Research

November 2020

The Critical Need for Confidential Computing

A J.Gold Associates Research Report

"This report will discuss what is needed for enterprises to move towards a Confidential Computing model. It will identify why this is such an important methodology, how it will benefit organizations in limiting costly data breaches, and how it will create a more effective security posture for cloud, data center and personal computing environments...."

Contents

Introduction **2**

What is Confidential Computing? **3**

Figure 1: The Three States of Data Protection 3

Why CC is Different *4*

Figure 2: Typical Computing Stack 4

Figure 3: Multiple Computer Attack Surfaces 5

Figure 4: Typical Computing Stack with Protected Enclave for Confidential Computing 5

Why now? *6*

The Negative Effects of Data Breaches **6**

The High Cost of Data Breaches *6*

Table 1: The Average Cost of a Data Breach: US and Worldwide 7

The Time Needed to Identify a Data Breach is Costly and Growing *7*

Where Are Attacks Coming From? *7*

Security vs. Privacy **7**

Limiting Virtual Machine Vulnerabilities **8**

What's Needed in Compute Engines **9**

CC: What's Inside? *9*

Migrating to CC Systems *9*

Examples of CC Benefits Beyond Security **10**

What Should Companies Do to Deploy CC? **10**

Recommendations **11**

Conclusions **11**

Introduction

Enterprises and smaller businesses are very aware that bad actors are targeting their data to harvest information for financial gain, intellectual property exploitation, or simply to cause damage. Indeed, security remains one of the top 3 concerns for nearly every organization large or small. This is true both for those using primarily in-house data centers as well as those who have migrated to the cloud, regardless if using fully public cloud deployment or a hybrid/multi-cloud approach. And the use of virtualization techniques so prevalent in today's computing environment has further complicated security efforts by potentially exposing "cross VM" security risks running on the same piece of physical hardware.

> ***TREND**: Security remains an ongoing battle with companies trying to keep ahead of bad actors. And while there are many security techniques and products available, one area that has been lacking is in protecting data while it's being processed in computing systems, exposing it to possible data breaches. The Confidential Computing initiative aims to mitigate this risk. We expect that Confidential Computing will become a major component of all computing, whether for cloud, local data center servers, personal computing devices or even mobile devices. In the next 2-3 years, companies must adopt a confidential computing first model that specifies those capabilities across all cloud deployments, as well as procure computing equipment for corporate and/or personal use.*
> *J.Gold Associates LLC.*

Application solution providers, hardware suppliers and cloud players have all gone to great lengths to enhance security, and have made significant improvements over the years. But most efforts at security in the past have centered on protecting data at rest or in transit through application of data encryption and/or transmission encryption techniques. Indeed, encryption of data while in a database, over a LAN/WAN or moving through a 5G network, is a key component of nearly every such system. And for the most part, this has been a successful protection strategy.

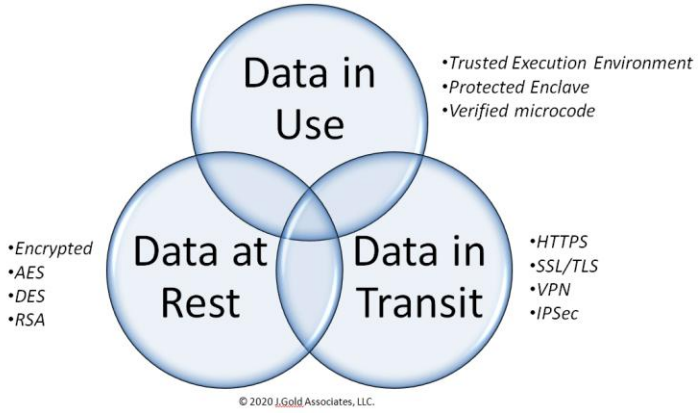
Today, nearly every computing system, even personal devices like smartphones, have data encryption built in, often enhanced by specialized compute engines built into the processor chips. But one area that has been relatively ignored is the ability of all of this encryption to be defeated if a bad actor can access the device hardware through either a malicious app, corrupted "root access" to the OS or other means. Encrypted data needs to be in the clear when being processed, and this is a real vulnerability. If you can get to the machine memory at this point, all data is available for easy viewing/copying. And if you can exploit this vulnerability on a machine running multiple virtual machines, then the amount of data available is greatly multiplied. Eliminating this risk is the vision of Confidential Computing.

