PROTECTING THE PHYSICAL BOUNDARIES OF VIRTUAL MACHINES IN CLOUD COMPUTING

Ethan Hadar
CA Technologies,16 Shenkar, Herzelia, Israel

Keywords: Access control, Agile architecture, Cloud computing, Cloud security.

Abstract: Cloud computing is currently focused on providing utility computing, such as storage and computing power, and Software as a Service (SaaS.) However, use of these offerings in an enterprise environment requires security compliance to be enforced on managed servers deployed in the cloud in order to prevent un-authorized administration of deployed images. This requirement may apply to the initial deployment, upgrades, or any physical roaming due to the virtualization of the cloud. This paper presents a conceptual reference architecture, that defines a set of conceptual components that are responsible for intercepting calls, managing administrative commands, controlling roaming and portability, and providing enterprise compliance needs. It inherently supports delegation of the enforcing policies, and consequently it provides a scalable solution.

SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

Cloud computing is one of the latest paradigms in computing technology, gaining considerable popularity within the past year. Cloud computing consists of next-generation datacenters providing reliable "on-demand" and "pay-per-use" services (Weiss, 2007; Wikipedia, 2009). This paradigm has implications for application providers who usually develop secured solutions for internal deployment (local) or within enterprise datacenters (near), rather than using remote web-services of third-party datacenters (far).

An example of a complex deployment topology in a cloud environment is a far deployment of web servers, load balancers and front end interfaces in an external public cloud, while the application database remains local. A complex topology like this is often mandated when regulations or contracts require that data be stored on premises. Such cases require secure communication between the local database and the remote front end, as well as typical enterprise security policies and credentials for the remote front end.

An example is a transient workload system that offloads some jobs that do not contain sensitive information to a public cloud. However, these jobs executed on a public Infrastructure as a Service (IaaS) domain may participate in an overall business transaction and are consequently part of a workflow with business logic that must be protected within the enterprise firewall. In a further complication, part of the business transaction might involve a service offered by a third party, such as identity validation, that is utilized as Software as a Service (SaaS) and is not part of the consumer infrastructure. In such a case, the third-party service must be monitored as well, and be part of a secured end-to-end transaction. This highlights a third aspect of cloud security: the way in which transactions must be secured may vary as loads shift from cloud to cloud.

Virtualization technologies (Vouk, 2008) introduce a new set of identity, access and auditing problems that could hamper the adoption of cloud computing. These technologies increase the risk of security exposure in several ways, one of which is the *Protecting the physical boundaries of virtual machines*. Since A Virtual Host (VH) image can be deployed and run on many possible physical machines. In some cases (such as legal, governance or privacy) there is a need to restrict geographical locations of the deployed VH.

There is a gap between the cloud offerings and the required secure solution. A compelling characteristic of the cloud paradigm is the power of abstraction, which frees consumers from concern with the physical management of the underlying infrastructure, but the absence of limitations on the

Hadar E..
PROTECTING THE PHYSICAL BOUNDARIES OF VIRTUAL MACHINES IN CLOUD COMPUTING.
DOI: 10.5220/0003583306650670
In *Proceedings of the 1st International Conference on Cloud Computing and Services Science* (CLOSER-2011), pages 665-670
ISBN: 978-989-8425-52-2
Copyright © 2011 SCITEPRESS (Science and Technology Publications, Lda.)

