PATHFINDER REPORT

Trusting in the CPU: Getting to the Roots of Security

How chip-level security is taking on an increasingly important role in the securing of workloads, particularly when virtualized and cloud-based infrastructure is involved.

By John Abbott, cofounder and distinguished analyst, 451 Research

JUNE 2017

COMMISSIONED BY:

AMD

About this paper

A Pathfinder paper navigates decision-makers through the issues surrounding a specific technology or business case, explores the business value of adoption, and recommends the range of considerations and concrete next steps in the decision-making process.

About 451 Research

451 Research is a preeminent information technology research and advisory company. With a core focus on technology innovation and market disruption, we provide essential insight for leaders of the digital economy. More than 100 analysts and consultants deliver that insight via syndicated research, advisory services and live events to over 1,000 client organizations in North America, Europe and around the world. Founded in 2000 and headquartered in New York, 451 Research is a division of The 451 Group.

NEW YORK
1411 Broadway
New York, NY 10018
+1 212 505 3030

SAN FRANCISCO
140 Geary Street
San Francisco, CA 94108
+1 415 989 1555

LONDON
Paxton House
30, Artillery Lane
London, E1 7LS, UK
+44 207 426 1050

BOSTON
75-101 Federal Street
Boston, MA 02110
+1 617 598 7200

Introduction

Security is the IT problem that won't go away. As we move toward the digital enterprise, and more of the daily actions of businesses, governments and individuals depend on the use of data and application services – often delivered from the cloud – the number of successful cyberattacks goes up as well, as does the attackers' motivation to find and exploit the vulnerabilities, whether for criminal or state purposes.

Defense against these attacks is only as strong as the weakest link. It follows that there is no simple way to provide security. The hardware and software infrastructure is important, as is the network; the secure design and operation of applications and systems is critical, as is the training and vigilance of each user.

That said, security at the CPU and system-on-chip (SoC) level is starting to play a perhaps unexpectedly large role in all of this. That's because of changes in system and application structure (particularly the rise and ubiquity of virtualization over the past decade), because of the importance of defining trust and limiting attack surfaces, and because of the foundational role that only secure hardware can play. Chip-level security is likely to become even more top of mind as the Internet of Things distributes processing power toward edge infrastructure, where security concerns become even more critical.

Context

The impact of changes in the IT landscape over the past decade – triggered by the introduction of smartphones and the rise of social media, and backed by elastic cloud-based applications and services – has been huge. All this complex consumer activity led to rapid advances in wireless networking and cloud-based data and services, and drove the market for highly efficient, scale-out datacenters of massive proportions to support it.