This report will discuss what is needed for enterprises to move towards a Confidential Computing model. It will identify why this is such an important methodology, how it will benefit organizations in limiting costly data breaches, and how it will create a more effective security posture for cloud, data center and personal computing environments.

What is Confidential Computing?

In 2019, the Linux Foundation launched the Confidential Computing Consortium. Its stated goal is to define standards for Confidential Computing (CC) as well as support and propagate the development of open-source CC tools and frameworks. Members include Alibaba, AMD, Arm, Facebook, Fortanix, Google, Huawei, IBM (Red Hat), Intel, Microsoft, Oracle, Swisscom, Tencent, and VMware. More companies are expected to join this preliminary list of members.

Figure 1: The Three States of Data Protection



The foundation has stated in its goals that: “The Consortium is concentrating on the area of “data in use,” with the confidentiality of “data in transit” and “data at rest” as outside the scope of the Consortium. Contributions to the Confidential Computing Consortium, according to their website, already include:

- Software Guard Extensions (Intel SGX) SDK, designed to help application developers protect select code and data from disclosure or modification at the hardware layer using protected enclaves in memory.
- Open Enclave SDK, an open source framework that allows developers to build Trusted Execution Environment (TEE) applications using a single enclaving abstraction. Developers can build applications once that run across multiple TEE architectures.
- Enarx, a project providing hardware independence for securing applications using TEEs.

As this is an ongoing “work in progress” standardization effort, there will likely be many more projects that come up in the future. Each of the member companies has approached the notion of CC from their own expertise and perspective (e.g., chips, machines, interconnects, apps). This diversity is important to the growth of CC, but while several companies already have proprietary tools available, it’s likely that in the future they will become integrated into a more open sourced framework for CC, given the Linux Foundation background. Longer term, standardization will be critical both for the acceptance and the growth of CC.

Why CC is Different

Unlike data encryption that cryptographically protects individual sets of data, CC uses hardware-based functions to create a "Trusted Execution Environment" (TEE), also referred to as an "Enclave", which is used for the manipulation of data, specific computational functions, or executing an entire application. CC isolates this vaulted area from access by the OS, hypervisor, VMs, or other applications, and thus protects against any potential for cross contamination as no one can gain access that isn't assigned to this TEE. Any attempt to alter the app code or tamper with the data will be prevented. This is especially critical in multi-user systems, such as virtualized and public cloud systems, where cross contamination of data is a real risk. Indeed, some potential users of public cloud compute have resisted moving for this specific reason. While a slight chance of a "side channel" attack against CC is theoretically possible if you have physical access to the system hardware, sophisticated toolsets, and significant time to mount the attack, the risk is trivial compared to the potential risks associated with non-CC systems.

Figure 2 demonstrates the major blocks of a typical computing stack, while Figure 3 indicates the potential attack surfaces. There is no shortage of possible attack surfaces in the modern computer and indeed, malware creators have been very creative at exploiting them. What's needed is a way to reduce, as much as is possible, the many attack surfaces.

