NETRONOME

WHITE PAPER

Accelerating Micro-segmentation

THE INITIAL CHALLENGE WAS THAT TRADITIONAL SECURITY INFRASTRUCTURES WERE CONCERNED WITH SECURING THE NETWORK BORDER, OR EDGE, WITHOUT BUILDING IN EFFECTIVE SECURITY GATES ON THE INSIDE OF THE NETWORK.

CONTENTS

BACKGROUND AND INTRODUCTION TO MICRO-SEGMENTATION

The concept of Microsegmentation was based on a security method called segmentation, initially introduced in a Forrester research paper that described "Zero-trust Network Architecture". The initial challenge was that traditional security infrastructures were concerned with securing the network border, or edge, without building in effective security gates on the inside of the network. This method of infrastructure suggests that traffic inside of the network is assumed to be trusted, and all threats exist on the outside of the network and emphasis should be on securing the network border. Over time, this method of security has been proven to be ineffective as more threats are initiated and growing from inside of the network and can potentially have unlimited access to all of the network due to lack of network security inside of the infrastructure.

This is where the security measure referred to as segmentation has become relevant in solving the inner-network security challenge. The Forrester group suggested a network architecture where all traffic sourced from and destined to endpoints contained within the same network would traverse a device called a "segmentation gateway". This device (or cluster of devices) would be centrally located in the network, and be responsible for applying security policy for isolation, access control, and threat detection.

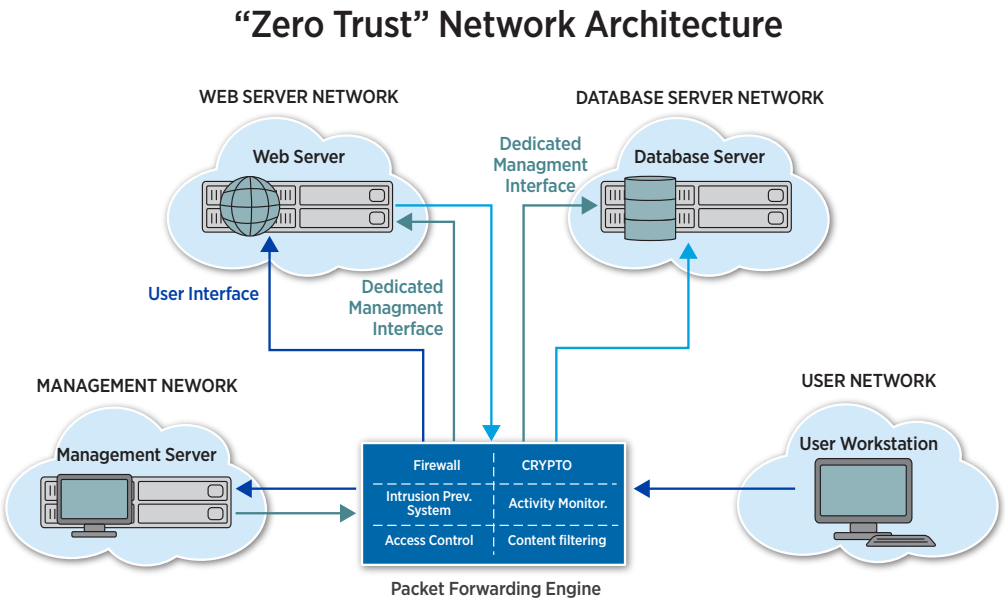

Figure 1. Segmentation Gateway Location in Network

THE TREND WITH LARGE- AND HYPER-SCALE DATA CENTERS IS TO SIMPLIFY THE PHYSICAL NETWORK AND WIDELY DISTRIBUTE NETWORKING AND SECURITY TASKS AMONGST THE THOUSANDS OF ENDPOINTS THAT EXIST WITHIN THE NETWORK. THIS STRATEGY ENABLES GROWTH THROUGH SCALE-OUT, INCREASES FLEXIBILITY,

The network would need to be architected in such a way to support transporting all traffic through a centrally located segmentation gateway. For networks of all sizes, applying security to every packet is in general a solid security strategy and good practice. From a networking perspective, however, this can create other problems. For small- to medium-sized enterprise networks, and even some percentage of large enterprise networks, forcing all traffic inside of the network to a centralized location will not hinder performance assuming the performance of the segmentation gateway is matched the size of the network. However for most medium-, large- and hyper-scale date centers, forcing all traffic inside of the network to a centralized location severely limit performance for the following reasons:

- This practice extends network paths and conflicts with modern goals of maximizing band-width between all endpoints
- It is unreasonable to assume you can get a cost-effective, centralized solution that can handle the performance/scale required of modern data centers
- Management is of such a device (or cluster of devices) is unrealistic at the scale of these networks

The trend with large- and hyper-scale data centers is to simplify the physical network and widely distribute networking and security tasks amongst the thousands of endpoints that exist within the network. This strategy enables growth through scale-out, increases flexibility, and makes the network environment more agile. The process of taking the functions of a seg-mentation gateway, virtualizing them through software, and distributing them widely across the entire server infrastructure is the core definition of micro-segmentation. Rather than forcing all traffic through a centralized location in the network, Micro-segmentation allows security to be stitched into the virtual fabric of the network by attaching security functions to every VM instance inside of the network without losing the ability to tune policies to a specific VM, group of VMs, or network.

MICRO-SEGMENTATION RECIPE

Micro-segmentation has various varying degrees of sophistication in security enforcement. In general there are four underlying security measures that are used within micro-segmentation, and the combination of services are dependent on the security requirements of each unique environment.

Network Isolation

Network isolation ensures that traffic belonging to different tenants does not intersect, interact, or overlap with one another. This is achieved through network tunneling using VXLAN, NVGRE, or GENEVE protocols for overlay networks. The overlays allow the underlay network to be abstracted away from the virtual network and also allow tenants and their IP addressing scheme to be virtually separated by using a virtual network identifier (or equivalent) to uniquely identify a tenant's traffic and differentiate its traffic from other tenants. A host-based vSwitch is often used to instantiate these overlays on a host server and handle the entunnel/detunnel in the vSwitch data path.

Access Control

The definition and objectives of network access control is unchanged for virtual machines and networks. With network access control, policies dictate how different resources (VMs, services, networks) can be accessed within a network environment. This is achieved by applying L2-L4 filters, access control lists (ACLs), or stateless firewall rules for each VM in a virtual network. In the case of micro-segmentation this function is executed on the host server. The host-based vSwitch or built-in firewall (IPtables) is the typical component used to enforce access control. The management and configuration of policies is typically handled with a higher-level, centrally located controller or orchestration system.

Stateful Firewalling

The next logical progression in security from ACLs is the ability to apply stateful filtering on a per-connection basis. Stateful firewalling considers the state of a flow in combination with matching on L2-L4 header fields to determine if a packet should be permitted or denied. Stateful inspection requires every connection passing through the firewall to be tracked so the state information can be used for policy enforcement. In IPtables, for example, a connection-tracking table is used to report flows as NEW, ESTABLISHED, RELATED, or INVALID to the firewall. The policy applied to the system then dictates how packets are evaluated within the stateful firewall. Today, this task is typically executed by the IPtables subsystem built into Linux, however some vSwitches are beginning to incorporate this functionality.

Advanced Security Services

Lastly, some administrators may require security spanning the entire stack (up to L7) and therefore require application level inspection. These services may include intrusion detection (IDS), intrusion prevention (IPS), anti-virus (AV), data-loss prevention (DLP), malware detection, application firewall, user identification, and other services. In a virtual environment utilizing micro-segmentation, these services are deployed within VMs and traffic is plumbed into and out of the service using a host-based vSwitch. Like the previously mentioned security functions, these L7 services are distributed across the network and deployed on the same servers hosting the VMs. By distributing to servers, per-tenant policies can be applied to each

endpoint virtual machine and the service is attached VM, allowing inspection to occur as soon as the VM packets enter the network.

