is permitted to request in his/her slice. The measurement capabilities permitted are based on pre-defined resource policies in the measurement topology. The policies enforced for measurements scheduling include: (a) semantic priorities based on user roles to resolve measurement scheduling conflicts, and (b) measurement level restrictions (e.g., allowable measurement bandwidth and measurement duration) to regulate the amount of network measurement traffic permitted on network paths. Once a measurement request has been cleared by the Policy Authority, the measurement request will be processed to generate slivers in experiment slices with installs of Node and Root Beacon software. In addition, the user can add/remove a set of measurement tasks for the OnTimeMeasure scheduler to process and generate measurement timetables for Node Beacons. Further, the user can start, stop and also check status of the various measurement service components. The measurement service components whose status is reported include: (a) Slice Accessibility, (b) Root Beacon Scheduler, (c) Node and Root Beacon Communications, (d) Measurement Data Collector, (e) Analysis and Publish Authority, and (f) Measurement Data Visualization. We remark that the "Policy Authority" can be integrated within any enterprise policy-based management system [29] in order to implement any business policies and procedures to configure/control compute, network and storage resources,

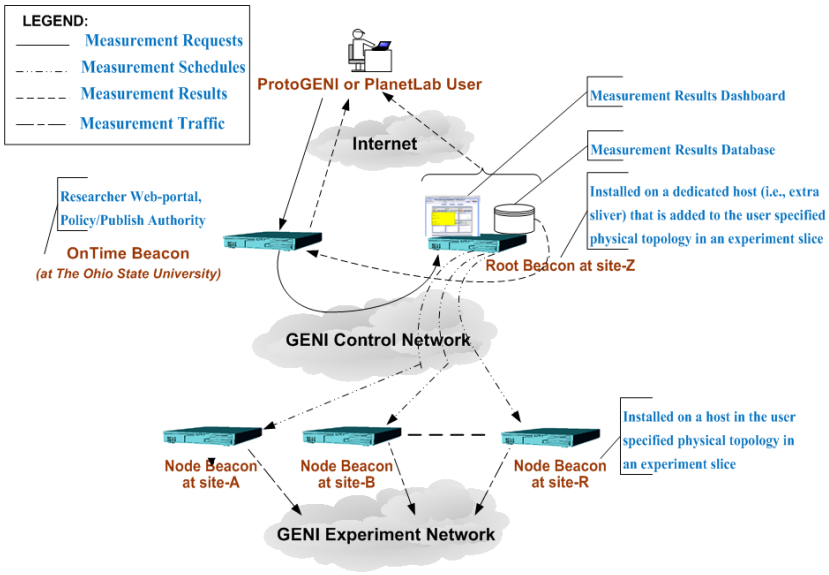


Figure 5: Centralized OnTimeMeasure Deployment

and their associated user-application services. Further, the performance intelligence provided by OnTimeMeasure can also be used reversely as stimuli to the policy-based management systems to modify the business policies and procedures. The "Publish Authority" is also part of the OnTime Beacon web-portal that provides raw and processed measurement results back to the user. The processed measurement results could correspond to: (a) time series of active measurements over a time range, (b) time series of active measurements over a time range with associated network performance anomaly events, and (c) time series of active measurements over a time range with associated network weather forecasts.

3.2 Software Deployment

The OnTimeMeasure software modules can be deployed either using a "Centralized Orchestration" architecture, or a "Distributed Orchestration" architecture. The centralized orchestration allows measurement scheduling for continuous monitoring, persistent measurements storage in an user's experiment slice and processed network measurement feeds. Such functionality will generally be useful for "network weathermaps" and long-standing experiments with advanced measurement analysis capabilities. In comparison, the distributed orchestration allows measurement scheduling of on-demand measurement requests without need for persistent measurements storage. Such functionality will generally be useful for users needing one-off or occasional raw measurement tool outputs.

Figure 5 shows the OnTimeMeasure deployment for centralized orchestration in GENI. We can see that the GENI user interacts with the Policy Authority and Publish

