Gartner Research

Publication Date: 24 May 2007 ID Number: G00147785

Server Virtualization Can Break DMZ Security

Neil MacDonald, Greg Young

This Research Note explores the specific risks of the use of virtualization in an organization's demilitarized zone (DMZ) and recommendations to mitigate the risks. Information security professionals should use this information in a discussion of the risks of DMZ virtualization before the decision to virtualize is made and use the cost savings of virtualization to offset the costs of the security tools and process changes needed to ensure overall DMZ security levels are not reduced.

Key Findings

- Virtualization and consolidation of similar trust levels in a DMZ introduces manageable risk for most organizations today.
- A full collapse of the DMZ, including workloads of different trust levels and externally based network security enforcement points (for example, firewalls), introduces a level of risk that most organizations will find unacceptable.
- Some DMZ virtualization risks can be removed with mitigating controls; other risks can only be reduced.
- With extreme consolidation, the cost of maintaining the needed level of security may exceed the savings.

Recommendations

- Just because an organization can consolidate servers and network security devices in a DMZ using virtualization doesn't mean it should.
- The decision of how much to virtualize in the DMZ must be made with a full understanding of the additional risks that are incurred.
- The cost of implementing mitigating controls must be factored into the return on investment (ROI) decision.
- Ultimately, the decision to implement mitigating controls or live with increased risks of DMZ virtualization must be made by the virtualization decision owner.

STRATEGIC PLANNING ASSUMPTIONS

By year-end 2010, 70% of large organizations will collapse all or part of their DMZ by using risk-managed virtualization techniques, up from 10% at year-end 2006 (0.8 probability).

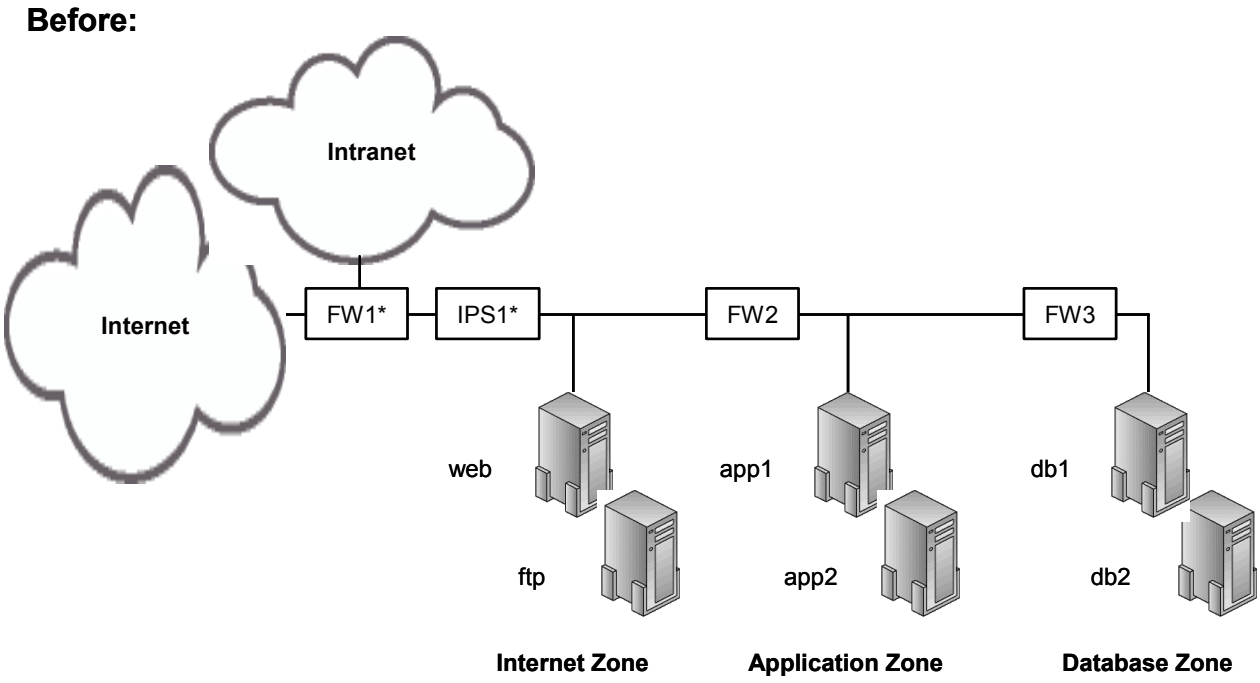
However, fewer than 25% of large organizations will collapse the entire DMZ (0.7 probability).

