ZERO TRUST E IL CLOUD

Come cambia l'approccio al Cloud e all'Edge computing

Autore: Aldo Pedico – Enterprise Security & Privacy Architect

Contatto: pedicoaldo@gmail.com

1. PREMESSA

Ho voluto in questo articolo evidenziare come i principi Zero Trust possano trovare un'applicazione pratica ed ho ritenuto opportuno condividere delle "pillole" che possano essere di interesse.

Tali «pillole» possono essere oggetto di approfondimento nel manuale "Draft NIST IR 8320 Hardware-Enabled Security: Enabling a Layered Approach to Platform Security for Cloud and Edge Computing Use Cases".

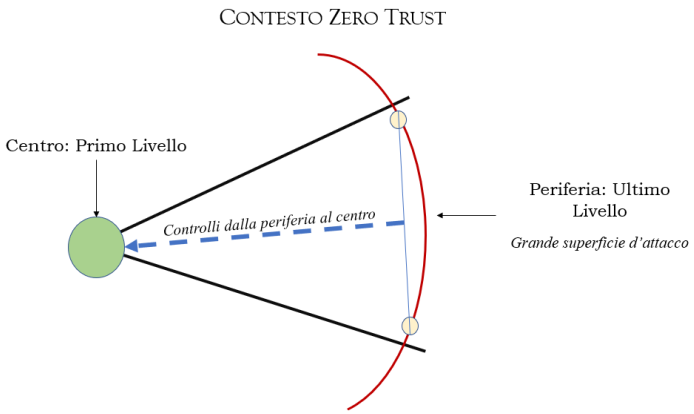
Considerando che siamo inesorabilmente parte integrante dei sistemi digitalizzati, perché essi svolgono un ruolo essenziale nella società e sono diventati i pilastri della crescita economica (vedi considerando 1, 8, 13 del Reg. (UE) 2019/881 Cybersecurity Act), è indispensabile che la cibersicurezza sia uno dei pilastri fondamentali per la realizzazione di un Sistema Qualità e la cui gestione rientri sia nel comportamento sia nella mentalità umana. Di conseguenza, come dichiarato nel considerando 8 del Reg. (UE) 2019/881: "... è opportuno promuovere energicamente l'«igiene informatica», vale a dire semplici misure di routine che, se attuate e svolte regolarmente da cittadini, organizzazioni e imprese, riducono al minimo la loro esposizione a rischi derivanti da minacce informatiche".

In riferimento all'argomento di questo articolo, il considerando 15 del Reg. (UE) 2019/881 cita la direttiva (UE) 2016/1148 in cui i gestori dei servizi di cloud computing hanno l'obbligo di adottare tutte le misure di sicurezza e le notifiche degli incidenti. Ed è logico pensare che la Protezione dei Dati Personali (Reg. (UE) 2016/679) sia parte integrante della Cibersicurezza.

Zero Trust, annunciato dal NIST in bozza agli inizi del 2020 e pubblicato a giugno dello stesso anno (vedi NIST SP 800-207), presenta l'inizio di una concretizzazione dei suoi principi nel Draft NIST IR 8320 di Maggio 2021 con informazioni atte a realizzare nel contesto Cloud quanto enunciato precedentemente.

ZT richiede che la concentrazione della sicurezza si trasferisca dalla periferia al centro, ovvero dalla rete, sempre più difficile da gestire, verso controlli interni alla Unità Centrale di Elaborazione o alla sua componente hardware.

Di seguito riporto un mio schema che non ha la presunzione di voler rendere chiaro il principio di trasferimento delle garanzie di sicurezza dalla periferia (rete) al centro (HW) ma di dare un piccolo contributo dei principi ZT.



2. ANALISI DEL CONTESTO ATTUALE

Sempre di più assistiamo al trasferimento dei carichi di lavoro delle aziende negli odierni data center cloud e edge computing, le cui superfici di attacco sono notevolmente aumentate, l'hacking è diventato industrializzato e la maggior parte delle implementazioni di controllo della sicurezza non sono coerenti. Ciò implica che la base di qualsiasi strategia di sicurezza per data center o edge computing dovrebbe essere la protezione della piattaforma su cui saranno eseguiti i processi.

