

# Design and Implementation of the Control Plane in a Self-service Cloud Platform

**Abstract**—Public cloud platforms, such as Amazon EC2 and Microsoft Azure, have recently become a popular means to host services. However, there are still a number of hurdles that hinder their wide-spread adoption. Two such hurdles are: (1) ensuring the security and privacy of client data on the cloud, and (2) giving clients flexible control over their own VMs. The first problem arises because on contemporary platforms, cloud operators have the privileges to access and modify client VMs at will. The second problem arises because the cloud provider typically decides what services (*e.g.*, intrusion detection, traffic shaping) are available to a client's VMs. Clients must choose from these services, and typically do not have the liberty to deploy a new service over their VMs.

We focus on Self-service Cloud Computing (SSC), which is a recently-proposed solution to these problems. An SSC platform modifies the cloud administrative interface to prevent cloud operators from snooping on or modifying client VMs, while also giving clients the flexibility to deploy services on their own VMs. Prior work described the hypervisor privilege model to enable this administrative interface.

In this paper, we focus on designing and developing the control plane of an SSC platform. The control plane is the layer that facilitates interaction between hosts in the cloud infrastructure as well as between the client and the cloud. We describe a number of novel features in SSC's control plane, such as its ability to allow specification of VM dependencies, flexible deployment of network middleboxes, and new VM migration protocols. We report on our design and implementation of SSC's control plane, and present experimental evaluation of services implemented atop the control plane.

## I. INTRODUCTION

In recent years, an increasing number of enterprises have migrated their applications to the cloud. Cloud providers host the physical infrastructure to execute these applications, and use virtualization to multiplex physical resources and isolate applications from one another. Migrating applications to the cloud therefore offers enterprises a number of attractive benefits, such as freeing them from procuring computing infrastructure, allowing elasticity in the use of computing resources, and offloading a number of management and maintenance tasks to cloud providers.

In this paper, we are interested in *public* cloud infrastructures, such as Amazon EC2 and Microsoft Azure. Despite their apparent popularity, many enterprises hesitate to migrate their applications to such infrastructures, opting instead to use private or in-house cloud offerings. Doing so often negates many of the benefits that public cloud computing has to offer, *e.g.*, in an in-house cloud infrastructure, the enterprise must still procure and manage the computing infrastructure.

We see two main reasons why enterprises are often reluctant to use public cloud infrastructures. The first reason is se-

curity and privacy of client data. Enterprises that host sensitive or proprietary data (*e.g.*, banks and pharmaceutical industries) may wish to protect this data from outsiders. On public cloud infrastructures, the cloud provider controls the virtual machine (VM) management interface. This management interface typically has the privileges to inspect the state of individual VMs hosted on the cloud. For example, the management VM on Xen-based platforms can inspect the memory contents and network traffic of all the work VMs hosted on the platform. Thus, sensitive client data is vulnerable to attacks by malicious cloud operators or exploits directed against vulnerabilities in the management interface [1, 2, 3, 4, 5, 6, 7].

The second reason is the inability of clients to flexibly control their VMs on public clouds. Delegating management of VMs to the cloud provider is often a two-edged sword. While doing so can reduce operating costs in some cases, it also has the disadvantage of making the cloud provider responsible for services that the enterprise client may wish to avail. For example, a client may wish to use a custom network intrusion detection system (NIDS) or a memory introspection service on its VMs. On public cloud infrastructures, such services must be deployed by the cloud provider. It is often unusual for cloud providers to deploy custom services for individual clients. In the unlikely case that this does happen, each client request to configure or modify these services must go through the cloud provider.

These two problems have garnered significant recent interest in the research community, which has proposed a number of solutions to address one or both of the problems [8, 9, 10, 11, 12, 13]. We overview these solutions in Section VII.

Our focus is on one such recent solution called *Self-service Cloud (SSC)* Computing [14]. SSC aims to address both the problems discussed above by modifying the privilege model of the hypervisor. Traditional hypervisors endow the management VM (which we will henceforth refer to as *dom0*, following Xen terminology) with unrestricted privileges over client VMs. In contrast, an SSC hypervisor employs privilege separation to offer two kinds of management VMs: (1) a deprived system-wide management domain, used by cloud operators, and (2) per-client management domains, used by clients to administer and manage their own VMs. This new privilege model, discussed in further detail in Section II, simultaneously prevents cloud operators from snooping on or modifying client VM state, and allows clients to flexibly control their own VMs.

The original SSC paper [14] studied the design and implementation of the aforementioned privilege model within the hypervisor, and developed protocols to operate VMs on such a hypervisor. The main technical contribution of the present paper is the design and implementation of the control plane of the SSC platform. The control plane facilitates the

interaction between hosts in the cloud infrastructure as well as between the client and the cloud. It presents a unified administrative interface to clients, and transparently manages all of a client's VMs running across different hosts on the cloud. Simultaneously, it also protects clients from the cloud by preserving SSC's twin objectives of security/privacy and flexible VM control.

SSC's control plane includes several novel features that are not usually seen in contemporary cloud infrastructures:

(1) **Specifying VM dependencies.** SSC's privilege model allows clients to create VMs that have privileges over other VMs. For example, a client can create a memory introspection VM that has the privileges to inspect the memory contents of a work VM for malware infection (*e.g.*, [15, 16, 17, 18]). SSC's control plane presents an enhanced *dashboard* interface via which clients can specify inter-VM dependencies (*e.g.*, that the work VM depends on the memory introspection VM). SSC's control plane uses these dependencies to automatically determine how VMs must be placed on individual hosts. Thus, the memory introspection VM, for instance, will be co-located on the same host as the work VM(s) that it monitors.

(2) **Flexible middlebox deployment.** Existing cloud infrastructures do not offer a flexible way to deploy network middleboxes. SSC's control plane offers clients the flexibility of plumbing the I/O path of the work VMs, thereby allowing easy deployment of networked services, such as NIDS, trace anonymization services, and network metering. Clients simply specify VM dependencies using the interface described above, and the control plane automatically configures the I/O path to enable middlebox placement.

(3) **New VM migration protocols.** VM migration is an essential component of elastic cloud infrastructures. Contemporary VM migration tools focus on migrating individual VMs from one host to another. In SSC, VM migration must respect inter-VM dependencies. SSC's control plane includes algorithms to safely migrate VMs in the presence of inter-VM dependencies.

We begin the rest of this paper by presenting background material on SSC in Section II, following which we discuss the main components of SSC's control plane in Section III. In Section IV, we present the language used to specify inter-VM dependencies, and discuss in Section V how these dependencies affect VM migration. We present an experimental evaluation in Section VI and discuss related work in Section VII.

## II. THE SELF-SERVICE CLOUD PLATFORM

In this section, we present background material on the SSC platform and motivate the need for a control plane. We begin with a discussion of SSC's threat model.

### A. Threat Model

The cloud platform is a challenging one in which to protect the security of client data. SSC uses a practical threat model that we believe applies well in the real world. Similar threat models have also been adopted in other projects on cloud security [10, 11, 12, 19].

In SSC, the *cloud provider* is assumed to be an entity such as Amazon, Microsoft, or Rackspace, and is *trusted*. The client

| VM   | Tasks  |
|--|--|
| Platform-wide VMs, <i>i.e.</i> , one instance per physical machine |  |
| <b>Sdom0</b>   | Manages physical hardware and executes device drivers. Pauses/unpauses and schedules client VMs. Manages I/O quotas.   |
| <b>domB</b>  | Creates new VMs in response to client requests. Virtualizes the TPM and interacts with physical TPM.   |
| Per-client VMs, <i>i.e.</i> , at least one instance per client     |  |
| <b>Udom0</b>   | Administrative interface for all of the client's VMs. Delegates privileges over UdomUs to SDs. Hosts client's SSL private keys for secure communication with client. |
| <b>UdomU</b>   | The client's work VM, possibly multiple UdomUs per client.   |
| <b>SD</b>  | Service domains, each performing a specific administrative task over designated UdomUs. Possibly multiple SDs per client.  |

Fig. 1. Various VMs on an SSC platform.

believes that the cloud provider has a vested interest to protect its reputation, and that it will not deliberately compromise the security and privacy of the client's code and data. The threat model also assumes that physical machines on the cloud platform are equipped with trusted hardware (*i.e.*, TPM chips) and IOMMU units to enable I/O virtualization.

Thus, SSC assumes that the cloud provider will supply a trusted computing base (TCB) equipped with a hypervisor that implements the privilege model described in Section II-B. The original SSC paper included TPM-based protocols that would allow the client to verify the contents of the TCB before starting its VMs on the platform. SSC's TCB also consists of various components in its control plane. We defer a detailed discussion of these components to Section III, where we introduce these components and describe their role in the TCB.

On the other hand, *cloud operators*, who may include human administrators employed by the cloud provider, are *untrusted*. By extension, we assume that all actions originating from the cloud platform's administrative interface, including relaying I/O on behalf of client VMs, are untrusted. Note that even if cloud operators are not overtly malicious in intent, the administrative interface (*i.e.*, dom0) may contain remotely-exploitable vulnerabilities. SSC's threat model protects the security of client VMs against the effects of such exploits as well.

One of the key implications of this threat model is that SSC cannot defend against attacks launched by the cloud provider itself. This may include denial of service attacks, *e.g.*, failure to schedule the client's VMs for execution, or explicit monitoring of the client's activities, *e.g.*, via subpoenas or PRISM-style surveillance programs. Protecting against such threats may require heavyweight cryptographic techniques.

### B. Components of SSC

The hypervisor on an SSC platform uses privilege separation to partition the responsibilities traditionally entrusted with dom0 to two new kinds of administrative domains (see

Figure 1). *Sdom0*, short for System-dom0, is the administrative domain that manages the resources of the physical platform, schedules VMs for execution, and manages I/O quotas. *Sdom0* also executes the device drivers that control the physical hardware of the machine. Each physical platform is equipped with one *Sdom0* instance. Each client gets its own *Udom0* instance, short for User-dom0, which is an administrative interface that the client can use to monitor and control its VMs executing on that physical platform. While *Sdom0* holds the privileges to pause/unpause client VMs, access their read-only state (e.g., number of vCPUs assigned or their RAM allocation), and manage their virtual I/O operations, the hypervisor disallows *Sdom0* from performing privacy-sensitive operations on the client's VMs. Thus, for instance, *Sdom0* does not have the privileges to map the memory and vCPU registers of any of the client's VMs (i.e., *Udom0*, and the two families of user VMs introduced below).