DMZ consolidation using virtualization will be a "hot spot" for auditors, given the greater risk of misconfiguration and lower visibility of DMZ policy violation. Through year-end 2011, auditors will challenge virtualized deployments in the DMZ more than nonvirtualized DMZ solutions (0.9 probability).

ANALYSIS

Most server consolidations are driven by operations and the desire to improve efficiencies and reduce costs without necessarily considering the security implications (see "Security Considerations and Best Practices for Securing Virtual Machines"). In security-critical areas of the network, such as a typical DMZ (see Figure 1), or an organization's data center, workloads of similar trust levels may be consolidated today (see Figure 2) if the recommendations of our previous research are implemented. However, some consolidation scenarios will result in the consolidation of workloads of different trust levels (see Figure 3), and in the highest-risk case, trusted and untrusted workloads may end up being combined on the same server. Furthermore, with more security vendors supporting virtualization of their security functionality, such as firewalls and intrusion prevention systems (IPSs) (see "The Need for Virtual Security Partitions Is Growing"), a full collapse of the DMZ (a "DMZ in a box") that consolidates all workloads and security protection is now possible (see Figure 4). However, the real question organizations should ask is not, "Can virtualization be used in the DMZ?" — which it can — but instead ask, "Should this be done?" The answer to the latter requires a careful consideration of the unique risks that such a configuration might introduce. In this Research Note, we discuss four risks of server consolidation that uses virtualization, which results in the mixing workloads of different trust levels, and ways in which these risks might be managed. Overall, the mixing of trusted and untrusted workloads in the DMZ introduces a level of risk that most IT security organizations will not find acceptable.

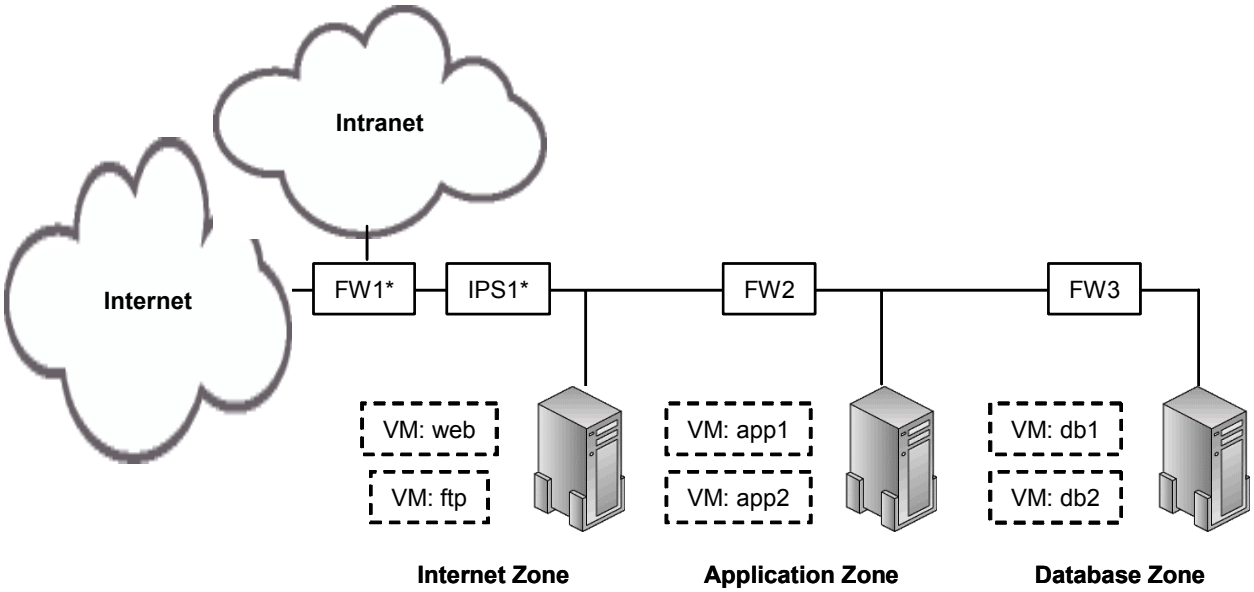
Figure 1. Typical DMZ



app1 = application 1, and so forth
db1 = database 1, and so forth
FW = firewall
*FW1 and IPS1 may be a single next-generation firewall device.