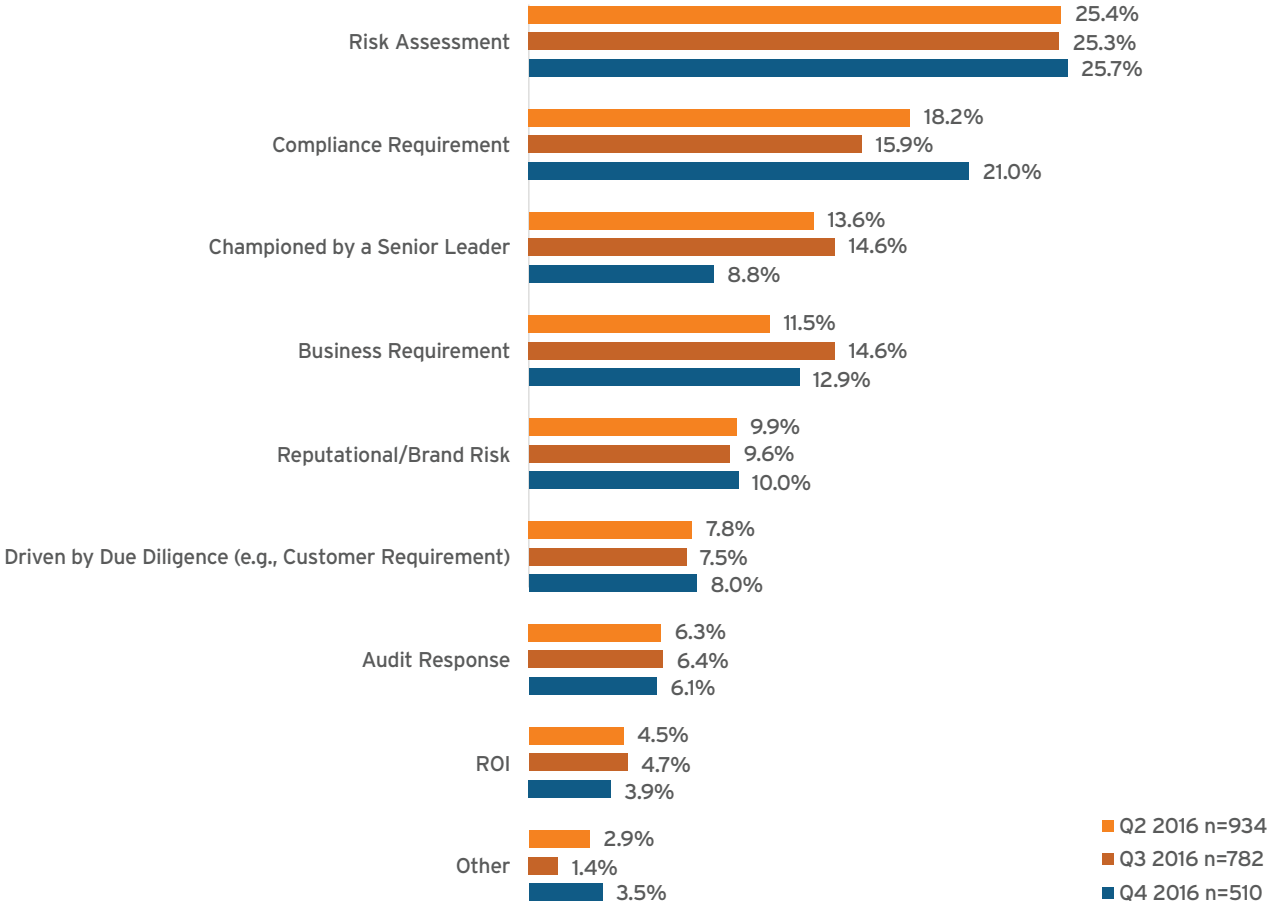
Enterprise IT teams had to respond in kind – providing the appropriate infrastructure to support the increased diversity of devices employees wanted to use, replicating the efficiency of web-scale architectures with their own 'private clouds,' and incorporating the use of public cloud services within the context of their internal IT policies and control. Today, companies are required to offer flexibility to their users while protecting company data, employee identity and securing their infrastructure – all against a backdrop of ongoing and advanced cybersecurity threats and increasing regulatory requirements.

Customer Concerns

According to a 451 Research survey of IT end-user decision-makers focused on security *(Voice of the Enterprise: Information Security, Budgets and Outlook 2016),* about 51.6% of respondents said they anticipate a slight increase in their budgets during 2017, while 15.6% said they have plans for a significant increase. This 67.2% that are planning to spend more on security in 2017 is slightly off of the 70.1% who said the same for 2016, but not significantly so. The average percentage increase over the next 12 months is 21%, indicating a 'project implementation' level budget for information security.

In the same survey, we asked the following question: For the top information security projects currently being implemented within your organization, what was the key determinant in their approval? The responses showed that more than one-quarter of active projects capturing security spending are still driven by some manner of risk assessment, an accounting of the probability and impact of a potential security problem.

