DEMYSTIFYING SERVER ROOT OF TRUST

UNDERSTANDING TECHNOLOGIES OFFERED TO COUNTER THREATS TO INFRASTRUCTURE

EXECUTIVE SUMMARY

Secure data begins with secure infrastructure. Protecting infrastructure begins with not only making sure it will operate as expected but also with the confidence that all the necessary firmware needed to run the system remains secure.

Root of Trust (RoT) is ideally based on a hardware-validated boot process to ensure the system can only be started using code from an immutable source[1]. This involves an anchor for the boot process rooted in hardware that cannot be updated or modified in any way. When combining this foundation with a cryptographically secured signature, there are no easily accessible gaps for hackers to exploit. Similar to a proof by induction, the entire server state will stay well known, as long as that initial element tests the next element, and so on.

Any variety of methods may start a root of trust, with some more comprehensive than others. In principle, a chain can be any length, but there must be an initial trusted image ("root of trust") that can start the process.

The root of trust's location determines the degree of protection. This paper will explain the various technologies and approaches that can achieve complete protection when the anchor is in the silicon.

SECURING THE DATACENTER IS MORE COMPLEX THAN EVER

Most industry experts believe that the Internet of Things (IoT) will account for more than 25 billion devices generating 600 Zettabytes of data per year. This breaks down to 18.9 Petabytes of data *per second*. Less than 20 percent of this data would be considered "secure". Of the 18.9 PB of data generated every second, 15.7 PB would be considered exploitable.

The cloud adds complexity to any enterprise security strategy. Virtual machines, containers, microservices and data routinely move from on-premises to off-premises

[1] http://www.uefi.org/sites/default/files/resources/UEFI%20RoT%20white%20paper_Final%208%208%2016%20%28003%29.pdf

clouds with an assumed fidelity of security from IT administrators. And while protections are built and deployed on the perimeter, infrastructure hardware (i.e., the servers that run apps and store data) are often overlooked attack surfaces ripe for exploitation.

In addition to the explosion of IoT data, organizations of all sizes face what may seem like an unmanageable number of devices open to new attack surfaces for hackers, using ever more sophisticated tools. These factors have led experts to predict that hacking will cost the global economy $8 trillion in 2022, increasing fourfold from a forecasted $2 trillion in 2019.

YOUR SECURITY STRATEGY IS LACKING

The U.S. National Security Agency experiences 300,000,000 hack attempts per day. While a staggering number, it runs consistent with what organizations of all sizes are experiencing in every country on a daily basis. This overwhelming burden has caused many organizations to harden re-harden their perimeter security strategy with a regular cadence.

Infrastructure security has lagged in both consideration and deployment. While pockets of security exist, the underlying server infrastructure still sits exposed in many of today's organizations, leaving gaping holes for hackers to exploit through rootkit and firmware attacks.

The key to a successful IT security strategy is to reconsider and refresh scenario planning on a regular basis. Given the breadth, depth and complexity of this ever-evolving threat, it is logical to think the question becomes not "if," but "when."

A response plan that accounts for detection, isolation, destruction and recovery is critical to mitigation.

THE THREATS ARE NOT ALWAYS EXTERNAL

A 2016 Harvard Business Review cybersecurity study found that 60 percent of security breaches in the datacenter were the result of employees or contractors. Of those internal breaches, 75 percent were conducted with malice or criminal intent and 25 percent from negligence.

The list of publicized internal U.S. government breaches is long, affecting all sectors. A Ponemon Institute study found that nearly 90 percent of healthcare organizations in the

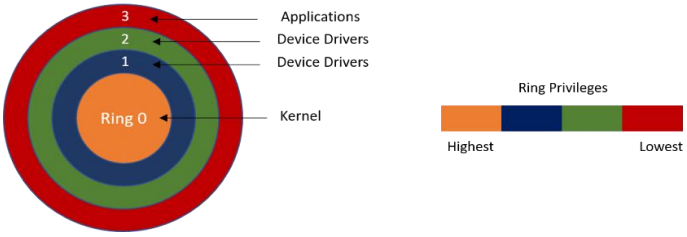
U.S. suffered a data breach. Nearly half of those breached suffered five or more cyber breaches during the year.