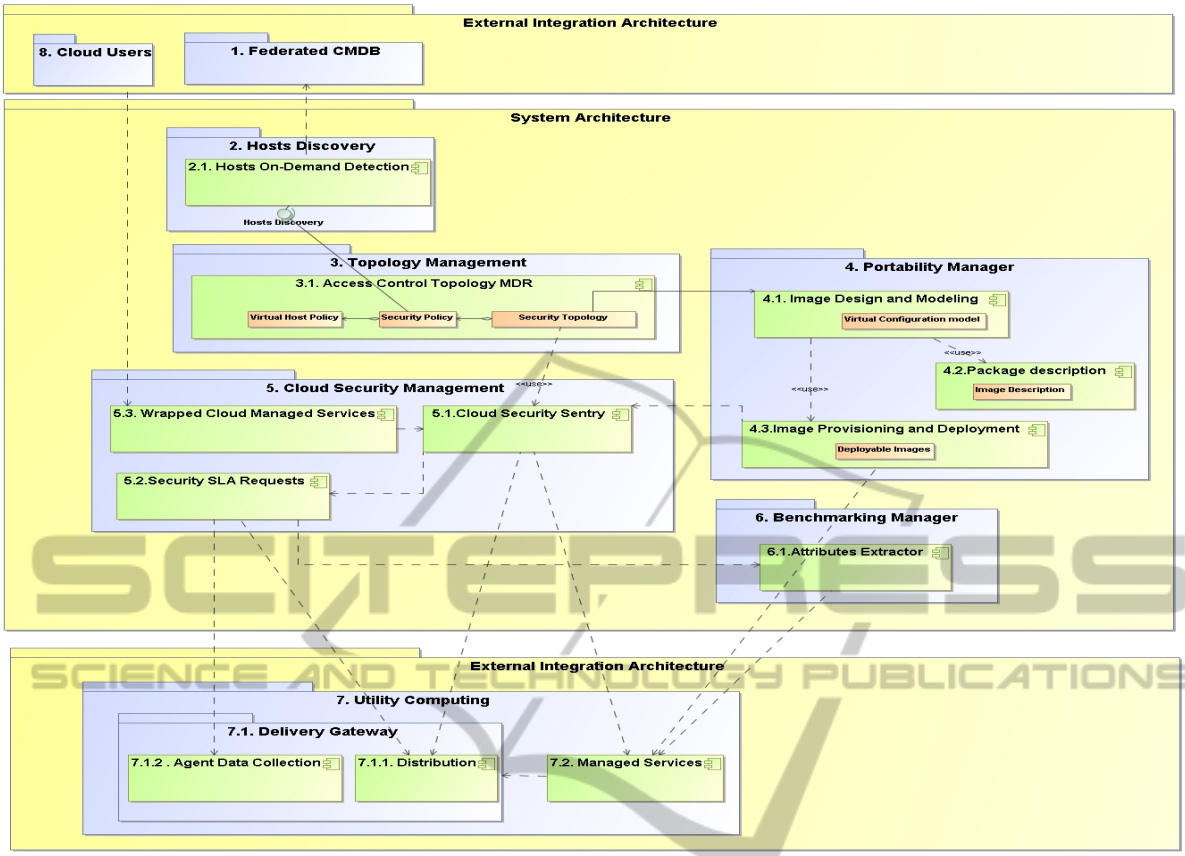

Figure F-1: Access Control Reference Architecture for Cloud Computing.

physical server hosting the customer's sensitive data, in some cases, conflicts with privacy and other regulations. In these cases, failover and backup of data must occur within prescribed geographic limits, and the data must be secured according to law within restricted, typically national, boundaries. Even more stringent regulations, such as those typical of banking and other financial activities, can require full transparency into the system structure to prove that the sensitive data is secured with a known set of technologies, and the location of the data is clearly circumscribed to a limited zone. Healthcare privacy regulations present similar requirements

An adaptation layer that wraps the cloud offering, restricts roaming, and protects physical boundaries is one means of providing transparency and control while seamlessly connecting the cloud to the common enterprise IT environment.

This paper proposes a conceptual solution for Access Control in the form of a security enforcement layer and architecture stack for cloud computing. The solution is expressed as a reference architecture that employs the C3A architecture methodology (Hadar and Silberam, 2008) to aggregate security needs into conceptual components, as well as to define the dependency and utilization of such abstract modules.

Section 2 of this paper details the conceptual architecture components. Section 3 discusses and concludes this study's observations and provides some future directions.

2 ACCESS CONTROL REFERENCE ARCHITECTURE FOR CLOUD COMPUTING

Figure F-1 details an architectural blueprint for Access Control and Cloud computing adaptation layer, using the C3A architecture methodology (Hadar and Silberam, 2008). The responsibilities, and thus the requirements for the solution, are contained within separate components. In the detailed description below, we present part of the components that are relevant to the protecting the physical boundaries of virtual machines problem

with brief indications on other components for contextual framework purposes.