Figure 1: Reasons for implementing security projects

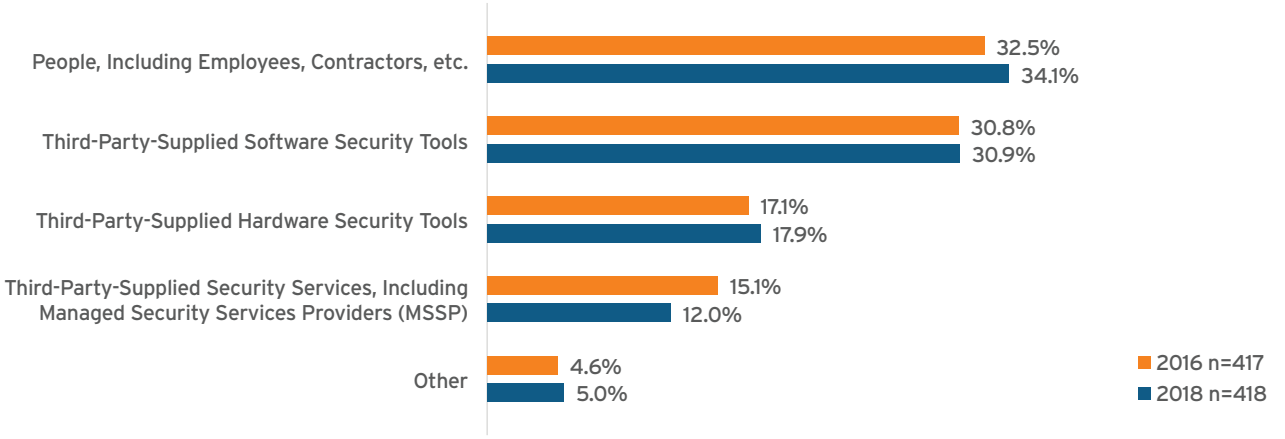


Source: 451 Research, Voice of the Enterprise: Information Security, Budgets and Outlook 2016

That said, compliance as a primary project driver also increased significantly in this latest study, from 15.9% to 21%. Industry-specific compliance regulations, such as HIPAA and PCI, were cited by half of respondents as the key aspect that's driving their overall compliance concerns. We expect security products to continue to appeal to and pivot off of compliance requirements over the next 12 months and beyond.

The more dramatic shift, though, is in hardware-based security tools, which made up 20% of the security budget in the survey in 2015, 17.1% in 2016, and is predicted to be 17.9% in 2018. The reason behind this shift is that the portability of security products remains key in a hybrid on-premises/hosted operating environment, so proprietary appliances, hardware modules and adapter cards have been losing out to more flexible software tools. However, this trend masks a contrary increase in the adoption of processor-level security, which sits below the OS and hypervisor layers and doesn't require any application changes further up the stack. Processor-level security is foundational and can prevent the exploitation that software-based products can be prone to.

Figure 2: Security spending distribution: now and in two years' time



Source: 451 Research, Voice of the Enterprise: Information Security, Budgets and Outlook 2016

It takes both hardware and software to provide comprehensive security. But OS and application-layer security built in by the developer community, or external tools such as antivirus protection, are no longer sufficient to keep a system secure on their own. CPU and SoC makers can provide many of the key foundational elements that will help – ranging from isolation and trusted execution environments to more transparent and granular data and process encryption.

Software: the Good, the Bad and the Ugly

In today's environment, as Marc Andreessen, Netscape cofounder and general partner at VC firm Andreessen Horowitz, has colorfully pointed out, "software is eating the world." In the big picture, the advances in integrated-circuit technology have made the cost of computational hardware a small part of the overall cost of IT solutions – the value is in the software.

But the software is unquestionably also the Achilles' heel of security. Modern operating systems and applications have tens of millions of lines of code with many undiscovered vulnerabilities (testing can demonstrate the presence of defects but never prove the absence). From a security point of view, software is the 'attack surface.' And the smaller the attack surface, the easier it is to make it secure.

One key element that has changed the security landscape over the past decade is server virtualization – the ability to run multiple disparate (different software stacks) workloads on the same shared infrastructure. Virtualization has played a key enabling role in the establishment of cloud computing, both on- and off-premises, and continues to influence the evolution of security and the server CPU. However, virtualization has introduced a whole new set of security challenges around data privacy and the risks of consolidation.