Aside from these administrative domains, each platform also hosts the client's work VMs. We call these *UdomUs*. Each client's *Udom0* has the privileges to administer that client's *UdomUs*. For example, it can map the memory of a *UdomU*, thereby allowing it to inspect the contents of the *UdomU*. This feature can be used to implement memory introspection tools, such as rootkit detectors. Likewise, *Udom0* can set I/O backends for the *UdomU*, thereby intercepting the I/O path of the *UdomU*. This feature can be used to implement intrusion detection by inspecting network traffic or disk I/O.

Although *Udom0* has the privileges to perform the aforementioned tasks, SSC's design aims to keep *Udom0* stateless for the most part. Thus, SSC also supports *service domains* (*SDs*), which can perform these administrative tasks on client VMs. An *SD* can be started by the *Udom0*, which delegates specific privileges over a target *UdomU* (or a set of *UdomUs*) to the *SD*. For example, *Udom0* can create a rootkit detection *SD*, and endow it with the privileges to map the memory contents of all the client's *UdomUs* on that platform.

Note that all the physical hardware on the platform is controlled by *Sdom0*. Thus, all I/O performed by a client's VMs is ultimately relayed to and performed by *Sdom0*. By our threat model, *Sdom0* is untrusted and potentially malicious, and can therefore inspect/modify the client's I/O traffic. However, clients can still protect their I/O traffic via interception. For example, the *Udom0* could set a *UdomU*'s storage and network I/O backends to be an *SD* that encrypts all outgoing traffic. Thus, I/O from the *UdomU* first traverses the *SD*, which encrypts the traffic before relaying it to the *Sdom0*, which is therefore unable to inspect the traffic.

The final component of SSC is a platform-wide *domain builder* (*domB*). The sole responsibility of *domB* is to create VMs in response to client requests. Building a VM on a platform requires accessing the VM's memory and registers onto the physical platform. Because this task is privacy-sensitive, it cannot be performed by *Sdom0* although it involves mapping the client VM's state onto the physical platform. The TCB therefore includes *domB*, which is entrusted with this responsibility. *DomB* also interacts with the TPM, and hosts the code to virtualize the TPM for individual clients [20].

In addition to these components, SSC also supports *mutually-trusted service domains* (*MTSDs*), which are a re-

stricted kind of service domain. We elide a detailed description of *MTSDs* in this paper, but provide a brief overview in Section VI-A3.

A client session to create VM instances on an SSC platform begins with a request to create a *Udom0* instance. As discussed above, *Udom0* instances are stateless for the most part, and it is therefore possible to use off-the-shelf OS distributions to create *Udom0s*. *Udom0* creation happens via a protocol (described in the original paper [14]) that also installs the client's SSL private key in its memory. Note that *Udom0* memory is inaccessible to *Sdom0*, thereby protecting the private key from cloud operators. This key enables an SSL handshake with the client, thereby allowing the creation of an encrypted communication channel. The client then transmits future requests for VM creations (e.g., *UdomUs* and *SDs*, which may contain sensitive code and data) via this channel to *Udom0*, which then creates these VMs for the client on the physical host.

### C. The Need for a Control Plane

The control plane of traditional cloud platforms is responsible for monitoring resource usage of individual hosts, and suitably placing or migrating client VMs based upon the cloud provider's load balancing policies. It is also responsible for deciding where any client-chosen services offered by the cloud provider (e.g., an intrusion detection service) will be placed on the network. Examples of such software include VMWare's vCloud [21], Amazon Web Services [22], as well as open-source systems such as OpenStack [23], OpenNebula [24], CloudStack [25], and Eucalyptus [26].

The control plane of an SSC-based cloud platform must have two key features:

(1) **Allowing clients to specify inter-VM dependencies.** By introducing per-client administrative domains (*Udom0s*), SSC gives clients the ability to create *SDs* that can have specific administrative privileges over *UdomUs*, and the ability to interpose on their I/O paths. Thus, clients must be able to specify how the VMs that they propose to create are inter-related. For example, a client may wish to specify that an *SD* that he creates has the privileges to map the memory of one of its *UdomUs*. A consequence of this dependency is that these two VMs, i.e., the *SD* and the *UdomU*, must reside on the same physical host. Likewise, the client can specify that an *SD* is to serve as the network middlebox for a collection of *UdomUs*. The control plane must respect this dependency in placing VMs, ensuring that all outgoing and incoming network connections to any *UdomU* in that collection will first traverse the *SD*. This invariant must be maintained even if any of the *UdomUs* or the *SD* itself migrates from one physical host to another.

While traditional control planes do support middleboxes, the key difference is that these middleboxes are offered as services by the cloud provider. Both client-defined middleboxes and privileged *SDs* are unique to the SSC platform. The control plane of an SSC platform must be aware of and enforce these dependencies during execution.

(2) **Presenting a unified administrative interface to the client.** On SSC, each physical platform that hosts a client VM must also have a *Udom0* instance executing on it. This is because

the Udom0 hosts the client's SSL private keys, which are described for the SSL handshake with the client, as described in Section II-B.

Naïvely using traditional control plane software with such a setup can lead to security holes. For example, a client may strategically create VMs in a manner that forces the cloud provider to place them on different physical hosts *e.g.*, by creating a large number of VMs that place large resource demands on individual hosts. The client could leverage the fact that there is a Udom0 instance on each physical platform that executes at least one of the client's VMs to estimate the number of physical hosts in the cloud, or to map the cloud provider's network topology.

SSC's control plane must provide the illusion of a single administrative interface to the client, while hiding the presence of individual Udom0 instances. The control plane must suitably relay administrative operations from this interface to the corresponding Udom0 instances. It is also responsible for transparently handling VM migrations across physical hosts.

### III. SSC'S CONTROL PLANE

The control plane of most traditional cloud platforms has three key components. The first component is a platform-wide *cloud controller*. This component has a global view of the provider's infrastructure, and is tasked with allocating resources to clients elastically, provisioning and scheduling virtual machines, and in general, monitoring the overall health of the platform. Cloud providers may choose to partition their infrastructure into multiple zones in the interest of fault tolerance and scalability, and execute multiple instances of cloud controllers in each zone. The second component is a per-host *node controller*. This component typically executes as a daemon within dom0, and interacts with the hypervisor on the platform to monitor local resource consumption, and reports these statistics to the cloud controller. The final piece is a *dashboard*, which is the interface via which clients interact with the cloud. Each client's dashboard instance reports the state of the client's slice of the cloud, *e.g.*, the number of VMs executing, the resources consumed, and the current bill.

#### A. Components of SSC's Control Plane

SSC's control plane functionally enhances each of the components discussed above and introduces new protocols for inter-component communication. It introduces new client-centric features, such as the ability to specify relationships among VMs and manage client-deployed middleboxes. From the cloud provider's perspective, the new features include VM migration protocols, and the ability to provision resources while respecting client VM dependencies. Figure 2 summarizes the components of the control plane, and their interactions.

(1) **The Cloud Controller.** One of the key features of SSC is that it allows clients to specify inter-VM dependencies (discussed in more detail in Section IV). These dependencies may imply that certain VMs must be co-located on the same physical host. They may also specify how the I/O path of a client's work VM must be routed through the cloud. For example, an SD that serves as the back-end for a UdomU must lie on the path into and out of that UdomU, irrespective

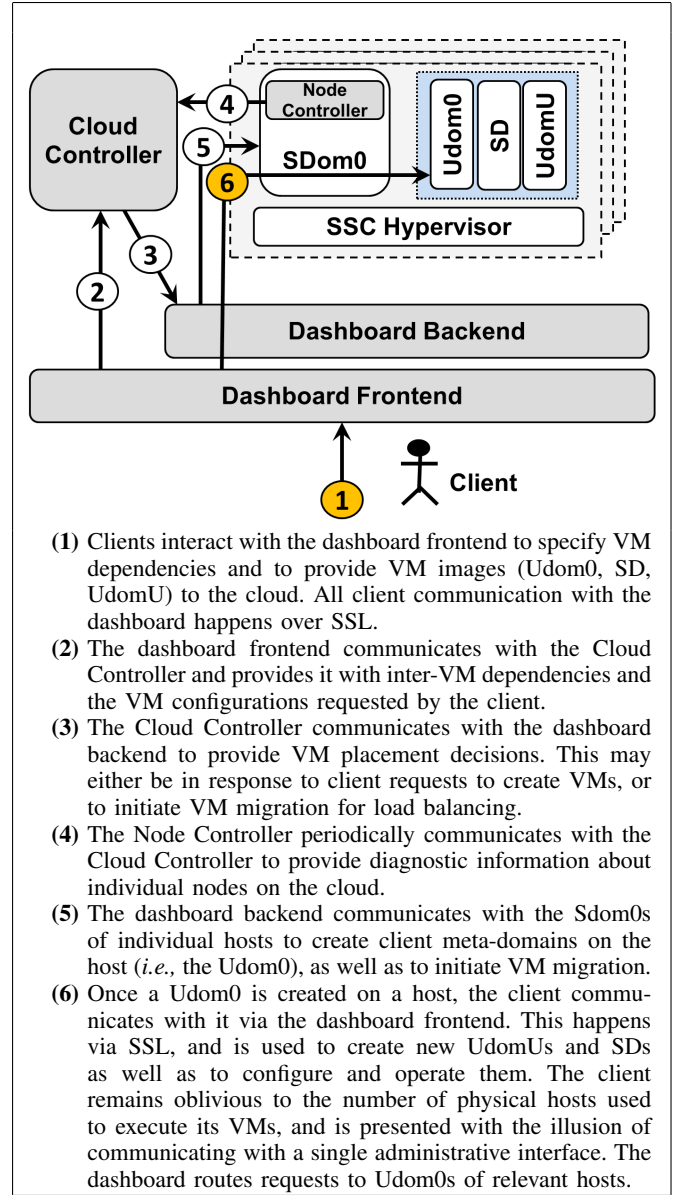


Fig. 2. Components of the control plane and their interactions. Communications labeled with shaded circles are secured by SSL.

of the machines on which the SD and UdomU are scheduled for execution.