1. Strategic Federated CMDB (Configuration Management Database), records the planned and authorized configuration of the system and tracks the updated and synchronized (with the real infrastructure) servers that are discovered and reconciled with existing information from the federated repositories.
2. Hosts discovery is linked to Federated CMDB web services and can extract a list of hosts according to criteria. Hosts On-Demand discovery sends requests to the comparable Federated CMDB. Security managers can use it to select specific IT sectors or groups.

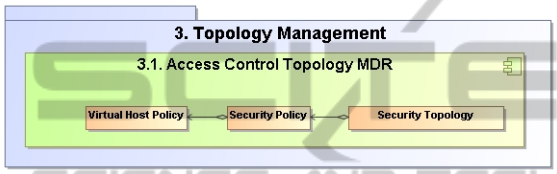

Figure F-2: Topology Management component.

3. Topology Management (Figure F-2) triggers request to detect a list of specific hosts as needed.

Portability Manager transfers deployed images across virtual domains or provisions new images on the utility computing vendors sites (Figure F-3).

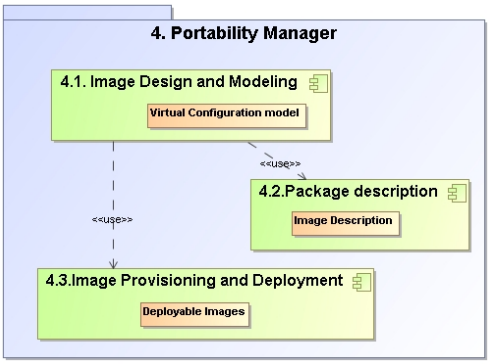

Figure F-3: Portability Manager.

3.1. Image Design and Modeling defines a virtual configuration model of packaged entities. It updates the Package Description component with added definitions on upgrades and changes.

3.2. *Package Description* generates specific vendor packaging instructions and configuration descriptions by using vendor defined description files, such as the Amazon Machine Image description file (Amazon web, 2009) used by the Amazon Elastic Cloud (EC2) service.

4. *Image Provisioning and Deployment Manager* deploys and manages the integrity of the Access Control Agents installed on a host image deployed in the cloud
5. *Cloud Security Management* - Figure F-4

5.1. Cloud Security Sentry provides a managed guarded-call pattern that intercepts calls intended for the existing published Management Services of the cloud vendor. The guarded call prohibits the direct activation of the cloud web services and prevents a non-authorized access. Even though in this context we refer to the protection of accessing servers (and hosts), intercepting web calls in a well-known pattern, and the sentry can also wrap any external call to Platform as a Service (PaaS), or Software as a Service (SaaS). Call protection that interrupts the natural transaction, inevitably slows down system performance, but security has its price, as do regulatory requirements, high-availability, redundancy, and any other non-functional needs of a solution. The sentry only intercepts administrative calls that modify the system deployment, which may be a insignificant relative to the overall performance when deploying a full image, or a major interuption if the payload is a small policy file. The same rule applies to PaaS and SaaS, where the payload size transmission and the regular call overhead, relative to the call interception. The cloud environment is web-services based, over WAN, and by definition, is considerably slower than regular on-premise, LAN solutions. In our Access Control implementation, receiving the credentials for an operation while intercepting kernel calls does not exceed 5% of the CPU consumption on the supported platforms. However, in the case of WAN and web-services, one must consider network latency on the overall calls, which depends on the network configuration. Still, only administration calls are intercepted, which are not frequent for privileged users. For regular SaaS applications, a different solution is employed, beyond the scope of this paper, known as federated security. Thus, the managed aspects of the cloud are

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

enforced, either from calls originating from within the customer's enterprise sites or from any other external location. It provides transient security entitlements to the Provisioning and Benchmarking components and enables these components to access the cloud for changing images. Moreover, it intercepts security policy distribution that is activated directly from the topology component.

5.2. *Security SLA Requestor* sets metrics and mediates requests for delegating the need to deploy Access Control policies within the cloud. It can (1) forward and monitor requests for access policy, (2) insert restrictions on moving the physical location of images, and (3) produce reports and monitoring information. In order to provide capabilities to limit and control the Managed Services of the cloud, it uses the Attributes Extractor component to extract the data from the cloud and thus provides limitations on the roaming capabilities. It uses the Agent Data Collection component for retrieving reports and monitoring data, and the Distribution component for propagating security enforcement requests to the cloud vendor.