Most pressing are the issues with full isolation between co-resident virtual machines running on the shared infrastructure that cloud providers operate. Hypervisors have unrestricted access to the contents of their guest operating systems and the applications running inside them. A host hypervisor has a complete view of all the data and code in each guest virtual machine, making it possible for administrators to scrape the memory of guest data areas, or even inject code into a guest virtual machine (VM). Any bug in the hypervisor could enable a hosted guest to steal data from other guests. In short, current virtualization architectures put too much reliance on the security of the hypervisor. And as private, shared-infrastructure clouds gather momentum, the problem is becoming visible in on-premises enterprise IT, as well as at cloud companies.[1]

The Importance of Hardware

Even in the context of (or perhaps because of the) millions of lines of code, the CPU or SoC can play a very powerful role in system security. As soon as software is involved, issues of trust become very complex, very quickly.

Traditional computing systems have used a ring-based security model, in which high-privilege code has full access to the resources at its level and of all lower-privileged levels. By its nature, hardware can provide a more secure root of trust. A microprocessor can be designed to provide basic authentication and secure communications capability without depending on any additional software. It can include embedded logic that enables an authenticated and protected communication link to be created over a network connection to the device, including a robust means of authenticating the device's identity (providing some response that only it could make, based on data stored within the device) and untampered-with state.

Implementing security at the chip level is typically faster and fits more naturally into existing practices and tasks (such as kernel/user separation and data encryption) where the CPU is already heavily trusted. Once the CPU has demonstrated it can be trusted, it can then be used to validate that the initial software loaded onto the CPU is correct (i.e., what is desired). Trust in all cases must be physically 'rooted' through the use of specialized hardware, or in this case by a silicon functionality in a server CPU.

Baking in Security – Evolution and the Current Market

As the x86 architecture became the ubiquitous CPU platform for enterprise computing and the datacenter, building in security became a priority. In fact, Intel pioneered some of its chip-level security technology in the high-end Itanium RISC architecture in the late 1990s. But by mid-2005, the focus had shifted almost exclusively to x86. Around then, both Intel and AMD introduced the first hardware extensions to support native virtualization (Intel VT and AMD-V™), boosting security through logical isolation, as well as improving the performance by reducing the need for software emulation. Alongside that, Intel introduced its Trusted Execution Technology, which validates key components as they are launched at system startup, and Identity Protection Technology, which allows the hardware to store an authentication token.

Intel's big plan to bring a broader security stack in-house and bake it into silicon – through its $7.7bn acquisition of McAfee in 2010 – didn't pan out as expected, and the McAfee unit has been spun back out to a private equity firm. But Intel continues to invest in security, an example being its longtime Enhanced Privacy ID project – now open source – which signs each device with a unique signature at point of manufacture and ties to Intel's broader digital rights management scheme.

However, the x86 market for enterprise infrastructure is once again a two-horse race, with Intel and AMD each taking somewhat different approaches. AMD hadn't fielded a server chip family against Intel for half a decade. But in June 2017, it reentered the space with the launch of a server version of the 'Zen' micro-architecture, code-named 'Naples' and officially branded EPYC™. With EPYC, the aim has been to take a fresh look at its SoC technology in the light of modern system and application design and implementation.

Key additions to EPYC, alongside a new design that focuses on the improvement of memory bandwidth and memory capacity, are new capabilities that enable the chip to contribute more to system security through the real-time encryption and decryption of data in memory, in hypervisors and in virtual machines. The rest of this paper will focus in on AMD's newly-thought-out approach and strategy for chip-level security in these areas.

Main Memory Encryption and Encrypted Virtualization

Just as Intel has its Management Engine, AMD has included a dedicated security processor within its server SoC products. Initially known as the Platform Security Processor, this has recently been rebranded as the AMD Secure Processor. It's an integrated ARM® Cortex® A5 that sits alongside CPU cores. It provides a dedicated secure space to run multiple security-related functions that require full isolation and can be off-loaded from the main CPU core. One of these functions is a secure OS, and that's the starting point for the AMD Secure Root-of-Trust Technology, one of the three basic categories of security features offered by the EPYC SoC. The other two are AMD Secure Run and AMD Secure Move technologies. These are summarized in Figure 3.

