illumio

The Truth About Micro-Segmentation

Alan S. Cohen, Chief Commercial Officer at Illumio

WP201903

Micro-segmentation is fast becoming a foundational layer of the security architecture for today's data center and cloud computing environments. It has seen a big push by a range of vendors as well as growing recommendations from leading analyst firms such as Gartner, ESG, and the 451 Group. There are three reasons why organizations of all sizes are considering this technology in their data center and cloud security plans:

1. It is a core compensating control that complements patching, vulnerability scanning, and identity in reducing attack surface, particularly for east-west traffic;
2. Because it is based on a zero-trust or least-privilege model, it actually reduces the number of false positives pinning down security operations teams today;
3. If implemented correctly, it can keep up with increasing the heterogeneous, hybrid, and dynamic nature of today's computing.

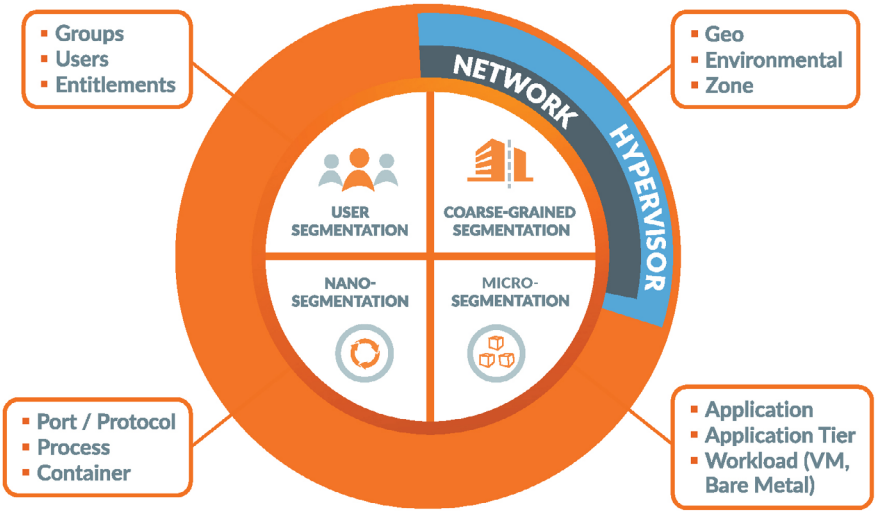
Since the first micro-segmentation technology for data center security was introduced about five years ago by Nicira, the security and networking industries have been racing to introduce competing approaches to reduce the lateral spread of bad actors in the data center and cloud. One of the key insights of VMware's NSX team, to which my team fully subscribes, is that traditional networking technology presents a boatload of limitations to implementing micro-segmentation at scale. As they noted in August 2014:

> The idea of using network segmentation to limit lateral traffic isn't new, but until recently it was never feasible. Even if you blanketed your data center with a legion of hardware firewalls, there'd be no way to operationalize them, and the costs would be astronomical. Until now.

The paper outlines the key components to look for with micro-segmentation vendors:

Don't Sell Me Micro When You Mean Macro

The original segmentation model for the data center was the network security firewall. Because it was a single chokepoint that can process a blacklist model at line speed, it has been manifested in hardware devices with increasing levels of capacity and throughput. Network devices do a good job of coarse-grained segmentation—not only for the perimeter—but for well-defined zones that include environmental separation in relatively static and well-defined boundaries.

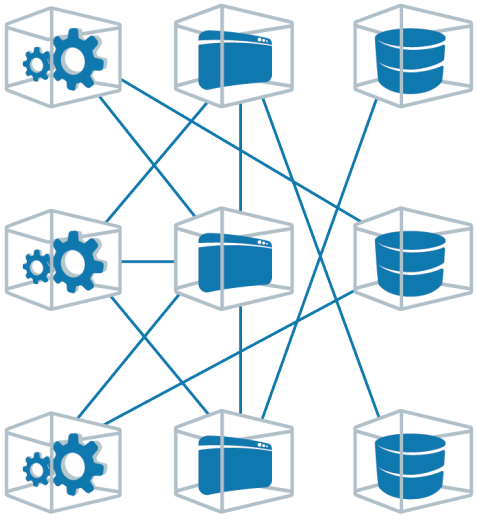


Where networking approaches fail—and this includes the network stack in the hypervisor or containers—is where you need the more granular security segmentation of micro-segmentation. As you move to ringfence applications, tiers of applications, or individual workloads, the network and hypervisor model both lacks the context and the flexibility to do the job. What happens if an application spans several data centers? Would you hairpin traffic back to an enforcement point?

