The Computer Journal Advance Access published June 26, 2012

doi:10.1093/comjnl/bxs060

Enterprise-Ready Virtual Cloud Pools: Vision, Opportunities and Challenges†

TIMOTHY WOOD[1], K.K. RAMAKRISHNAN[2,*], PRASHANT SHENOY[3] AND JACOBUS VAN DER MERWE[2]

[1]*George Washington University, Washington, DC 20052, USA*
[2]*AT&T Labs Research, Florham Park, NJ, USA*
[3]*Department of Computer Science, University of Massachusetts Amherst, AM, USA*
[*]*Corresponding author: kkrama@research.att.com*

Cloud computing platforms such as Amazon EC2 provide customers with flexible, on-demand resources at low cost. However, while existing offerings are useful for providing basic computation and storage resources, they have not provided the transparency, security and network controls that many enterprise customers would like. While cloud computing has a great potential to change how enterprises run and manage their IT systems, a more comprehensive control over network resources and security needs to be provided for such users. Toward this goal, we propose a virtual cloud pool (VCP) abstraction to logically unify cloud and enterprise data center resources, and present the vision behind CloudNet, a cloud platform architecture that utilizes virtual private networks to securely and seamlessly link cloud and enterprise sites. It also enables the pooling of resources across data centers to provide enterprises the capability of having cloud resources that are dynamic and adaptive to their needs. We describe several usage scenarios for VCPs and discuss the benefits of using this abstraction in enterprise settings.

Keywords: cloud computing; IT management; data centers; virtualization; virtual private networks; security

Received 16 January 2012; revised 2 May 2012
Handling editor: Bharat Jayaraman

1. INTRODUCTION

Cloud computing enables enterprises large and small to manage resources better—some because they no longer need to invest in local IT resources and instead can lease cheaper, on-demand resources from providers, and others because they can utilize the flexibility of cloud resources to dynamically meet peak demand without investing in in-house resources. Cloud computing is a natural fit for enterprise customers since it enables outsourcing of another set of non-core competencies: IT infrastructure selection, ordering, deployment and management. Cloud computing allows enterprises to obtain as much computation and storage resources as they require, while only paying for the precise amount that they use. Since cloud platforms rely on virtualization, new resources can be quickly and dynamically added to a customer's resource pool within minutes. From a cloud computing service provider's perspective, server virtualization allows them to flexibly multiplex resources among customers without needing to dedicate physical resources individually. These features have driven the growth of commercial cloud computing services, making them increasingly popular and economical.

However, current cloud computing services need to further evolve to fully meet the needs of businesses. It is highly desirable that cloud resources be seamlessly integrated into an enterprise's existing infrastructure without having to deal with substantial configuration, address management or security concerns. Instead, current commercial solutions present cloud servers as isolated entities with their own IP address space that is outside the customer's control. The separation of cloud and enterprise resources increases software and configuration complexity when deploying services that must communicate with an enterprise's private network. This can lead to security concerns since enterprise customers must utilize IP addresses on the *public* Internet in order to link application components in the cloud to their own sites. It is left to the customer to manage security on the cloud resources and the enterprise network through firewall rules. Finally, existing cloud services

†A preliminary version of this paper appeared at the 2009 USENIX HotCloud workshop.

focus on storage and computation resources, and do not allow for control over network resources either within the cloud or the network linking enterprise and cloud sites. The lack of coordination between network and cloud resources leaves the customer again responsible for independently arranging for traffic isolation and bandwidth guarantees with a separate network service provider. We propose enhancing the cloud computing framework to seamlessly integrate virtual private networks (VPNs) and overcome these deficiencies. To this end, we propose a new abstraction called virtual cloud pools (VCPs) that unify resources across different enterprise and cloud data center sites into logical pools of resources that are easy to manage. We describe the vision behind CloudNet, a platform that implements the VCP abstraction by combining cloud computing platforms and VPN protocols to provide secure communication channels and to allow customer's greater control over network provisioning and configuration.

We discuss several challenges that must be dealt with when combining VPNs and cloud platforms in this manner. First, creating VPN endpoints requires coordination between the network operator and cloud service provider; existing cloud services do not provide sufficient 'hooks' to allow cloud resources to be securely attached to a VPN endpoint. Next, provider-operated VPNs typically extend only between edge routers within the provider network. Cloud operators must ensure that network isolation extends through any local network infrastructure, e.g. switches and routers, within the cloud site itself. Finally, VPNs have traditionally been provisioned at slow management timescales mainly because the endpoints of a VPN were expected to remain static for long periods of time. In cloud computing, however, flexibility and rapid provisioning are key requirements, and it is essential that the network transparency and secure communication channels provided by VPNs remain effective despite rapid changes in configuration.

We discuss the vision, opportunities and challenges in realizing VCPs by presenting an overview of the CloudNet platform architecture, which enables the creation of flexible, secure resource pools transparently connected to enterprises via VPNs. CloudNet achieves these goals by automating the creation and management of VPN endpoints and allowing for explicit coordination between the cloud platform and the network service provider. Finally, we present several usage scenarios for VCPs and discuss the benefits of using the VCP abstraction and CloudNet platform in enterprise settings.

2. BACKGROUND & RELATED WORK

Commercial Platforms: Cloud computing has rapidly grown in popularity over the last few years due to the reemergence of virtualization as an efficient method of flexibly sharing resources. There are many different types of cloud computing services ranging from web-based word processors and email clients to application development platforms like Google App Engine [1] to virtual infrastructure providers like Amazon EC2 [2] that lease full virtual machines to customers. Armbrust *et al.* [3] provide a good overview of the various types of cloud computing platforms, as well as many of the challenges and benefits of cloud computing. In this work we focus on *Infrastructure as a Service* (IaaS) providers since they provide the greatest flexibility for enterprise users who already have large software systems that they would like to move 'to the cloud' with minimal changes.