Figure 3: The three basic categories of security features supported by EPYC

Secure Root-of-Trust (AMD Secure Processor, Secure Boot)	Prevents the use of rootkits/bootkits that inject malicious code prior to an OS loading. Creates hardware root of trust, enabling only known and trusted software to be loaded and run – from initial boot load through BIOS load.
Secure Run[2] (Secure Memory Encryption, Secure Encrypted Virtualization)	Memory scraping and cold boot attacks are both modes of attack that can be used when data running in the main system memory is not encrypted. Secure Memory Encryption (SME) encrypts system memory. Secure Encrypted Virtualization (SEV) isolates the hypervisor and guest VMs to prevent access to data in shared guest data areas.
Secure Move[2] (SEV-enabled servers, APIs, third-party key management)	Migrating VMs within the datacenter, to off-premises private cloud, or to the public cloud can be insecure. SEV-enabled servers can establish a secure channel between them and send memory encryption keys to the remote platform.

Source: AMD

- SECURE ROOT-OF-TRUST is a well-known security requirement for bare-metal system boots – Intel's Boot Guard is the obvious comparison. But in addition to securely booting the native system (including authenticating AMD Secure Processor code and OEM BIOS code), AMD also supports secure boot for SEV virtual machines, and can prove that the boot image for those machines has not been tampered with by the hypervisor or another third party. The system authenticates and loads code for the AMD Secure Processor to perform key management. And there's an additional benefit for AMD's OEM customers, which can set their systems up so that only their own BIOS can be loaded.
- SECURE RUN is all about 'data at work' – that is, data in the main system memory – an expansion beyond the usual coverage of encryption technologies that secure 'data at rest' (on a disk or solid-state drive) or 'data in motion' (on a network). Data at work is the focus of the two central security additions to EPYC: Secure Memory Encryption (SME) and Secure Encrypted Virtualization (SEV).

  Memory-encryption technologies have been used before – in Microsoft's Xbox, for example, while Intel's Software Guard Extensions, first introduced in 2015, included new instructions to enable user-level code to specify protected memory regions called enclaves. AMD's approach has been to integrate technology in the CPU, making it more broadly applicable, scalable (from embedded systems to high-end servers) and, crucially, requiring no application software modifications. Some OS- and hypervisor-enablement is required for both SME and SEV, and AMD is working with the leading OS and hypervisor vendors on the required modifications.

  SME uses dedicated hardware in the on-die memory controllers with an Advanced Encryption Standard Engine to encrypt data written to DRAM and decrypt it when read. Encryption keys, managed by the AMD Secure Processor, are randomly generated on each system reset and not visible to software running on the main CPU cores.

SEV directly addresses the problem of the unrestricted access that the hypervisor has to data and code, mentioned above. The requirement for virtual machines to fully rely on the security of the hypervisor is avoided through the cryptographic isolation of code running at different levels (namely hypervisor or guest). When using SEV, neither has access to the resources of the other, meaning that lower-privileged code is protected and no longer dependent on high-privileged code for startup and execution. The hypervisor and guest layers are still able to communicate, but only through much more tightly controlled communication paths.

For public cloud providers, the implications of this approach are particularly important. They can provide customers with VMs that are better isolated and protected on a server, enabling them to offer secure, dedicated VMs without having to provide a physical dedicated server, thus increasing server utilization. All virtual machines have unique keys, so a compromised VM can no longer access the hypervisor or any other VMs. And the providers themselves will not be able to access any client data because they don't control the keys – more specifically, the security processor owns the keys, and the administrator does not have access to them. This prevents the hypervisor from being able to see what the VMs are running. The net benefit to this is that clients are protected from the cloud providers, and the cloud providers have isolation from what the clients are running on their servers.