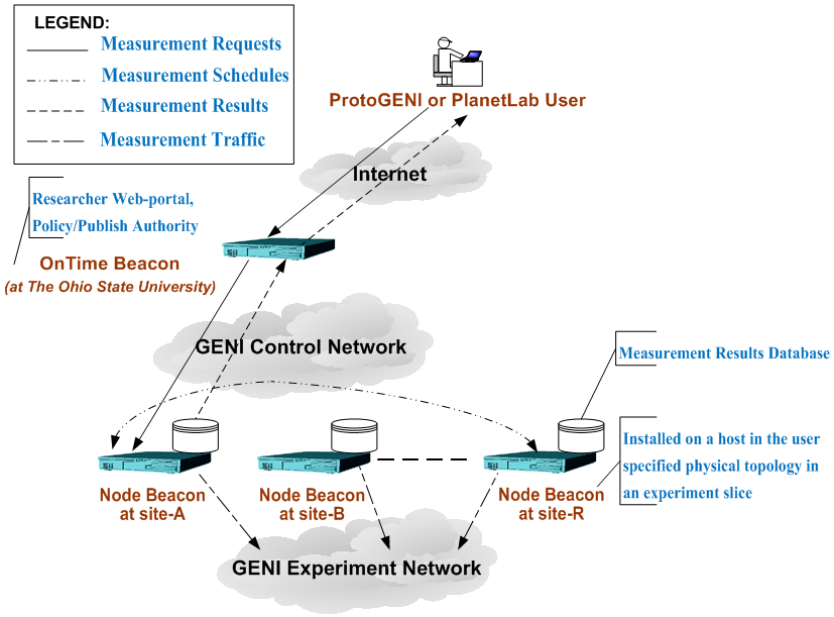


Figure 6: Distributed OnTimeMeasure Deployment

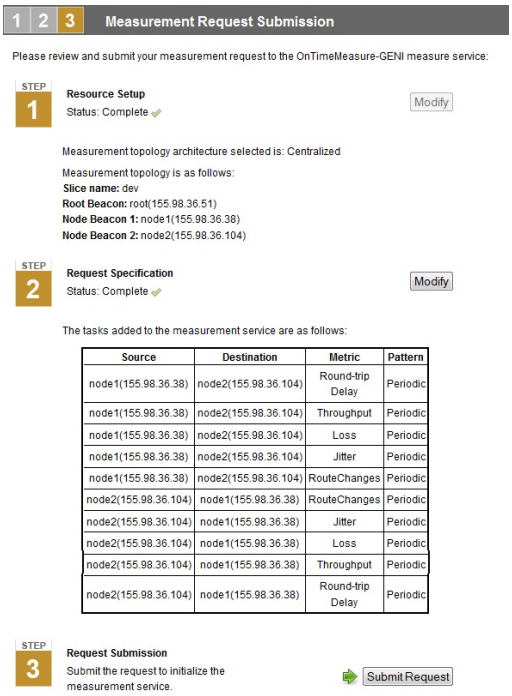


Figure 7: Centralized Measurement Request Submission

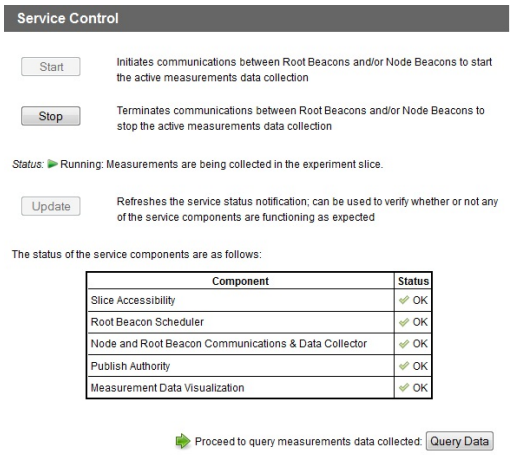


Figure 8: Service Control

Authority in GENI using the Internet. The Root Beacon does not rely on experimental network that carries the FI application traffic to interact with the Node Beacons, but instead uses the GENI Control Network. The reason is that the FI application traffic can become unstable and lead to network partitions since a GENI user sometimes may be exploring with experimental protocols and software. The active measurements collected will be on the experimental network paths to which the Node Beacons will be connected. Figure 6 shows the GENI deployment for distributed orchestration. No Root Beacon will be installed when a user demands a distributed orchestration for the measurement service. The Node Beacons again do not rely on experimental network to interact with each other for exchanging measurement schedules, but instead use the GENI Control Network. The measurements at Node Beacons will be collected on the experimental network paths.

OnTimeMeasure web services allow GENI users to interact with the measurement service functions. Both "system-specific" web-services with our service definitions, and "GENI-specific" web services with GENI compliant schemas i.e., those standardized within the GENI community can be supported. The measurement functions supported by OnTimeMeasure can be broadly grouped under: (i) Measurement Request Submission (shown in Figure 7), (ii) Measurement Service Control (shown in Figure 8), and (iii) Measurement Results Query (shown in Figure 9). In the case of centralized orchestration, dashboards are supported, whereas in the case of distributed orchestration, GENI user measurement requests are handled in real-time as shown in Figure 10.

3.3 GENI Interoperability

We now describe the ability of OnTimeMeasure to interoperate with other GENI components at the substrate plane, application plane, and measurement plane. OnTimeMeasure has been developed to discover and configure *substrate resources* in the GENI's ProtoGENI [19] and PlanetLab [20]. Both these substrates allow GENI users to create

