2012 IEEE/ACM Fifth International Conference on Utility and Cloud Computing

BlueShield: A Layer 2 Appliance for Enhanced Isolation and Security Hardening among Multi-tenant Cloud Workloads

Saurabh Barjatiya (Author)
IBM Research and IIIT Hyderabad
Email: saurabh.barjatiya@iiit.ac.in

Prasad Saripalli (Author)
IBM Cloud CoE and IBM Research
Email: prasadsar@in.ibm.com

***Abstract*—Enhanced Isolation and Security (EIS) in a cloud are of significant concern. Many organizations are hesitant in migrating to a cloud based infrastructure due to the perceived limitations with EIS. Earlier, we had presented the quantitative risk and impact assessment framework (QUIRC) [1]. QUIRC can be used to assess the security risks associated with the cloud computing platforms. In the present work, design and implementation of BlueShield is presented. BlueShield is a Layer 2 appliance for an EIS hardening among multi-tenant cloud workloads. BlueShield architecture provides EIS, significantly reducing the threats faced by the tenants in a cloud environment. EIS provided by BlueShield is validated using a proof of concept implementation. Then shortcomings of the various present approaches in addressing the identified security threats are explained. It is shown that the present security applications, deployed in a non-cloud environment, do not require modification during migration to BlueShield based clouds. Furthermore, the proposed design provides high level of protection among the VMs in the same VLAN.**

***Index Terms*—Cloud; Security; Enhanced isolation; Multi-tenant isolation; BlueShield; VM agent; Network; Auditing; Echelon**

I. Introduction

Cloud security can be judged with the help of network security risks faced by the tenants. Overall security risk can be calculated with the help of QUIRC framework. Probability and impact values are used for calculating risk based on QUIRC. Since these values are specific to an application, they help in assessing the risks faced by the application, when deployed in a cloud environment. QUIRC framework identifies six main security objectives for cloud as Confidentiality, Integrity, Availability, Multi-party trust, Mutual Auditability and Usability (CIAMAU). The present security and workload isolation architectures deployed in the clouds do not meet all of these security objectives within a single design. Therefore, a new security framework which is specifically targeted to address these security objectives is created.

In this paper, first the EIS requirements of the multi-tenant clouds are discussed. Then the existing literature in cloud security is reviewed. It is shown that the available architectures do not meet all the EIS requirements of multi-tenant workloads. Then the BlueShield architecture is proposed, and its design and implementation details are explained. Later, the results obtained in the test-bed, which validate the functionality of BlueShield implementation, are presented.

The present BlueShield design is targeted to mitigate threats related to network security. Other threats due to storage, malicious cloud provider, hypervisor bugs, etc. are not affected by BlueShield architecture. Hence, the other threats have not been evaluated in the test-bed. Future work in extending BlueShield in order to reduce other threats is possible.

II. EIS Architecture for Cloud

A. EIS Requirements from Clouds

One of the most important requirements of the cloud providers is the unrestricted VM to VM bandwidth within a cloud. VM to VM bandwidth within a cloud is limited because of over-subscription of network links near the root of a network tree. Generally, cloud networks get deployed in tree topology and most inter-leaf (inter-VM) traffic flows through the nodes near the root. Such designs are used because the security devices are deployed near the root of a network tree and it is desired that most of the inter-VM traffic flows through them. This leaves very limited VM to VM bandwidth available to be shared by all tenants. Thus, it is possible for a few tenants to use large proportion of network resources unfairly, leaving less bandwidth available for other tenants. This can also lead to performance and availability problems.

Even if a large number of security devices are deployed in transparent mode to scan traffic between base machines, VM to VM traffic within a same base machine will not get secured. This will result in security blind-spots.

Apart from the bandwidth, cloud providers require that the hypervisors used are well tested over the years so that the probability of discovering new bugs in such hypervisors is less. The providers also desire EIS solution to be flexible and open source, so that they can modify the solution without requiring authorization from hypervisor or security solution providers.

Furthermore, cloud providers should not get tied to a particular hypervisor. The security solution must allow the use of multiple hypervisors within same cloud without affecting the EIS.

Along with this, it should be possible for the tenants to test EIS features of a new application in a small test-bed. Tenants should be able to re-use the knowledge and experience gained in securing a traditional enterprise network within the cloud

978-0-7695-4862-3/12 $26.00 © 2012 IEEE
DOI 10.1109/UCC.2012.21

environment. It should be possible to re-use the applications and scripts written for the enterprise networks within a cloud environment with little or no modification.