Existing IaaS platforms such as EC2 already allow customers to lease storage and computation resources on demand. While EC2 allows control over the type of CPU and storage available to each virtual machine instance, it has more limited controls over the network setup, particularly for enterprise customers looking to securely connect cloud resources to their existing infrastructure. EC2 allows for the specification of either public or internal (cloud only) network interface for each virtual machine (VM), although the precise placement and IP details are determined by the provider. EC2 also allows for Security Groups to be created which specify firewall rules for each VM. While firewalls provide very fine grain access controls, a higher level of abstraction is useful when trying to cleanly link different enterprise and cloud sites. Managing complex firewall rules as virtual machines are dynamically created and moved between sites can be very difficult. Instead, we propose that VPNs be used to provide not only stronger security, but additional features such as network reservation controls and seamless integration of cloud and local resources.

Two additional types of cloud computing environments have been developed to help with these concerns: private and overflow clouds [3]. In a private cloud, the customer is given exclusive access to a portion of a data center which then manages the resources using cloud computing techniques. Overflow clouds are used as backup service pools that are only used when an enterprise's own resources are completely saturated. These techniques allow for enterprises to obtain some of the benefits of cloud computing while reducing the security concerns related to using public clouds. However, the network transparency and resource control issues remain. Our work attempts to make any type of cloud more transparent, secure and flexible.

Compute as a Service and Storage as a Service: Enterprise-ready IaaS platforms also come in two flavors: one that is primarily a Compute as a Service (CaaS), offering computation resources; another is to have Storage as a Service in the cloud.

CaaS offers processing power that can be used when running the enterprise's applications, for 'cloudbursting' (which we describe in more detail below) of in-house applications, or as a testing and development platform. The service can run as a public cloud, or as a private cloud on a cloud provider's infrastructure, connected to a customer data center by a network. This enables customers to create a provider-based private cloud accessed either via the public Internet or 'private' network connections.

Enterprises retain more and more data which are growing all the time. Compliance regulations also require more and more retention of data. Regardless of the data's location, there is also a need to provide for disaster recovery (DR). If the current storage system goes down, the expectation is that the data are immediately available forever. Storage as a Service provides secure, virtualized storage in the cloud that is accessible from virtually anywhere. Enterprises load their data into the secure, virtual container and pay only for what they use. The virtual container automatically expands as data are added, and, as the data are deleted, it automatically contracts. Storage can be presented to developers as an object store or a network attached disk [2], or directly to users as an online backup program [4]. Enterprises get as much or as little capacity as they need for as long as they need it, providing tremendous flexibility.

For all these types of services, it is key for the enterprises to have control over their data and to have some assured performance and security. Enterprise customers may be particularly interested in the ability of the network to defend against denial of service attacks. Accessing compute and storage services in private clouds housed in a service provider's data centers has the potential to ease the concern's of enterprise customers.

VPNs: In this work, we focus on VPNs provided by a network operator, as opposed to technologies such as IPSec VPNs that create software tunnels between each end host. Network-based VPNs are typically realized and enabled by multiprotocol label switching (MPLS) provider networks, following the 'hoses model' [5] and are already commonly used by many enterprises. Provider-based VPNs can provide either layer-3 VPNs following RFC 2547, or layer-2 virtual private LAN Service (VPLS) VPNs according to RFC 4761. CloudNet relies on network-based VPNs since they require no endhost configuration, have lower overheads, and can provide additional services from the network provider such as resource reservation. While these VPNs currently require long setup periods, our work is investigating methods for automating their deployment to allow for rapid changes in VPN topologies.

Related Work: In this work, we focus on providing networking support for enterprise cloud platforms with VPNs. While they do not target enterprise applications, both the Virtuoso and VIOLIN projects address a similar problem of managing networks of virtual machines hosted across multiple grid computing sites [6, 7]. Both systems use overlay networks to intercept VM traffic and tunnel it between sites. This requires additional software to be run at each site to create the overlay network. In our work, we leverage existing technology available from network providers such as Layer 2 and 3 VPNs to not only create seamless connections between sites, but to provide greater security and resource control as well.

The InterCloud system explores the potential for federated cloud platforms to provide highly scalable services [8]; CloudNet seeks to build a similar environment, but focuses on providing seamless and secure connectivity between both data center and cloud platform sites. While our work focuses primarily on the interface between the network and cloud providers, others have argued for the need of more flexible control within cloud platforms themselves [9].

We use the VCP abstraction provided by CloudNet for a variety of scenarios, including live migration between data centers and clouds. Transitioning applications to the cloud is a difficult problem because of the need to reconfigure applications and infrastructure [10]. Live wide area network (WAN) migration of VMs simplifies this process, but it is only beginning to be supported by commercial virtualization platforms [11]. Researchers have demonstrated live migration over the Internet [12, 13], but a key remaining challenge has been building a network infrastructure that supports dynamic redirection of traffic after an application is moved.

3. VISION, CHALLENGES AND INSIGHTS

Cloud computing has seen widespread adoption for public web services and infrequent batch-style applications, but has not yet been accepted as a viable choice for many enterprise uses. To illustrate some of the reasons, consider an enterprise accounting application that consists of a front-end interface, a processing tier and a back-end database. This is an application that contains private business data, and thus is traditionally run in a secure private network environment within the enterprise. Let us examine the challenges that appear when the enterprise desires to move the processing component of this application to a cloud computing data center in order to achieve greater scalability or to reduce IT costs.