Figure 2: Typical Computing Stack

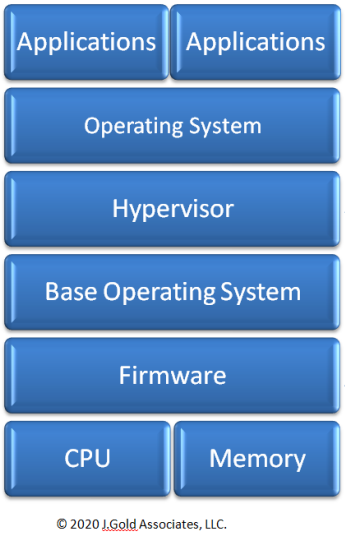
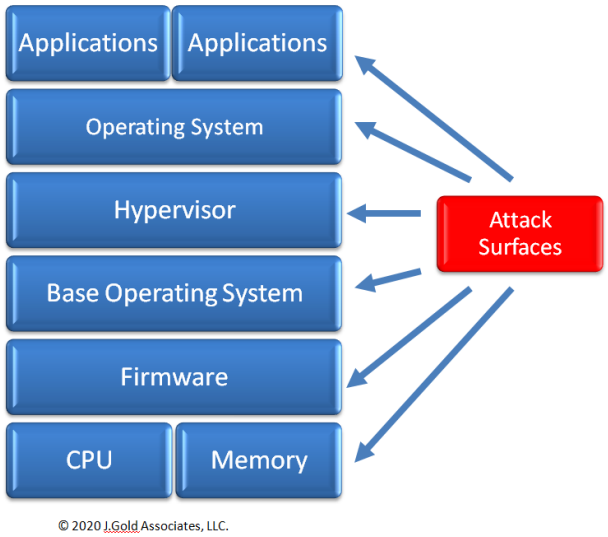
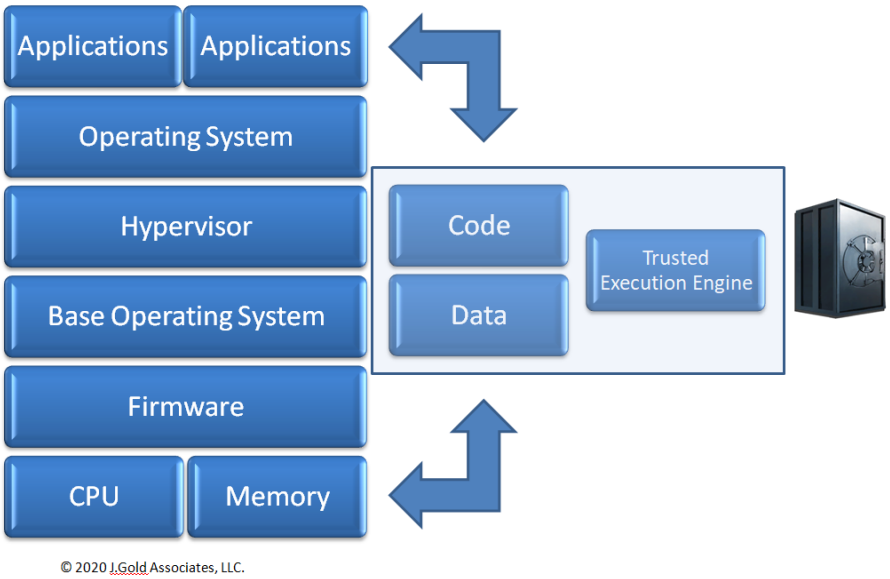


Figure 3: Multiple Computer Attack Surfaces



What CC is designed to do is bypass many of the potential attack surfaces by placing any discoverable data or code in current use into a hardened "vault" where processing can safely take place. Figure 4 indicates the general flow of this process at a high level. Of course there is more to it than just this simple diagram, but it does provide a conceptual view of the process. It is this concept of protection that CC provides.

Figure 4: Typical Computing Stack with Protected Enclave for Confidential Computing



Why now?

With the move to a more virtualized processing environment, especially in the public and private cloud, enterprises need a way to ensure that a failure of one virtual machine, often being operated for the benefit of a totally different organization, can't find its way to compromising my organization's processes and data. But this is not just an issue for cloud based compute. Traditional on premises data center computing systems are subject to similar vulnerabilities. And this is not just as issue with data center servers. Even PCs and mobile devices ultimately need to create a CC environment to prevent data breaches. While some devices have implemented more rudimentary TEE capabilities, most personal and mobile devices will likely implement a more complete CC capability at a later date, given the current emphasis being placed on the more critical nature of large data processing systems.