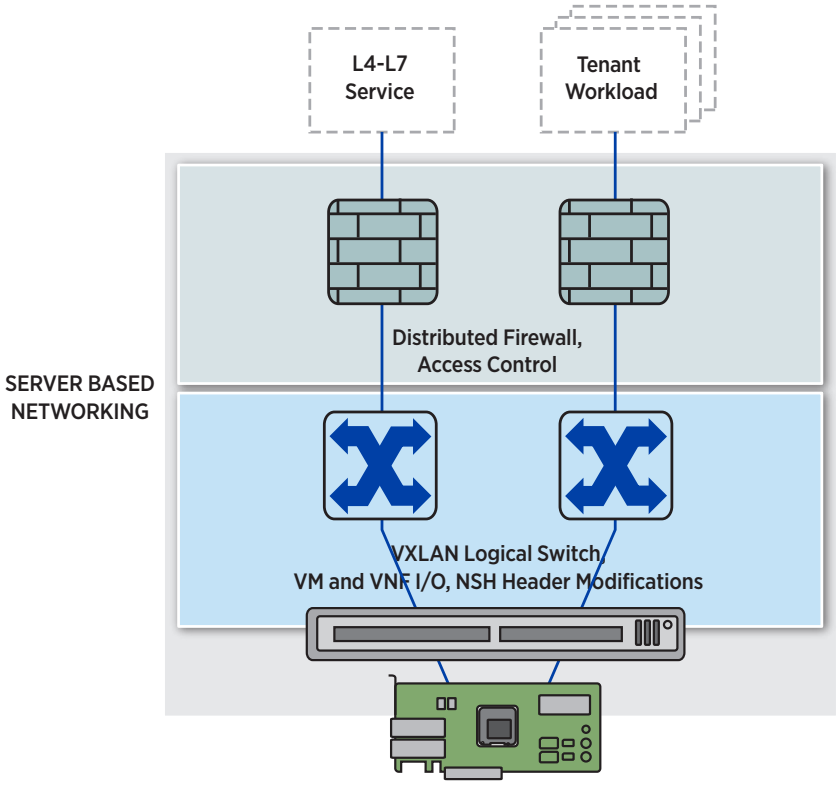

Figure 2. Breakdown of Micro-segmentation Components

While an administrator may choose any combination of these security techniques to secure one's network, it is important to note that the value of micro-segmentation is that any security policy can be effectively enforced in a large network environment by widely distributing the functions across virtualized servers. Micro-segmentation allows each of these functions to establish an attach point in the server hosting the VM, allowing security to be pushed to the VMs entry point into the virtual network.

PERFORMANCE CONSIDERATIONS

In modern data centers, the goal is to achieve scale by leveraging the computing power of thousands to tens and hundreds of thousands of servers by breaking data center tasks up into smaller, more manageable workloads and executing them on the server. This means that a data center's service capacity will grow at the same pace (linearly) as the server footprint. In general this is a cost effective way to construct a network rich in services, however it also relies on a minimum set of resources available on the server to effectively host the target application, service, or virtual machine. These resources include networking connectivity and policy, CPU cores and frequency, and memory availability. These resources are critical to the operation of a data center as they are the execution platform for the VMs, which drive in revenue. Introducing micro-segmentation threatens this operation, however, as micro-segmentation by its basic definition suggests that security functions should widely distributed across the network and executed on servers. Deploying security on servers creates a scenario

where these services are competing for the same server resources (network, CPU, memory) and is counter-productive to the revenue goals of the data center. An additional side effect of micro-segmentation is that executing these security functions on an x86 CPU results in low performance and can create unforeseen bottlenecks in the system.

Consider some of the characteristics of micro-segmentation and x86 compute and the performance challenges become clearer:

- **Network Isolation through Tunneling:** Network tunneling is an intense process requiring lookups on several headers (inner and outer headers) as well as inserting new headers in the case of encapsulation. Assuming a single x86 core can execute the lookups and actions at an average cycle count (a few hundred cycles), network tunneling becomes a bottleneck at high network speeds (multi 10G, 40G) due to lack of compute parallelism. On higher capacity networks, the packet arrival time exceeds the packet processing time and x86 CPUs cannot support the core or thread capacity to parallelize these network tasks to keep up with arrival rate.
- **Access Control with 10s of Thousands of Rules:** Using a host-based vSwitch or filtering table for access control exposes a major flaw in using x86 for networking the CPU cache. the x86 architecture relies heavily on various levels of CPU cache to achieve performance. The rule capacity requires of ACLs for 10s to 100s of VMs per server deployed forces a CPUs cache to be in a state of continuous thrash and significantly reduces performance.
- **Stateful Tracking of Millions of Sessions:** Keeping state on millions of flows exacerbates the cache-thrash problem described with access control. Per-connection state is an additional data structure that must be accessed with every packet received, putting pressure greater pressure on caching and memory accesses. x86 CPUs achieve networking performance through caching, and stateful firewalling significantly reduces (or eliminates) any effectiveness of an x86 cache.