Transparent Cloud Connections: Using current commercial cloud offerings, the enterprise could easily create a set of virtual machines within the cloud to run the application, but would quickly encounter difficulties when trying to link the different application components in and out of the cloud. Both the VM in the cloud and the components still in the enterprise would need to be allocated public IP addresses in order to establish connectivity. Even ignoring the security concerns in this situation, the enterprise will have to make significant changes to its own infrastructure to enable external access to formerly private resources, plus it must modify the application code to handle the new network topology, especially if the existing code assumed all components were within a LAN, for example, to utilize broadcasting. This demonstrates a critical limitation of current cloud offerings: the inability to create seamless connections between cloud and enterprise resources.

Security and Isolation: Simply providing the appearance of cloud resources being attached to an enterprise's local network is insufficient (and in fact potentially dangerous) unless the network connections are made via secure channels that safely link only authorized cloud nodes to the enterprise network. To secure the processing component moved to the cloud in our example, an enterprise would need to create firewall rules both

within the cloud and at the gateways to its own network in order to securely limit connectivity. While firewall rules can be used to provide fine grained access controls, they can easily expose security holes if misconfigured, and are particularly vulnerable to this when resources are frequently being added or moved as is the case in dynamic cloud computing environments. Managing the set of dynamically changing firewall rules across potentially multiple enterprise sites can be complex. Current cloud systems are ill-suited for dealing with enterprise applications because they cannot provide secure communication channels or strict network isolation in and out of the cloud.

Flexible Cloud & Network Resources: A key motivation for moving applications into the cloud is the ease with which new resources can be allocated or moved. The enterprise may have moved the processing component into the cloud so that new replicas could easily be added on demand. This action may be unreasonably difficult if it requires further reconfiguration of the application code to handle the new VM's IP address, as well as adjustments to both the cloud and enterprise firewalls to support the new network topology. Instead, cloud operators must coordinate with network providers to offer dynamic configuration of server, storage and network resources to meet enterprise demands.

Additionally, the enterprise may be willing to pay for quality of service guarantees that ensure low latency between the front-end component hosted at the enterprise site and the processing component in the cloud. Alternatively, it may require a large amount of dedicated bandwidth on the link between the processing component and database. Existing cloud services are unable to provide these guarantees because they do not support coordination with network operators to provision resources within the network that links the cloud and enterprise sites.

4. VIRTUAL CLOUD POOLS

To address these challenges, we propose the abstraction of a VCP. A VCP is a combination of cloud computing and storage resources with a VPN infrastructure to give users the abstraction of a private set of cloud resources that are transparently and securely connected to their own infrastructure. VCPs are created by taking dynamically configurable pools of cloud resources and connecting them to enterprise sites with VPNs. Figure 1 shows a pair of VCPs connected to two different enterprises, each composed of multiple sites. A VCP can span multiple enterprise and cloud data centers, but presents a unified pool of resources to the enterprise.

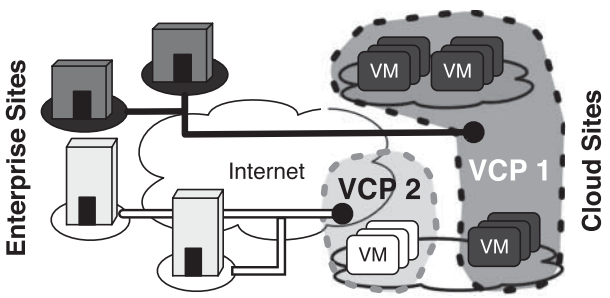

FIGURE 1. Two VCPs isolate resources within the cloud sites and securely link them to the enterprise networks.

VPNs can be leveraged to provide *seamless network connections* between cloud and enterprise sites. VPNs create the abstraction of a private network and address space shared by all VPN endpoints. Since addresses are specific to a VPN, the cloud operator can allow customers to use any IP address ranges that they like without worrying about conflicts between cloud customers. The level of abstraction can be made even greater with VPLS that bridge multiple VPN endpoints onto a single LAN segment. If the cloud provider in the previous section's example used VCPs, a VPLS could be set up so that the processing component could be easily run within the cloud without requiring any modifications since the cloud resources would appear indistinguishable from existing compute infrastructure already on the enterprise's own LAN.

VPNs are already used by many large enterprises to enable *secure any-to-any communication*, and cloud sites can be easily added as new secure endpoints within these existing networks. VCPs use VPNs to provide secure communication channels via the creation of secure, 'virtually dedicated' paths within the provider network. This eliminates the need to configure complex firewall rules between the processing component in the cloud and the enterprise, since all sites would be connected via a private network inaccessible from the public Internet. The VCP solution must guarantee that the secure VPN links extend to the virtual machines that compose the VCP. In order to ensure that different cloud customers are kept on isolated networks, the cloud provider must segment the LAN for each VCP. These techniques can provide strong security guarantees at a convenient level of abstraction.

VCPs enable *flexible resource control* by utilizing resource reservation mechanisms provided by VPNs. By coordinating with the network provider, the cloud service can offer enterprise users quality of service guarantees along the full path from the enterprise to the cloud site. The virtual network abstraction offered by VPNs also allows for flexibility in response to dynamic VM allocation and placement changes. The seamless LAN environment created by a VPLS service can be exploited to automatically handle routing changes as VMs are moved between sites, allowing them to maintain their identities on the LAN. Once two cloud sites are bridged with VPLS, existing LAN migration techniques can be used to move virtual machines across the WAN. This allows a single VCP to span multiple cloud sites as well as multiple enterprise sites, presenting a pool of geographically distributed servers as a flexible resource seamlessly attached to the enterprise's own network.