Source: Gartner (May 2007)

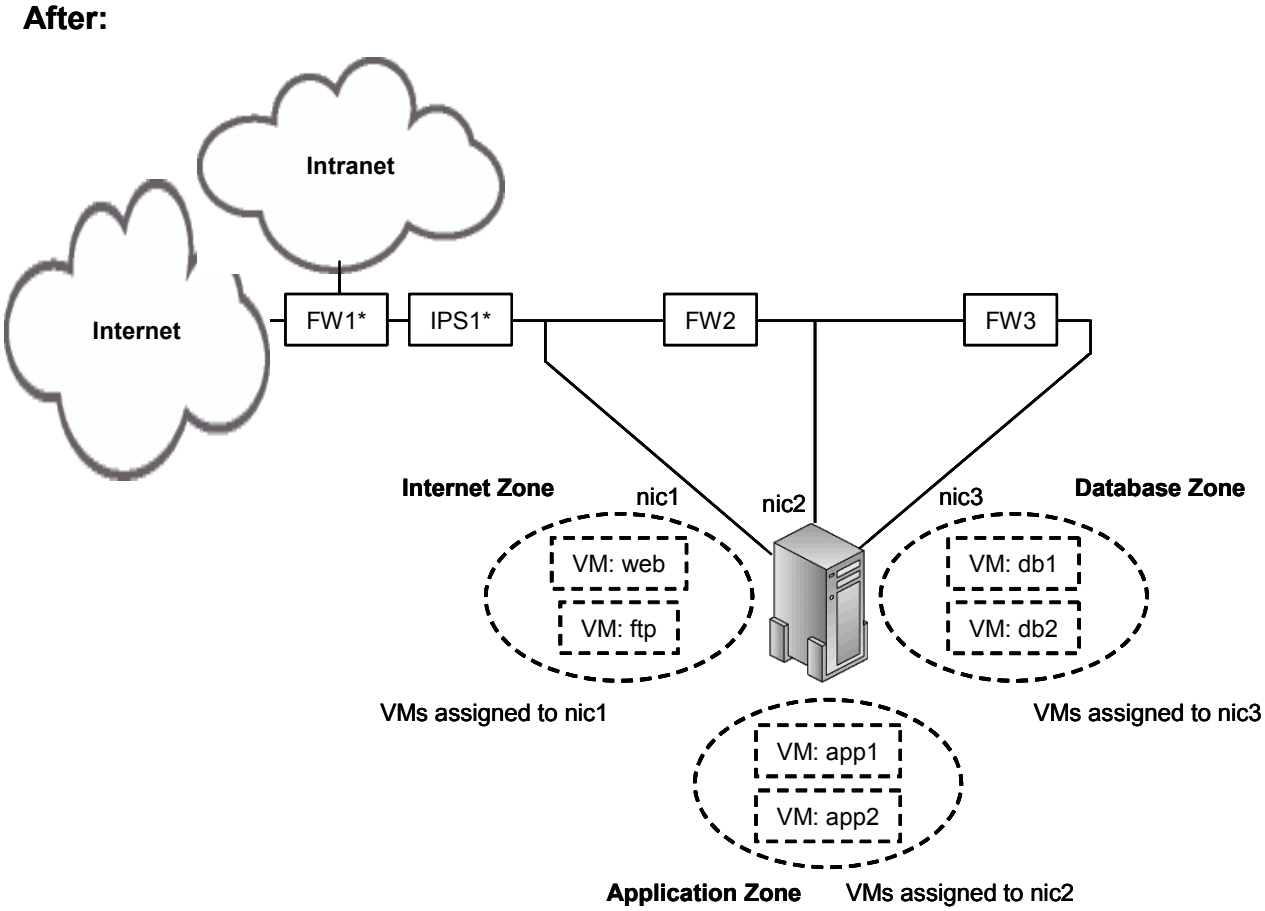
Figure 2. Partially Collapsed DMZ Without Mixed Trust Zones



VM = virtual machine
*FW1 and IPS1 may be a single next-generation firewall device.