B. Isolation of Network Groups: Layer 2 versus Layer 3 Architectures

In an enterprise network, different categories of users (employees, guests, and administrators) can be isolated from each other using techniques such as static L2 VLANs; Network Admission Control (NAC) and dynamic VLAN allocation using 802.1X; L3 VLANs with Virtual Routing and Forwarding (VRF), and Access Control Lists (ACL) based filtering etc. All these methods suffer from one or the other drawbacks. For example static VLAN assignment is easy to understand and also widely supported by devices. But it is very hard to maintain in a dynamic environment. Similarly, NAC and dynamic VLAN allocation using 802.1X are not easy to understand and configure. In the case of VRF based filtering, creation of Switched Virtual Interfaces (SVI) on all trunk ports is required, so that a VRF table can be assigned to each VLAN.

Isolation of network groups using static VLANs has a major limitation in terms of number of VLANs. All 802.1q headers have only 12-bits for the 802.1q VLAN tag. VLAN tag is required for associating a packet with a given VLAN. Thus there is scope of only $2^{12} = 4096$ VLANs in a network, and such VLAN based isolation is not feasible for clouds which provide services to very large number of tenants. This problem can be solved by having nested child VLANs inside parent VLANs. But such complex solutions are very hard to maintain. L2 VLANs also cause problems in the sharing of resources like DNS, Internet etc. which need to be accessible from all the VLANs. Considerable complexity arises in ensuring that the access to common resources like Internet, does not results into a possibility of undesired inter-VLAN communication.

It is worth noting that, about 15% of data centers cost is due to network [2]. Further, conventional approaches to network data center affect agility due to static assignments and cause fragmentation of resources. Such approaches also provide poor server to server connectivity due to use of proprietary hardware that scales up and does not scale out [2].

In order to have agility, a data center network should have location independent addressing, coupled with uniform bandwidth and latency. It should also have security and performance isolation [2].

BlueShield based network design does not require VLANs. Hence, it does not suffer from the drawbacks of L2 or L3 VLAN based isolation. BlueShield architecture provides security, performance isolation and location independent addressing. Hence, networks based on the BlueShield design, address the main problems that arise in a data center network. The BlueShield also provides security and auditing features that are flexible, scalable and feature intensive.

C. Earlier Work

VMWare vShield [3] runs on the top of x86 hardware to provide isolation between the VMs. This protection also exists for VMs co-located on the same physical host. vShield provides isolation with the help of a modified hypervisor, which ensures that all VM traffic passes through a vShield virtual security appliance.

TABLE I
COMPARING VSHIELD WITH BLUESHIELD AGAINST EIS REQUIREMENTS

EIS Requirement	vShield	BlueShield
Unrestricted VM to VM bandwidth	×	√
Absence of security blind-spots	√	√
Use of unmodified hypervisor code	×	√
Open source	×	√
Freedom for building new solutions and extending existing ones	×	√
Support for multiple hypervisors	×	√
Allows live migration of protected VM	√	√
Small test-bed support	√	√
Support for reuse of existing enterprise security applications in cloud	×	√
Support for very large number of VMs	×	√

Table I compares various features of vShield and BlueShield against the mentioned requirements.

III. BLUESHIELD ARCHITECTURE AND IMPLEMENTATION

BlueShield is a layer 2 EIS appliance. It comprises of one or more networked VMs. It achieves high level of tamper-proof isolation among multi-tenant workloads in a cloud. It only allows required communication and prevents all undesired traffic between VMs.

BlueShield design requires broadcast and multi-cast to be blocked using virtual switches (vSwitches). It replaces ARP based address resolution with directory lookup based resolution.

BlueShield has provision of deploying special VMs that monitor and regulate traffic between other VMs. These special VMs are referred as **Echelon VMs**. Echelon VMs allow network security and isolation at many different layers of a network model.

BlueShield also provides the EIS among VMs co-located on a same base machine. This makes the system highly scalable. The blocking of broadcast and multicast is possible on most vSwitches. This can be done easily on Linux based hypervisors which use software bridge. For BlueShield implementation and functionality validation, 'ebtables' firewall has been used to block all broadcast and multicast. BlueShield components within a base machine and the overall BlueShield architecture are shown in Figures 1 and 2.

If address resolution is not working and static hardware address entries are not added, then machines cannot communicate with each other. This fact is exploited to provide complete isolation in BlueShield design.

BlueShield uses directory lookup to provide full isolation and Echelon VM based protection to provide enhanced isolation.