ROOT KIT ATTACKS TAKE MANY FORMS

The longer malware resides on a server, the more damage it inflicts. On average, malware goes undetected for more than 99 days. The lower malware sits in a server's operating environment, the greater chance it has of avoiding detection.

Most enterprise IT professionals are familiar with root kit attacks. However, the level of exploitation at the lowest levels is not as well known. Computer architectures implement a security mechanism known as "rings." Ranging from 0 - 3, each ring provides protected and privileged status at the lowest levels. Ring 0 contains the OS Kernel, while rings 1 and 2 contain device drivers and hypervisor (if virtualization is deployed). With virtualization technologies from Intel and AMD, hypervisors now access ring 0. Root kit attacks may gain access and privileges in ring 0 and even below.

FIGURE 1: RING ARCHITECTURE SHOWN



Source: Moor Insights & Strategy

Bootkit and *firmware attacks* allow for bad actors to gain access to a server before an OS is even present. This level of access allows for malware to remain present, yet virtually undetectable to even modern security technologies deployed in the datacenter.

Firmware attacks are perhaps the more difficult to combat, and they will likely remain undetected. As difficult as it is to maintain control over server firmware in the enterprise, the vast majority of enterprise IT organizations are ill prepared. In a study by the Information Systems Audit and Control Association (ISACA), only 8 percent of enterprises had adequate measures in place to control and manage the firmware in their environment. Managing firmware security is a time and resource-intensive effort because the technologies used to secure firmware to date have been immature.

ROOT OF TRUST SIMPLIFIED – IT ALL STARTS HERE

Root of Trust (RoT) is a means by which a system validates the authenticity and state of its boot process. Normally implemented with a compute engine (co-processor), RoT helps prevent many exploitation techniques, including rootkit and bootkit attacks.

Not all RoT implementations are created equal. The depth and completeness of protection are tied to where RoT is established, and the integrity of the RoT supply chain. The following paragraphs will describe some of the more common RoT implementations found in server technology.

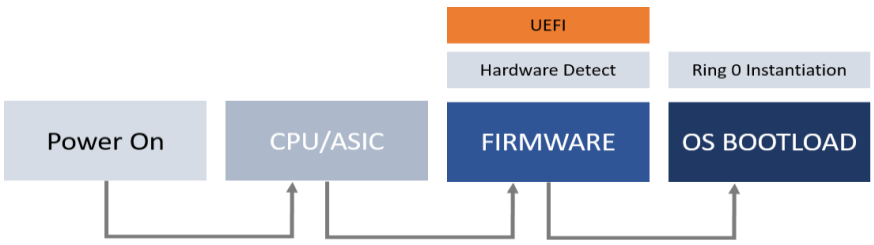
UEFI SECURE BOOT

UEFI Secure Boot is an optional element of the Unified Extensible Firmware Interface (UEFI), a firmware specification which defines a common format for code the firmware can execute.

UEFI Secure Boot helps ensures that the server boots using only software that is trusted by the manufacturer. When the server starts, the firmware checks the signature of each piece of boot software, including the bootloader, firmware drivers operating the system kernel and operating system drivers. If the signatures are good, the server boots and the firmware gives control to the operating system.

Although UEFI provides a boot method with a level of security against rootkits and bootkits, it assumes the integrity of the server's essential firmware. Stated another way, the server firmware is still exploitable during the boot process.

FIGURE 2: TRUSTED BOOT VIA UEFI



Source: Moor Insights & Strategy

TRUSTED PLATFORM MODULE

Trusted Platform Module (TPM) is a separate processor that monitors system state. TPM is a passive component needing to be updated and cannot lock down any component in the system except access to its own memory. It also provides some cryptographic operations, such as creating and working with RSA keypairs.

The first verification of signatures happens by code on the CPU, which can be intercepted and replaced. Emulating a "properly" booted system is possible by sending the right values to the TPM.

The bootblock, the part of the firmware that contains the first instructions executed by the CPU, comes first and anchors the root of trust. However, if you cannot trust the bootblock to send a truthful state into the TPM, the system is vulnerable[2].

Another concern with TPM is integrity because there is no guarantee of it in the supply chain. What does this mean for the IT administrator deploying servers in the datacenter? There is no guarantee that their server vendor has generated and maintained possession of keys while implementing the TPM chip on the server they are about to rack.