Source: Gartner (May 2007)

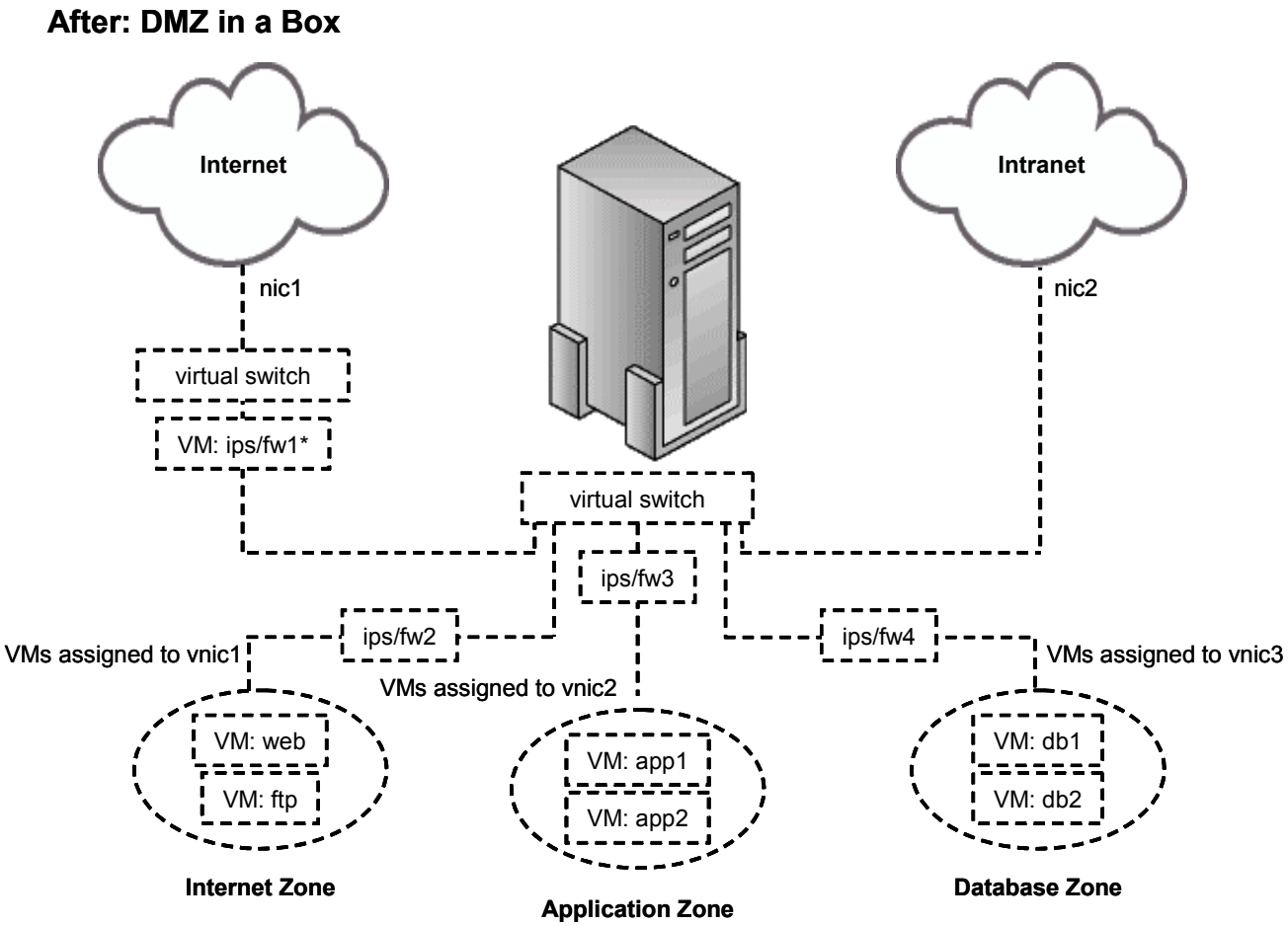
Figure 3. Partially Collapsed DMZ With Mixed Trust Zones



nic1 = network interface card 1 (physical), and so forth
*FW1 and IPS1 may be a single next-generation firewall device.