Our VCP approach is similar to Amazon's virtual private cloud (VPC) product [14]. Amazon's system enables the cloud

to support basic VPN connectivity between the enterprise and the cloud. This allows customers to assign custom IP addresses to their cloud systems and securely connect them to their own networks; however, Amazon's VPC does not yet support the layer 2 VPNs used in CloudNet to enable live migration nor does it allow for dynamic network resource management.

5. CLOUDNET: TOWARD ENTERPRISE CLOUDS

We have developed a system called CloudNet that attempts to meet the requirements of an enterprise ready cloud computing environment by implementing the VCP abstraction. CloudNet is designed based on the three goals laid out previously: (1) seamless connections between enterprise and cloud resources, (2) secure communication channels and well isolated cloud resources and (3) flexible resource allocation capable of handling dynamic changes of cloud and network resources. Since the focus of this paper is on describing the vision and benefits of VCPs for enterprises, we provide a brief overview of the CloudNet architecture to provide sufficient context for our discussion; further details of the CloudNet design may be found in [15, 16]. A commercial implementation with many of CloudNet's ideas can be found in AT&T's CaaS offering [17].

CloudNet leverages existing virtualization technologies at the server, router and network levels to create dynamic resource pools that can be transparently connected to enterprises. The CloudNet architecture is composed of two intelligent controllers that automate the management of resources in the provider network and in the cloud computing data centers, as shown in Fig. 2.

The *Cloud Manager* dynamically partitions the cloud computing data centers into VCPs for use by its customers. The Cloud Manager handles the creation of new virtual machines and manages performance within each VCP. The Cloud Manager utilizes several forms of virtualization so that physical resources can be multiplexed across many customers. In our current prototype, Xen is used to virtualize servers, and VLANs are used to partition the local area networks within each cloud data center. The Cloud Manager uses virtual (or 'logical') routers to dynamically configure the customer edge (CE) routers associated with each VCP. Logical routers are a means to partition physical routers into slices, each with independent control planes. This means that full, physical routers do not need to be dedicated to each VCP, and allows them to be created and reconfigured more rapidly.

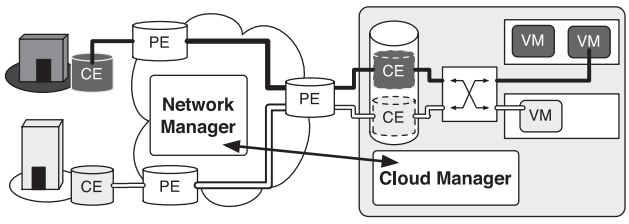

FIGURE 2. The *Network Manager* is responsible for controlling the PE routers at each enterprise and cloud site. This creates the secure, transparent connectivity between locations. The *Cloud Manager* handles the networking within the cloud site and manages the individual VM and storage devices.

The *Network Manager* is run by the network provider and is responsible for the creation and resource provisioning of VPNs. CloudNet utilizes MPLS VPNs that span between the provider edge (PE) routers. The Network Manager dynamically configures the PE routers to create VPN endpoints associated with each VCP. The Network Manager can also be used to specify fine grain access controls that restrict which systems within a single VPN are able to communicate, or to reserve network resources along VPN paths.

Although the Network and Cloud Managers may be controlled by separate entities, possibly different business enterprises, communication between them is required. This is necessary to coordinate the link between the network and customer edges, as well as when a virtual machine is migrated between cloud sites. While standardization and specification of many aspects of Cloud Computing [18] have been ongoing, an important aspect that remains outstanding is the need for a clear specification of the interfaces between the Cloud Manager and the Network Manager. It is highly desirable to have a standardized interface to communicate with the entity that manages and orchestrates the resources within a data center, as well as a standardized interface for the network to expose for the purposes of establishing connectivity between the dynamically created cloud-based resources and the specific endpoints in the network that it needs to communicate with. We see the next phase of development of a 'Networked Cloud' framework as a specification between the entities that implement the Cloud Manager and Network Manager abstractions.

6. VCP AND CLOUDNET BENEFITS AND USAGE SCENARIOS

In this section, we discuss several usage scenarios for VCPs and discuss the benefits of using this abstraction in enterprise settings.

Hybrid cloud: Adding Seamless Cloud Resources: Our first usage scenario is that of a hybrid cloud platform, where an enterprise deploys applications that combine resources from local data centers (also known as a private cloud) with those from a public compute or storage cloud. Doing so involves the creation of a new VCP on the public cloud and connecting it to the local data center, and then exploiting the combined (hybrid) pool of resources to deploy the application(s) or connect to the storage resources. Figure 3 illustrates the process of adding a new secure VCP and transparently connecting it to a customer's VPN. First the Cloud Manager uses a placement algorithm to determine a host with sufficient spare capacity to run the VM

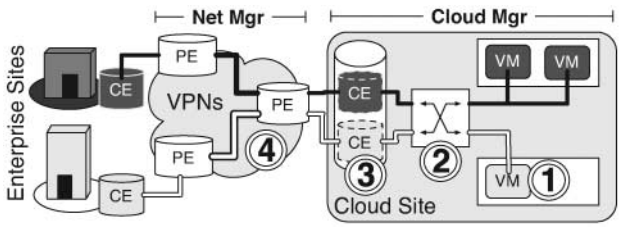

FIGURE 3. Adding a VM to a new site has four steps done by the Cloud Manager (1. VM creation, 2. VLAN configuration on switch, 3. logical CE creation) and the Network Manager (4. VPN setup on PE routers).