INTEL TRUSTED EXECUTION TECHNOLOGY (INTEL TXT) & BOOTGUARD

Intel TXT uses a TPM and cryptographic techniques to provide measurements of software and platform components. This enables system software and local and remote management applications to use those measurements to make trust decisions.

Intel TXT technology attempts to provide a trusted way for loading and executing system software, e.g. operating system kernel or Virtualization Machine Monitor (VMM). Intel TXT performs software measurements and stores them in particular TPM registers[3].

Intel TXT does not make any assumptions about the state of the system before loading the software, thus making it possible for a user to ensure secure load of an OS or VMM, even in a potentially compromised machine. Boot sector viruses and BIOS rootkits

[2] https://www.hpe.com/h20195/v2/getpdf.aspx/a00008520enw.pdf?ver=1 p10

[3] https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/trusted-execution-technology-security-paper.pdf

could exist and TXT would still allow the load of a clean VMM (or OS kernel) in a secure manner, despite malware present in the system.

Introduced with Intel's 4th generation core processor platforms, Intel Boot Guard is a hardware-based technology designed to prevent tampering with the UEFI firmware. Intel Boot Guard works upstream in time helping to protect the system before UEFI Secure Boot's protections kick in[4].

Boot Guard has three separate modes:

1. ***Verified Boot Mode*** cryptographically verifies and initial boot block.
2. ***Measured Boot Mode*** uses a measuring process (footprint) as a means for verification.
3. **Combination Boot Mode** employs cryptographic verification along with measuring for absolute verification.

Boot Guard configurations vary somewhat across OEMs, as each OEM is responsible for configuring a public key for the verified boot and establishing boot policies. The security of the verified boot is rooted to the OEM's key pair. The OEM generates a 2048-bit key that is only used for verifying the initial boot block, the private portion of which must be kept securely. The public portion of the key is then programmed into field programmable fuses during the manufacturing process. These fuses cannot be updated once written, providing a solid starting point for UEFI Secure Boot.

However, Intel's Boot Guard functionality only authenticates a small portion of the BIOS (the BIOS bootblock). Implementations may or may not extend to authenticate the entire BIOS. Boot Guard does not authenticate BMC firmware. Also, Intel's Boot Guard presents supply chain challenges as it requires permanently installing the hashes and public keys in the factories manufacturing the system to achieve a hardware root of trust. Some server manufactures leverage this technology to create a root of trust, but unfortunately, it still leaves the BMC firmware exposed to compromise.

[4] http://www.uefi.org/sites/default/files/resources/UEFI%20RoT%20white%20paper_Final%208%208%2016%20%28003%29.pdf

AMD Secure Root-of-Trust Technology

AMD Secure Root of Trust Technology stems from its roots in the gaming console market. As the processor technology behind Microsoft Xbox and Sony PlayStation, AMD implemented AMD Secure Root of Trust Technology to prevent the hacking and pirating of gaming consoles. AMD claims absolute integrity of Secure Root of Trust Technology through millions of gaming consoles shipped worldwide.

Built on ARM TrustZone, Secure Root of Trust Technology is an AMD-specific form of RoT that roots the trust to hardware in a Platform Security Processor (PSP) on-chip ROM and verifies the integrity of the system BIOS. The PSP ROM contains the initial "impenetrable" PSP code. The PSP ROM validates a secure boot key and then uses the key to validate the PSP firmware, which it reads from system flash. The PSP firmware loads and starts the system application execution.

With Secure Root of Trust Technology, OEMs can choose whether the PSP validates the BIOS platform-initialization code. The PSP then initiates BIOS execution. The PSP completes its own initialization and enters steady state while the BIOS and OS finish booting on x86[5].

It is important to note that none of these technologies have been designed to provide runtime protection. They only attempt to provide the boot-time protection when the server is started, ensuring that the code loaded validated only one time.

Additionally, one should consider that each RoT implementation discussed requires proper implementation of key management to a third party. The supply chain of trust is interrupted, adding a weakness and opportunity for potential exploitation.