In the past, trying to build out a CC capability in processing systems created a bottleneck that had a major impact on performance. This is a non-starter for many computing applications that require high performance and/or throughput. With the new generation of CC systems now coming to market, this challenge has been met, and we expect that the impact is well below 5% in processing overhead and/or performance degradation. For most processing, this is a trivial and an acceptable amount of performance degradation, especially given the positive security benefits obtained by a CC approach.

Finally, with all the emphasis on the Zero Trust computing model – which assumes no user or device should ultimately be trusted as being safe and any systems allowing access must take measures to assume the worst possibilities – computing system have had to expand their security requirements to maintaining a trusted environment for data security. CC has an important role to play in this zero trust model, as it provides a mechanism for isolating any potentially "bad actors" from gaining access to an entire computing center by only allowing them to process selected data within a CC enclave.

The Negative Effects of Data Breaches

The cost of a data breach has been increasing for most organizations – not only in terms of dollars, but also in terms of negative impact on customer views of the organization's credibility. Further, in many countries and/or individual states, any data breaches run afoul of the increasing number of regulations being imposed that can not only result in fines, but also potentially in criminal penalties. As a result, any approach that improves overall security to eliminate data breaches is not only welcome, but can have a positive effect both in money saved and reputation protected.

The High Cost of Data Breaches

While each data breach is unique and has its own costs associated with finding the breach, remediating the breach and paying any potential penalties, there is information available that assesses the average costs of breaches from many companies' actual results. According to the IBM/Ponemon Cost of a Data Breach Report 2020, the average worldwide cost of a data breach was $3.86M, while the US had the highest cost at $8.64M, as shown in Table 1

Table 1: The Average Cost of a Data Breach: US and Worldwide

	USA	Worldwide
Average Cost of Data Breach	$8.64M	$3.86M

The Time Needed to Identify a Data Breach is Costly and Growing

Even with all of the security tools like SIEM available to organizations, the time to identify a data breach is growing. The IBM/Ponemon Cost of Data Breach Report 2020 indicated that the average time to identify and contain a data breach was 280 days in 2020. Further;

- The potential savings of containing a data breach in less than 200 days was $1M, compared to containment in more than 200 days.
- And the report found that 39% of the total cost of the data breach was spent more than a 1 year after the breach occurred.

Where Are Attacks Coming From?

There is a wide array of attack methodologies and attack vectors that are being used to compromise computer systems. According to the IBM/Ponemon Cost of Data Breach Report 2020;

- The share of data breaches as a function of malicious attacks were 59% of the total breaches,
- Compromised credentials accounted for 19% of attacks.

And according to the 2020 Verizon Data Breach Investigations Report;

- 45% of all data breaches were caused by Hacking, while 22% were caused by errors, and 17% involved Malware.
- 72% of breaches were in large businesses, while 28% involved small businesses.
- 70% of breaches were caused by external actors, while 30% were caused by internal actors.
- Physical breaches are only about 5% of the total, but once a bad actor has access to devices, they can be highly damaging.
- Servers account for approximately 70% of breaches, which makes sense given the massive move to web based applications and “as a service” offerings, while personal devices account for approximately 25% of breaches.
- By industry, Healthcare led with the greatest number of breaches, followed by Financial, Professional, Information and Public verticals.

These results show that data breaches are a persistent and even growing problem for organizations large and small, and need to be addressed urgently.

Security vs. Privacy

While related, security and privacy are not the same when it comes to the utilization of organizational data assets. Indeed, enterprises have added many layers of security products at all levels of their infrastructure to insure they are keeping their data assets secure. But with the current growing emphasis on keeping data private, especially personal data used internally or sensitive data to be shared across corporate boundaries, it’s become

increasingly important for companies to insure data is kept private and not just secure. Indeed, nearly every state, country and governmental agency has implemented regulations to protect privacy, and any breach of such privacy can have major business impacts and may also have significant financial and legal penalties.

One of the basic advantages of using a CC approach to data processing is that by using its capabilities to isolate and protect data from exposure, CC can dramatically add to the privacy of processing data assets in both internal and shared environments. This has major benefits for nearly all industries, but especially in highly regulated industries like banking, finance, insurance, healthcare, public sector, education, etc.