(Fig. 3, step 1). Once the VM is created, it must be connected to a VLAN to isolate the customer's traffic within the cloud data center (Fig. 3, step 2). Finally, the Cloud Manager must configure a logical CE router that will be used to connect the new VCP to a VPN endpoint created by the Network Manager (Fig. 3, step 3). This creates an isolated partition of server and network resources within the cloud site that is dedicated to the VCP.

The Network Manager then is responsible for seamlessly connecting the new VCP to the enterprise's network. It does this by defining a new VPN endpoint in the PE router connected to the cloud site. The router advertises the new VPN endpoint via BGP sessions shared with other provider PE routers connected to the enterprise's sites (Fig. 3, step 4). If a Layer 3 VPN is used, then the VCP will be attached to the enterprise network as a new IP-routed domain equivalent to any other enterprise site. Alternatively, VPLS can be used to attach the VCP to an existing extended LAN segment within the enterprise, thus enabling the VCP to seamlessly connect into the enterprise's IT infrastructure. In either case, the VCP network will be securely attached to the enterprise VPN so that its resources are transparently linked to the enterprise infrastructure, but inaccessible from potentially malicious users on the public Internet. This style of approach is starting to be used by cloud platforms such as Amazon EC2, which now both allows VPN connectivity to cloud computation and offers a storage gateway to safely access cloud-based data [14, 19].

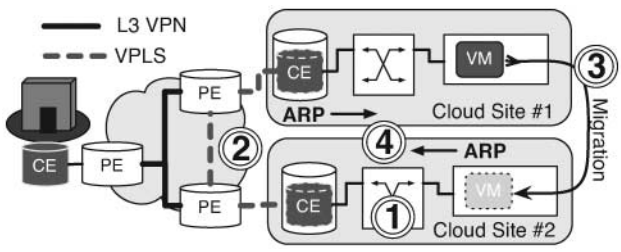

FIGURE 4. CloudNet's VCP infrastructure enables live WAN migration of VMs. Network reconfiguration is supported through the use of dynamically configured VPLS tunnels.

Cloud-to-Cloud Migration: We envision future clouds as flexible resource pools that seamlessly span multiple data center and enterprise sites. With this kind of architecture, cloud providers and enterprise customers are able to transparently migrate VMs between sites. This can allow a data center operator to perform load balancing between sites without impacting customer applications, or customers to change placement decisions based on, for example, WAN latency.

CloudNet takes a step in that direction by simplifying the network reconfiguration for migration of virtual servers between cloud sites. Current virtualization software supports transparent VM migration between physical servers on the same LAN [20, 21], but WAN migration has been a challenge due to the need for network reconfiguration and storage migration. CloudNet exploits the benefits of VPLS to tie networks across the WAN into a single LAN, making transitions across the WAN function as if within a LAN, except for higher network delays during the migration.

Figure 4 depicts the steps to prepare for a VM migration across sites. CloudNet first performs the steps necessary to initialize the VLAN endpoint for the destination site (Fig. 4, step 1). In addition, a VPLS VPN is created to link together the source and destination VLANs (Fig. 4, step 2). At this point, the virtual machine can be migrated between the two sites (Fig. 4, step 3). When the VM is transferred to its new host, it will emit an unsolicited ARP message (Fig. 4, step 4) [20]. The local switch will use this ARP message to establish the mapping between the VM's MAC address and its switch port. The ARP message will also be forwarded through the VPLS to the VM's original site. At that site, the old switch will replace its existing MAC address mapping with the new entry, allowing data to be forwarded through the VPLS to the VM's new site. We discuss the full details of how CloudNet can efficiently migrate both VM storage and memory state between data centers in a separate paper [15].

Cloud bursting: Cloud bursting is a technique where an enterprise normally employs local servers to run enterprise applications and dynamically harnesses cloud servers to enhance capacity during periods when additional resources are needed to manage a significant increase in the workload. For example, enterprises in the retail sector see a substantial increase in their workloads during the holiday peak periods. Investing in the necessary compute resources to handle this peak locally on a year-round basis is economically undesirable.

The availability of computing 'on-demand' from the cloud is a much more attractive way for enterprises to meet their resource requirements dynamically without in-house investment. As shown in Fig. 5, the stress on local IT servers can be mitigated by temporarily migrating a number of overloaded applications to the cloud or by instantiating new application replicas in the cloud to absorb an appropriate amount of the workload increase. These cloud resources are de-allocated once the workload peak has ebbed. Cloud bursting eliminates the need to pre-provision for the peak workload locally, since cloud resources can be

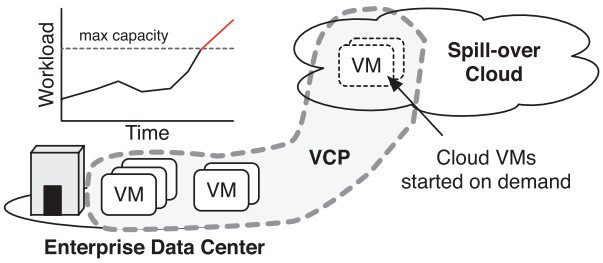

FIGURE 5. When the local enterprise data center reaches full capacity, applications can 'cloud burst' to obtain additional resources on demand.

provisioned dynamically when needed, yielding cost savings due to the cloud's pay-as-you go model. Basic cloud bursting support is already being included in popular cloud platforms and software [22–24].

The key challenge in cloud bursting is that the enterprise needs to be able to bring these compute (and possibly storage) resources into their computing environment in a seamless and secure manner. VCPs and CloudNet simplify the process of cloud bursting by allowing applications from the local data center to be moved to cloud servers in a seamless fashion. Transparency is achieved because the applications do not need to be reconfigured to accommodate the additional IP addresses as is typical in a 'public cloud' scenario. Because of the use of VPNs, the cloud resources appear as an integral part of the enterprise's intra-net environment.