HPE Silicon Root of Trust[6]

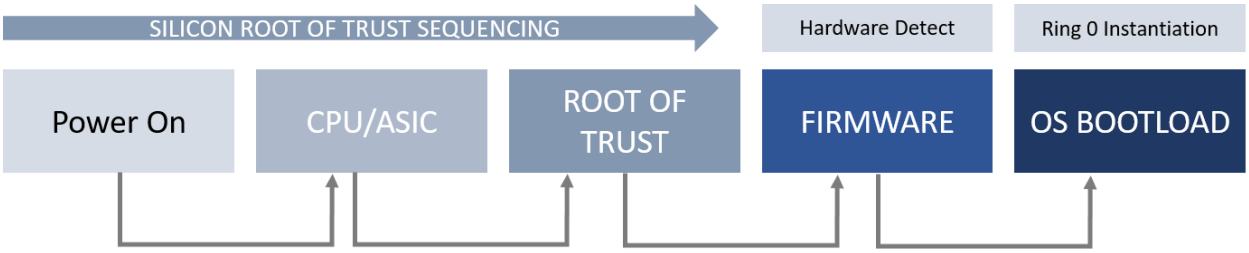
HPE's silicon root of trust provides protection as soon as the server is powered on and the iLO firmware comes alive. As it initiates, HPE Integrated Lights-Out 5 (iLO 5) firmware looks into the silicon for the immutable fingerprint that verifies all the firmware code is valid and uncompromised. More than a million lines of firmware code run before the operating system starts, making it essential to confirm that all server essential firmware is free from malware or compromised code.

[5] http://academlib.com/24869/computer_science/secure_technology

[6] https://www.hpe.com/h20195/v2/getpdf.aspx/a00008520enw.pdf?ver=1, pp 10, 11

The iLO 5 (HPE's BMC) startup code includes a cryptographic algorithm (hash) that is permanently burned into the silicon. The silicon validates the iLO 5 firmware code before it is fetched and executed. If any malware or compromised code has been inserted in the iLO 5 firmware, the silicon will detect it because any infected firmware code would be altered and, therefore, not match-up with the hash burned into the silicon. From there, the iLO 5 firmware validates the rest of the server firmware, namely the UEFI, SPLD, IE and ME. The UEFI then validates the connection to the operating system through secure boot, thus completing a complete root, or chain, that is anchored into the silicon.

FIGURE 3: HPE SILICON ROOT OF TRUST



Source: Moor Insights & Strategy

During operation of the server, HPE has a new technology that conducts run-time firmware verification to check the firmware stored in the server. At any point during operation, if compromised code or malware is inserted in any of the critical firmware, an iLO audit log alert is created to notify the customer that a compromise has occurred. This is achieved by storing iLO 5 and UEFI firmware in non-volatile Flash memory which is thoroughly scanned at regular user-determined intervals. The contents of the firmware stored in memory must be exactly right down to the individual bit or else it is flagged as compromised and a recovery will begin.

In the unlikely event of a breach into the HPE server firmware, after detection has been completed, the customer may then securely recover the firmware automatically to a previous known good state. HPE provides this function through a new HPE license called HPE iLO Advanced Premium Security Edition.

ROOT OF TRUST IS THE FOUNDATIONAL BUILDING BLOCK TO A SECURE DATACENTER

An IT organization's comprehensive plan includes deploying technology from vendors that locks down the entire datacenter, including devices in the wild, the perimeter and infrastructure that houses, computes and manipulates data. And that data, whether at rest, in use or in flight, requires the highest levels of protection. Absolute protection has to begin with absolute trust. And absolute trust must be rooted in RoT implementations that protect against the lowest level of attacks.

Because nearly 25 percent of users in an enterprise have responded to phishing attacks, it is likely that security breaches will occur in every IT organization. While RoT cannot prevent these breaches, technologies like HPE's Silicon Root of Trust can expedite a recovery process. Additionally, a comprehensive approach to security can identify, mitigate and remove threats. And the best RoT implementations can assure that your servers detect issues quickly and are restored to a known good state. Meaning, malware that was introduced into the environment has not injected firmware, bootkit, rootkit or other attacks.

ORCHESTRATION CLOSES GAPS

Root of Trust technologies like HPE Silicon Root of Trust work best when deployed in conjunction with solutions that address security throughout the enterprise – further up the server stack, across devices, on the edge and server-to-server. To assure tightest levels of integration and coordination, an IT organization should look to minimize their reliance on too many point solutions from too many suppliers. Rather, look to leverage the breadth of offerings from known vendors with deep roots.