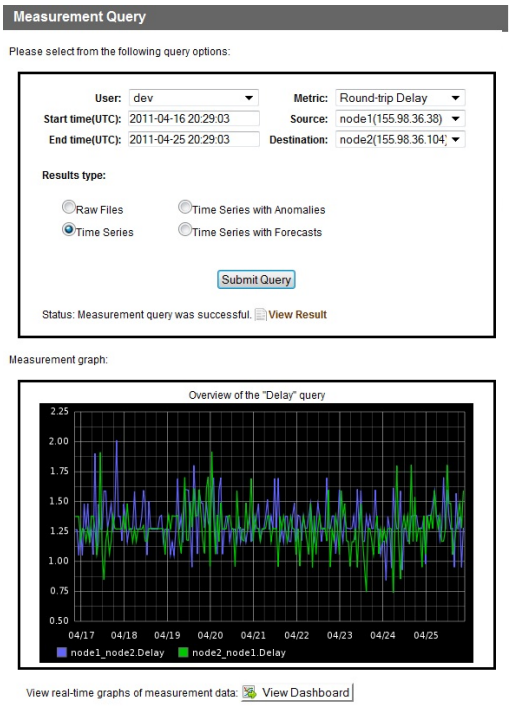


Figure 9: Centralized Measurement Query

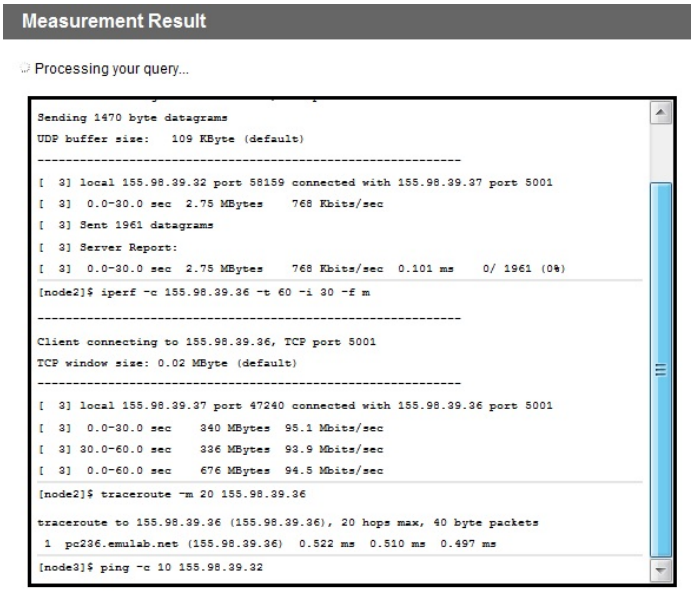


Figure 10: Distributed Measurement Result

slices that can be registered with OnTimeMeasure, which inturn allows GENI users to control the measurement service functions with OnTime Beacon or OnTime Control. In addition, OnTimeMeasure has been integrated with an experimenter workflow tool viz., GENI User Shell (Gush) [21] that allows a user to control *experiment applications* within ProtoGENI and PlanetLab via a user-shell commands. Further, OnTimeMeasure has been integrated with several other *measurement frameworks* such as the GENI Instrumentation Tools (InsTools) [22], as well as the VMware power shell tools (PowerTools) [23]. Each of these tools access several measurement data sources. For e.g., InsTools allows a GENI user to obtain passive packet-captures, SNMP and NetFlow measurements; PowerTools allows a GENI user to monitor virtualized environments in terms of resource consumption at the virtual machine level, as well as the host level. In the GENI measurement plane context, OnTimeMeasure has also been integrated into the Digital Object Repository [24] that allows archiving experiment slice measurement datasets along with meta-data collected by OnTimeMeasure into the GENI Measurement Data Archive. Use Cases supported with the Digital Object Repository integration include: (a) archive and share subsets of experiment results, (b) archive and share entire experiment slice measurement results, and (c) backup/restore entire experiment slice measurement results.

Given a FI possiblity where OnTimeMeasure interoperates with several measurement frameworks, there will arise a situation where there will be large number of measurement requests that need to be orchestrated in FI slices. The measurement orchestration needs to be such that it does not affect application (e.g., CPU cycles needed for monitoring of a resource's performance should not impact the CPU cycles needed for actual application) and also measurement collection should not produce conflicts (e.g., two Iperf tools invoked at the same time along common hosts or network paths) [25]. Orchestration should also consider minimizing "cycle time" so that more frequent real-time network status updates can be provisioned to application. Cycle time is defined as the time taken to complete one round of measurements execution over a measurement topology. Smaller the cycle time, larger the number of measurement rounds, and thus more detailed network status can be sampled. To reduce the cycle time, it is essential to reduce the idle periods in measurement schedules for each cycle time, while still not affecting the application and avoiding measurement conflicts. However, existing works [25] [26] [27] use "open loop" principles for scheduling measurements that conservatively guess measurement execution times. This conservative guessing results in many idle periods in measurement execution schedules. Consequently, existing open loop methods do not scale to accommodate FI measurement loads.