Source: Gartner (May 2007)

Figure 4. Fully Collapsed DMZ With Mixed Trust Zones



vnic1 = virtual network interface card 1, and so forth
*ips/fw1 may be a single next-generation firewall device.

Source: Gartner (May 2007)

Risk: Loss of separation of duties between security/network security and operations

The issue: When external network-based security systems are used to enforce separation between DMZ trust zones (also referred to as DMZ tiers), a level of separation of duties (SOD)-by-default is achieved, because the functions are hosted on separate physical systems that are managed and configured by separate teams. In a traditional DMZ architecture, it is difficult, if not impossible, for a server manager or application manager to alter or change security settings. A separate security control plane forms the foundation for defense-in-depth and enables rapid response in the event of an outbreak. This SOD-by-default is lost in the move to a collapsed DMZ. Within the virtual server, the server administrator who has access to the "root" ID of the privileged software levels and configuration can now alter or disable security settings, creating the potential for conflict of interest.

Organizations could take several steps to mitigate, but not eliminate, this risk:

- Use virtualization for the consolidation only of similar trust levels of workloads (as shown in Figure 2), so that SOD of security configuration and controls is minimally impacted.

- If workloads of different trust levels are consolidated, all virtual server traffic could be internally grouped by network in each trust zone (see Note 1), bound to dedicated physical (NICs) sent for external inspection and security policy enforcement, and then returned to the virtual server to the appropriate destination zone (see Figure 3 and Note 2).
- Require strong authentication for privileged account access, with full auditing and logging.
- A check-in/check-out process for administrative IDs, including privileged virtual server "root" administrators, with full auditing and logging of activities performed — This doesn't prevent the alteration or disablement of security software, but the detailed auditing and logging help to discourage casual unauthorized tampering.
- Depending on the architecture, it may be possible to control the ability to change security settings in the virtual server. However, there is a risk that the "root" administrator would still retain the ability to disable the software entirely, leaving the risk that the layer or VM containing security software could be disabled (thus, fail closed if tampering is a concern).
- In the event of an outbreak, security professionals must retain the ability to quickly change settings and policies, potentially affecting the operation of critical production workloads. Tools for understanding application and service dependencies are immature, compounding the problem.

Risk: Privileged software layers, such as hypervisors, virtual machine monitors, host operating systems, parent partitions and drivers, will contain embedded vulnerabilities that may be exploited to breach zone isolation

The issue: All software will contain embedded security vulnerabilities that, when compromised, could lead to a breakdown in isolation. This problem is exacerbated with newer-and-not-yet-battle-tested hypervisor technologies that are granted the highest system privilege levels. A compromised hypervisor is a worst-case security scenario, placing all workloads at risk (see "Secure Hypervisor Hype: Myths, Realities and Recommendations" and "Building Blocks for Trusted, Secure Hypervisors"), and is undetectable to any security software running above it.

Organizations could take several steps to mitigate, but not eliminate, this risk:

- Bigger is not better from a security perspective; thus:
  - Favor layered hypervisor-based architectures in which privileged software layers are as thin as possible (see "A Blueprint for Hypervisor Implementation")
  - Use virtualization-enabled processors and virtualization software that supports the hardware-based capabilities to reduce size and complexity of privileged virtualization code.
- Require the hypervisor vendor to subject the code to open source or external third-party review for security vulnerabilities.
- Favor virtualization technologies that have been shown to be secure in real-world deployments over time. Newer is not better from a security perspective.
- Extend the organization's vulnerability and patch management processes to encompass the privileged layers of software in a virtual server (see "Understanding Vulnerability Management Life Cycle Functions").

- Use an externally based in-line network IPS that can shield the privileged layers of virtualization software from network-based attacks.
- Prohibit and disable the loading of arbitrary code in privileged partitions,
- Require the privileged software layers in the virtualization vendor's offering to provide the ability to detect and ideally prevent corruption, including:
    - Placing the hypervisor in nonvolatile storage (for example, basic input/output system [BIOS])
    - Measuring the hypervisor and other trusted boot elements during the boot process to detect tampering
    - Providing built-in tripwire-like functionality to detect tampering
    - Supporting the native hardware-based security protection capabilities (for example, the No Execute [NX] and Execute Disable [XD] capabilities of AMD and Intel processors, respectively) of the underlying processor
- Require the vendor to provide secure update mechanisms for privileged layers.
- Develop strict internal processes and controls for configuration changes of these layers, including device drivers.
- At a minimum, disable promiscuous mode on NICs and disable microprocessor hyperthreading for guest VMs that enable potential eavesdropping attacks for trusted information, such as encryption keys. In security-critical deployments, disable these capabilities for the entire virtual server.