The goal of the cloud controller is to accept these specifications from the client (via the dashboard) and produce a VM placement that satisfies these specifications. In doing so, it has to account for the current load on various hosts on the network (which it obtains by communicating periodically with node controllers), the resource requirements of the client's VMs, as well as any additional constraints imposed by the inter-VM dependencies. The cloud controller's scheduler therefore solves a constraint satisfaction problem, and produces a placement decision that is then communicated back to the dashboard. In turn, the dashboard interacts with individual hosts to schedule VMs for execution.

The cloud controller also initiates VM migration decisions. Based upon the resource usage information received from node controllers, the cloud controller may decide that a client's VMs

need to be migrated in order to balance load on the platform. It produces a new VM placement, and communicates this to the dashboard. The dashboard then communicates with the source and target hosts to initiate the actual migration of the VMs.

The client never directly interacts with the cloud controller, nor does the cloud controller interact with the client's VMs. It simply produces VM placement decisions and communicates them with the dashboard. The client only trusts the cloud controller to produce a fair VM placement decision. Violation of this trust can potentially lead to denial of service attacks, which are not covered under SSC's threat model (as discussed in Section II-A).

(2) **The Node Controller.** As on traditional cloud platforms, node controllers in SSC's control plane execute on individual hosts. Each node controller executes as a daemon within that platform's Sdom0. It can therefore monitor the resource utilization on the host and possibly control VM scheduling decisions, but *cannot* create user VMs (that is done by domB) or read/modify individual client VMs.

The client never interacts directly with the node controller. In fact, as an entity that executes within the Sdom0, it is untrusted. The node controller cannot compromise the security of client VMs in any manner besides launching denial of service attacks by failing to schedule the VM for execution, or by reporting false resource utilization to the cloud controller, thereby triggering frequent VM migrations.

(3) **The Dashboard.** In SSC, the dashboard serves as the layer between the client and the cloud platform and has two responsibilities: to interact with the client, and to interface with various components of the cloud platform. Accordingly, we logically split the dashboard into a *frontend* and a *backend*, as shown in Figure 2.

The dashboard frontend is an interface presented to the client (e.g., a Web interface) via which the client can enter VM images to be booted, and specify inter-VM dependencies. The dashboard also presents diagnostic information to clients, such as the resource consumed by the client's VMs. The frontend communicates any inter-VM dependencies specified by the client to the cloud controller. Likewise, the configurations of the client's VMs are also passed to the cloud controller, which uses this information to create a placement decision. However, the contents of the VM images itself are never passed to the cloud controller. The dashboard directly passes these images to the end hosts via an SSL channel (we discuss the setup of this SSL channel in Section III-B).

The dashboard backend orchestrates all communication between components of the cloud platform on behalf of the client. It obtains placement decisions from the cloud controller (whether to create VMs or to migrate them from one host to another) and transmits the client's VM images to the hosts on which they must be created. All communication between the dashboard and the end hosts is secured by the aforementioned SSL channel to protect it from any network snooping attacks launched by malicious cloud operators.

Because the dashboard handles sensitive information (client VM images), it is part of the TCB. While it is possible to implement the dashboard in any manner that allows the client to trust it, we chose to implement the dashboard itself as a VM that executes atop an SSC hypervisor. We assume that the cloud provider will dedicate a physical host (or a

|           |          |  |
|-----------|----------|--|
| 1. client | → Sdom0  | : $n_{TPM}$ , $Enc_{AIK}(freshSym  n_{SSL})$ , $Sig_{client}$              |
| 2. Sdom0  | → domB   | : $CREATE\_DASHBOARD(n_{TPM}, Enc_{AIK}(freshSym  n_{SSL}), Sig_{client})$ |
| 3. domB   | → client | : $TPMSign(n_{TPM}  PCR)$ , ML   |
| 4. domB   | → Sdom0  | : Schedule the dashboard for execution                                     |
| 5. DashVM | → client | : $n_{SSL}$  |
| 6. client | → DashVM | : $Enc_{freshSym}(SSLpriv)$  |
| 7. client | ↔ DashVM | : SSL handshake  |

Fig. 3. Protocol to create a dashboard VM instance.

set of physical hosts) simply to execute dashboard VMs for its clients. We call this machine the *dashboard host*. Each new client gets a dashboard VM instance that executes as a Udom0 on the SSC hypervisor. This ensures that even the cloud operator on these physical hosts cannot tamper with the execution of the client's dashboard VM instance.

### B. Overall Operation of the Control Plane

A client begins its interaction with an SSC platform by first requesting a dashboard VM instance. During the creation of this VM instance, the client configures it so that the client can communicate with the dashboard VM via SSL. It then provides VM images to this dashboard instance over the SSL channel, which then starts the VMs on physical cloud hosts on behalf of the client. In this section, we present and analyze the protocols that are used for these steps.

1) **Dashboard Creation and Operation:** A client creates a dashboard VM instance using the protocol shown in Figure 3. It communicates with the "cloud provider," in this case the Sdom0 VM of the dashboard host. The first message of the protocol consists of a nonce ( $n_{TPM}$ ), together with a piece of ciphertext ( $freshSym||n_{SSL}$ ) encrypted using the AIK public key of the TPM of the dashboard host.<sup>1</sup> Here  $freshSym$  is a fresh symmetric key produced by the client, while  $n_{SSL}$  is a nonce; we explain their roles below. The client also digitally signs the entire message (denoted in Figure 3 as  $Sig_{client}$ ).

The Sdom0 of the dashboard host communicates these parameters to its domB via the  $CREATE\_DASHBOARD$  command, which is a new hypercall to the underlying hypervisor. This command instructs domB to use the VM image used by the cloud provider for dashboard VMs, and create an instance of the VM for the client. DomB does so after verifying the client's digital signature. In this step, domB also communicates with the TPM to decrypt the message encrypted under its AIK public key, and places  $freshSym$  and  $n_{SSL}$  in the newly-created dashboard VM instance. Because the dashboard host executes an SSC hypervisor, the contents of this dashboard VM instance are not accessible to Sdom0, which therefore cannot read  $freshSym$  and  $n_{SSL}$ .

At this point, the client can verify that the dashboard VM has been created correctly. DomB sends a digitally-signed measurement from the TPM, containing the contents of its

<sup>1</sup>We assume that the TPM is virtualized [20], and that the corresponding vTPM drivers execute in the dashboard host's domB. Thus, the AIK public key used in this message is that of the virtual TPM instance allocated to this client. We assume that the client interacts with the cloud to obtain this AIK public key, and do not show this step explicitly in the protocol. In this paper, we used TPM-based attestation protocols that use code hashes [20, 27]. However, we can also support alternative property-based TPM protocols, e.g., [12, 28, 29, 30].

PCR registers and  $n_{\text{TPM}}$ , together with the measurement list (as is standard in TPM-based attestation protocols [20, 27]). We assume that cloud provider will make the measurements of dashboard VM instances publicly-available. This is a reasonable assumption because the dashboard VM does not contain any proprietary code or data. All information proprietary to the cloud provider is in other components of the platform, such as the cloud controller. The dashboard VM simply encodes the protocols discussed in this paper to interact with the client as well as with components of the cloud’s control plane. Thus, the cloud provider can even make the code of the dashboard VM available for public audit.

In the last two steps of the protocol, the dashboard VM instance interacts with the client to obtain the private key portion of the client’s SSL key pair (SSLPriv). This key is used during the SSL handshake phase between the client and the dashboard VM, thereby allowing the establishment of an SSL channel between the two. The dashboard VM sends the nonce  $n_{\text{SSL}}$  to the client, who sends in turn SSLPriv encrypted under *freshSym*, which is known only to the dashboard VM. This allows the dashboard VM to retrieve SSLPriv.

The protocol is designed to prevent attacks whereby a malicious cloud operator attempts to impersonate the client to start a dashboard VM instance as the client. The use of *freshSym* in the protocol ensures that the dashboard VM that is created has a secret value known only to the client. Two key features of SSL prevent the cloud operator from learning the value of *freshSym*: (a) the fact that the ciphertext sent in the first message can only be decrypted by the TPM (via domB), and (b) the fact that Sdom0 cannot obtain values stored in the memory of a dashboard VM instance. These features together allow the client to bootstrap the SSL handshake in a secure fashion. Finally, the nonces  $n_{\text{TPM}}$  and  $n_{\text{SSL}}$  prevent attempts by a malicious cloud provider to replay the protocol.

2) *Creation and Operation of Client VMs*: Once the dashboard VM is set up using the protocol discussed above, the client can create its VMs. It provides these VM images via the SSL channel to the dashboard. However, to boot these VMs on a physical host, the dashboard VM must still communicate with the Sdom0 on that host, which is untrusted.

To protect client VMs, which may contain sensitive code and data, from untrusted Sdom0s, we require that the first client VM that is booted on a physical platform be its Udom0. This is also the case during cloud controller-initiated VM migration, where client VM instances are moved from a source host to a target host that does not have any of the client’s VMs. As discussed in Section II-B, Udom0 is a stateless administrative interface for the client. It could therefore run a standard OS distribution, and not contain any sensitive client code or data. As a result, its image can be provided to the Sdom0 on the physical host. The dashboard does so on behalf of the client. The protocol to start a Udom0 on a host resembles Figure 3, with the major difference that the Udom0 image is also supplied in the first step. The original SSC paper [14] provides the steps of this protocol together with a security analysis, and we elide a description in this paper.

As with Figure 3, this protocol also establishes an SSL channel between the dashboard and the Udom0. With this secure channel in place, the Udom0 can accept VM images

from the dashboard to create UdomUs and SDs on the physical host.

