RdS
RICERCA DI SISTEMA

ENEA
AGENZIA NAZIONALE PER LE NUOVE TECNOLOGIE, L'ENERGIA E LO SVILUPPO ECONOMICO SOSTENIBILE

MINISTERO DELL'AMBIENTE E DELLA SICUREZZA ENERGETICA

Ricerca di Sistema elettrico

Implementazione di una infrastruttura di calcolo a basso consumo per il controllo informatico delle reti elettriche intelligenti cyber-resilienti

Paolo Palazzari, Luigi Acampora, Michele Casà

ENEA

RdS_PTR 22-24_PR 2.1_LA3.4_138

Implementazione di una infrastruttura di calcolo a basso consumo per il controllo informatico delle reti elettriche intelligenti cyber-resilienti

P. Palazzari (ENEA), L. Acampora (ENEA), M. Casà (ENEA)

Dicembre 2024

Report Ricerca di Sistema Elettrico

Accordo di Programma Ministero dell'Ambiente e della Sicurezza Energetica - ENEA Piano Triennale di Realizzazione 2022-2024

Obiettivo 2: Digitalizzazione ed evoluzione delle reti

Progetto 2.1: Cybersecurity dei Sistemi Energetici

Linea di attività: 3.4

Responsabile del Progetto: M. Valenti (ENEA)

Responsabile Linea di Attività: P. Palazzari (ENEA)

Mese inizio previsto: luglio 2023

Mese inizio effettivo: luglio 2023

Mese fine previsto: dicembre 2024

Mese fine effettivo: dicembre 2024

Il presente documento descrive le attività di ricerca svolte all'interno dell'Accordo di collaborazione: Si ringrazia per la collaborazione alle attività svolte

Indice

Indice delle figure

1 Risultati attesi

I risultati attesi della LA3.4 riguardano la realizzazione di:

- Infrastruttura di rete dati che possa ospitare al suo interno un sistema di sensori/attuatori, degli apparati di rete, un firewall, un sistema di calcolo e un sistema di storage. La rete dati dovrà essere modellata e programmata per chiudere in una sottorete protetta i sensori con un unico punto di ingresso e uscita dei dati. Questo punto di ingresso dovrà essere connesso in maniera diretta con un firewall e con il sistema informatico per la raccolta ed elaborazione dei dati. Il firewall dovrà essere di ultima generazione, cosiddetti intelligenti, per poter essere in grado di discriminare le sottoreti, la sensoristica presente ed essere già dotato di alti livelli di sicurezza anche basati su intelligenza artificiale.
- Il sistema HPC, progettato nella LA3.2, dovrà essere connesso direttamente al firewall e alla sottorete dei sensori per poter raccogliere e analizzare in maniera diretta i dati che vengono generati dai sensori. Il sistema HPC dovrà avere ampia disponibilità di storage per la gestione dei dati della rete di comunicazione, dei dati generati dalla smart-grid e dagli applicativi di analisi dati.
- Il sistema HPC, potenziato da acceleratori FPGA, sarà essere installato in un ambiente di produzione reale e sarà integrata nell'infrastruttura ENEA già disponibile e con la quale dovrà condividere protocolli e sistemi di comunicazione e calcolo. Sono previsti anche apparati di rete e relativi sistemi software per realizzare infrastrutture di rete che gestiscano in sicurezza e rapidità il flusso crescente di dati all'interno di un centro ENEA e tra diversi centri ENEA (Casaccia e Portici).
- La gestione dei dati in conformità con i requisiti di sicurezza e privacy dei dati.