Follow the sun: In a globalized economy, with enterprise locations spanning the globe, 'follow the sun' is a new IT strategy that is designed for project teams that span multiple continents. The scenario assumes multiple groups spanning different geographies that are collaborating on a common project and that each group requires low-latency access to the project applications and data during normal business hours. Computer-aided design (CAD) and very large scale integration (VLSI) design are typical of such applications. One approach is to replicate content at each site—e.g. a data center on each continent—and keep the replicas consistent. While this approach may suffice for content repositories or replicable applications, it is often unsuitable for applications that are not amenable to replication. Especially with project teams dealing with VLSI designs etc., the ability to work off a single common version of the data is desirable. In such a scenario, it may be simpler to migrate one or more VM containers with applications and project data from one site to another every evening; the migration overhead can be reduced by transferring only incremental state and applying it to the snapshot from the previous day to recreate the current state of the application. Figure 6 illustrates how by unifying resources across enterprise data centers on different continents, CloudNet and VCPs simplify the automatic implementation of follow-the-sun IT policies.

Enterprise IT Consolidation: Many enterprises with multiple data centers have attempted to deal with data center 'sprawl' and cut costs by consolidating multiple smaller sites into a few large data centers. Such consolidation requires applications and data to be moved from one site to another over a WAN; a subset of these applications may also be moved to cloud platforms if doing so is more cost-effective than hosting locally. Similarly, when a new data center is built and brought online by an enterprise, existing applications may be moved to servers at the newly built location. Typically such transformation projects have incurred application down-times, often spread over multiple days. Hence, the ability to implement these moves with minimal or no down-time is attractive due to the corresponding reduction in the disruption seen by a business.

Logical IT consolidation: While CloudNet and VCPs simplify physical consolidation of IT resources as described above, thay also enable a simpler form of consolidation, namely logical consolidation. By logically consolidating IT resources, an enterprise can unify computing and storage pools across different geographic sites and manage them as a single resource pool, as if these resources were actually physically consolidated at a single site, as shown in Fig. 7. Logical consolidation does not allow an enterprise to extract savings of economies of scale that can be achieved by physical consolidation of resources, but it still allows an enterprise to reduce IT management costs by simplifying administration of resources as if they were physically consolidated in one site. Logical consolidation may be the preferred choice when IT budgets constrain capital expenditures necessary to carry out a more drastic physical consolidation of resources. With the ability to construct VCPs that span different data center sites, CloudNet can easily permit logical IT consolidation. Doing so also enables an enterprise to retain smaller data center sites that are geographically separate, which provides better resiliency against site-wide

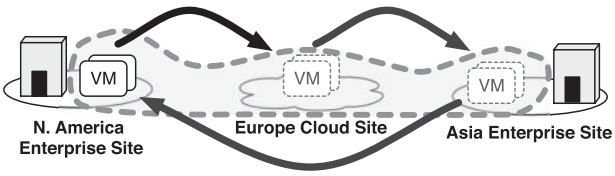

FIGURE 6. VCPs allow global enterprise and cloud sites to be treated as one omnipresent data center, allowing applications to 'follow the sun' between sites to be used by the active workforce.

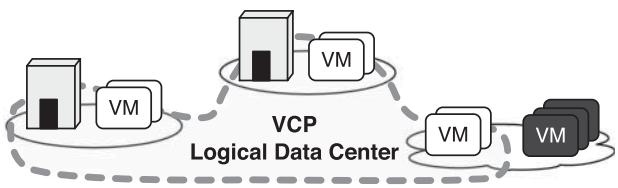

FIGURE 7. CloudNet's VCPs can be used to logically consolidate multiple data center and cloud sites into a single resource pool.

disaster outages when compared with a physically consolidated large data center.

DR using Cloud platforms: CloudNet also enables an enterprise to rely on a remote cloud platform as its backup disaster site. Business and Government services utilize DR systems to minimize data loss and the downtime incurred by catastrophic system failures. Given the need for enterprises to have their backup data safe, CloudNet can enable secure, efficient replication of disk state from local enterprise sites to storage at a remote cloud site. A key challenge in providing DR services is the need to support business continuity (BC), allowing applications to rapidly come back online after a failure occurs. CloudNet's automation enables an enterprise to quickly start up applications in the cloud when a disaster causes a site-wide outage at a specific enterprise data center. As a consequence, the overall goal of achieving BC is largely met by having a good DR capability at the remote cloud data center.

Current DR services come at very high cost both in terms of the dollar cost as well as the performance impact. In terms of the dollar cost, the present mode of having a colocation center requires dedicated resources that are present year round for DR, and to enable the business to restart operations from the colocation site. This is often a significant barrier for enterprises to adopt DR solutions. With virtualization and the cloud data center's pay-as-you-go pricing model for both compute and storage services, we find that it can be a very attractive DR solution. During normal operation, customers only pay for the resources used for the backup at the cloud DR site. This requires the disk storage for replicating data and a minimal server that is used to monitor and respond to the remote disk write requests as shown in Fig. 8. As a consequence, the 'replication mode' resources needed during normal operations is relatively small, and can be provided at the cloud site at low cost. Moreover, the server that is involved in monitoring and responding to the remote disk write requests can be shared across enterprises, further lowering the cost by taking advantage of statistical multiplexing of these resources. It is only during 'failover mode' that additional resources are required at the remote DR site. These resources can be rapidly provisioned and scaled up after the disaster is detected, and it is only during this time that the expense for the cloud resources is incurred (which may be a relatively small number of days in a year). Thus, the overall cost of providing a DR solution in the cloud can be much lower, and therefore can remove a significant barrier for its adoption by enterprises [25]. The VCP framework used by CloudNet ensures that data is sent to the backup site securely and simplifies the network redirection after a failure occurs.