Note that the client never interacts directly with the Udom0. Rather, the dashboard VM serves as a trusted intermediary. The dashboard presents the client with the illusion of a unified administrative interface, regardless of the number of Udom0 instances executing on various physical hosts. The dashboard interfaces with each of the physical hosts to start VMs and verifies TPM measurements after the VMs boot. The only virtual TPM key that is exposed to the client is the AIK public key of the dashboard host.

#### IV. SPECIFYING INTER-VM DEPENDENCIES

Clients on an SSC platform can avail of the power of SDs by using them to implement middleboxes offering novel security and system services. These middleboxes can either take the form of SDs that hold specific privileges over the client’s VMs, or serve as I/O backends for these VMs. SDs can be chained to flexibly combine services. For example, an SD that checks network traffic for intrusions can be chained with one that decrypts SSL traffic bound for the client’s UdomU. This allows the client to check inbound encrypted packets for attacks before processing them. Figure 6 presents another example of chaining.

To enable these features, SSC’s control plane provides a language, shown in Figure 4 for clients to specify inter-VM dependencies. The client specifies these dependencies via the dashboard, which forwards it to the cloud controller. SSC’s control plane allows two kinds of inter-VM dependencies:

(1) **GRANT\_PRIVILEGE dependencies**. This rule allows the client to specify that an SD must have specific privileges over a UdomU. These privileges may include mapping the user- or kernel-level memory of the UdomU, or reading vCPU registers. In the example program shown in Figure 5, *memscan\_vm* is given the privileges to map the kernel memory of *webserver\_vm*, e.g., to detect rootkit infections.

This dependency implicitly places a co-location constraint. The SD and the UdomU must be started on the same physical node. This is required because the operations to assign privilege (e.g., mapping memory) are local to a node’s hypervisor and can only be performed when both the SD and the UdomU run atop the same hypervisor.

(2) **SET\_BACKEND dependencies**. This rule specifies that one VM must serve as the I/O backend for another VM for a specific device type (we currently support storage and network devices). It also allows the client to specify whether the two VMs must be co-located (MUST\_COLOCATE), or whether the cloud controller can possibly place them on different physical hosts (MAY\_COLOCATE).

For example, a network back-end SD that runs an encryption/decryption service for a client’s UdomU must ideally be colocated on the same physical host as the UdomU. This is necessary to protect the secrecy of the client’s data before it reaches the device drivers hosted in Sdom0. If on the other hand, the client is not concerned about the secrecy of his network traffic, but wishes only to check for intrusions, the SD that performs network intrusion detection can potentially be placed on another machine, provided that all inbound traffic traverses the SD first before being routed to the UdomU.

```

Program := (Decl)* (Init)* (VMDep)*
Decl := VM vm // vm is a variable, VM is a keyword
Init := vm.name = String; vm.image = VM image
VMDep := GrantRule | BackRule
GrantRule := GRANT_PRIVILEGE(sd, udomu, PrivType)
BackRule := SET_BACKEND(backvm, frontvm, Device, Location)
PrivType := USER_MEM | KERN_MEM | VCPU
Device := STORAGE | NETWORK | ...
Location := MUST_COLOCATE | MAY_COLOCATE

```

Fig. 4. Language used by the dashboard to specify VM relationships.

```

VM webserver_vm; // Client's Web server
VM empDB_vm; // Client's employee database
VM memscan_vm; // Memory introspection SD
VM enc_vm; // SSL proxy for the employee DB
VM Snort_vm; // SD running the Snort NIDS

webserver_vm.name = "MyWebServer";
webserver_vm.image = ApacheVM.img;
empDB_vm.name = ...; empDB_vm.image = ...;
memscan_vm.name = ...; memscan_vm.image = ...;
enc_vm.name = ...; enc_vm.image = ...;
Snort_vm.name = ...; Snort_vm.image = ...;

GRANT_PRIVILEGE(memscan_vm, webserver_vm, KERN_MEM);
SET_BACKEND(Snort_vm, webserver_vm, NETWORK, MAY_COLOCATE);
SET_BACKEND(enc_vm, empDB_vm, NETWORK, MUST_COLOCATE);

```

Fig. 5. An example program showing inter-VM dependencies.

Figure 5 presents an example of such a scenario. The client specifies that the network traffic to *webserver\_vm* must be checked using Snort. The client does not consider the traffic to and from the Web server to be sensitive, so it may potentially be exposed to the cloud provider. However, any interactions with its employee records database *empDB\_vm* must happen over SSL. The *enc\_vm* SD serves as the network I/O backend for the *empDB\_vm* database, encrypting all outgoing traffic and decrypting all incoming traffic, while residing on the same host as *empDB\_vm*.

Note that on traditional cloud platforms, customers usually do not control how their VMs are placed. Rather, the cloud controller determines VM placements based upon the global state of the cloud platform, the configurations requested by the client, and the cloud provider's load balancing policies. However, SSC's control plane gives clients *some* control over how their VMs are placed. For example, a client can specify that a rootkit detection SD that inspects the memory of its UdomUs must be placed on the same physical host as the UdomUs.

The cloud controller's scheduling algorithm uses these requests as additional constraints. It processes the entire program specified by the client to determine VM placements. For example, consider the dependencies shown in Figure 6. The client has a *webserver\_vm* that receives data over an encrypted channel. The client dedicates the *enc\_vm* VM to handle encryption and decryption, while *memscan\_vm* scans the kernel memory of *webserver\_vm*. Additionally, the client has specified that all packet headers destined for the *webserver\_vm* must be inspected by *firewall\_vm*. These rules imply that *memscan\_vm*, *webserver\_vm* and *enc\_vm* must be colocated on the same machine. However, *firewall\_vm*, which only inspects packet headers and is set as the backend for *enc\_vm*, can be located on a different host.

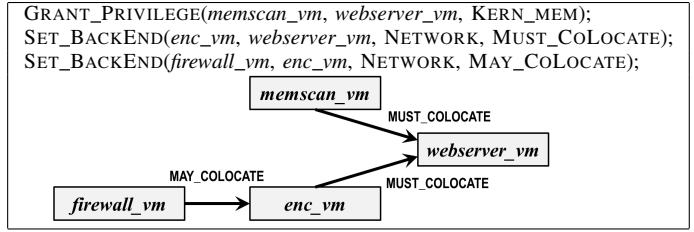


Fig. 6. Example showing inter-VM dependencies that require certain VMs to be co-located. We have elided VM declarations and initializations for brevity. Also shown is the VM dependency graph for this example.

In general, it is possible to depict these dependencies as a *VM dependency graph*, as shown in Figure 6. In this graph, an edge  $vm1 \rightarrow vm2$  depicts that *vm1* either serves as the backend for *vm2* or that *vm1* has privileges over *vm2*. Edges are also annotated with *MUST\_COLOCATE* or *MAY\_COLOCATE* to denote co-location constraints.

In some cases, it may not be possible to resolve the client's VM dependencies, given hardware constraints or the current load on the cloud infrastructure. In such cases, the client is suitably notified, so that it can modify the dependencies. The dashboard also performs certain sanity checks on the program input by the client. For example, it checks to determine that the VM dependency graph implied by the program is acyclic (the acyclic property is required to make migration decisions, as discussed in Section V). In such cases, the dashboard raises a warning, akin to a compile-time error, so that the client can correct the program.

Once the specifications are accepted, and the cloud controller produces a placement, the dashboard orchestrates the actual setup of the VMs. For a pair of dependent VMs  $vm1 \rightarrow vm2$  that are located on the same physical host, enforcing the dependencies is relatively straightforward. In case they are dependent via a *GRANT\_PRIVILEGE*, the dashboard simply instructs the Udom0 to assign suitable privileges over *vm2* to *vm1*. Likewise, the dashboard can instruct Udom0 to set *vm1* as the device backend for *vm2* in case of a *SET\_BACKEND* dependency.

However, *vm1* and *vm2* could be located on different physical hosts (*S* and *T*, respectively) if they related via a *SET\_BACKEND*(*vm1*, *vm2*, ..., *MAY\_COLOCATE*). In this case, the dashboard configures switches on the network to route traffic destined for *vm2* via *vm1*. More concretely, SSC uses the Open vSwitch [31] software switch for this purpose. In this case, the dashboard instructs the Udom0s on *S* and *T* to create SDs on these hosts running the Open vSwitch software. On *S*, traffic from *vm1* is sent to the Open vSwitch SD running on that host, which routes it to the Open vSwitch SD on *T* via a Generic Routing Encapsulation (GRE) tunnel. On *T*, this Open vSwitch SD is configured to be the backend of *vm2*, thereby routing traffic to *vm2*. Outbound traffic from *vm2* is routed via *vm1* in a similar fashion. Figure 7 illustrates the setup using the VMs for Figure 6 as an example.

As is clear by now, SSC deviates from traditional cloud platforms in that it allows clients to specify VM co-location



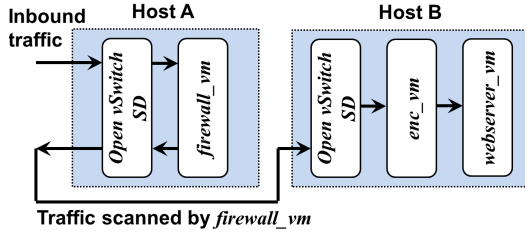


Fig. 7. Open vSwitch setup showing the path followed by inbound network traffic to the web server example from Figure 6. Outbound network traffic follows the reverse path. The Udom0 instances on the two hosts and the *memscan\_vm* instance on Host B are not shown.

constraints.<sup>2</sup> The reader may observe that such constraints could possibly give malicious clients the opportunity to infer proprietary information about the cloud platform using *probe-response* attacks. In such an attack, the client provides a program with a sequence of GRANT\_PRIVILEGE dependencies, requiring a certain number of VMs (say,  $n$  VMs) to be co-located on the same physical host. The client could repeatedly probe the cloud with programs that successively increase the value of  $n$ . When the cloud controller is no longer able to accommodate the client's requests, the client can use this failed request to gain insight into the limitations of the hardware configurations of the cloud platform's hosts or into the current load on the cloud.