OnTimeMeasure's web-services architecture easily allows measurement tools at MPs to provide feedback i.e., in a "closed loop" manner to the MO service when their measurement execution is completed. Leveraging this intelligence, MO could schedule other queued measurements on MPs in the idle periods that do not affect the application or cause conflicts with other scheduled measurements. Figure 11 and Figure 12 show how OnTimeMeasure handles full-mesh, tree and hybrid measurement topologies for increasing number of measurement requests at MPs, and produces significant improvements in cycle times. The results were obtained by running an instance of OnTimeMeasure in a GENI slice comprising of 21 servers in ProtoGENI. The full-mesh corresponds to the case where measurements are collected between all the servers over

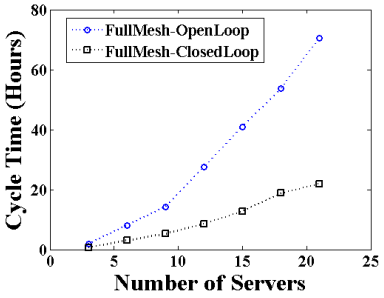
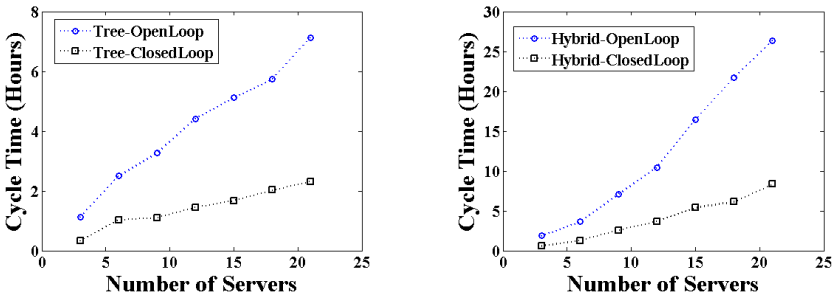


Figure 11: Full Mesh Measurement Topology Comparison

the entire measurement topology. The tree case is where measurements are collected only between neighboring servers in the measurement topology. Lastly, the hybrid case is where measurements are collected in a full-mesh manner for a subset of servers in the measurement topology, and tree manner for the other servers that are not part of the full-mesh subset. In all these 3 cases, we can clearly see that OnTimeMeasure's closed loop implementation can be leveraged to notably improve cycle times in MO in comparison to existing open loop implementations, and thus OnTimeMeasure is suited for handling large number of measurement requests that will need to be orchestrated in FI slices.



Tree Measurement Topology

Hybrid Measurement Topology

Figure 12: Comparison of Open Loop and Closed Loop Measurement Orchestration

3.4 GENI Customization

We now describe how OnTimeMeasure can be customized via web services to work with any application-specific metrics using a feature called "*Custom Metric Integration*". This feature is similar to a capability in the Amazon CloudWatch measurement service [30], where customers of Amazon Web Services can integrate their custom application metrics into their monitoring activities using web services. Such a cus-

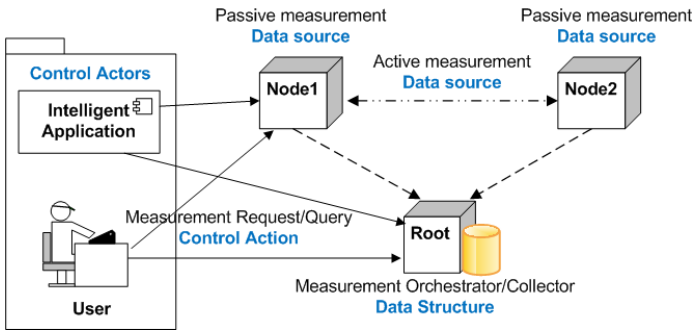


Figure 13: OnTimeMeasure' "Custom Metric Integration" Feature Primitives

tomization is a key requirement since each GENI experiment is unique; often, users want to not only use off-the-shelf tools (e.g., Ping, Netflow) within measurement services in GENI, but also other tools they have developed or others have developed that are relevant to their experiment. Figure 13 shows the primitives (*Control Actors*, *Data Sources*, *Data Structures*, *Control Actions*) that need to be configured by the users within OnTimeMeasure in the custom metric integration process.

To better understand the primitives, let us consider a case study of dynamic resource allocation within thin-client based virtual desktop clouds (VDCs), where allocations are based on OnTimeMeasure' feedback of network health from the client, and load on the server (for details, please refer to Section IV). The specification of primitives in the VDC experiment can be illustrated as follows:

1. *Control Actors*: Actors refer to VDC experiment users that access the application data or share the application data with other actors.

2. *Data Sources*: These are resource monitors or data generation tools deployed in Node Beacons within an experiment slice; the tools communicate with each other to perform active measurements or collect passive measurements within a host in an on-going or on-demand basis. A specific data source example is PowerTools that monitors memory, CPU, and network health of each host.

3. *Data Structures*: These refers to the sets of measurement data types that would be stored in the Root Beacon database. Examples of data types can be 'Host: `string`', 'Resource Pool: `string`', 'Sample Time: `timestamp`', 'CPU Load: `float`'.

4. *Control Actions*: They relate to the actor controls such as start/stop times of the data generation tools, or sampling frequency. Example of a control action for sampling can be to configure PowerTools to query memory usage every 10 seconds.