2 Risultati ottenuti

La linea di attività ha conseguito con successo gli obiettivi prefissati. Nello specifico, è stata realizzata un'infrastruttura di rete dati contenente: un gruppo di sensori/attuatori, degli apparati di rete, un firewall, un sistema di calcolo e un sistema di storage. I sensori sono stati allocati su una VLAN dedicata in modo tale da isolarli logicamente dal resto della rete ENEA. La sottorete protetta così definita è terminata su un firewall di ultima generazione che funge da default gateway per questa rete. Il firewall selezionato per il progetto è un dispositivo Palo Alto, dettagliato nelle sezioni dedicate di questo documento. Questo firewall si distingue per le sue avanzate funzionalità di sicurezza, alcune delle quali sfruttano tecnologie basate sull'intelligenza artificiale. I dati di traffico generati dai sensori, così come quelli relativi al traffico in ingresso ed in uscita dalla rete ENEA prodotti da un qualsiasi altro utente, possono essere visualizzati su DarkTrace che opera come Threat AI Analyzer sulla base del mirroring dei flussi dati. I raw data collezionati da DarkTrace sono quindi trasferiti in formato JSON ad un nodo ElasticSearch che permette di indicizzare, storicizzare e pseudo-anonimizzare, in conformità alla normativa vigente in materia privacy, i dati. Il nodo ElastiscSearch è stato installato sul sistema HPC, descritto nel deliverable della Linea di Attività (LA) 3.2, integrato nell' ambiente di produzione della rete ENEA e condivide con esso protocolli e sistemi di comunicazione e calcolo. Grazie all'integrazione all'interno dell'ambiente ICT-HPC di ENEA, i server beneficiano delle procedure di backup erogate dai centri di calcolo ENEA. Per quanto riguarda la ridondanza delle parti critiche, questa è stata prevista nella realizzazione del

sistema di storage che supporta vari livelli di RAID - Redundant Array of InexpensiveIndependent Disks (1, 5, 6, e 10). Il nodo ElasticSearch è funzionale sia all'indicizzazione, alla storicizzazione, alla pseudo-anonimizzazione del traffico dati che all'accessibilità degli stessi nelle fasi di training degli algoritmi di Machine Learning e di applicazione di tali algoritmi nella fase di elaborazione in Real Time. Per quanto riguarda l'elaborazione in streaming dei dati di traffico, sul sistema HPC è stato installato Apache Kafka come backbone per l'elaborazione dei dati in tempo reale, come meglio descritto nella Linea di Attività 3.7. L'infrastruttura Hw/Sw nel suo complesso, associata alle tecniche di Machine Learning elaborate sviluppate dalle università co-beneficiarie, consente di fornire un valore aggiunto in termini di sicurezza rispetto a possibili vulnerabilità delle reti smart grid.

Le funzionalità dell'infrastruttura sono state verificate utilizzando gli applicativi sviluppati dalle università co-beneficiarie, la realizzazione di attacchi cyber controllati avvenuti nella rete ENEA durante la sua piena operatività. I risultati sono riportati nei report delle LA3.7 e LA3.8.

Queste sessioni di test non solo hanno consentito ad ENEA di verificare il raggiungimento degli obiettivi di progetto ma anche di generare dataset completi, rappresentativi di molteplici pattern di traffico e pseudo-anonimizzati in conformità alle normative vigenti, che rappresentano validi campioni reali di dati di alta qualità. La qualità dei dati e la numerosità degli stessi permette lo sviluppo di modelli numerici accurati e approfondite analisi per gestire eventuali intrusioni.

2.1 Avanzamento della ricerca rispetto allo stato dell'arte internazionale con riferimento ai risultati ottenuti

L'intrusion detection gioca ormai un ruolo fondamentale nei processi di difesa delle reti, assicurando agli amministratori di rete un potente strumento di analisi rispetto ad intrusioni informatiche. Per questa ragione, nel corso degli anni molti ricercatori si sono focalizzati nell'elaborazione di algoritmi che permettessero di cogliere queste anomalie di sicurezza nel traffico dati. Presupposto fondamentale per lo sviluppo di questa linea di ricerca è la disponibilità di un dataset che abbia caratteristiche adeguate all'addestramento degli algoritmi. La mancanza di tali dati in passato ha comportato l'addestramento mediate dati sintetici, polarizzati rispetto a specifici pattern di traffico, creati artatamente e quindi non coerenti con il traffico di una rete "reale".

2.1.1 Dataset proposti in articoli precedenti

DARPA (Lincoln Laboratory 1998-99): il dataset proposto non si dimostra adeguato in quanto non rappresentativo del traffico di una rete "reale" ed inoltre contiene dell'irregolarità quali l'assenza di falsi positivi. Il dataset che riflette i flussi dati relativi a posta elettronica, navigazione, FTP, Telnet, SNMP ed attacchi quali DoS, Brute Force, SynFlood e NMAP è stato creato in modo artificioso ed è ormai obsoleto sia rispetto agli attacchi informatici in costante evoluzione che rispetto all'architettura di rete analizzata. McHugh, J. (2000). "Testing Intrusion Detection Systems: A Critique of the 1998 and 1999 DARPA Intrusion Detection".