Such threats are definitely a possibility, and we view them as the cost of giving honest clients the flexibility to control VM placements to enable useful services. Nevertheless, there are defenses that the cloud provider could use to offset the impact of such threats. For example, the provider could pre-designate a cluster of machines to be used for clients with larger-than-usual VM co-location constraints, and try to satisfy the client's requests on this cluster. This would ensure that the effects of any probe-response attacks that give the client insight into the provider's proprietary details are constrained to that cluster alone, and not the entire cloud platform.

Finally, we note that much recent work in cloud security has focused on the possibility of attacks enabled by VM co-location, enabling a variety of malicious goals (e.g., [33, 34, 35, 36]). These results are not directly applicable in our setting, because SSC allows a client to specify VM co-location constraints for *its own VMs*. In contrast, the works cited above require co-location of a victim VM with the attacker's VM. In SSC, co-location of a client's VMs with another client's VMs is purely co-incidental, and a result of a placement decision made by the cloud controller.

Does allowing malicious clients to co-locate their own VMs on a host ease the possibility of launching attacks (e.g., like those in prior work [33, 34, 35, 36]) on *other* VMs co-located on that host? We don't know the answer to this question, but it would be prudent to acknowledge that such attacks may be possible. Exploring the full extent of such attacks and defenses against them is an active area of research, and is beyond the

scope of the current paper. Any defenses that are developed against VM co-location attacks would be applicable to and can benefit the SSC platform as well.

## V. VM MIGRATION

During the lifetime of a client VM, the cloud controller may choose to migrate it to balance resource utilization across hosts. Migration usually involves pausing the VM on the source host, copying its image to the target host, and starting the VM on the target. Because this operation can result in significant downtimes during migration, most cloud platforms employ *live migration* [37].

Live migration typically involves an *iterative push phase* and a *stop-and-copy phase*. When the cloud controller initiates the migration of a VM, the source host commences the iterative push phase, copying pages of the VM over to the target host, even as the VM continues to deliver service while executing on the source. As the name suggests, this phase has several iterations, each of which pushes VM pages that have been dirtied since the previous iteration. Once the bulk of the VM's pages have been copied over to the target (e.g., when the number of dirty pages falls below a threshold), a *stop-and-copy phase* pauses the VM on the source, copies any remaining dirty pages of the VM on the source to the target, and resumes the VM on the target. Because the number of dirty pages is significantly smaller than the size of the VM itself, live migration usually has downtimes on the order of a few seconds (sub-second delays in several cases).

In SSC, the decision to migrate is made by the cloud controller, which produces a new VM placement and communicates it to the dashboard. In turn, the dashboard coordinates with the source and target hosts to actually perform the migration. The decision to migrate a client VM can also result in the migration of other VMs that are related to it. When a UdomU is migrated, all the SDs that service the UdomU and must be co-located with it must also be migrated. The cloud controller uses the dependency specifications supplied by the client when producing a new placement decision after migration.

Dependency specifications also implicitly dictate the order in which VMs must be migrated. Consider the example in Figure 6. Both *memscan\_vm* and *enc\_vm* must be collocated when *webserver\_vm* is migrated (in fact, the three of them must always be migrated together). However, because *memscan\_vm* and *enc\_vm* service *webserver\_vm*, they must both continue to work as long as *webserver\_vm* is not paused. Thus, during the stop-and-copy phase of live migration, *webserver\_vm* must be paused before *memscan\_vm* and *enc\_vm* are paused. Likewise, on the target host, *memscan\_vm* and *enc\_vm* must be resumed before *webserver\_vm*. In general, the order in which VMs must be paused and resumed can be inferred using the VM dependency graph, which must be acyclic. All of a VM's children in the graph must be paused before it is paused, and all of a VM's parents must be resumed before the node itself is resumed.

Figure 8 summarizes the steps followed during migration. The dashboard orchestrates migration and checks TPM attestations when VMs resume after being migrated. It also parses the VM dependency graph and identifies the order in

<sup>2</sup>Some public cloud providers do offer clients the option of hosting their VMs on a "private" cloud [32], which does imply certain co-location constraints. However, in this case, the client is given exclusive use of the hosts in this "private" cloud. SSC's control plane allows clients to specify co-location constraints even if the client does not have exclusive use of a host.



- (1) Cloud controller decides to migrate a group of MUST\_COLOCATE VMs ( $vm_1, vm_2, \dots, vm_n$ ) from source  $S$  to target  $T$ .
- (2) Dashboard uses VM dependency graph to determine the order in which VMs must be paused on  $S$  and resumed at  $T$ .
- (3) Dashboard checks whether client has a Udom0 instance running on  $T$ . If not, starts it as described in Section III-B2, and specifies order in which VMs received must be started.
- (4) Dashboard requests Udom0 on  $S$  to initiate migration to  $T$ , and specifies the order in which to pause the VMs.
- (5) Udom0 on  $S$  establishes an encrypted channel to communicate with Udom0 on  $T$ .
- (6) Udom0 on  $S$  initiates the iterative push phase, copying VM pages to  $T$ .
- (7) Udom0 on  $S$  pauses VMs (stop-and-copy phase) and sends VM pages to Udom0 on  $T$ .
- (8) Dashboard obtains TPM measurements from domB on  $S$ , containing hashes of paused VMs.
- (9) Dashboard identifies any May\_Colocate VM backends with dependencies on ( $vm_1, vm_2, \dots, vm_n$ ), and instructs switches to update network routes from these backends to  $T$  instead of  $S$ .
- (10) Udom0 on  $T$  resumes the VMs, and forwards TPM measurements obtained from domB on  $T$  to dashboard.
- (11) Dashboard checks the TPM measurements obtained from  $S$  and  $T$  for equality. If not equal, raises security alert.
- (12) Dashboard determines whether there are any remaining client VM instances on  $S$ . If not, it initiates a shutdown of the Udom0 on  $S$ .

Fig. 8. Steps followed to migrate a group of co-located VMs.

which VMs must be migrated and paused. The actual task of copying VM pages is carried out by the Udom0 of the source host. When a set of co-located VMs with MUST\_COLOCATE dependencies is migrated from a source to a target, there may be VMs with MAY\_COLOCATE dependencies that must be suitably updated. For example, switches on the cloud must be updated so that traffic to and from *firewall\_vm* are routed to the target machine to which the VMs *webserver\_vm*, *enc\_vm* and *memscan\_vm* are migrated. In SSC, this is accomplished by suitably modifying the configurations of the Open vSwitch SDs running on the machines that host *firewall\_vm* and the target machine to which the three VMs are migrated.

## VI. EVALUATION

The main goal of our experimental evaluation is to understand the performance overhead introduced by the components of SSC’s control plane. Other aspects of SSC, such as the overheads involved in creating VMs on an SSC hypervisor and executing system services as SDs (rather than within dom0), were evaluated in the original SSC paper. Those experiments showed that an SSC-based platform was competitive with a traditional Xen-based platform, and focused on evaluating overheads incurred on a *single host*. Performance overheads imposed by SSC were under 5% for system services implemented as SDs in most cases.

We conducted two classes of experiments to evaluate SSC’s control plane:

(1) **Networked services.** SSC’s control plane allows networked services to be implemented as SDs. We evaluate the cost of

running these SDs both when they are co-located with the UdomUs they serve, and when they are located on a different host.

(2) **Migration.** We evaluate the cost of migrating VMs and report VM down times for various configurations of the migration algorithm.

### A. Networked Services

All hosts in our setup execute an SSC-based hypervisor, which we implemented by modifying the privilege model of the Xen-4.3 hypervisor. In all our experiments, we compare the overheads of SSC by comparing the performance of a networked service implemented as an SD against the same network service implemented in a traditional setup (*i.e.*, within dom0 executing on an unmodified Xen-4.3 hypervisor). The hosts in our setup are Dell Poweredge R610 systems equipped with 24GB RAM, eight 2.3GHz Xeon cores with dual threads (16 concurrent executions), Fusion-MPT SAS drives, and a Broadcom NetXtreme II gigabit NIC. All virtual machines in the experiments (dom0, domU, Sdom0, Udom0, UdomU, SD, domB) were configured to have 2GB RAM and 2 virtual CPUs.

1) **Baseline Overhead:** Our first experiment aims to evaluate the baseline overhead of running networked services atop the SSC control plane. For this experiment, we create a network SD *netbsd* VM and set it as the backend of a work VM *udomu* using

SET\_BACKEND(*netbsd*, *udomu*, NETWORK, MAY\_COLOCATE)

The SD *netbsd* does no additional processing on the packets that it receives, and simply forwards them to *udomu*. (The remainder of this section talks about network SDs to achieve various goals, and their overhead must be compared against this baseline, which reports the best performance achievable for a networked service implemented as an SD). Under this setup, *netbsd* may either be co-located on the same host as *udomu*, or be located on a different host, depending on the placement decision of the cloud controller. Each setup results in a different network topology, as illustrated in Figure 9. We evaluate both setups, and use the keywords SAMEHOST and DIFFHOST to differentiate the cases when *netbsd* and *udomu* are co-located or not. We compare these against a traditional setup, where the networked service executes within dom0, either on the same host or on a different host.

We dedicated a separate *measurement host* in the same local network as our experimental infrastructure to transmit and receive network packets to the *udomu*. We measured the network throughput to *udomu* using the iperf3 [38] tool, and used ping to measure the round-trip time (RTT) of network traffic to and from the *udomu*. Our results report average over five executions along with the standard deviations.