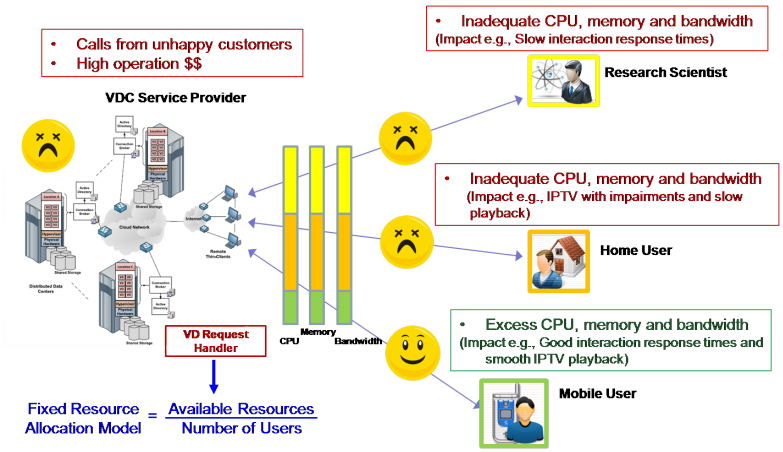


Figure 14: Illustration of F-RAM limitations in Virtual Desktop Clouds

4 FIPA Case Study: Dynamic Resource Allocation in Virtual Desktop Clouds

We now present an application-adaptation case study in a GENI slice that is based upon a demonstration we presented during the 10^{th} GENI Engineering Conference (GEC10) [2]. The demonstration uses OnTimeMeasure-enabled performance intelligence in the context of dynamic resource allocation within thin-client based virtual desktop clouds (VDCs). In the following, we first provide an overview of the resource allocation problem. Next, we describe application-adaptation that uses OnTimeMeasure-enabled performance intelligence information within a VDC. Lastly, we present our GENI slice setup for our VDC experiments and the obtained experiment results.

4.1 The Resource Allocation Problem

We consider a futuristic scenario, where users at home or at enterprises subscribe for virtual desktops (VDs) with a VDC service provider (CSP), who ships to subscribers set-top boxes that function as thin-clients. The drivers for users to transition from traditional desktops to VDCs are obvious: (i) desktop support in terms of operating system, application and security upgrades will be easier to manage, (ii) the number of underutilized distributed desktops unnecessarily consuming power will be reduced, and (iii) mobile users will have wider access to their applications and data.

Currently, CSPs allocate resources to VD requests based primarily on CPU and memory measurements [12] - [15]. There is surprisingly sparse work [16] [17] on resource adaptation coupled with measurement of network health and user quality of experience (QoE). It is self-evident that any cloud platform's capability to support large user workloads is a function of both the server-side desktop performance as well as the remote user-perceived QoE. In other words, a CSP can provision adequate CPU and memory resources to a VD in the cloud, but if the thin-client protocol configuration

does not account for network health degradations and application context, the VD will be unusable for the user. Also, the CSP has to allocate resources to individual Virtual Desktops (VDs) in a way that decreases the resource consumption cost, while at the same time maximizes the user QoE. Hence, lack of proper performance intelligence in terms of "network-and-human awareness" in cloud platforms inevitably results in costly guesswork and over-provisioning while managing physical device and human resources, which consequently annoys users due to high service cost and unreliable quality of experience.

Figure 14 illustrates the above drawback of current resource allocation schemes in a typical VDC provider. We can see that the fixed resource allocation models (F-RAM) used by CSPs results in situations where a Mobile user is provisioned excess resources that results in satisfactory QoE (Q_{excess}), whereas a Scientist or Home user are provisioned with inadequate resources, which results in unsatisfactory QoE ($< Q_{min}$), and the CSP incurs a high cost of operation. This problem is compounded by the fact that users use a wide range of applications (e.g., Web browsing, Video playback) with different resource consumption profiles, and connect to data centers within a VDC over network paths with different network health conditions.

4.2 Application Adaptation

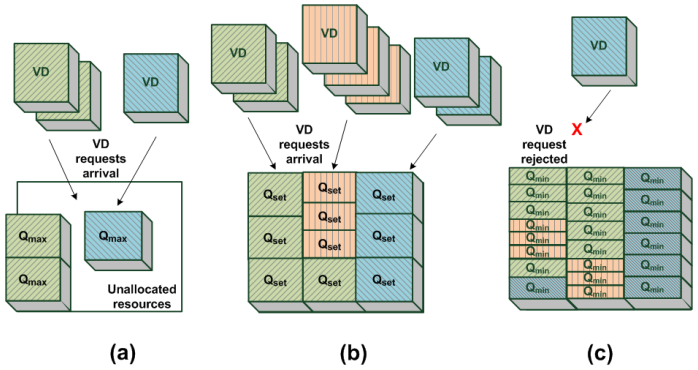


Figure 15: U-RAM allocations when there are: (a) New VD requests handling with freely available resources, (b) New VD requests handling with all available resources allocated, and (c) New VD request rejected when SLA violation situation occurs

To overcome this problem, we compare F-RAM with a utility-directed[1] resource allocation model (U-RAM) that builds upon earlier works such as QoS-based resource allocation model (Q-RAM), originally proposed in [28]. The U-RAM adapts the resource (i.e., CPU, memory and network bandwidth) allocation dynamically based on the real-time performance intelligence feedback from system, network and user QoE measurements that can be collected using OnTimeMeasure.

[1]A utility function indicates how much of application performance can be increased with larger resource allocation. Beyond a certain point, application utility saturates and any additional resource allocation fails to further increase application performance.