Server-Based Networking Compute Impacts

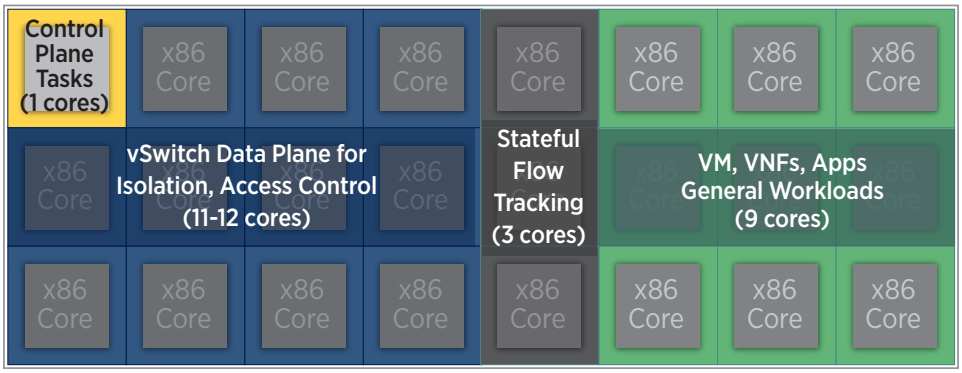

Figure 3. Illustration of the CPU Resources Consumed with Micro-segmentation

When considering security posture and data center scale-out, microsegmentation is an effective strategy in practice. However, there are severe penalties on each server in the way of low performing functions creating bottlenecks and increased resource consumption by security leaving little resource left for revenue-generating VMs. The upcoming benchmark section will illustrate how microsegmentation significantly limits the "scale-up" of a server.

BENCHMARKS

Network Isolation using Basic Overlay and Small- to Moderate-Tunnel Capacity

Graphic???

This data shows measured performance of executing network isolation via VXLAN tunnels and Open vSwitch (OVS). The tunnel capacities used are 1, 10, 100, and 1000 tunnels to illustrate the performance degradation that occurs as the tunnel capacity increases. This degradation is due to the reliance on CPU caching for network processing when using an x86 for such workloads. As a result, the networking subsystem becomes the bottleneck in the server, significantly limiting the network I/O of each VM deployed on the system. An additional and equally negative side effect is the CPU consumption. A high percentage of CPU resources are required for vSwitch processing. These CPUs cycles are critical to revenue generation and also have an impact on data center cost. Consequently, [x%] of the server resources are lost due to host-based networking in the case of network isolation.

Access Control using Traditional L2-L4 Filtering

Graphics??

This data shows measured performance of executing network access control in addition to isolation by configuring L2, L3, and L4 filters with VXLAN tunnels in OVS. The rule capacity is increased significantly to represent what would typically be deployed on a server hosting 10s to 100s of VMs. The performance degradation in this case is even more significant due to the rule capacity. Additionally, the CPU cores required to support the reported rate has increased to [x%], worsening the total output and system effectiveness of the server.

USING AGILIO TO ACCELERATE AND SCALE-UP SERVERS REQUIRING MICRO-SEGMENTATION

The Netronome Agilio line of intelligent server adapters aims to accelerate host-based networking, and are particularly effective in environments requiring micro-segmentation. Agilio adapters fully offload the matching and action processing for OVS and support extremely high rule counts, with a typical configuration supporting up to 64K rules. Higher end configurations can extend well beyond 64K, if needed. There is a two-fold benefit to the server when offloading the micro-segmentation workload to Agilio adapters: First, the networking I/O bottleneck previously mentioned is eliminated, allowing VMs to reach their full processing potential. Second, there is significant CPU savings realized by taking the OVS workload and executing it on the Agilio adapter at high rule capacities (64K) and high flow counts (millions of flows). By offloading the vSwitch and its associated policies, between 25-75% of the CPU resources can be reclaimed and used for VM deployments while eliminating the networking bottlenecks. This “scale-up” of servers effectively doubles or triples VM capacity, which has an

even greater effect on revenue per server generated.