Figure 10 presents the results of our experiment. If *netbsd* is co-located with *udomu*, the throughput remains unaffected. However, the RTT drops as a result of having to traverse another element (*netbsd*) on the path to the *udomu*. When *netbsd* and *udomu* are on different hosts, the RTT overhead increases to ( $2.7\times$ ) as a result of new network elements encountered in the path to *udomu*. However, we observed that the throughput also reduces to nearly 50% compared to the traditional non-SSC-based setup. Upon further investigation, we found that the

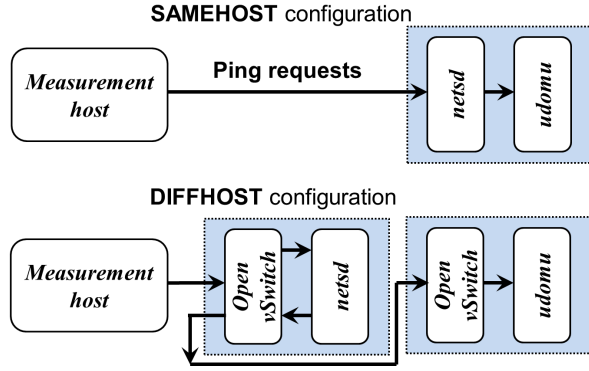


Fig. 9. The network topologies used to evaluate the baseline overhead of networked services executing atop SSC. We only show the inbound network path. The outbound path is symmetric.

| Setup                         | Throughput (Mbps) | RTT (ms)      |
|-------------------------------|-------------------|---------------|
| <b>SAMEHOST configuration</b> |                   |               |
| Traditional                   | 925.4±0.5         | 0.38±0        |
| SSC                           | 924.0±1.2 (0%)    | 0.62±0 (1.6×) |
| <b>DIFFHOST configuration</b> |                   |               |
| Traditional                   | 848.4±11.2        | 0.69±0        |
| SSC                           | 425.8±5.5 (49.8%) | 1.6±0 (2.3×)  |

Fig. 10. Baseline overhead of networked services.

reason for this is because the way Xen currently implements support for backend network drivers prevents concurrent bidirectional network transmission, as explained below.

On Xen, dom0 executes the network driver (called *netback*) that serves as the backend for all domU network communication. As of version 4.3, Xen uses only one kthread in *netback* to process domU’s transmit and receive queues [39]. The SSC hypervisor, which is based on Xen, inherits this limitation, and uses only one kthread in the *netback* drivers of SDs that serve as backends for UdomUs. Thus, if we consider the DIFFHOST configuration on a traditional Xen-based platform, where the functionality of *netsd* executes within dom0, the network driver simply receives inbound traffic from the measurement host, and forwards it to the dom0 of the machine that hosts *udomu*. In contrast, on an SSC-based platform, the *netback* driver within the Open vSwitch SD receives inbound traffic from the measurement host, forwards it to the *netsd* VM, receives traffic from the *netsd* VM, and forwards this to the Open vSwitch SD of the machine that hosts *udomu* (as shown using the arrows in Figure 9). As a result, even though the network hardware used in our experiments supports concurrent bidirectional network bandwidth of 1Gbps, the inability of the *netback* drivers to support concurrent bidirectional transmission cuts the throughput by approximately half.<sup>3</sup>

<sup>3</sup>An astute reader may observe that the throughput gap between the SAMEHOST case and the DIFFHOST case in SSC allows a malicious client to infer whether *netsd* and *udomu* are co-located. While this is definitely the case in our prototype, we observe that an enhanced implementation of the *netback* driver in Xen, with separate kthreads to process transmit and receive queues will address this attack (and improve network throughput in the DIFFHOST case!)

| Setup                         | Throughput (Mbps) |
|-------------------------------|-------------------|
| <b>SAMEHOST configuration</b> |                   |
| Traditional                   | 925.1±0.7         |
| SSC                           | 923.2±1.6 (0%)    |
| <b>DIFFHOST configuration</b> |                   |
| Traditional                   | 846.7±17.2        |
| SSC                           | 425.2±7.2 (49.7%) |

Fig. 11. Network Access Control Service.

2) *Network Access Control SD*: Network access control services (e.g., firewalls) are often the first layer of defense in any operational network. Traditional network access control services in the cloud, such as security groups, allow clients to specify rules on network packets. However, security groups are quite restrictive and only filter incoming packets. Our network access control SD is implemented as a middlebox that can be customized by clients. In our implementation, we used a set of rules that included a list of IP addresses and open ports for which packets should be accepted. The SD has a MAY\_COLOCATE dependency on the VM(s) it protects.

Figure 11 presents the performance of this SD, implemented both in the traditional setting within dom0, and atop SSC. The numbers here report overheads very similar to the baseline (Section VI-A1), thereby showing that the extra CPU processing overhead imposed by the SD is minimal. (RTT numbers were similar to the baseline numbers, and are not reported).

3) *Trustworthy Network Metering*: On traditional cloud platforms, clients trust the cloud provider to correctly charge them (i.e., as agreed upon in the service-level agreement) based upon the resources that they consume. If a client has reason to believe that a cloud provider is charging it for more than its share of resources consumed, it cannot prove that the cloud provider is cheating. What is therefore needed is a trustworthy service that performs resource accounting. Both the cloud provider and the client must be able to access the service and verify that the charges correspond to the resource utilization of the client. Unfortunately, it is impossible to design such a service on today’s cloud platforms. Recent work [40] has investigated the possibility of using nested virtualization to implement such a service.

SSC allows the creation of such trustworthy resource accounting services. The key mechanism to do so is SSC support for *mutually-trusted service domains (MTSDs)*, which were introduced in the original paper on SSC [14]. MTSDs resemble SDs in all aspects but two. First, unlike SDs, which are started and stopped by the client, the cloud provider and the client collaborate to start or stop an MTSD. Second, although an MTSD executes within a client’s meta-domain with privileges to access other client VMs, a client cannot tamper with or modify an MTSD once it has started.

MTSDs offer a powerful way for the cloud provider and client to collaborate. The cloud provider can use MTSDs to enforce regulatory compliance (e.g., ensure that the client is not hosting malware on the cloud platform) without compromising client privacy. The MTSD executes within the client’s meta-

| Setup                         | Throughput (Mbps) |
|-------------------------------|-------------------|
| <b>SAMEHOST configuration</b> |                   |
| Traditional                   | 924.8±1.1         |
| SSC                           | 924.1±0.4 (0%)    |
| <b>DIFFHOST configuration</b> |                   |
| Traditional                   | 845.4±11.1        |
| SSC                           | 424.3±3.1 (49.8%) |

Fig. 12. Trustworthy Network Metering Service.

domain, and can check the operation of the client’s VMs. However, the MTSD’s memory cannot be accessed from Sdom0, thereby protecting the client’s VMs from malicious cloud operators. At the same time, the MTSD simply reports back to the cloud provider whether the client is in compliance with the terms of the SLA. It simply reports a Boolean value to the provider, without disclosing any other details about the client’s VMs, thereby protecting client VM privacy.

The same MTSD mechanism can also be used to implement trustworthy network metering. The client and cloud provider agree upon the software that will be used to account for the client’s network bandwidth utilization. This metering software executes as an MTSD and serves as the network backend for all of the client’s network-facing VMs. The client and cloud provider can both verify that the MTSD was started correctly (using TPM attestations), and the SSC hypervisor ensures that neither the cloud provider nor the client can tamper with the MTSD once it has started execution. Thus, both the cloud provider and the client can trust the network bandwidth utilization reported by the MTSD.

Our network metering MTSD captures packets using the libpcap [41] library, simply counts the number of packets captured, and reports this number when queried. Because it measures network bandwidth utilization for *all* the client’s VMs, it must have a *May\_Colocate* dependency with all of them. Figure 12 shows the impact of network metering service on the network throughput of a single work VM (setup similar to Figure 9). As before, the additional overhead imposed by the network metering MTSD over the baseline is minimal.

4) *Network Intrusion Detection*: SSC allows clients to deploy and configure customized network intrusion detection systems as middleboxes. On traditional cloud platforms, this is not possible. Rather, they are forced to accept the offerings that the cloud provider has. Moreover, they cannot configure the placement of these middleboxes and must rely on the cloud provider to do so.

As an example, we used Snort [42] to set up an intrusion detection system as a middlebox before our work VMs. Snort uses libpcap[41] to capture network traffic. Our setup uses the Stream5 preprocessor that performs TCP reassembly and handles both TCP and UDP sessions. We used the latest snapshot of signatures available from the Snort website in our setup. The Snort SD has a *May\_Colocate* dependency on the UdomU(s) it monitors. Figure 13 presents the results of our experiments and shows that the overhead imposed by an SD implementation of Snort is minimal.

| Setup                         | IDS              |
|-------------------------------|------------------|
| <b>SAMEHOST configuration</b> |                  |
| Traditional                   | 922.8±1.1        |
| SSC                           | 920.9±1.9 (0%)   |
| <b>DIFFHOST configuration</b> |                  |
| Traditional                   | 841.2±14.2       |
| SSC                           | 422.6±7.1(49.7%) |

Fig. 13. Network intrusion detection (Snort) service.

| Setup       | Time (μsec)   |
|-------------|---------------|
| Traditional | 1014±6        |
| SSC         | 1688±31 (66%) |

Fig. 14. Time to establish a TCP connection in VMWall.

5) *VMWall service*: Our last example is a virtualized application-level firewall, called VMWall, that was developed in prior work [43]. In the traditional setting, VMWall operates as a daemon within dom0, and intercepts network packets originating from the VM that it monitors. It then performs memory introspection of the VM to identify the process that is bound to the network connection. If the process belongs to a whitelist, then VMWall permits the flow, else blocks the connection. VMWall therefore performs fine-grained application-level filtering while also remaining tamper-resistant by executing outside the monitored VM.

When implemented as an SD, VMWall serves as the network backend for the UdomU that it monitors. It must also have the privileges to inspect the memory of the UdomU, so that it can identify the process from which the flow originates. Thus, it is created with the following dependency rules, which imply that it must be co-located with the UdomU.

```
GRANT_PRIVILEGE(vmwall_vm, udomu, KERN_MEM);
SET_BACKEND(vmwall_vm, udomu, NETWORK, MUST_COLOCATE);
```

Our re-implementation of VMWall uses libvmi [17, 44] for memory introspection. Figure 14 presents the results of our experimental evaluation of VMWall. We measured the TCP connection setup time using a simple client/server setup. Compared to the traditional setup, establishing a TCP connection with the VMWall SD incurs an overhead of 66%. The main reason for this overhead is that in the traditional setup, TCP connections are established within dom0 itself. In SSC, connection requests are forwarded from Sdom0 to the VMWall SD, resulting in overhead.