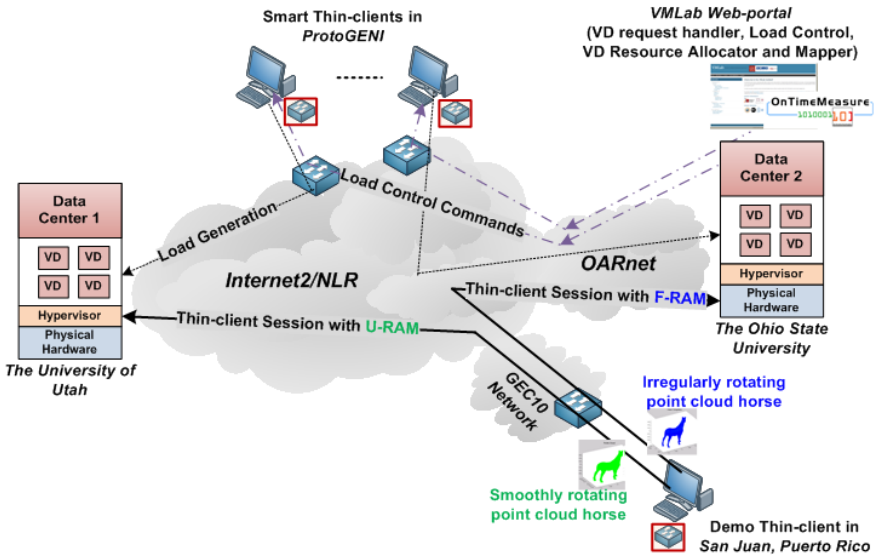


Figure 16: Demonstration of Virtual Desktop Cloud with U-RAM in GENI

The fundamentals that guide our U-RAM solution can be illustrated by the example shown in Figure 15 involving 3 desktop pools being provisioned at a data center site within a VDC. We can see that U-RAM allocates resources to a new VD request such that it results in utility saturation quality (Q_{max}) when there are freely available resources. When all available resources are allocated under light loads, U-RAM allocates resources to new VD requests such that all the VDs operate below Q_{max} but above Q_{min} i.e., Q_{set} range. Lastly, when all available resources are allocated under heavy loads, U-RAM allocates Q_{min} resources to both the already allocated VDs as well as in the case of the new VD. Once it is determined that provisioning a new VD request will cause violation of user service level agreement (i.e, resource allocation causes quality to drop below Q_{min}), all new VD requests arriving at the data center thereafter are either rejected or directed to alternate data center resource sites with the VDC. The key to improved performance in U-RAM is the performance intelligence feedback from OnTimeMeasure that enables U-RAM to continously *discover* the required resources in a VDC, *configure* desktop pools to suit different user application groups, *measure* and *analyze* their performance in order to re-discover and re-configure resources, and so on - to improve the overall cloud scalability, while delivering satisfactory user QoE.

4.3 Application Setup in GENI

The experiment setup on GENI is shown in Figure 16. We created two GENI slices for the experiment, one slice for the VDC data center environment and the other slice for distributed thin-client user sites. The 'data center' slice consisted of two data centers, one at The Ohio State University and the other at The University of Utah. The data

Algorithm	#VDs Requested/Allowed	Total /Minimum CPU	Allocation
U-RAM	12/10	8 GHz /800 MHz	800 MHz
F-RAM-1	12/12	8 GHz /800 MHz	500 MHz
F-RAM-2	12/5	8 GHz /800 MHz	1.4 GHz

Table 1: Comparison of VD connections handled by U-RAM and F-RAM in GENI Demonstration

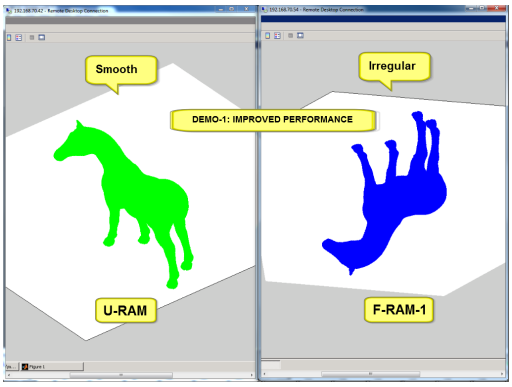


Figure 17: U-RAM versus F-RAM-1 to show Improved Performance

center at Utah ran the U-RAM algorithm and the data center at Ohio ran the F-RAM algorithm. A number of thin-clients were launched from geographically distributed locations within the 'user slice' to make VD requests to either data center for load generation. The 'user slice' included a demo site at the conference venue in Puerto Rico which initiated two separate thin-client connections to each data center. These two demo thin-clients were running identical Matlab-based point-cloud animations of a horse (but with different colors; green for U-RAM and blue for F-RAM), to compare their application performance. Node Beacon instances (integrated with PowerTools) of the OnTimeMeasure framework were deployed at all the thin-client sites in the 'user slice' to measure the network and system performance, and a Root Beacon instance of the OnTimeMeasure framework was deployed at the Ohio data center to collect the Node Beacon measurements.