Risk: Security isolation between different levels of trust depends on absolutely correct configuration of the internal virtual network, including any virtual LAN (VLAN) settings, NIC bindings and information flows. Incorrect configuration could result in a compromise of zone isolation

The issue: Most security vulnerabilities are the result of misconfiguration and mismanagement, and the virtualized world of collapsed servers is no exception. Incorrect configuration could result in a compromise of zone isolation. Incomplete and immature configuration tools, lack of industry standards for secure configuration, and lack of staff expertise with the security of virtual servers exacerbate the issue.

Organizations could take several steps to mitigate, but not eliminate, this risk:

- Purchase third-party tools to assess the correct configuration of the virtual server, including all privileged layers.
- Use third-party standards (for example, Center for Internet Security [CIS] and National Institute of Standards and Technology [NIST]) to baseline the configuration as the standards become available. Configuration standards are available in the physical world, but similar standards for the virtualized world are not yet available.
- Detect, log and ideally prevent unauthorized configuration changes. Alarm (or prevent) authorized changes that violate policy.
- To reduce the number of security configurations within the virtual server, the virtual server could be configured so that all traffic is sent for external inspection and policy enforcement, and then returned to the virtual server to the appropriate destination zone;

however, this is inefficient, and the success of this control still depends on absolutely correct configuration of the internal virtual network (see Note 2), network bindings and traffic flows.

- Pressure your established security and management tool vendors to provide solutions for virtualized environments so that third-party solutions are not required. Ideally, the same vendors and tools would work seamlessly across the physical and virtual environments.
- Because inter-VM traffic flows will be completely invisible to externally based network enforcement devices, virtual-server-based in-line IPS should be used to ensure no unexpected flows occur.
- Increased diligence in configuration management and change management processes provide additional oversight.

Risk: Virtualization technologies for sharing hardware among consolidated workloads increases the impact of a DoS attack

The issue: A compromised VM may inflict a resource denial of service (DoS) on other VMs on the same physical server. Certainly, external-facing workloads will be targeted for DoS attacks, but all partitions are at risk of DoS attacks launched from other partitions if configured incorrectly. Similar DoS attacks are possible on virtual server architectures in which a compromise of a "parent" partition places all of its "child" partition workloads at risk. Finally, the patching of privileged layers of software will impact all workloads, effectively taking down the entire DMZ at once.

Organizations could take several steps to mitigate, but not eliminate, this risk:

- Ensure that the virtual server is configured absolutely correct to prevent DoS:
  - Require dedicated/separate NICs for VM management
  - Require dedicated NICs for each trust zone
  - Implement processor quotas per VM
  - Implement disk space quotas
  - Protect system disk partitions from oversized logs and queues
  - Understand that compromise of a privileged software level will be harder to protect against DoS, because it is assumed to be privileged
- Plan for "hot" standby and transfer of DMZ workloads when patching is required.
- Implement behavioral profiling and monitoring of VMs (for example, Netuitive for VMware to detect potentially compromised partitions that are affecting other partitions on the virtual server).
- Avoid virtualization architectures in the DMZ that require a "parent" partition to host "child" security workloads. The parent partition becomes a single point of failure and target for attack.

What Does the Future Hold?

Many security limitations of virtualization will be addressed over time — as hardware platforms natively support virtualization; vendors improve their technologies; virtualization technologies, such as hypervisors, are battle-tested; industry best practices and secure configurations are established; organizations improve their processes; and operations and security professionals become more skilled in the configuration and maintenance of secure virtual server platforms.