Limiting Virtual Machine Vulnerabilities

Nearly all cloud implementations, and especially public cloud implementations, rely on having multiple virtual machines running on the hardware assets deployed within the cloud data center. Certainly, the use of VMs will continue to grow. VMs are what allow the cost of computing to be shared across many different customers, keeping the overall cost down and allowing for flexible load sharing to create true compute as a commodity, which is what drives the cloud.

But securing the individual VMs, where much of the security efforts are targeted, is not sufficient. An individual VM can only be as secure as the hardware and OS/Hypervisor it is running on. If either of these base platforms is compromised, the VM, indeed all of the VMs running on that system, are also compromised. So it's critical to make sure that the base level hardware as well as the hypervisor be fully protected separately from the individual VM. To add to this security, CC protects the VM or application from a malicious OS or hypervisor, so that an exploited or hacked OS still cannot access the data.

VMs also can have some specific types of vulnerabilities. Some specific VM vulnerabilities include:

- VMs "ghosts" – VMs deployed with a lack of control and policies, so that many VMs stay in play even if they are not in use.
- Hyperjacking – by taking control of the hypervisor mechanisms in the machine, hackers could gain access to all of the various VMs and their features, data sets and controls
- VM breakouts – where a guest OS can subvert the VM and allow access directly to the machine hardware assets
- Poor VM Isolation – keeping each VM fully isolated from the others running on the host is critical to preventing cross VM contamination and avoiding breaches.
- Host vulnerabilities – a VM can only be as secure as the host it's running on, so ensuring the host does not continue to any potential security risks is critical.

To minimize risks and maximize security, CC can go a long way to ensuring that VMs remain isolated, independent and secure.

What's Needed in Compute Engines

To enhance machine security and comply with many of the Confidential Computing concepts, machines should include at a minimum:

- A method for execution environments to interact directly with CPU and memory by bypassing the host OS
- Full memory encryption to protect against certain types of physical attacks such as memory freezing and DIMM removal
- Cryptographic acceleration to minimize the impact of encodes and decodes. Any impact from such operations should be kept to a minimum – preferably in the 2-5% range
- Firmware resilience to prevent breaking into the machine through a firmware exploit that allows detection and correction of any firmware intrusions

CC: What's Inside?

Trusted Execution Environments are the key to making CC work. We've had TEE's for some time, including on Arm based chips (Trust Zone) as well as in x86 chips (e.g., Intel SGX). Indeed, early versions of this concept go back more than a decade to TPM modules that were available in many PCs. The difference with modern versions of TEE is that they are built into the core of the processors and capable of running full general-purpose workloads rather than just protected storage for small, special-purpose functions.

Despite the fact that we've had TEE enabled systems available for some time, few enterprises have sought to use them, and many app providers don't support them either. The reason is that they have always been difficult to implement and you needed specific code enabled with the app to enforce the use of a TEE environment. Further, TEEs were not universally available on all processors (e.g. some Intel Xeon chips support SGX and some don't), nor were TEEs compatible across chip families. The result is that many organizations have not yet implemented what could be a very important security method.

Migrating to CC Systems

With a move to off premises and multi-tenant cloud computing, there is now a greater need to protect the processing integrity of customer data, as well as protect certain proprietary algorithms running in the processes. As a result cloud providers (e.g., Azure Confidential Computing & IBM Confidential Computing) are making it easy to spin up new CC instances for customers to utilize. This eliminates the need for organizations to have their own CC enabled systems to run in. This is a win-win situation, as the customers get what they need to protect their data assets, and cloud providers bring in the necessary hardware assets that customers don't necessarily own themselves. This new availability is being brought about by an increasing number of cloud-based processors that include the CC concepts built in. And as cloud providers generally obtain new high end processing capability early in their stages of availability, this makes access for the user community much more rapid than if they had to acquire it on their own. Further, it enables app providers to quickly design CC into their

product given the availability of hardware and toolkits running in the cloud, and further, allows them to have a more ready market to recover their development investment.