5.3. *Wrapped Cloud Managed Services* component provides the same functions of the managed services of the cloud. However, they are wrapped with security parameters for preventing un-authorized administration calls; thus they serve as interception guards.

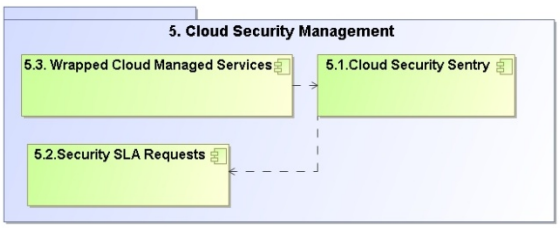

Figure F-4: Cloud Access Security Manager.

5.4. *Benchmarking Manager* Contains the *Attributes Extractor* that extracts and displays all the non-functional characteristics of the utility computing offerings for furtherer evaluation. In the case of Access Control, in order to provide limitations on the roaming capabilities, the information that is required is the virtual host name, id: MAC addresses or other strong identifiers and VM configuration (network interfaces, external disks, and its Input/Output ports).

6. *Utility Computing* (Figure F-5)

6.1. *Delivery Gateway* is deployed within the cloud vendor's site, and is able to accept and delegate the need for enforcements of policies on physical servers and virtual hosts supporting the cloud. The *Distribution* component is responsible for receiving and delegating the Access Control policy. It propagates them to internal enforcement agents that monitor the operating systems and applications that run the cloud virtualization software. The *Agent Data Collection* component is able to collect internal reporting information about the system's status, and delegate this information to the cloud consumers.

6.2. *Managed Services* is the vendor's proprietary services that provide interactions with the utility computing solutions.

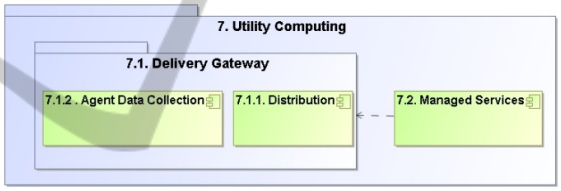

Figure F-5: Utility Computing.

7. *Cloud Users* are the regular administrators of the IT enterprise infrastructure that interact with the cloud using the supervised services. This component includes Single-Sign-On capabilities, as well as Identity and Access Management (such as LDAP).

3 DISCUSSION AND CONCLUSIONS

As presented in this conceptual paper, the collection of the abstract components encapsulates the needs and possible solution to overcome security gaps in the cloud offerings.

The reference architecture defines the components' stack for handling security interceptions (enforcement) of the managed web services; limitations on image roaming and portability; delegation of enforcement needs to the underlying cloud core, and aggregated reports. This

set of requirements, when added to the cloud offerings, enables to use the cloud as if it were deployed within a local datacenter.

Consider the example in which an enterprise implements a store front on a public cloud but keeps the backend database for the store front in their datacenter. In this simple example, the datacenter administrator is faced with several security problems. In the reference architecture these issues are addressed. The architectural definition of the solution is to deploy the front-end of the shop on a public cloud, and the sensitive data as well as the application financial processing servers on the local enterprise datacenter. High-availability and transient (on-demand) scalability requirements define a minimum of a 3 front-end server cluster, 2 application servers, and 2 replicated database servers. The maximum allowed provisioned servers are 15 for the front-end, 5 for application logic, and during known heavy usage periods (holidays), to add another single replicated database.

According to this architecture definition, the Strategic Federated CMDB will contain the planned structure of 3 front-end servers mounted to be deployed on Amazon EC2 cloud, 2 application and 2 database servers will be deployed as virtual guests on a VMWare host within the internal datacenter. All these images (VMWare guest and server, and Amazon Image) are configured according to the Package Description component, after being designed in the Image Design and Modeling component.