The risks discussed in this Research Note are real and don't go away, but the ability to mitigate the risks will continue to improve to the point at which most organizations will find the mitigated risks acceptable during the next four years. By year-end 2010, 70% of large organizations will collapse all or part of their DMZ by using risk-managed virtualization techniques, up from 10% at year-end 2006 (0.8 probability). However, fewer than 25% of large organizations will collapse the entire DMZ (0.7 probability). DMZ consolidation using virtualization will be a "hot spot" for auditors, given the greater risk of misconfiguration and lower visibility of DMZ policy violation. Through year-end 2011, auditors will challenge virtualized deployments in the DMZ more than nonvirtualized DMZ solutions (0.9 probability).

Recommendations

Every organization must determine its own tolerance for risk and assess the risk/benefit of a full or partial DMZ collapse using virtualization. Risk acceptance for the security of the DMZ is an information security decision, not operations. Thus, while consolidating servers saves money, this consolidation must be weighed against the cost of implementing the mitigating controls discussed in this Research Note. If technologies and controls are not purchased (or, in some cases, not available), a decision to accept the increased risk must be made. For most organizations, virtualization should not be used to consolidate trusted and untrusted workloads in the DMZ. Other recommendations are as follows:

- Consolidation efforts using virtualization should be focused on workloads of similar trust levels in the DMZ, where externalized security policy enforcement mechanisms should be used to separate different trust levels in the DMZ (as shown in Figure 2). However, don't mix dissimilar asset sensitivity within the same VM. For example, data servers hosting high-sensitivity customer information are not a good match for sharing a VM with data servers for partner sales data (even though both are in the same database DMZ layer/tier).
- For enterprises that do mix trusted and untrusted workloads in the DMZ, use the mitigating controls in this Research Note as a guideline to reduce risks to an acceptable level.
- Virtualization can be used to consolidate workloads of different trust levels within an enterprise network if none of the workloads are untrusted.
- Modify your information security processes, and pressure your standard security vendors to understand and manage virtual environments seamlessly with physical environments using consistent policy.

Note 1
Virtual Networks and VLANs

VLANs alone are not sufficient to separate workloads of different trust levels (see "Findings From the 'Client Inquiry': VLAN Separation Is Not Security Separation"). Also, the term "virtual LAN" is confusing in VM scenarios. The confusion comes over the Institute of Electrical and Electronics

Engineers (IEEE) naming convention with its use of the term "VLAN." In IEEE terms, a VLAN is a private LAN or lower-layer LAN (a Layer 2 LAN) and not a virtual LAN. It is a LAN running at Layer 2. A virtual LAN is the LAN that runs in a virtual environment, such as within VMware's ESX server. The entire LAN exists inside computer memory without the involvement of any physical infrastructure, such as wires, switches and network adapters. To avoid this confusion, in this Research Note, we used the term "virtual network" to describe the internal network inside a virtual environment. However, most virtualization vendors' virtual switches inside their virtual network also support IEEE VLAN configuration, which potentially creates more confusion if the terms are used interchangeably without clarification.

Note 2
Routing Internal Virtual Server Traffic for External **Inspection and Policy Enforcement**

This routing increases the number of NICs required and the complexity of the configuration, and virtualized servers are not well-suited for the heavy input/output streams that this would likely require. Furthermore, this mitigation scenario also depends on absolutely correct configuration of the internal virtual network, network bindings and traffic flows. Dependency on precisely correct configurations, especially in complex configurations where tools are lacking, is in of itself a risk and is discussed in more detail in this document.

REGIONAL HEADQUARTERS

Corporate Headquarters
56 Top Gallant Road
Stamford, CT 06902-7700
U.S.A.
+1 203 964 0096

European Headquarters
Tamesis
The Glanty
Egham
Surrey, TW20 9AW
UNITED KINGDOM
+44 1784 431611

Asia/Pacific Headquarters
Gartner Australasia Pty. Ltd.
Level 9, 141 Walker Street
North Sydney
New South Wales 2060
AUSTRALIA
+61 2 9459 4600

Japan Headquarters
Gartner Japan Ltd.
Aobadai Hills, 6F
7-7, Aobadai, 4-chome
Meguro-ku, Tokyo 153-0042
JAPAN
+81 3 3481 3670

Latin America Headquarters
Gartner do Brazil
Av. das Nações Unidas, 12551
9° andar—World Trade Center
04578-903—São Paulo SP
BRAZIL
+55 11 3443 1509

Gartner