BlueShield design requires that a software agent, referred as **BlueShield Agent**, is installed in all the VMs. BlueShield agent is required for inter-VM communication to work. Blue-

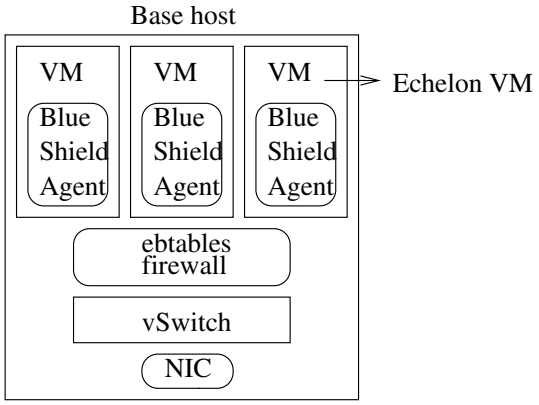

Fig. 1. BlueShield components within a base machine

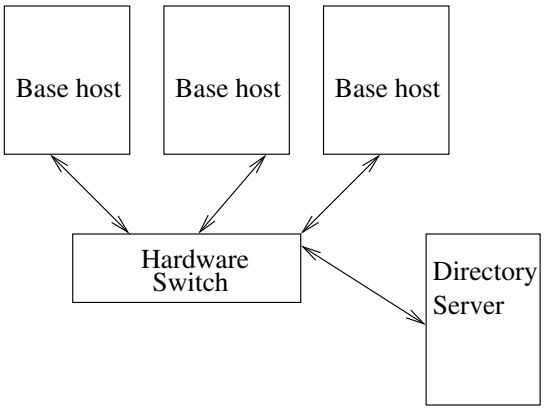

Fig. 2. Overall BlueShield architecture

Shield agent converts an ARP request sent by a VM to a directory look-up request.

BlueShield agent captures an outgoing ARP request from the VM, before it is dropped by a vSwitch. It then converts the request from broadcast to an unicast query, and sends it to directory-lookup servers. This requires BlueShield agent to be configured with the location (MAC address) of directory-lookup servers. BlueShield agent sends query to more than one directory-lookup servers for redundancy. Directory servers reply to a query with a target MAC address only if it is permitted by the policy.

Let us consider a case where it is required for two VMs, say $H1$ and $H2$, to be completely isolated from each other. In this case the directory server is configured such that it does not reply when $H1$ or $H2$ send a query for each others MAC address. **This ensures that $H1$ and $H2$ cannot resolve each others MAC address.** This results into complete isolation between $H1$ and $H2$. The same concept is scaled to provide complete isolation among VM Groups.

This allows creation of several isolated VM groups. VM belonging to a particular group communicates only with other VMs in the same group. This provides an illusion that VMs of the same VM Group are in one separate VLAN.

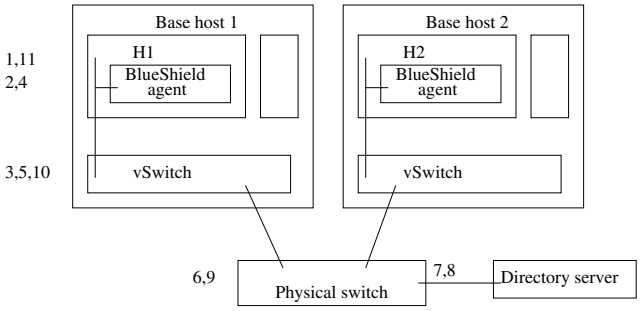

Fig. 3. Steps in the address resolution process using directory server

The communication among VMs belonging to different VM groups is completely blocked, unless an exception for such a communication is defined in the directory server policy.

The overall query and resolution process is described below with the aid of Figure 3.

Consider Figure 3 to understand various steps involved in the address resolution process:

1) $H1$ generates an ARP query for resolving $H2$'s MAC address. This ARP query is a broadcast packet with the destination MAC address as FF:FF:FF:FF:FF:FF.
2) BlueShield agent captures the ARP packet before it leaves the VM.
3) vSwitch receives this packet and then drops it because it is a broadcast packet.
4) BlueShield agent generates a unicast query by replacing the destination MAC address on the packet from broadcast address (FF:FF:FF:FF:FF:FF) to one of the directory servers unicast MAC address.
5) vSwitch receives this packet and forwards it if it satisfies following properties:
   a) Packet is not multicast packet
   b) Packet is sent from $H1$'s own MAC address
   c) If packet contains an ARP query, then destination MAC address is in the directory server MAC address range and target MAC address in ARP data section is 00:00:00:00:00:00.
6) The physical switch forwards the packet based on the destination MAC address of the packet to one of the directory servers.
7) The directory server receives the query packet.
8) The directory server checks the policy, and if required sends an appropriate response.
9) The physical switch forwards the packet to the base machine 1 based on destination MAC address of the packet.
10) vSwitch forwards the packet to the appropriate VM, if it satisfies following properties:
    a) The packet is neither a multicast and nor a broadcast packet
    b) If the packet is an ARP response packet, then source MAC address should be in the directory server MAC address range.

11) The VM OS receives the packet and learns the MAC address for $H2$.