KDD'99 (University of California, Irvine 1998-99): si tratta di una versione aggiornata del dataset DARPA che integra ulteriori attacchi quali Neptune-DoS, pod-DoS, Smurf-DoS. Il dataset contiene un significativo numero di record corrotti e ridondanti che inducono a risultati distorti. Tavallaee et al., 2009. "A detailed analysis of the KDD CUP 99 data set".

CAIDA (Center of Applied Internet Data Analysis 2002-2016):

I dataset CAIDA si focalizzano solamente su attacchi specifici ed inoltre non sono benchmarking efficaci a causa di una serie di lacune. Ali Shiravi e Ghorbani, 2012. "Toward developing a systematic approach to generate benchmark datasets for intrusion detection".

CICIDS2017: Il dataset, come altre precedenti proposte, è generato in maniera artificiale e non riflette un traffico "reale". Iman Sharafaldin (2018). "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization".

ENEA, nell'ambito della Linea di Attività 3.4, è riuscita a dare una risposta all'esigenza di generare, storicizzare ed avere facile accesso ad un dataset completo, rappresentativo di molteplici pattern di traffico, pseudo-anonimizzato in conformità alle normative vigenti e con adeguati volumi di traffico ai fini della rilevazione di traffico malevolo.

3 Prodotti attesi

Infrastruttura Hardware di Infrastruttura di calcolo HPC a basso consumo per il controllo informatico delle reti elettriche intelligenti cyber-resilienti

Il prodotto della ricerca consiste in 1 infrastruttura di calcolo HPC a basso consumo per il controllo informatico delle reti elettriche intelligenti cyber-resilienti, caratterizzata da:

a. almeno 1 nodo di calcolo dotato di 1 scheda accelerata (FPGA) e di almeno 8 GB di memoria RAM DDR4;
b. almeno 1 server di storage (fisico o virtuale) con disco di 5 TB.

Rapporto Tecnico che descrive l'implementazione di una infrastruttura di calcolo HPC a basso consumo per il controllo informatico delle reti elettriche intelligenti cyber-resilienti

Il Rapporto deve fornire:

1. descrizione dell'infrastruttura logica della rete;
2. dettagli sull'implementazione dell'infrastruttura con particolare riferimento ai seguenti componenti hardware:
    a. sensori di temperatura e consumo elettrico interrogabili da remoto;

b. dispositivo dotato di scheda FPGA (Field Programmable Gate Array) per la raccolta dei dati provenienti dai sensori;
c. server di storage per l'immagazzinamento dei dati;
d. firewall per la protezione dell'infrastruttura basati su algoritmi AI;
e. rapporto disponibile sul sito ENEA dedicato alla Ricerca di Sistema a valle del processo di rendicontazione.

4 Prodotti ottenuti

Il prodotto sviluppato nella Linea di Attività 3.4 è una infrastruttura di rete con un server di calcolo per il controllo cyber di sottoreti e apparati. Lo schema logico dell'infrastruttura e di come avviene la raccolta dei dati è riassunto nella Figura 1.