TRUST THE EXPERTS

All organizations can benefit from an outside firm that looks at the direct and indirect impacts of digital transformation and cloud adoption. From workload and data migration to security and beyond, a successful strategy is driven by a fully considered plan coupled with flawless execution.

The complexity of today's threats requires organizations to be complete in deploying and orchestrating security across the enterprise. When possible, work with organizations that have technical depth in assessing, developing and deploying

enterprise security solutions. Organizations like HPE's Pointnext exist to help companies deal with the challenges of today and tomorrow.

MI&S Perspective

Enterprise security is a complex topic that becomes increasingly complex as the IoT landscape explodes and cloud adoption continues to skyrocket. Threats will continue to attack more surfaces across more vectors. The tools and resources available to aid in hacking have created a virtual cybercrime supply chain.

Highly publicized ransomware events and other forms of exploitation continue to dominate headlines. As a result, IT organizations must focus on protecting the perimeter and ensuring efficacy at identifying, isolating and removing threats based on policies and technologies that may not be equipped to meet these new threats.

The best security strategy begins at the cornerstone of the cloud – the server. And the best server security begins at its cornerstone – the silicon. When considering server deployments, strongly consider adopting platforms that assure the greatest level of security at the lowest levels of the platform.

When servers are protected at the BMC silicon level, IT managers should deploy complementary security technologies that can integrate from silicon to server to perimeter and to the edge. This comprehensive approach will allow for greater coordination across the enterprise and allow for the most complete response to any threat.

HPE is unique because its security portfolio is perhaps the most comprehensive when combining HPE Silicon Root of Trust with solutions such as Aruba ClearPass Policy Manager and Niara (Niara, a recent purchase of HPE, integrates the HPE security portfolio and drives advanced analysis on devices, users and systems).

Organizations of all sizes would be well served to engage companies like HPE to assist in the consulting, designing and deployment of security policies and strategies that will protect them today and prepare them for tomorrow.

CALL TO ACTION

IT organizations of all sizes should review and recommit to developing and executing a sound cyber security strategy. Even those organizations that feel fully prepared should engage with experts and ensure technologies in the datacenter enjoy full protection against exploitation.

IT decision makers should consider deploying technologies from companies that have the most comprehensive portfolios of security technologies. Likewise, IT decision makers should consider a company like HPE that can secure datacenter environments from the server to the perimeter.

Develop holistic and living plans that are refreshed with regular cadence. As the threat landscape becomes more complex, security plans and associated technologies will soon fall out of date.

IMPORTANT INFORMATION ABOUT THIS PAPER

AUTHOR
Matthew Kimball, Senior Analyst at Moor Insights & Strategy

PUBLISHER
Patrick Moorhead, Founder, President, & Principal Analyst at Moor Insights & Strategy

INQUIRIES
Contact us if you would like to discuss this report, and Moor Insights & Strategy will respond promptly.

CITATIONS
This paper can be cited by accredited press and analysts but must be cited in-context, displaying author's name, author's title, and "Moor Insights & Strategy". Non-press and non-analysts must receive prior written permission by Moor Insights & Strategy for any citations.

LICENSING

DISCLOSURES
This paper was commissioned by Hewlett Packard Enterprise. Moor Insights & Strategy provides research, analysis, advising, and consulting to many high-tech companies mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

DISCLAIMER
The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. Moor Insights & Strategy disclaims all warranties as to the accuracy, completeness, or adequacy of such information and shall have no liability for errors, omissions, or inadequacies in such information. This document consists of the opinions of Moor Insights & Strategy and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice.

Moor Insights & Strategy provides forecasts and forward-looking statements as directional indicators and not as precise predictions of future events. While our forecasts and forward-looking statements represent our current judgment on what the future holds, they are subject to risks and uncertainties that could cause actual results to differ materially. You are cautioned not to place undue reliance on these forecasts and forward-looking statements, which reflect our opinions only as of the date of publication for this document. Please keep in mind that we are not obligating ourselves to revise or publicly release the results of any revision to these forecasts and forward-looking statements considering new information or future events.