Agilio Adapters Improve Compute Usage

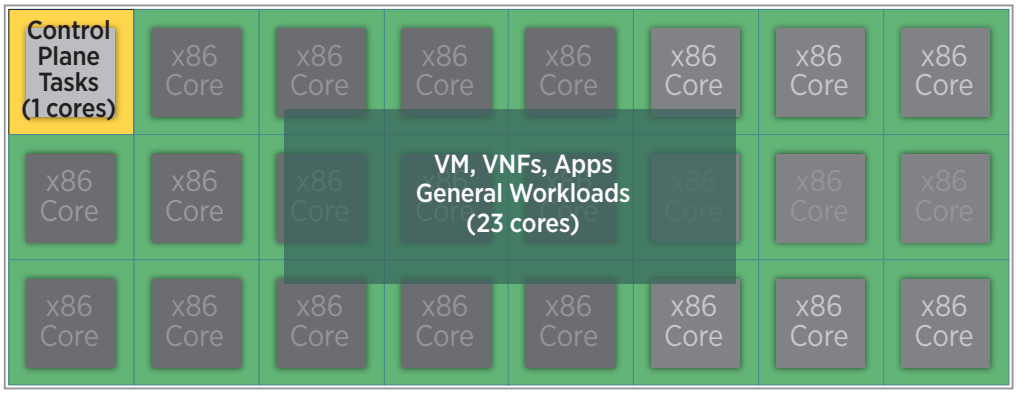

Figure 4. Consumption is Drastically Improved with Agilio Offload

Agilio adapters are purpose built for the networking and security processing required of micro-segmentation. Unlike x86 CPUs, Agilio adapters are typically configured with 48 cores optimized for OVS processing, with each core being 8-way threaded. This makes the adapter highly parallelized, which is critical for keeping up with 40G line rate. The Agilio software seamlessly integrates with host-based OVS. Rules and policies programmed into OVS are also programmed to the Agilio adapter, which allows pre-existing orchestration and controller systems to be used without modification.

BENCHMARKS WITH INTELLIGENT NETWORK ADAPTERS

Network Isolation using Basic Overlay and Small- to Moderate-Tunnel Capacity

Need a Graph that shows PPS performance (PPS as a percentage of 40GbE line rate possibly) and % CPU utilization at the following rule capacities (1, 10, 100, 1000), and Overlay CPU utilization

[Possibly include a hyphenated or grayed out version of the previous graphs to show contrast and improvement]

[Text to Match Graph]

This data shows measured performance of offloading network isolation via VXLAN tunnels to the Agilio adapter. The tunnel capacities used are 1, 10, 100, and 1000 tunnels to illustrate performance as the tunnel capacities increase. No performance degradation is observed due to the highly parallelized Agilio architecture. In this configuration, network bottlenecks are eliminated, giving VMs native networking performance without sacrificing functionality or security. The data also shows significant CPU resources being recovered, enabling increased VM density on the server and increased revenues.

Access Control using Traditional L2-L4 Filtering

Need a Graph that shows PPS performance (PPS as a percentage of 40GbE line rate possibly) and % CPU utilization at the following rule capacities (10K, 24K, 36K, 48K, 64K), and Overlay CPU utilization

[Possibly include a hyphenated or grayed out version of the previous graphs to show contrast and

improvement]

[Text to Match Graph]

This data shows measured performance of offloading network access control in addition to isolation by configuring L2, L3, and L4 filters with VXLAN tunnels on the Agilio adapter. The rule capacity is increased significantly to represent what would typically be deployed on a server hosting 10s to 100s of VMs. More impressive in this case, no performance degradation is observed at the highest of rule capacities (64K). When compared to the software-only vSwitching for access control, there are even more CPU savings and network performance gains when using the Agilio adapter for offloading.

CONCLUSION AND FUTURE WORK

In addition to network isolation and rules processing, the Agilio adapters and software include an exact-match flow tracker that keeps state on every micro-flow (or session), which traverses the offloaded vSwitch. This flow tracker can be used for connection tracking in the stateful firewall function. Stateful firewall functionality is current in research both at Netronome and in the community. Open vSwitch version 2.5 promises integration of connection tracking for stateful firewall, and Netronome is tracking this capability in research. While connection tracking will improve overall security, data center operators are already bracing for the performance impact and are looking for solutions. When that functionality is available for general consumption, Netronome will issue an update to this paper.

The Agilio Intelliegnt Server Adapters have proven to be the optimal solution for offloading micro-segmentation. By offloading isolation, access control, and in the future stateful firewalling, Agilio adapters and software prove to eliminate networking bottlenecks at high rule and flow capacities while simultaneously restoring more than 50% of the CPU cycles which can be repurposed for VM deployments and applications.

NETRONOME

Netronome Systems, Inc.
2903 Bunker Hill Lane, Suite 150 Santa Clara, CA 95054
Tel: 408.496.0022 | Fax: 408.586.0002
www.netronome.com