Examples of CC Benefits Beyond Security

CC can bring many benefits to companies deploying systems that are more secure and confidential in their computing models. Indeed, CC not only protects against data breaches, but can also be used as an "isolation ward" for shared data that needs to be kept confidential, opening up new methods for cross-organizational sharing of private data. In such situations, the increased usage of ML/AI, Multi-Party Compute, Multi-Party Analysis, shared data sets and federated learning can all benefit. For example;

- In Healthcare ML/AI provides collaborative learning models between hospitals on early disease detection or fast-track drug trials, while maintaining isolation and privacy of personal data
- In Education in a Healthcare setting, a research hospital may have several thousand archived brain scans for training a tumor diagnostic model. The model's accuracy and scope could be increased using scans from other hospitals, but privacy regulations may not allow for sharing the data. With a confidential computing approach, many hospitals can contribute their data to train the model, while maintaining patient privacy from the other institutions.
- In Banking and Finance, multi-party compute and federated learning are agreed upon analytics on the combined, sensitive dataset, providing insights without giving access. This allows financial institutions to meet confidentiality requirements, while benefiting through increased fraud detection rates, reduced false positives, and providing Iterative learning.
- Government exchanges of personally identifiable information (PII) and intellectual property (IP) must be protected even when shared across branches of government, such as Healthcare agencies (Medicare & Medicaid) at both the state and federal levels (State ID's, Voter ID's, birth & death records, and national cancer registries to name a few).

What Should Companies Do to Deploy CC?

The concepts behind CC are not new, but the availability of CC in the cloud makes it much more attractive to deploy by organizations that need to secure their data from application vulnerabilities and/or privacy. Most on-premises servers stay in service for 3-5 years or more. So upgrading to CC on premises will likely take a long time as installed machines can't be easily upgraded to CC, and replacements need to be acquired after the useful life of current equipment expires. The best path to quickly implement CC is to employ public cloud based services that can provide any run time services on up to date Confidential Computing compliant hardware. In today's competitive cloud environment, this should be easy to accomplish as all the major cloud players offer such capabilities as part of their offerings. If yours doesn't, you should ask when they will, or consider switching to a cloud provider that does.

Recommendations

We strongly recommend organizations take the following actions:

- Enterprises must explore the use of Confidential Computing techniques in the next 6-12 months, and specify to their key application solution providers that they expect them to comply with the CC strategy and offer technology implementations within the same time period.
- Require that any public cloud computing implementations specify that the cloud instance provided be conformant with Confidential Computing standards.
- Determine the replacement cycle for on premises data center equipment and specify that any new machines purchased be Confidential Computing complaint.
- Long term, explore a strategy that includes migrating to a Confidential Computing capability for all internal devices deployed not just at the server, but also for personal computing and mobile devices.
- Those organizations that fail to adopt a Confidential Computing model for deploying security enhancing computing systems will be at significantly increase risk of having a security breach that can easily cost the company millions of dollars as well as being highly disruptive its operations, or even cause a business failure.

Conclusions

We recommend that enterprises explore the use of CC techniques in the next 6-12 months, and specify to their key application solution providers that they expect them to comply with the CC strategy and offer technology implementations within the same time period. CC can significantly enhance enterprise security by virtually eliminating the ability of data in use to be exploited. While there is no 100% sure thing when it comes to security, CC is a major step forward and should be implemented whenever possible, particularly for those organizations deploying applications in the cloud. We expect CC to become the defacto standard approach to compute, especially in the cloud, within the next 1-2 years.

About J.Gold Associates

J.Gold Associates provides insightful, meaningful and actionable analysis of trends and opportunities in the computer and technology industries. We offer a broad based knowledge of the technology landscape, and bring that expertise to bear in our work. J.Gold Associates provides strategic consulting, syndicated research and advisory services, and in-context analysis to help its clients make important technology choices and to enable improved product deployment decisions and go to market strategies.

J.Gold Associates, LLC
6 Valentine Road
Northborough, MA 01532 USA
+1 508 393 5294
www.jgoldassociates.com