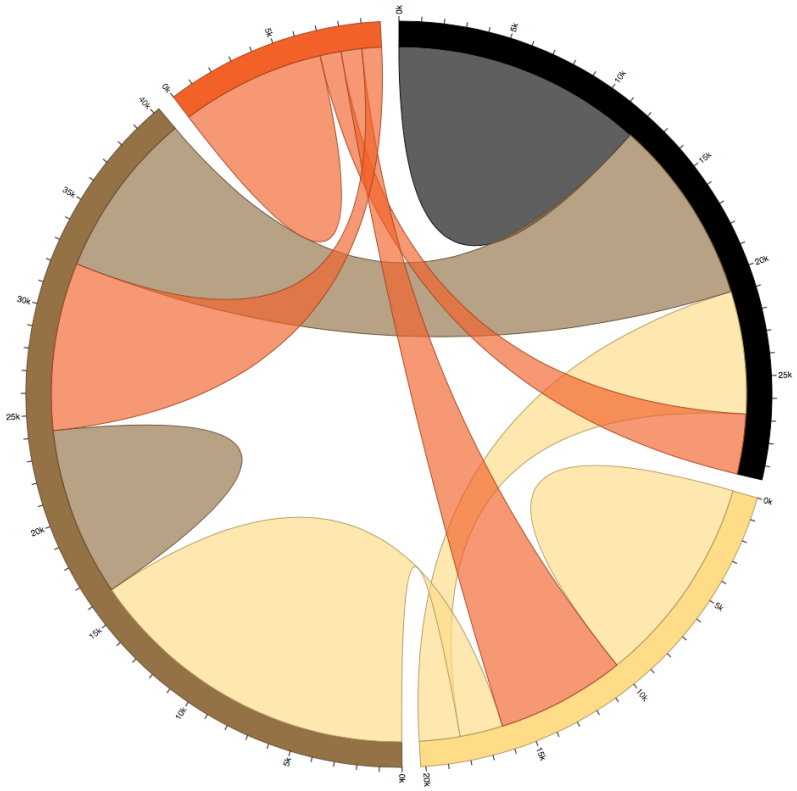
There's the Neat and Clean World; and Then There's the Real World

You cannot build granular segmentation without understanding how applications communicate. With the increasing spread of attacks inside the data center and cloud—malware, insider threats, or simply application or communications vulnerabilities exploited by bad actors—strong micro-segmentation approaches cannot be implemented unless IT operations and security have clear visibility into how their applications are communicating so that they can quickly determine what should be communicating.

Traditionally, application dependency maps are built manually, as network flow and one server at a time. This approach is nearly unworkable in the largest data center and cloud environments. The marketectures have a symbolic icon view to describe a perfect three-tier application:



When you get past the marketing side of things, there is a strong movement in the industry to use D3 JavaScript diagrams to create stronger visibility into application and network environments. Automated data collection and new visualization tools do a better job of creating the map. However, most vendors still offer simple, stylized views of application dependencies, which are not particularly useful at scale. D3 "chord diagrams" can offer directed relationships among a set of entities. "Sunburst" diagrams go a step further to show relationships as well as application groups.