## B. Evaluating VM Migration

We measure the performance of VM migration using two metrics: *VM down time* and *overall migration time*. Recall from Section V that migration happens in two phases, an iterative push phase, and a stop-and-copy phase. The VM down time metric measures the time taken by the stop-and-copy phase, while the overall migration time measures the time from the initialization of VM migration to its completion.

We perform three sets of experiments to evaluate VM migration. In the first experiment, we migrate a single VM

| Setup       | Time (seconds)        |
|-------------|-----------------------|
| Traditional | 23.27 $\pm$ 0.11      |
| SSC         | 23.81 $\pm$ 0.03 (2%) |

Fig. 15. Total migration time for one virtual machine.

| # of VMs | Sequential (seconds) | Parallel (seconds) |
|----------|----------------------|--------------------|
| 2        | 47.29 $\pm$ 0.18     | 27.91 $\pm$ 0.16   |
| 4        | 128.89 $\pm$ 0.76    | 57.78 $\pm$ 0.49   |

(a) Udom0 configured to have 2 virtual CPUs.

| # of VMs | Sequential (seconds) | Parallel (seconds) |
|----------|----------------------|--------------------|
| 2        | 47.41 $\pm$ 0.29     | 28.01 $\pm$ 0.26   |
| 4        | 103.96 $\pm$ 0.20    | 39.21 $\pm$ 0.50   |

(b) Udom0 configured to have 4 virtual CPUs.

Fig. 16. Migrating multiple virtual machines using sequential and parallel migration policies.

in SSC and compare it against migration on a traditional Xen platform. In our second experiment, we consider the case in SSC where a group of co-located VMs must be migrated together. In this experiment, we evaluate the performance implications of two migration policies. Third, we evaluate how the length of a dependency chain in the VM dependency graph affects the performance of migration. The first two experiments report the overall migration time only, while the third experiment explores VM down time in detail. For all the experiments reported in this section, we assume that the VMs to be migrated are 1GB in size, and are configured with 1 virtual CPU. The setup is otherwise identical to the one used in Section VI-A except when otherwise mentioned.

1) *Migrating a Single VM*: Figure 15 reports the time to migrate a single VM from one host to another in a traditional setting and on an SSC platform. The small overhead (2%) in the SSC setting can be attributed to the extra steps involved in migrating a VM in SSC, in particular, setting up a Udom0 at the target host. Note that because migration is live, the VM is still operational on the source as it is being migrated to the target. The down time in this case is approximately 100ms (as discussed in more detail in Section VI-B3).

2) *Migrating a Group of VMs*: When a group of dependent co-located VMs ( $vm_1, vm_2, \dots, vm_n$ ) is migrated from one host to another, there are two options to implement the iterative push phase of live migration. The first option is to perform iterative push for  $vm_1, vm_2, \dots, vm_n$  sequentially, *i.e.*, perform one iteration of page migration for  $vm_1$ , one for  $vm_2$ , and so on before starting again at  $vm_1$  for the second iteration. The second option is to perform the iterative push phase for all  $n$  VMs, using all the available parallelism in the underlying physical platform. The tradeoff is that while parallel migration approach can lead to lower migration times overall, it can saturate the network link and increase overall CPU utilization.

Figure 16(a) presents the overall time required to migrate a group of 2 VMs and 4 VMs, respectively, using the sequential and parallel migration policies. Naturally, the overall time to migrate using the parallel policy is smaller than for the sequential policy. We also used the iftop utility to measure the peak network utilization during VM migration. We found that

| Chain length | Down time (ms) |
|--------------|----------------|
| 1            | 97 $\pm$ 4     |
| 2            | 308 $\pm$ 3    |
| 3            | 528 $\pm$ 8    |
| 4            | 778 $\pm$ 7    |

Fig. 17. Down time for migrating VMs.

with the sequential policy network utilization never exceeded 40%, while for the parallel migration policy, peak network utilization never exceeded 70%, even when four VMs are migrated in parallel. For this experiment, the Udom0 (which performs migration) is configured to have 2 virtual CPUs and 2GB RAM, as discussed before.

To determine whether increasing the number of virtual CPUs assigned to the Udom0 can increase network utilization (and thereby reduce overall VM migration time), we repeated the experiments with the Udom0 configured to have 4 virtual CPUs and 2 GB RAM. Figure 16(b) shows that in this case, the time to migrate 2 VMs remains relatively unchanged and so does the network utilization (at 40%). When 4 VMs are migrated, the Udom0 is able to exploit the underlying parallelism in the host to complete migration faster. However, this comes at a cost, and the network utilization of the host shoots to 100%.

3) *VM Downtime*: Recall from Section V that dependent co-located VMs are migrated in the order specified by the VM dependency graph, *i.e.*, all children of a VM must be paused before it is paused, and vice versa for resumption. Typically, this means that a client's UdomUs that are serviced by several SDs (which may themselves be serviced by other SDs) must be paused before the SDs, and must be resumed on the target host only after all the SDs have been resumed. Thus, the length of a *dependency chain* (*i.e.*, the longest path) in this graph affects the performance of the stop-and-copy phase.

To evaluate the down time of a UdomU serviced by several SDs, we created dependency chains of varying length using the SET\_BACKEND rule, *i.e.*, we created a chain of SDs  $sd_1 \rightarrow sd_2 \rightarrow \dots \rightarrow sd_n \rightarrow udomu$  each of which was the backend for another. We migrated these VMs from one host to another, and measured the down time of the *udomu*. We only used the parallel migration policy for this experiment. Figure 17 presents the result of this experiment, showing the number of VMs in the dependency chain. As expected, the result shows that the down time increases with the length of this chain, adding approximately 200ms for each VM in the chain.

## VII. RELATED WORK

We classify prior work most closely related to SSC into two broad categories, those based on nested virtualization, and those based on software-defined networking (SDN).

### A. Techniques based on Nested Virtualization

The Turtles project [45] demonstrated that nested virtualization is possible with relatively low overheads: it demonstrated an overhead of approximately 7% on some I/O bound

workloads for one level of nesting. Since then, nesting has been implemented in popular virtualization platforms such as KVM. A number of recent projects have exploited this trend to provide many of the features that SSC does.

The CloudVisor project [10] uses nested virtualization to protect the contents of user VMs from the cloud operator. It does so using a small bare-metal hypervisor atop which the cloud’s hypervisor executes. The bare-metal hypervisor only exports an encrypted view of the client’s VMs to dom0, thereby achieving the goal of security. CloudVisor’s approach has the advantage of presenting a small TCB and using strong cryptographic techniques to protect security. In contrast, it is unclear from the CloudVisor paper whether the architecture can be extended to implement SD-like services on a client’s VMs.

The XenBlanket project [8] uses nested virtualization to allow clients to implement their own services on their VMs. XenBlanket is a hypervisor-like layer that executes atop the cloud’s hypervisor. The client can deploy SD-like services and administer its VMs, all of which execute atop XenBlanket. While XenBlanket provides clients flexible control over their VMs, it does not aim to protect the security of client code and data from cloud operators.

It may be possible to achieve the goals of both CloudVisor and XenBlanket using two levels of nested virtualization. However, research has shown that the overheads of nested virtualization grow exponentially with the number of nested levels [46].

Aside from these projects, whose goals are directly related to SSC, there are also other projects that use nesting to implement specific services that can be implemented as SDs within SSC. Alibi [40], for instance, is one such effort that uses nesting to implement trustworthy resource accounting services (such as network metering).

### B. Techniques based on SDN

Software-defined networking (SDN) technologies allow programmatic control over the network’s control elements. The client implements its policies using a high-level language, and the SDN system configures individual network elements to enforce these policies. The beginnings of the current OpenFlow-based SDN revolution can be traced back to projects that aimed to control middlebox placement in the cloud. This topic remains an active area of current research in the networking community (e.g., [47, 48, 49, 50, 51]).

Among these projects, closely related to SSC is Cloud-NaaS [47], which develops SDN-based techniques to configure network elements so that certain client policies are enforced, e.g., a network packet to a work VM always traverses an IDS middlebox. Recent work on SIMPLE [51] enhances this basic model to allow composition of middleboxes, among other new features.

We view this line of work as being complementary to SSC. SSC enables a number of new features that cannot be implemented using SDN alone—protecting client VMs from cloud operators, endowing SDs with specific privileges over client VMs via GRANT\_PRIVILEGE, specifying rich inter-VM dependencies, and offering VM dependency-aware migration

policies. SSC currently uses Open vSwitch-based VMs to suitably route traffic to client VMs that have a network middlebox hosted on a different physical machine. It may be possible to leverage SDN technology to enable such routing.

### C. Other Work Related to SSC

Aside from work in the two broad areas discussed above, SSC is also related more broadly to work on improving the client VM security in cloud environments. Efforts in this direction include reducing the size of the cloud’s TCB (e.g., using tiny hypervisors [52, 53, 54] and hardware-only techniques [11]). SSC takes a practical approach by working with a commodity hypervisor. While this has the implication of having a TCB of relatively large size, it also ensures that popular hypervisor features, often developed in response to customer demand, are supported by SSC. In taking this approach, SSC follows the lead of the work by Murray *et al.* [55] and Xoar [9].

## VIII. SUMMARY

The SSC platform, originally introduced in prior work [14], offers a new hypervisor privilege model that protects client code and data from cloud operators while simultaneously giving clients greater and more flexible control over their VMs running in the cloud. While the original paper on SSC focused on the hypervisor privilege model, it left largely unaddressed the question of how to build a distributed cloud computing platform using such hypervisors.