4.4 Application Adaptation Results from GENI

The collected measurements in the Root Beacon were used by the U-RAM algorithm to make dynamic resource allocation decisions. We illustrate this decision making using two scenarios we demonstrated at GEC10, one where the U-RAM delivers "improved performance" than F-RAM, and in the other where U-RAM achieves "increased scalability" (i.e., more number of VDs are handled) while guaranteeing the same level of user QoE performance compared to F-RAM.

The performance comparison results between U-RAM and F-RAM at GEC10 are summarized in Table 1. Each of the two data centers we deployed allocates 8 GHz

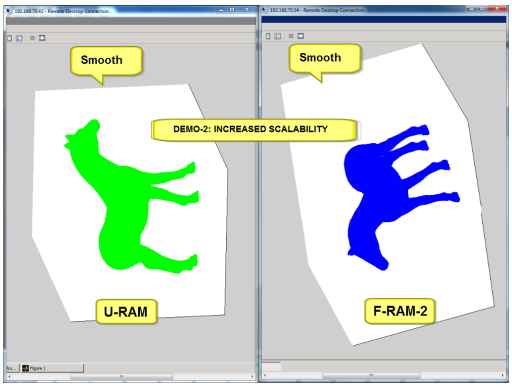


Figure 18: U-RAM versus F-RAM-2 to show Increased Scalability

of total CPU resources and each user needs at least 800 MHz allocation for satisfactory user QoE performance; we characterized satisfactory user QoE performance as the "smooth" rotation of the point-cloud horse animation in Matlab, and unsatisfactory user QoE performance as the "irregular" rotation. The U-RAM algorithm with such performance intelligence functions as follows: consider a case when there are 12 user VD connection requests, it will accept 10 of them and rejects 2 requests. The 10 users will be allocated 800 MHz each and all of them will deliver satisfactory user QoE performance. U-RAM's rejection of 2 VD connections is justified, since these 2 connections will anyway not deliver satisfied user QoE performance even if they were allocated resources and will actually affect the QoE performance of the already allocated VDs. On the other hand, for the same case, two different F-RAM allocations are possible, one in which F-RAM under-allocates (defined as F-RAM-1 in Table 1) i.e., 500 MHz per VD, and the other in which F-RAM over-allocates (defined as F-RAM-2 in Table 1) i.e., 1.4 GHz per VD. In the under-allocation possibility, F-RAM-1 accepts all 12 connections, but none of them deliver satisfactory user QoE since all of them are not allocated adequate resources, while 2 GHz CPU capacity is still available for allocation. This results in an irregular point-cloud animation rotation as illustrated in Figure 17. In the over-allocation possibility, F-RAM-2 delivers satisfactory user QoE performance for only 5 VD users, while rejecting 7 connections. U-RAM compared to F-RAM-2 supports more VDs while guaranteeing the same level of user QoE performance, and hence achieves better scalability, as illustrated in Figure 18. Thus, the OnTimeMeasure framework enables performance intelligence in novel adaptive applications such as dynamic resource allocation in VDCs in the FI.

5 Conclusions and Outlook

The FI, that is slowly but steadily evolving, is likely to be application centric, where the network and application are tightly coupled to deliver improved performance. Application service providers will own and operate 'slices' on the FI, which can be configured

to suit user needs at Internet-scale. In this paper, we envisioned a Future Internet Performance Architecture viz., FIPA that is extensible, fault-tolerant, standards-compliant and secure. In addition, we described our reference implementation of FIPA called 'OnTimeMeasure' that supports services to 'measure' the performance within a user slice, 'analyze' and derive 'intelligence' that can be used in timely manner for application adaptations to improve performance and scalability.

We evaluated the effectiveness of OnTimeMeasure in the GENI infrastructure, which is an FI platform being developed by academia and industry partnerships. We demonstrated the effectiveness of closed-loop orchestration used in OnTimeMeasure, which is critical for interoperability with other existing measurement services, particularly when large number of application specific measurements need to be orchestrated. Lastly, we used OnTimeMeasure-enabled performance intelligence to compare utility-driven resource allocation schemes in virtual desktop clouds. Our results from the GENI infrastructure experiments demonstrated how performance intelligence enables autonomic nature of FI applications to mitigate the costly resource overprovisioning and user QoE guesswork, which are common in the current Internet.

References

[1] L. Peterson, S. Shenker, J. Turner, "Overcoming the Internet Impasse through Virtualization", *Proc. of Hot Topics in Networks*, 2004.

[2] GENI: Global Environment for Network Innovation. `http://www.geni.net`

[3] A. Feldmann, "Internet clean-slate design: what and why?", *ACM SIGCOMM Computer Communication Review*, Vol. 37, No. 3, 2007.

[4] J. Allen, "Driving by the Rear-View Mirror: Managing a Network with Cricket", *Proc. of USENIX Network Administration Conference*, 1999.