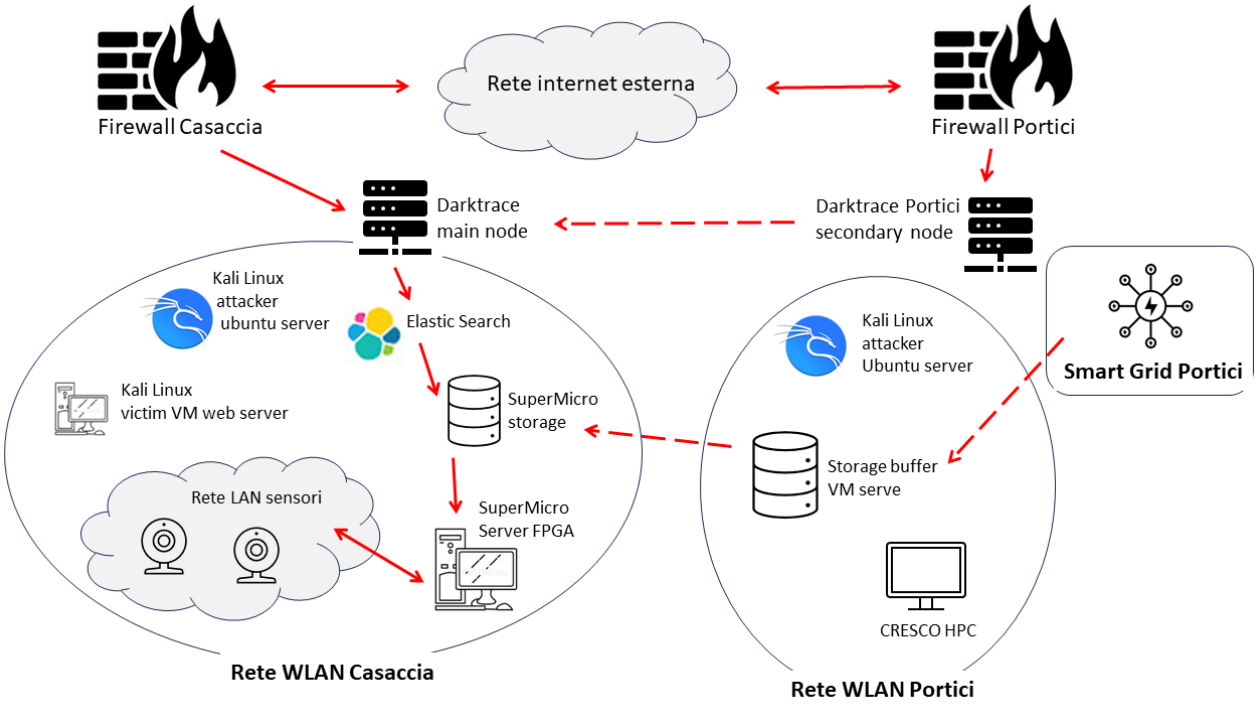

Figura 1: Schema logico dell'infrastruttura di calcolo HPC

4.1 Infrastruttura Hardware di Infrastruttura di calcolo HPC

È stata implementata un'infrastruttura di calcolo e comunicazione, costituita da:

- Due firewall PaloAlto di nuova generazione modello PA-3260.
- Server DarkTrace configurato per duplicare il traffico proveniente dai firewall ed inviarlo allo storage server e al server di calcolo.
- ElasticSearch.
- Server di calcolo SuperMicro con 2 schede FPGA U280 di Amd (8 GB HBM + 16 GB DDR4).
- Storage server SuperMicro con 248 TB di spazio disco.

- Set di sensori/attuatori (temperatura, tensione, corrente, relais).

4.1.1 Firewall Palo Alto

I firewall Palo Alto Networks di nuova generazione (NGFW) Machine Learning Powered PA-3260 (Figura 2), installati presso i centri di Casaccia e di Portici, offrono una protezione completa per tutto il traffico, inclusi i dati crittografati. Grazie all'uso di elaborazione e memoria dedicate per rete, sicurezza, gestione e prevenzione delle minacce, garantiscono prestazioni elevate e affidabilità.
Il sistema operativo PAN-OS, che governa i firewall, classifica nativamente tutto il traffico, comprese applicazioni, minacce e contenuti, associandolo direttamente all'utente, indipendentemente dalla sua posizione o dal dispositivo utilizzato. Questa funzionalità avanzata permette un controllo granulare e una sicurezza ottimizzata.
L'applicazione, il contenuto e l'utente, ossia gli elementi portanti dell'Ente, vengono quindi utilizzati come punto di partenza per le policy di sicurezza, migliorando il livello di protezione e riducendo i tempi di risposta agli incidenti.

Figura 2: Firewall

Il Next Generation Firewall di Palo Alto:

- Incorpora l'apprendimento automatico (ML) nel firewall per fornire una threat prevention inline senza firma per gli attacchi basati su file, identificando e bloccando immediatamente tentativi di phishing mai osservati prima.
- Sfrutta i processi di apprendimento automatico basati su cloud per trasmettere istruzioni e firme a ritardo zero all'NGFW.
- Tramite un servizio distribuito nel cloud e integrato in modo nativo nell'NGFW, utilizza l'analisi comportamentale per rilevare i dispositivi IoT e formulare raccomandazioni di policy.
- Automatizza le raccomandazioni delle policy, favorendo un risparmio di tempo e riducendo la possibilità di errori umani.
- Utilizza le applicazioni, e non le porte, come base per tutte le decisioni relative alle policy di abilitazione sicura (consenso, rifiuto, pianificazione, ispezione e applicazione di traffic shaping).

- Si integra facilmente con un'ampia gamma di repository per sfruttare le informazioni sugli utenti: controller LAN wireless, VPN, server di directory, SIEM, proxy e altro ancora.
- Tramite l'autenticazione a più fattori (MFA) a livello di rete per qualsiasi applicazione, senza che essa venga modificata, impedisce che le credenziali aziendali trapelino su siti Web di terze parti e previene il riutilizzo di credenziali rubate.