La piattaforma fisica rappresenta il primo livello per qualsiasi approccio alla sicurezza a più livelli e fornisce le protezioni iniziali per garantire l'affidabilità dei controlli di sicurezza al livello superiore (verso la periferia).

Nei data center cloud e nell'edge computing odierni, ci sono tre forze principali che incidono sulla sicurezza:

1. *l'introduzione di miliardi di dispositivi connessi e la maggiore adozione del cloud hanno aumentato significativamente le superfici di attacco;*
2. *l'hacking è diventato industrializzato con tecniche sofisticate e in continua evoluzione per compromettere i dati; e*
3. *le soluzioni composte da più tecnologie di diversi fornitori comportano una mancanza di implementazioni coerenti e coerenti controlli di sicurezza.*

Date queste forze d'attacco, la base per una strategia di sicurezza dovrebbe avere un approccio consolidato per proteggere in modo completo l'intera piattaforma hardware su cui sono gestiti i carichi di lavoro.

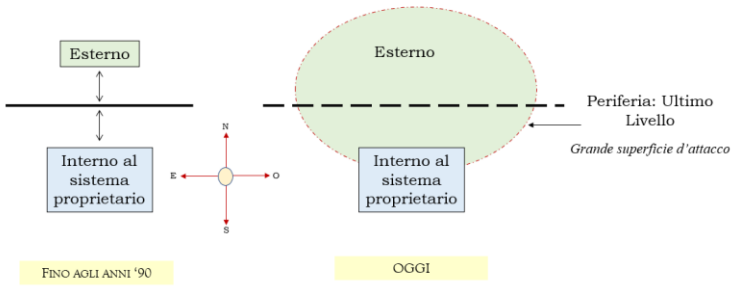
La piattaforma hardware rappresenta la prima parte dell'approccio alla sicurezza a più livelli.

La sicurezza abilitata per l'hardware può fornire una base più solida di quella offerta dal software o dal firmware, che può essere modificata con relativa facilità.

La radice di attendibilità hardware presenta una superficie di attacco più piccola a causa della piccola base di codice.

Di seguito, riporto un mio schema che dovrebbe permettere di comprendere l'evoluzione delle interconnessioni della comunicazione tra un sistema infrastrutturale proprietario e la rete o il mondo esterno.

SCHEMA DEL SISTEMA D'INTERCONNESSIONE TRA L'INFRASTRUTTURA INTERNA E LA RETE ESTERNA



3. RISCHI

Con una maggiore attenzione applicata alla sicurezza del software di alto livello, gli aggressori stanno spingendo più in basso nello stack della piattaforma, costringendo gli amministratori della sicurezza ad affrontare una serie di attacchi che minacciano il firmware e l'hardware della piattaforma.

Queste minacce possono comportare:

- ✓ *Accesso non autorizzato e potenziale estrazione di dati sensibili della piattaforma o dell'utente, incluso l'accesso fisico diretto a Moduli di Memoria Doppi In Linea (Dual In-line Memory Modules DIMM);*
- ✓ *Modifica del firmware della piattaforma, come quello appartenente all'Unified Extensible Firmware Interface (UEFI)/Basic Input Output System (BIOS), Board Management Controller (BMC), Manageability Engine (ME), dispositivo Peripheral Component Interconnect Express (PCIE), e varie carte acceleratrici;*
- ✓ *Intercettazione della catena di approvvigionamento attraverso la sostituzione fisica di firmware o hardware con versioni dannose;*
- ✓ *Accesso ai dati o esecuzione di codice al di fuori dei confini geopolitici regolamentati o di altro tipo;*
- ✓ *Elusione dei meccanismi di sicurezza basati su software e/o firmware.*

Questi attacchi possono essere devastanti per gli ambienti cloud perché spesso richiedono ricostruzioni o sostituzioni server per server, che possono richiedere settimane.