Similar steps are used by $H2$ to resolve $H1$'s MAC address so that both $H1$ and $H2$ can communicate with each other.

A. EIS using Echelon VMs

To meet the EIS requirements of a cloud, Echelon VMs are proposed as part of BlueShield architecture. **Echelon VM is a VM with hardened operating system and EIS software similar to the firewall and IPS.** On Echelon VM EIS software is configured to provide the desired protection between the VMs. Echelon VMs do not depend on any particular flavor of Operating System. They can use any secure operating system which provides a proper support for network security software.

Directory servers ensure that traffic between protected VMs passes through an Echelon VM. **They achieve this by replying with the Echelon VMs MAC address, when protected VMs send an address resolution query for each other.** The Echelon VM can then filter the received traffic before forwarding it to the destination VM. This filtration can be directed with the help of an EIS policy.

An Echelon VM based protection does not suffer from security blind spots. An Echelon VM can scan all traffic going to or coming from a protected VM. This protection works even if the communicating VMs are co-located on the same base machine.

The proposed Echelon VMs are hypervisor independent. They can be deployed on top of any hypervisor without requiring modification in the hypervisor code. Both protected VM and Echelon VM can be migrated from one base machine to another without affecting the ongoing connections.

IV. Implementation Results

BlueShield can be implemented in a small test-bed to verify its working and usefulness. A small test-bed was created to verify the various features provided by BlueShield.

Details of BlueShield architecture and test results are mentioned in longer version of paper. Longer version would be hosted at http://www.sbarjatiya.com/ after conference.

V. Future Work

Basic proof-of-concept presented here can be extended with additional features to further improve the BlueShield design. Interesting future work possible is summarized in this section.

An Echelon VM can be made highly available. Multiple Echelon VMs can be configured to share firewall connection state.

An Echelon VM can help in SSL offloading by receiving a SSL client connection, and forwarding its plain-text to a protected VM. Conversely, an echelon VM can receive plain-text communication from a protected VM and forward Its encrypted counterpart to a remote machine. An Echelon VM can also pass received traffic through an anti-virus or an anti-spam engine to protect from virus and spam.

An Echelon VM can be used for load balancing by distributing the incoming connections among a specified set of VMs. An Echelon VM can also mirror the incoming network data to multiple destinations for redundancy. Packets related to an attack or port scan can be captured on an Echelon VM for further analysis. This capture can also serve as evidence of an attack.

An Echelon VM can be used for mutual multi-party trust by giving its control to a trusted third party. This trusted third party can monitor logs on the Echelon VM to verify that the communication from a protected VM is based upon agreed protocols [5]. Logs captured on an echelon VM can be used for mutual auditability. To ensure that an Echelon VM is not tampered, in multi-party cases, access to the Echelon VM can be restricted through a specific application or API which is trusted by all parties.

Selective multicast and broadcast can be enabled by converting them to unicast. List of machines which should receive a particular multicast or broadcast can be maintained by directory servers. This would ensure EIS while providing broadcast and multicast functionality. In this manner DHCPv4 can be made to work with the help of BlueShield agent, which can convert broadcast packets to special unicast packets.

VI. Conclusion

A flexible and scalable EIS framework called BlueShield has been developed. The key contribution of the presented work is multi-tenant EIS with auditing. The key advantage of an L2 implementation is its ability to scale. This provides uniform server to server bandwidth with performance isolation. BlueShield features can be tested and verified within an enterprise boundary with minimal resources. BlueShield design is hypervisor independent and does not suffer from blind-spots. Echelon VMs allows possibility of tightly controlled EIS at various layers of network protocol stack.

References

[1] Prasad Saripalli, Ben Walters. QUIRC: A Quantitative Impact and Risk Assessment Framework for Cloud Security. 2010 IEEE 3rd International Conference on Cloud Computing (CLOUD)

[2] Albert Greenberg, James Hamilton, David A. Maltz, and Parveen Patel. The cost of a cloud: research problems in data center networks. SIGCOMM Comput. Commun. Rev. January 2009

[3] Debashis Basak, Rohit Toshniwal, Serge Maskalik, and Allwyn Sequeira. Virtualizing networking and security in the cloud. SIGOPS Oper. Syst. Rev. December 2010

[4] Hanqian Wu; Yi Ding; Winer, C.; Li Yao. Network security for virtual machine in cloud computing 2010 5th International Conference on Computer Sciences and Convergence Information Technology (ICCIT)

[5] Bleikertz, Matthias Schunter, Christian W. Probst, Dimitrios Pendarakis, and Konrad Eriksson. Security audits of multi-tier virtual infrastructures in public infrastructure clouds. Proceedings of the 2010 ACM workshop on Cloud computing security workshop