4.1.2 DarkTrace

Darktrace ha un ruolo fondamentale nella gestione della sicurezza informatica di ENEA in quanto rappresenta un punto di osservazione centralizzato e privilegiato delle reti di Casaccia e Portici. I dati di traffico generati dagli apparati di rete presenti in ENEA sono collezionati dal Darktrace che, quindi, riveste il ruolo di traffic log nell'architettura logica della soluzione.

Darktrace adotta un approccio tecnologico fondamentalmente diverso rispetto ad altre tecnologie presenti sul mercato: utilizza tecniche di intelligenza artificiale basate su algoritmi non supervisionati sui dati unici dell'organizzazione, piuttosto che applicare tecniche supervisionate allenate su dati preesistenti generate da fonti diverse. Facendo in modo che l'IA impari l'attività, non l'attacco o l'aggressore, può identificarsi minacce nuove e interne che superano le altre tradizionali difese informatiche.

L'utilizzo di questo approccio di autoapprendimento consente ai prodotti Darktrace di fornire una risposta rapida ad un panorama di minacce in continua evoluzione.

4.1.3 ElasticSearch

ElasticSearch è un motore di ricerca e analisi distribuito, nonché un archivio dati scalabile e un database vettoriale basato su Apache Lucene (Figura 3). Progettato per essere uno strumento estremamente flessibile, consente di cercare, indicizzare, archiviare e analizzare dati di qualsiasi tipo e dimensione con prestazioni quasi in tempo reale, rendendolo ideale per una vasta gamma di applicazioni, dalla ricerca full-text all'analisi complessa dei dati.

Figura 3: Logo ElasticSearch, motore di ricerca e analisi distribuito, archivio e database vettoriale

Nell'ambito del progetto è stato utilizzato come interfaccia di comunicazione tra DarkTrace e la piattaforma di elaborazione Machine Learning sia nella fase di training che nella fase di stream analytics, installata sul sistema di calcolo realizzato con il server SuperMicro.

4.1.4 Server SuperMicro

Il server di calcolo installato è un sistema biprocessore AMD EPYC 9124, con 256 GByte di memoria DDR, 2 TB di spazio disco e dotato di due schede FPGA Alveo U280. Ogni processore è dotato di 16 core fisici (32 core logici), ha 32 istanze di memoria cache privata di primo livello (L1d: 32KB dedicata ai dati e L1i:32KB dedicata alle istruzioni), 32 istanze di memoria cache di secondo livello (L2, ogni istanza privata misura 1 MB) e 8 istanze di memoria cache di terzo livello condivisa (L3: 16 MB è la dimensione di ogni istanza).

4.1.5 Storage Supermicro

Il sistema di storage è progettato per ospitare il database realizzato nella LA3.8 e i vari log dei dati di traffico. Con l'obiettivo di omogeneizzarsi ai sistemi già presenti nel centro di calcolo di ENEA Casaccia, all'interno del quale è ospitata l'infrastruttura oggetto del progetto, si è deciso di acquisire il seguente sistema SuperMicro SSG-6049P-E1CR24H che include:

- 2 Processori Intel Xeon Bronze 3204, 6 core con clock a 1,9 GHz, TDP = 85W.
- 96 GB di memoria DDR4, organizzati in 6 moduli di memoria da 16GB ECC Registered 2933 MHz.
- Controller RAID, con 2 GB di cache dedicata,
    - SAS3 (12 Gb/s) livelli RAID 0,1, 5,6, 10, 50, 60
    - SATA (6 Gb/s) livelli RAID 0,1,5,10
- 2 SSD da 240 GB SATA per sistema operativo.
- 24 Dischi rigidi 3.5" da 12 TB,7200 rpm,NL-SATA 6Gb/s inseriti nei 24 slot hot-swappable presenti.
- Connettività LAN tramite 2 RJ45 10GBase-T LAN ports e 1 RJ45 Dedicated IPMI LAN port.

Le configurazioni RAID, integrate con la tecnologia hot-swappable, assicurano elevati livelli di sicurezza, affidabilità e continuità operativa richiesti dallo storage server, garantendo la protezione dei dati e la disponibilità del servizio anche in caso di guasti hardware.

4.1.6 Set di sensori/attuatori

Sono stati acquisiti dal System Integrator Promente due rack, ognuno accessibile via rete tramite protocollo http.