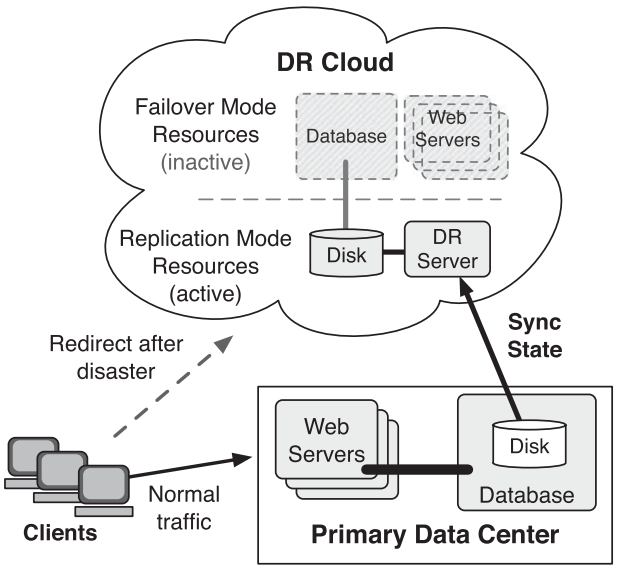

FIGURE 8. Applications can be replicated to the cloud using only a small set of cloud resources. After a failure occurs, additional cloud resources are brought online to run the full applications.

It is also desirable that the DR site be far enough away from the enterprise data center so as to not share the same risks for failure. Using the cloud as a backup DR site typically alleviates this risk since cloud platforms only offer resources at a few locations in each continent, e.g. the east or west coast of the US. However, this geographic separation also leads to higher network latency, which results in a significant performance impact if the remote replica needs to be kept consistent with the primary data center. Synchronous replication of writes results in a significant performance degradation because of the round-trip delay. Alternatively, the weak guarantees coming from Asynchronous replication results in a significant amount of data loss. Our proposed solution of Pipelined Synchrony [26] seeks to overcome this barrier as well.

Other benefits: Online gaming is an additional area that could benefit, due to the increasing demands of massively multiplayer games that require large amounts of compute power, but also need low latency between the server and players to achieve high performance. The logical consolidation supported by CloudNet would allow game operators to link many geographically distributed data centers into a single pool, providing easily expandable capacity and a high quality of service for a wider body of customers. CloudNet's communication between the cloud and network controllers would also enable greater control over network resource management, further enhancing application performance by allowing bandwidth to be dedicated to individual flows.

The global tendency toward renewable, but more variable, energy sources may also encourage the use of a larger number of small data centers rather than a single massive data center. The VCP abstraction would support dynamic relocation of tasks between seamlessly connected data centers to make heavier use of whichever regions currently have the lowest energy price.

CloudNet's use of VPNs to provide secure and seamless connectivity between data center and cloud sites is useful for a variety of other use cases. VCPs provide convenient abstractions

for programmers accustomed to LAN interconnects between application components, secure WAN links between both enterprise and cloud sites, and greater control over both cloud and network resources.

7. CONCLUSIONS

Cloud computing promises to revolutionize computing by providing cheap, flexible, on-demand resources. However, current commercial cloud platforms are incapable of meeting the requirements of enterprise customers. In this paper, we proposed the notion of VCPs that combine existing technologies like VPNs with automated controllers to meet three key requirements of enterprise users: (1) transparent connections between cloud and enterprise resources, (2) isolation within the cloud and secure communication channels between sites and (3) flexible resource allocation schemes capable of responding to dynamic changes of cloud and network resources. We described several challenges in instantiating the VCP abstraction and presented an overview of our CloudNet system that implements VCPs to provide secure and seamless cloud resources to enterprises. VCPs and its CloudNet instantiation have the ability to provide several benefits in today's enterprise settings. We discussed the benefits of VCPs for deploying hybrid public–private clouds, in supporting cloud-to-cloud migration, in implementing cloud bursting and follow-the-sun policies, in logical and physical IT consolidation, in designing DR policies and in scenarios such as online gaming. Our future work involves the design of specific mechanisms and systems to demonstrate the benefits of VCPs for these scenarios.

FUNDING

This work was supported in part by NSF grants CNS-0720271 and CNS-0720616 and an AT&T VURI award.

REFERENCES

[1] Google app engine. http://code.google.com/appengine/. (accessed October 25, 2010).

[2] Amazon web services. http://aws.amazon.com/. (accessed October 25, 2010).

[3] Armbrust, M. *et al.* (2009) Above the Clouds: A Berkeley View of Cloud Computing. Technical Report UCB/EECS-2009-28. EECS Department, University of California, Berkeley.

[4] Dropbox. http://dropbox.com/. (accessed October 25, 2010).

[5] Duffield, N.G., Goyal, P., Greenberg, A., Mishra, P., Ramakrishnan, K.K. and van der Merwe, J.E. (2002) Resource management with hoses: point-to-cloud services for virtual private networks. *IEEE/ACM Trans. Netw.*, **10**, 679–692.

[6] Sundararaj, A.I. and Dinda, P.A. (2004) Towards Virtual Networks for Virtual Machine Grid Computing. *Proc. 3rd Conf. on Virtual Machine Research And Technology Symposium*, Berkeley, CA, USA, pp. 14–14. USENIX Association.