Questo genere di attacchi stanno aumentando man mano che gli aggressori diventano più sofisticati e i carichi di lavoro soggetti a normative specifiche o contenenti dati sensibili presentano ulteriori sfide di sicurezza per i cloud.

Sebbene la virtualizzazione avvantaggia notevolmente l'efficienza, l'adattabilità e la scalabilità, queste tecnologie consolidano i carichi di lavoro su un numero inferiore di piattaforme fisiche e introducono la migrazione o proliferazione dinamica di carichi di lavoro e dati tra piattaforme, aumentando la superficie d'attacco.

Ne consegue una perdita di visibilità e controllo da parte dei consumatori sulle piattaforme che ospitano carichi di lavoro e dati virtualizzati e introduce l'utilizzo di amministratori di infrastrutture di terze parti.

La virtualizzazione simula l'hardware su cui sono eseguiti più carichi di lavoro cloud.

Ogni carico di lavoro è isolato dagli altri in modo da avere accesso solo alle proprie risorse e ogni carico di lavoro può essere completamente incapsulato per la portabilità.

Le macchine virtuali convenzionali (Virtual Machine - VM) hanno uno spazio kernel isolato che esegue tutti gli aspetti di un carico di lavoro insieme al kernel.

Oggi, l'ambiente virtualizzato è stato esteso per includere ***container*** *e motori di orchestrazione dei carichi di lavoro completi.*

I ***container*** *offrono la portabilità delle applicazioni condividendo un kernel sottostante, che riduce drasticamente le risorse consumate dal carico di lavoro e aumenta le prestazioni.*

Sebbene i ***container*** *possano fornire un livello di praticità, le vulnerabilità nello spazio del kernel e nei livelli condivisi possono essere suscettibili di uno sfruttamento diffuso, rendendo ancora più importante la sicurezza per la piattaforma sottostante.*

I fornitori di servizi cloud e gli utenti che adottano il cloud seguono un modello di responsabilità condivisa, in cui ciascuna parte ha la responsabilità di diversi aspetti dell'implementazione complessiva.

Inoltre, ad aggravare (teoricamente) i fornitori di servizi cloud spesso dispongono di data center che si estendono su più confini geopolitici, sottoponendo i proprietari del carico di lavoro a complicati requisiti di conformità legale e normativa di più paesi.

Le architetture cloud ibride, in particolare, utilizzano più provider di infrastrutture, ciascuno con le proprie configurazioni e gestioni dell'infrastruttura, senza il controllo fisico o la visibilità sulle

configurazioni della piattaforma, senza le migliori pratiche di sicurezza convenzionali ed, oltretutto, con requisiti normativi difficili o impossibili da implementare.

Con nuove strutture normative come il Regolamento europeo sulla protezione dei dati (GDPR) che introduce multe elevate per non conformità, avere visibilità e controllo su dove è possibile accedere ai dati è più importante che mai.

Le principali preoccupazioni tra i professionisti della sicurezza includono la protezione dei carichi di lavoro dai rischi generali per la sicurezza, la perdita o l'esposizione dei dati in caso di violazione dei dati e la conformità normativa.

Le mitigazioni esistenti delle minacce contro i server cloud sono spesso radicate nel firmware o nel software, rendendole vulnerabili alle stesse strategie di attacco.

Ad esempio, se il firmware può essere sfruttato con successo, molto probabilmente i controlli di sicurezza basati sul firmware possono essere aggirati allo stesso modo.

Di seguito evidenzio altri tipi di attacchi tipi di attacchi.

1. *Return Oriented Programming (ROP) and Call/Jump Oriented Programming (COP/JOP) Attacks*
2. *Address Transaction Attacks*

FURTO DELLE CHIAVI DI CIFRATURA

Le chiavi crittografiche sono risorse di alto valore, soprattutto in ambienti in cui il proprietario delle chiavi non ha il controllo completo dell'infrastruttura, come cloud pubblici, edge computing e implementazioni di virtualizzazione delle funzioni di rete (Network Function Virtualization – NFV).