This paper fills that gap, and presents the design and implementation of SSC’s control plane. Novel features of this control plane include giving clients the ability to specify inter-VM dependencies, allowing flexible placement of middleboxes, and new VM migration protocols to support these features. With the control plane in place, SSC can serve as an alternative to current cloud platforms, giving clients greater security and flexibility over their VMs in the cloud setting.

## REFERENCES

- [1] CVE-2007-4993, “Xen guest root escapes to dom0 via pygrub.”
- [2] CVE-2007-5497, “Integer overflows in libext2fs in e2fsprogs.”
- [3] CVE-2008-0923, “Directory traversal vulnerability in the shared folders feature for VMWare.”
- [4] CVE-2008-1943, “Buffer overflow in the backend of XenSource Xen paravirtualized frame buffer.”
- [5] CVE-2008-2100, “VMWare buffer overflows in VIX API let local users execute arbitrary code in host OS.”
- [6] K. Kortchinsky, “Hacking 3D (and breaking out of VMWare),” in *BlackHat USA*, 2009.
- [7] Gartner, “Assessing the Security Risks of Cloud Computing,” <http://www.gartner.com/DisplayDocument?id=685308>.
- [8] D. Williams, H. Jamjoom, and H. Weatherspoon, “The Xen-Blanket: Virtualize Once, Run Everywhere,” in *ACM EuroSys*, 2012.
- [9] P. Colp, M. Nanavati, J. Zhu, W. Aiello, G. Coker, T. Deegan, P. Loscocco, and A. Warfield, “Breaking Up is

- Hard to Do: Security and Functionality in a Commodity Hypervisor,” in *ACM SOSP*, 2011.
- [10] F. Zhang, J. Chen, H. Chen, and B. Zang, “CloudVisor: Retrofitting Protection of Virtual Machines in Multi-tenant Cloud with Nested Virtualization,” in *ACM SOSP*, 2011.
  - [11] E. Keller, J. Szefer, J. Rexford, and R. Lee, “Eliminating the hypervisor attack surface for a more secure cloud,” in *ACM CCS*, 2011.
  - [12] N. Santos, R. Rodrigues, K. Gummadi, and S. Saroiu, “Policy-sealed data: A new abstraction for building trusted cloud services,” in *USENIX Security*, 2012.
  - [13] A. Srivastava and V. Ganapathy, “Towards a richer model for cloud app markets,” 2012.
  - [14] S. Butt, A. Lagar-Cavilla, A. Srivastava, and V. Ganapathy, “Self-service cloud computing,” in *ACM CCS*, 2012.
  - [15] A. Baliga, V. Ganapathy, and L. Iftode, “Detecting kernel-level rootkits using data structure invariants,” *IEEE TDSC*, vol. 8, no. 5, 2011.
  - [16] L. Litty, H. A. Lagar-Cavilla, and D. Lie, “Computer Meteorology: Monitoring Compute Clouds,” in *HotOS*, 2009.
  - [17] B. Payne, M. Carbone, and W. Lee, “Secure and Flexible Monitoring of Virtual Machines,” in *ACSAC*, 2007.
  - [18] B. Payne, M. Carbone, M. Sharif, and W. Lee, “Lares: An architecture for secure active monitoring using virtualization,” in *IEEE Symposium on Security & Privacy*, 2008.
  - [19] S. Bleikertz, S. Bugiel, H. Ideler, S. Nurnberger, and A. Sadeghi, “Client-controlled cryptography-as-a-service in the cloud,” in *Applied Cryptography and Network Security Conference*, 2013.
  - [20] S. Berger, R. Caceres, K. Goldman, R. Perez, R. Sailer, and L. van Doorn, “vTPM: Virtualizing the Trusted Platform Module,” in *USENIX Security*, 2006.
  - [21] VMware vCloud, <http://www.vmware.com/products/datacenter-virtualization/vcloud-suite/overview.html>, last accessed Aug. 8, 2013.
  - [22] Amazon, “Elastic Compute CCloud (EC2),” <http://aws.amazon.com/ec2>.
  - [23] OpenStack, <http://www.openstack.org/>, last accessed Aug. 8, 2013.
  - [24] OpenNebula, <http://opennebula.org/>, last accessed Aug. 8, 2013.
  - [25] CloudStack, <http://cloudstack.apache.org/>, last accessed Aug. 8, 2013.
  - [26] Eucalyptus, [www.eucalyptus.com/eucalyptus-cloud/iaas](http://www.eucalyptus.com/eucalyptus-cloud/iaas), last accessed Aug. 8, 2013.
  - [27] R. Sailer, X. Zhang, T. Jaeger, and L. van Doorn, “Design and implementation of a TCG-based integrity measurement architecture,” in *USENIX Security*, 2004.
  - [28] A.-R. Sadeghi, C. Stubble, and M. Winandy, “Property-based TPM virtualization,” in *Information Security Conference*, 2008.
  - [29] A. Nagarajan, V. Varadarajan, M. Hitchens, and E. Gallery, “Property-based attestation and trusted computing: Analysis and challenges,” in *Intl. Conf. on Network and System Security*, 2009.
  - [30] E. G. Sirer, W. de Bruijn, P. Reynolds, A. Shieh, K. Walsh, D. Walsh, and F. B. Schneider, “Logical Attestation: An authorization architecture for trustworthy computing,” in *ACM SOSP*, 2011.
  - [31] B. Pfaff, J. Pettit, K. Amidon, M. Casado, T. Koponen, and S. Shenker, “Extending networking into the virtualization layer,” in *HotNets*, 2009, <http://openvswitch.org>.
  - [32] Y. Zhang, A. Juels, A. Oprea, and M. K. Reiter, “Home-alone: Co-residency detection in the cloud via side-channel analysis,” in *IEEE Symposium on Security and Privacy*, 2011.
  - [33] T. Ristenpart, E. Tromer, H. Shacham, and S. Savage, “Hey, you, get off of my cloud: Exploring information leakage in third-party compute clouds,” in *ACM CCS*, 2009.
  - [34] Y. Zhang, A. Juels, M. Reiter, and T. Ristenpart, “Cross-VM side channels and their use to extract private keys,” in *ACM CCS*, 2012.
  - [35] V. Varadarajan, T. Kooburat, B. Farley, T. Ristenpart, and M. Swift, “Resource-freeing attacks: Improve your cloud performance (at your neighbor’s expense),” in *ACM CCS*, 2012.
  - [36] B. Farley, V. Varadarajan, K. Bowers, A. Juels, T. Ristenpart, and M. Swift, “More for your money: Exploiting performance heterogeneity in public clouds,” in *ACM SOCC*, 2012.
  - [37] B. Cully, G. Lefebvre, D. T. Meyer, M. Feeley, N. C. Hutchinson, and A. Warfield, “Remus: High availability via asynchronous virtual machine replication. (best paper),” in *NSDI*, 2008.
  - [38] iperf3, <http://code.google.com/p/iperf/>, last accessed Aug. 8, 2013.
  - [39] This reference is a URL pointing to our discussion with the Xen developers on [lists.xen.org](http://lists.xen.org). Because the discussion at this URL reveals our identity, we have not provided it here. PC members interested in obtaining this URL can contact the authors via the PC chair.
  - [40] C. Chen, P. Maniatis, A. Perrig, A. Vasudevan, and V. Sekar, “Towards verifiable resource accounting for outsourced computation,” in *VEE*, 2013.
  - [41] TCPDump and libpcap, <http://www.tcpdump.org>, last accessed Aug. 8, 2013.
  - [42] Snort, <http://www.snort.org/>, last accessed Aug. 8, 2013.
  - [43] A. Srivastava and J. Giffin, “Tamper-resistant, application-aware blocking of malicious network connections,” in *RAID*, 2008.
  - [44] vmitools, <http://code.google.com/p/vmitools/>, last accessed Aug. 8, 2013.
  - [45] M. Ben-Yahuda, M. D. Day, Z. Dubitsky, M. Factor, N. Har’El, A. Gordon, A. Liguori, O. Wasserman, and B. Yassour, “The Turtles project: Design and implementation of nested virtualization,” in *USENIX/ACM OSDI*, 2010.
  - [46] B. Kauer, P. Verissimo, and A. Bessani, “Recursive virtual machines for advanced security mechanisms,” in *1st International Workshop on Dependability of Clouds, Data Centers and Virtual Computing Environments*, 2011.
  - [47] T. Benson, A. Akella, A. Shaikh, and S. Sahu, “Cloudnaas: a cloud networking platform for enterprise applications,” in *SoCC*, 2011.
  - [48] A. Greenlough, M. Handley, M. Hoerdt, F. Huici, L. Mathy, and P. Papadimitriou, “Flow processing and the rise of commodity network hardware,” *ACM Computer Communications Review*, 2009.



- [49] M. Casado, M. Freedman, J. Petit, J. Luo, N. McKeown, and S. Shenker, “Ethane: Taking control of the enterprise,” in *ACM SIGCOMM*, 2007.
- [50] L. E. Li, V. Liaghat, H. Zhao, M. Hajiaghayi, D. Li, G. Wilfong, Y. R. Yang, and C. Guo, “PACE: Policy-aware application cloud embedding,” in *INFOCOM*, 2013.
- [51] Z. Qazi, C. Tu, L. Chiang, R. Miao, V. Sekar, and M. Yu, “SIMPLE-fying middlebox policy enforcement using SDN,” in *ACM SIGCOMM*, 2013.
- [52] U. Steinberg and B. Kauer, “NOVA: A Microhypervisor-Based Secure Virtualization Architecture,” in *ACM Eurosys*, 2010.
- [53] J. M. McCune, Y. Li, N. Qu, Z. Zhou, A. Datta, V. Gligor, and A. Perrig, “TrustVisor: Efficient TCB reduction and attestation,” in *IEEE Symposium on Security & Privacy*, 2010.
- [54] A. Seshadri, M. Luk, N. Qu, and A. Perrig, “SecVisor: A Tiny Hypervisor to Provide Lifetime Kernel Code Integrity for Commodity OSes,” in *ACM SOSR*, 2007.
- [55] D. Murray, G. Milos, and S. Hand, “Improving Xen Security Through Disaggregation,” in *ACM VEE*, 2008.