- SECURE MOVE relies on the establishment of a secure channel between two SEV-enabled platforms so that the hypervisor can implement migration and snapshot functions securely. VMs might be migrated within the datacenter itself, to private off-premises cloud, or to the public cloud (assuming that the cloud provider has an SEV-enabled platform in place). With the secure connection established, SEV firmware sends the guest's memory encryption keys over to the remote platform, which can then run the guest. In this case, AMD is providing the APIs and methodology for third-party software vendors with key management, lifecycle management and other tools to use.

Looking Toward the Future

Technology infrastructure platforms continue to evolve at a breakneck pace. New forms of non-volatile main memory are starting to reach the marketplace, and they will significantly affect the design of hardware platforms and the system software that runs on them. If storing unencrypted data in DRAM is an issue today, leaving it vulnerable to rogue admin or hardware-probing attacks, it's set to become even more important as these new persistent memory options become more widely used. That's because NVDIMM chips can be physically taken out of a system with the data intact, leaving confidential data, passwords and secret keys accessible.

Other areas beyond traditional virtualization and VMs are likely to benefit from hardware-based virtualization and memory-encryption technologies for protecting workloads. Containers and micro-services are already having a big impact on how infrastructure is being built and consumed. SEV usage is already being extended to support memory-encrypted containers, increasing the granularity and flexibility of protection levels. There's also growing interest in hardware virtualization as an isolation mechanism (such as Microsoft's Virtualization Based Security, the basis for the Credential Guard and Device Guard features in Windows 10). Stanford's DUNE project uses hardware virtualization as the basis for process isolation, while security startup Bromium has used hardware virtualization for micro-isolation techniques.

THE 451 TAKE

Security functionality embedded in hardware has not had the greatest reception in the past. Witness Intel vPro, which had several very beneficial technologies for isolation and management, but which found limited adoption because no one wanted to pay the premium on the desktop. Other efforts floundered mainly because of the dependencies they might have on – and, conversely, could impose upon – the ecosystem within which the hardware was deployed. It's a brilliant idea, but too often hampered by cost considerations and the potential implications of dependence on the low-level execution model.

The big exception is mobile, where limited control over what happens at the processor level is extended to the user, with resulting security benefits. Why did that work? Perhaps because the total user experience turned out to be highly positive. People embraced the iPhone model (and later Android) because of that holistic experience – which includes end-to-end control of the software ecosystem, in Apple's case. Security was a beneficiary of that strategic and comprehensive user experience concept.

One could make much the same argument for the security benefits of IaaS in the cloud. It behooves cloud providers to deploy a consistent, highly automated, easy-to-manage environment to capitalize on economies of scale – and security is a beneficiary of that. When large cloud providers adopt a security measure – or more significantly, when they brand and market it – they are doing so mainly to reduce adoption friction, and rarely because they see a market opportunity for security per se. Here again, security is primarily a benefit of the comprehensive and holistic customer experience of IaaS.

Summary

Security is an essential part of modern IT, especially given the growing use of shared resources (e.g., the public cloud) and the creation of composite applications via network integration of distributed components, often resident in multiple datacenters. Consumers require easier access (via biometric authentication, for instance) even while security levels are increased. Businesses need their corporate data secured both centrally and remotely. There's a particularly strong motivation from the industry to provide additional security isolation for workloads in the cloud, equally applicable and important to cloud users and cloud service providers. And as intelligence redistributes from the cloud back toward the edge, embedded security will be required to play its part in securing the Internet of Things.

The server CPU plays a fundamental role in security because of performance, hardware-rooted trust, and the ability to provide security functionality, such as encryption, while exposing minimal attack surface area. The use of virtualization and cloud computing – which integrate disparate software on a shared server – shows additional value in hardware-implemented security functions that avoid the need to implement the same capabilities individually for all the different workloads.

1 The Cloud Security Alliance details the top cloud security risks in its February 2016 paper The Treacherous 12 - Cloud Computing Top Threats in 2016. Threat 12 details shared technology issues. https://downloads.cloudsecurityalliance.org/assets/research/top-threats/Treacherous-12_Cloud-Computing_Top-Threats.pdf

2 Some features may not be immediately available and may require OS and/or hypervisor enablement.