Le chiavi sono in genere fornite su disco come flat file o voci nei file di configurazione.

In fase di esecuzione, i workload leggono le chiavi nella memoria ad accesso casuale (RAM) e le utilizzano per eseguire operazioni crittografiche come la firma dei dati, la crittografia/decrittografia o la terminazione Transport Layer Security (TLS).

Le chiavi su disco e nella RAM sono esposte ad attacchi convenzionali come:

1) *l'escalation dei privilegi,*
2) *l'esecuzione di codice remoto (Remote Code Execution - RCE) e*
3) *la cattiva gestione del buffer di input.*

Le chiavi possono anche essere rubate da amministratori malintenzionati o divulgate a causa di errori operativi, ad esempio durante una snapshot VM se questa non è correttamente protetta.

4. CONTROMISURE

Le tecniche di sicurezza abilitate per hardware possono aiutare a mitigare le minacce stabilendo e mantenendo la fiducia della piattaforma, una garanzia dell'integrità della configurazione della piattaforma sottostante, inclusi hardware, firmware e software.

Fornendo questa garanzia, gli amministratori della sicurezza possono ottenere un livello di visibilità e controllo su dove è consentito l'accesso a carichi di lavoro e dati sensibili.

Le tecnologie di sicurezza della piattaforma che stabiliscono l'attendibilità della piattaforma possono fornire notifiche o persino auto-correzione degli errori di integrità rilevati.

Le configurazioni della piattaforma possono essere ripristinate automaticamente a uno stato affidabile e dare alla piattaforma resilienza contro gli attacchi.

Tutti i controlli di sicurezza devono avere una radice di attendibilità (Root Of Trust - RoT), un punto di partenza implicitamente attendibile.

I controlli basati su hardware possono fornire una base immutabile per stabilire l'integrità della piattaforma.

La riduzione al minimo dell'impronta di questo RoT si traduce nella riduzione del numero di moduli o tecnologie che devono essere implicitamente attendibili. Ciò riduce sostanzialmente la superficie di attacco.

Il firmware della piattaforma verificato può, a sua volta, verificare il boot loader del sistema operativo, che può quindi verificare altri componenti software fino al sistema operativo stesso e ai livelli di runtime del container o dell'hypervisor.

FIDUCIA TRANSITIVA

*La **fiducia transitiva** qui descritta è coerente con il concetto di catena di fiducia (Chain Of Trust - CoT), un metodo in cui ogni modulo software in un processo di avvio del sistema è richiesto per misurare il modulo successivo prima di passare al controllo.*

Inoltre, esistono altre tecnologie di sicurezza della piattaforma hardware in grado di proteggere i dati inattivi, in transito e in uso fornendo crittografia del disco con accelerazione hardware o isolamento della memoria basato su crittografia.

*Utilizzando l'hardware per eseguire queste attività, la **superficie di attacco viene mitigata**, impedendo l'accesso diretto o la modifica del firmware richiesto.*

*Isolare questi meccanismi di **crittografia su hardware dedicato** può consentire di indirizzare e migliorare le prestazioni separatamente anche da altri processi di sistema.*

*Un concetto chiave di trusted computing è la verifica dell'integrità della piattaforma (**Platform Integrity Verification**) sottostante.*

L'integrità della piattaforma è tipicamente composta da due parti:

- ✓ *Misurazione crittografica di software e firmware;*
- ✓ *Verifica del firmware e della configurazione.*

In alcuni casi, viene aggiunta una terza parte all'integrità della piattaforma:

- ✓ *Ripristino firmware e configurazione.*

MODULO PER LA SICUREZZA HARDWARE (HARDWARE SECURITY MODULE – HSM)

HSM è "un dispositivo di elaborazione fisica che protegge e gestisce le chiavi crittografiche e fornisce l'elaborazione crittografica".

Le operazioni crittografiche come la crittografia, la decrittografia e la generazione/verifica della firma sono in genere ospitate sul dispositivo HSM e molte implementazioni forniscono meccanismi con accelerazione hardware per le operazioni crittografiche.