Since scaling up and down is part of the cloud promise (dynamic, on demand scaling), we can define transient conditions of the applications based on adaptation needs of the solution (number of concurrent connections, averaged transactions performance latency, shoppers peaks, etc.). These are inherently part of the strategy definition, containing the adaptation configuration parameters, managed by the Deployment and Change management tools, rather than CMDB. The discovery tools that will detect the added servers, can compare it to the stable minimum strategy and the approved deviations, and will indicate the difference as allowed, as long as they are within the above boundaries. It can even show the differences on the same view (planned vs. actual).

The security policy rules are:

1. Deployment administration policy:
   1.1. The provisioning of this application can only be conducted by the authorized *Image Provisioning and Deployment Manager* only.
2. Cloud roaming policy:
   2.1. The front-end servers cannot be deployed in Europe due to privacy regulation limitations.
3. Datacenter co-existence policy:
   3.1. The database servers cannot be deployed on machines that contain the company employees' salary database server.
   3.2. The application and the database must reside on separate physical servers, namely, each component of the solution must be on another hardware.
4. Sizing and scaling policy:
   4.1. The front-end provisioning is bounded between 3 and 15 servers
   4.2. The Application servers provisioning is bounded between 2 and 5 servers
   4.3. The Application servers provisioning is bounded between 2 and 3 servers according to dates attached to the policy.
5. Secured communication policy:
   5.1. The application servers can only respond to incoming requests from the front-end servers.

The policies are defined and maintained in the Virtual Host policy (3.1, 3.2, 4.2, 4.3, 5.1), and the cloud Security Distributed Policy (1.1, 2.1, 4.1) within the Topology Management component. Accordingly, the Topology component dispatches these separate policies to the Security Sentry and Image Design components.

The Cloud Security Sentry intercepts any management call to Utility computing that does not originated from the Image Provisioning and Deployment component.

The Managed Services component contains accordingly specific application security provisioning policy (1.1, 4.1) that enables the provisioning and un-provisioning of images on the public cloud, governed by the Security Sentry. Thus, change function calls will be activated only if they are within the policy boundary. The exact amount of provisioned machines is maintained in the Topology MDR, which receives the correct status of the system from the federated reconciled CMDB.

The Image Provisioning and Deployment component manages the internal datacenter as well, where each of the deployed servers has an enforcement access control agent installed. This agent receives the co-existence (3.1, 3.2), sizing (4.2, 4.3) and communication (5.1) policy, bundled with other access control policies, hardening the operating systems (out of scope for this paper).

In terms of Cloud Roaming (2.1) of the front-end parts of the application, the cloud Security Sentry component delegates the policy to the Security SLA Requests component, which in turn, sends the policy to the Distribution component. This triggers a request for the cloud provider, which follows the instructions, and report back to the adaptation layer via the Agent Data Collection.

The modular structure enables additional integrations, such as federated security management over a grid. Delegating the security requests from the services supplier is done in a "supply chain management" approach. Consequently, This delegation enables to structure a network of managed access control units, in different areas of the globe, over different datacenters.

This adaptation layer abstracts the understanding of cloud security management from the IT users, enabling them to pursue their regular working practices as well as simplifying cloud management security aspects.

The reference architecture presented here may be replaced by any implementation components that provide a technological solution to these design specifications. It is aimed at structuring an adaptation layer, extending the cloud offerings to a secured one. Aggregating the needs into the conceptual high-level components by means of architecture design defines the future context for specific implementations.

Future evolution of this conceptual layer will involve its practical integration with enterprise IT management tools, such as Access Control tools, Identity Management, and CMDB as well as with Cloud vendors such as Amazon, Google, or Microsoft.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

REFERENCES

Amazon Elastic Compute Cloud (2009), Web Site: http://aws.amazon.com/ec2, last accessed on March 5, 2010.

Hadar, E. and Silberman, G (2008): Agile architecture methodology: Long term strategy interleaved with short term tactics. In: Proceedings of the International Conference on Object Oriented Programming, Systems, Languages and Applications (OOPSLA), Nashville, US.

Vouk, M.A.(2008): Cloud computing - issues, research and implementations. In: Proceedings of the ITI 2008 30th International Conference on Information Technology Interfaces, Cavtat, Croatia.

Weiss, A (2007), Computing in the clouds. netWorker 11(4) (2007) 16-25