Beyond the Neatly Presented View of Your Application Communications

In reality, most data centers or even simply applications do not remotely look like the earlier diagrams from an application dependency map. They actually look like this

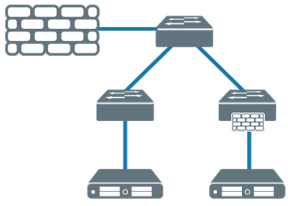
To bring visibility and application dependency mapping to micro-segmentation, both systems must have these five properties:

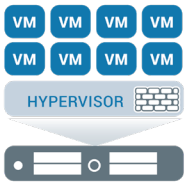
1. Work at scale, up to tens of thousands of workloads and hundreds of thousands of other objects in the map, including laptops;
2. Be precise enough to support a large whitelist model;
3. Adapt to changes in the environment;
4. Work in all environments and infrastructures during changes to applications or migration to the cloud; and
5. Provide the intelligence and automation to eliminate the manual model that scale deployments make nearly impossible.

From a security perspective, to understand application dependencies, you need not only to understand the flows and servers, you need to understand the ports and underlying processes. Most servers will have dozens of open, hence vulnerable, ports. More significantly, these maps must be living systems, not a one-time snapshot; otherwise, it is impossible to keep up with the dynamic and distributed nature of today's cloud and microservice architectures.

Getting Past Monoculture in Security and Computing Operations

For computing formats, we are in a post-virtualization era, where containers and "serverless" formats are rapidly gaining ground and forced to run alongside legacy approaches. So why would anyone choose a data center and cloud micro-segmentation solution that is boat-anchored to those fixtures of the client server era, the network, and the hypervisor? Today, there are different types of segmentation-centric hypervisor centers or distributed (e.g., host-centric) architectures:





In the Network

Segmentation, whether it is performed in a switch or a firewall, was designed during the era of static workloads and excels in north-south traffic flows where "big iron" plays an important role from a throughput or lookup consideration. High-capacity firewalls are terrific at filtering inbound traffic by providing granular flow analysis and are useful for clustered storage and aging legacy computing platforms, such as the IBM AS400.

The challenges of this model are reliance on proprietary and operationally complex hardware, where replacing hardware can prove daunting from a cost or availability perspective. Hardware solutions do not translate to cloud environments, such as Amazon Web Services or Microsoft Azure; and the "virtualized" versions of hardware solutions—for example, running on a virtual machine—have serious throughput limitations and create fragility through service chaining or traffic steering challenges.

In the Hypervisor

Segmentation in the hypervisor was designed to filter traffic through hypervisor-attached firewalls or increase network virtualization, such as VMware NSX. Each hypervisor has visibility into traffic flows and can enforce security policies locally. The limitations, however, are well understood:

- Poor support for legacy servers, NAS, bare-metal, containerized, or public cloud workloads;
- Limited support for heterogeneous virtualization environments; and
- No knowledge of processes or services initiating traffic.

Distributed, Host-Centric Segmentation

The newest form of micro-segmentation is derived from an overlay approach that is decoupled from the infrastructure yet takes advantage of packer filtering in the operating system (for example, Windows Filtering Platform, iptables, or Berkeley Packet Filters in Linux) or other devices (layer-4 firewall in load balancers, ACLs in network switches). This approach helps security professionals craft policy centrally and distribute enforcement for scale.

The benefits of this approach include the following:

- Complete application visibility, regardless of underlying infrastructure;
- Insight into processes and services establishing connections;
- Bare-metal, VMs, and containers on premises and/or in the cloud;
- Stops out-of-policy traffic before it hits the physical network; and
- Integration into heterogeneous environments.

If your world or future is increasingly distributed, dynamic, heterogeneous, and hybrid, the architectural choice is clear.

Follow Us

About Illumio

Illumio, the leader in micro-segmentation, prevents the spread of breaches inside data center and cloud environments. Enterprises such as Morgan Stanley, BNP Paribas, Salesforce, and Oracle NetSuite use Illumio to reduce cyber risk and achieve regulatory compliance. The Illumio Adaptive Security Platform® uniquely protects critical information with real-time application dependency and vulnerability mapping coupled with micro-segmentation that works across any data center, public cloud, or hybrid cloud deployment on bare-metal, virtual machines, and containers. For more information about Illumio, visit www.illumio.com/what-we-do or follow @Illumio.

Illumio, Inc. 160 San Gabriel Drive, Sunnyvale, California 94086 Tel (669) 800-5000 www.illumio.com