Un Trusted Platform Module (TPM) è un tipo speciale di HSM in grado di generare chiavi crittografiche e proteggere piccole quantità di informazioni sensibili, come password, chiavi crittografiche e misurazioni dell'hash crittografico.

CATENA DI FIDUCIA (CHAIN OF TRUST - COT)

CoT è un metodo per mantenere validi confini di fiducia applicando un principio di fiducia transitiva.

Ogni modulo firmware nel processo di avvio del sistema è necessario per misurare il modulo successivo prima di passare al controllo.

Una volta effettuata la misurazione del modulo firmware, si consiglia di estendere immediatamente il valore di misurazione a un registro HSM per l'attestazione in un secondo momento.

Il CoT può essere esteso ulteriormente nel dominio dell'applicazione, consentendo di misurare e attestare file, directory, dispositivi, periferiche, ecc.

ELABORAZIONE RISERVATA (CONFIDENTIAL COMPUTING)

Con la necessità di una protezione aggiuntiva nello spazio di lavoro virtualizzato, è stata posta enfasi sulla crittografia dei dati sia a riposo sia durante l'uso (vedi comma 2 dell'art. 32 del Reg. UE 2016/679).

La crittografia dei dati inattivi fornisce protezione per i dati su disco.

Questo si riferisce in genere a un archivio dati smontato e protegge da minacce come la rimozione fisica di un'unità disco.

La protezione e la sicurezza dei dati cloud durante l'uso, nota anche come ***elaborazione riservata (confidential computing)****, utilizza funzionalità abilitate all'hardware per isolare ed elaborare i dati crittografati in memoria in modo che i dati siano a minor rischio di esposizione e compromissione da carichi di lavoro simultanei o dal sistema e dalla piattaforma sottostanti.*

AMBIENTE DI ESECUZIONE AFFIDABILE (TRUSTED EXECUTION ENVIRONMENT - TEE)

TEE è un'area o un'enclave protetta da un processore di sistema.

Segreti o dati sensibili come chiavi crittografiche, stringhe di autenticazione o dati con problemi di proprietà intellettuale e privacy possono essere conservati all'interno di un TEE e le operazioni che coinvolgono questi segreti possono essere eseguite all'interno del TEE, eliminando così la necessità di estrarre i segreti al di fuori del TEE.

Un TEE aiuta anche a garantire che le operazioni eseguite al suo interno e i dati associati non possano essere visualizzati dall'esterno, nemmeno da software o debugger privilegiati.

La comunicazione con il TEE è progettata per essere possibile solo attraverso interfacce designate ed è responsabilità del progettista/sviluppatore del TEE definire queste interfacce in modo appropriato.

Una buona interfaccia TEE limita l'accesso al minimo indispensabile per eseguire l'attività.

TECNOLOGIE DI ISOLAMENTO

Di seguito sono riportate ulteriori tecnologie di isolamento con lo scopo di incrementare la fiducia del sistema.

1. *Supply Chain Protection*
2. *Memory Isolation*
3. *Application Isolation*
4. *VM Isolation*

GEOLOCALIZZAZIONE (TRUSTED LOCATION)

Trusted Geolocation è un'implementazione specifica di tag delle risorse attendibili utilizzata con l'attestazione della piattaforma. I valori degli attributi chiave che specificano le informazioni sulla posizione sono utilizzati come tag di asset e forniti all'hardware del server, come il TPM.

In questo modo, le informazioni sulla posizione possono essere incluse nei report di attestazione della piattaforma e quindi utilizzate da agenti di orchestrazione del cloud, applicazioni di gestione dell'infrastruttura, motori di policy e altre entità.

La geolocalizzazione può essere un attributo importante da considerare con ambienti cloud ibridi soggetti a controlli normativi come il GDPR. Violare questi vincoli consentendo l'accesso ai dati al di fuori di specifici confini geopolitici può innescare sanzioni sostanziali.