[5] A. Hanemann, J. Boote, E. Boyd, J. Durand, L. Kudarimoti, R. Lapacz, M. Swany, S. Trocha, J. Zurawski, "PerfSONAR: A Service Oriented Architecture for Multi-Domain Network Monitoring", *Proc. of Service Oriented Computing, Springer Verlag, LNCS 3826*, pp. 241-254, 2005. `http://www.perfsonar.net`

[6] C. Mingardi, G. Nunzi, D. Dudkowski, M. Brunner, "Event Handling in Clean-slate Future Internet Management", *Proc. of IEEE/IFIP Integrated Network Management*, 2009.

[7] S. Kim, M. Choi, H. Ju, M. Ejiri, J. Hong, "Towards Management Requirements of Future Internet", *Challenges for Next Generation Network Operations and Service Management*, Springer LNCS, Volume 5297, Pages 156-166, 2008.

[8] L. Mamatas, S. Clayman, M. Charalambides, A. Galis, G. Pavlou, "Towards an Information Management Overlay for the Future Internet", *Proc. of IEEE/IFIP NOMS*, 2010.

[9] OnTimeMeasure: Centralized and Distributed Measurement Orchestration Software. http://groups.geni.net/geni/wiki/OnTimeMeasure

[10] GENI Instrumentation and Measurement Architecture Community Resource. http://groups.geni.net/geni/wiki/GeniInstrumentationandMeasurementsArchitecture

[11] I. Baldine, Y. Xin, M. Anirban, C. Heermann, J. Chase, V. Marupadi, A. Yumerefendi, D. Irwin, "Networked Cloud Orchestration: A GENI Perspective", *IEEE Workshop on Management of Emerging Networks and Services (MENS)*, 2010.

[12] D. Gmach, S. Krompass, A. Scholz, M. Wimmer, A. Kemper, "Adaptive Quality of Service Management for Enterprise Services", *ACM Transactions on the Web*, Vol. 2, No. 8, Pages 1-46, 2008.

[13] P. Padala, K. Shin, et. al., "Adaptive Control of Virtualized Resources in Utility Computing Environments", *Proc. of ACM SIGOPS/EuroSys*, 2007.

[14] B. Urgaonkar, P. Shenoy, et. al., "Agile Dynamic Provisioning of Multi-Tier Internet Applications", *ACM Transactions on Autonomous and Adaptive Systems*, Vol. 3, No. 1, Pages 1-39, 2008.

[15] H. Van, F. Tran, J. Menaud, "Autonomic Virtual Resource Management for Service Hosting Platforms", *Proc. of ICSE Workshop on Software Engineering Challenges of Cloud Computing*, 2009.

[16] L. Grit, D. Irwin, A. Yumerefendi, J. Chase, "Virtual Machine Hosting for Networked Clusters: Building the Foundations for Autonomic Orchestration", *Proc. of Workshop on Virtualized Technology in Distributed Computing*, 2006.

[17] A. Berryman, P. Calyam, A. Lai, M. Honigford, "VDBench: A Benchmarking Toolkit for Thin-client based Virtual Desktop Environments", *Proc. of IEEE CloudCom*, 2010.

[18] N. Agoulmine, S. Balasubramaniam, D, Botvitch, J. Strassner, E. Lehtihet, W. Donnelly, "Challenges for Autonomic Network Management", *Proc. of Conference on Modelling Autonomic Communication Environment*, 2006.

[19] ProtoGENI: A GENI Wired and Wireless Substrate. http://groups.geni.net/geni/wiki/ProtoGENI

[20] PlanetLab: A GENI Wired Substrate. http://groups.geni.net/geni/wiki/PlanetLab

[21] Gush - GENI User Shell. http://groups.geni.net/geni/wiki/GushProto

[22] Intrumentation Tools: A GENI Instrumentation and Measurement Service. http://groups.geni.net/geni/wiki/InstrumentationTools

[23] VMware Power Tools: Virtual Infrastructure Administration Scripts. http://www.vmware.com

[24] Digital Object Repository: A GENI Measurement Data Archive Service. http://groups.geni.net/geni/wiki/DigitalObjectRegistry

[25] P. Calyam, C.-G. Lee, E. Ekici, M. Haffner, N. Howes, "Orchestrating Network-wide Active Measurements for Supporting Distributed Computing Applications", *IEEE Transactions on Computers*, 2006.

[26] E. Blanton, S. Fahmy, S. Banerjee, "Resource Management in an Active Measurement Service", *Proc. of IEEE Global Internet Symposium*, 2008.

[27] Z. Qin, R. Rojas-Cessa, N. Ansari, "Task-execution Scheduling Schemes for Network Measurement and Monitoring", *Elsevier Computer Communications*, Volume 33, Issue 2, Pages 124-135, 2010.

[28] R. Rajkumar, C. Lee, J. Lehoczky, D. Slewlorek, "A Resource Allocation Model for QoS Management", *Proc. of IEEE Real-Time Systems Symposium*, 1997.

[29] J. Strassner, "Policy-Based Network Management: Solutions for the Next Generation", *Morgan Kaufmann Series in Networking; ISBN: 1-55860-859-1*, 2004.

[30] Amazon CloudWatch: Monitoring Framework for Amazon Web Services Cloud Resources and Applications. http://aws.amazon.com/cloudwatch