[7] Ruth, P., Rhee, J., Xu, D., Kennell, R. and Goasguen, S. (2006) Autonomic Live Adaptation of Virtual Computational Environments in a Multi-domain Infrastructure. *Proc. IEEE Int. Conf. on Autonomic Computing*, San Jose, CA, June, pp. 5–14. IEEE Computer Society, Washington, DC.

[8] Buyya, R., Ranjan, R. and Calheiros, R.N. (2010) Intercloud: Utility-oriented Federation of Cloud Computing Environments for Scaling of Application Services. *Proc. 10th Int. Conf. on Algorithms and Architectures for Parallel Processing*, Busan, Korea, pp. 13–31. Springer, Berlin, Heidelberg.

[9] Williams, D., Elnikety, E., Eldehiry, M., Jamjoom, H., Huang, H. and Weatherspoon, H. (2011) Unshackle the Cloud! *Proc. 3rd USENIX Conf. on Hot Topics in Cloud Computing*, Portland, OR, pp. 16–16. USENIX Association, Berkeley, CA.

[10] Hajjat, M., Sun, X., Sung, Y.-W.E., Maltz, D., Rao, S., Sripanidkulchai, K. and Tawarmalani, M. (2010) Cloudward Bound: Planning for Beneficial Migration of Enterprise Applications to the Cloud. *Proc. ACM SIGCOMM 2010 Conf.*, New Delhi, India, pp. 243–254. ACM, New York, NY.

[11] Virtual machine mobility with VMware VMotion and Cisco Data Center Interconnect Technologies. http://www.cisco.com/en/US/solutions/collateral/ns340/ns517/ns224/ns836/white_paper_c11-557822.pdf. (accessed October 25, 2010).

[12] Harney, E., Goasguen, S., Martin, J., Murphy, M., and Westall, M. (2007) The Efficacy of Live Virtual Machine Migrations Over the Internet. *Proc. 2nd Int. Workshop on Virtualization Technology in Distributed Computing*, Reno, Nevada, pp. 8:1–8:7. ACM, New York, NY.

[13] Bradford, R., Kotsovinos, E., Feldmann, A. and Schiöberg, H. (2007) Live Wide-area Migration of Virtual Machines Including Local Persistent State. *Proc. 3rd Int. Conf. on Virtual Execution Environments*, San Diego, CA, USA, pp. 169–179. ACM, New York, NY.

[14] Amazon virtual private cloud. http://aws.amazon.com/vpc/. (accessed October 25, 2010).

[15] Wood, T., Ramakrishnan, K.K., Shenoy, P. and van der Merwe, J. (2011) Cloudnet: Dynamic Pooling of Cloud Resources by Live Wan Migration of Virtual Machines. *Proc. 7th ACM SIGPLAN/SIGOPS Int. Conf. on Virtual Execution Environments*, Newport Beach, CA, USA, pp. 121–132. ACM, New York, NY, USA.

[16] Wood, T., Gerber, A., Ramakrishnan, K. K., Shenoy, P. and Van der Merwe, J. (2009) The Case for Enterprise-ready Virtual Private Clouds. *Proc. 2009 Conf. on Hot Topics in Cloud Computing*, San Diego, CA. USENIX Association, Berkeley, CA.

[17] AT&T Cloud Services. https://www.synaptic.att.com/. (accessed October 25, 2010).

[18] Distributed Management Task Force, Inc. (DMTF) (2009) Open virtualization format white paper version 1.00. http://www.dmtf.org/sites/default/files/standards/documents/DSP2017_1.0.0.pdf (February).

[19] Amazon storage gateway. http://aws.amazon.com/storagegateway/. (accessed October 25, 2010).

[20] Clark, C., Fraser, K., Hand, S., Hansen, J.G., Jul, E., Limpach, C., Pratt, I. and Warfield, A. (2005) Live Migration of Virtual

Machines. *Proc. 2nd Conf. on Symp. on Networked Systems Design & Implementation*, Boston, MA, pp. 273–286. USENIX Association, Berkeley, CA.

[21] Nelson, M., Lim, B.-H., and Hutchins, G. (2005) Fast Transparent Migration for Virtual Machines. *Proc. Annual Conf. on USENIX Annual Technical Conference*, Anaheim, CA, pp. 25–25. USENIX Association, Berkeley, CA.

[22] Case Study: USA.gov Achieves Cloud Bursting Efficiency Using Terremark's Enterprise Cloud. http://bit.ly/ua5Qq2. (accessed October 25, 2010).

[23] Open Nebula: The Open Source Toolkit for Data Center Virtualization. http://www.opennebula.org. (accessed October 25, 2010).

[24] VMware: Public & Hybrid Cloud Computing. http://www.vmware.com/solutions/cloud-computing/public-cloud/products.html. (accessed October 25, 2010).

[25] Wood, T., Cecchet, E., Ramakrishnan, K.K., Shenoy, P., van der Merwe, J. and Venkataramani, A. (2010) Disaster Recovery as a Cloud Service: Economic Benefits & Deployment Challenges. *Proc. 2nd USENIX Conf. on Hot Topics in Cloud Computing*, Boston, MA. USENIX Association, Berkeley, CA.

[26] Wood, T., Lagar-Cavilla, H.A., Ramakrishnan, K.K., Shenoy, P. and Van der Merwe, J. (2011) Pipecloud: Using Causality to Overcome Speed-of-light Delays in Cloud-based Disaster Recovery. *Proc. 2nd ACM Symp. on Cloud Computing*, Cascais, Portugal, pp. 17:1–17:13. ACM, New York, NY.