Ogni rack contiene, come meglio rappresentato in Figura 4:

- 2 sensori di temperatura
- 4 prese di corrente controllate da relais, di ognuna delle quali è possibile conoscere tensione e corrente

I sensori sono interrogabili tramite il protocollo http e, sempre tramite lo stesso protocollo, si possono controllare gli stati delle singole prese cui è attaccato un carico.

Figura 4: Sensori Promente di temperatura, tensione e corrente

5 Analisi degli scostamenti su attività e risultati

Sono stati raggiunti con successo gli obiettivi prefissati, in coerenza con le attività dichiarate in sede di sottoscrizione del progetto. Di conseguenza, non sono presenti scostamenti da dichiarare.

6 Sintesi delle attività svolte

L'implementazione dell'infrastruttura di calcolo si è articolata in diverse attività quali:

- **Acquisizione della componentistica**: necessaria alla realizzazione della infrastruttura di calcolo e comunicazione.
- **Installazione e Migrazione dei Firewall Palo Alto**: si tratta degli elementi di sicurezza perimetrale di ultima generazione che garantiscono un più elevato grado di sicurezza al traffico della rete di ENEA Casaccia e di ENEA Portici.

- **Installazione e configurazione Server di Calcolo e Storage SuperMicro**: che rappresentano il cuore dell'infrastruttura di calcolo ha previsto un'attività di installazione e configurazione per consentire una loro comunicazione con la rete di produzione ENEA.
- **Installazione e configurazione dei sensori/attuatori**: i sensori/attuatori sono stati installati e configurati per comunicare con la rete di produzione di Casaccia su una VLAN dedicata.
- **Installazione di server Kali Linux:** utili alla generazione, anche se in maniera controllata, di un traffico malevolo che si sovrappone a quello normalmente presente sulla rete ENEA.
- **Generazione Attacchi Malevoli**: al fine di effettuare dei test sul corretto funzionamento dell'infrastruttura di calcolo ed al fine di generare un dataset utile all'addestramento degli algoritmi di Machine Learning, sono stati simulati, grazie ad i tool disponibili su Kali Linux, attacchi informatici rivolti ai sensori e ad un webserver.
- **Installazione di un Webserver:** il webserver apache è stato installato e configurato al fine di simulare la reazione di un server web che riceve traffico malevolo.
- **Configurazione DarkTrace per comunicare con ElasticSearch**: la soluzione Darktrace, già ampiamente utilizzata in ENEA per la segnalazione di eventi anomali, ha svolto nell'ambito del progetto l'elemento chiave che ha permesso di collezionare i dati per poi trasferirli nel formato opportuno ad ElasticSearch.
- **Installazione e Configurazione di un nodo ElasticSearch**: è stato installato e configurato un nodo ElasticSearch al fine di permettere una veloce ed efficace indicizzazione dei dati nonché una loro storicizzazione.
- **Supporto alla definizione del framework privacy**: al fine di garantire la conformità nel processo di elaborazione dei dati di traffico alle norme europee vigenti in tema privacy.

7 Dettaglio delle attività svolte

7.1 Acquisizione della componentistica

Attività propedeutica all'implementazione della infrastruttura di calcolo è l'acquisizione dei seguenti apparati:

- 2 Firewall Palo Alto modello PA-3260.
- 1 Server di calcolo SuperMicro con 2 schede FPGA U280 di AMD (8 GB HBM + 16 GB DDR4).
- 1 Storage server SuperMicro con 248 TB di spazio disco.
- 2 sensori/attuatori.

7.2 Installazione e configurazione Firewall Palo Alto

L'attività di installazione e la relativa configurazione dei firewall di ultima generazione Palo Alto ha richiesto a sua volta l'implementazione di una serie di attività preliminari che permettessero di minimizzare i rischi di disservizio:

- **Strategia di migrazione**: attività propedeutica ad ogni step relativo alla migrazione alle soluzioni fornite da Palo Alto è stata la definizione di un piano di migrazione condiviso tra i centri di ENEA.
- **Assessment delle configurazioni esistenti**: presenti sui firewall Fortigate precedentemente installati sui siti di Casaccia e di Portici.
- **Portabilità delle policy di sicurezza:** diversamente dei firewall Fortigate che implementano policy di sicurezza basata sulle porte, i firewall Palo Alto adottano regole basate sul riconoscimento delle applicazioni. Questa differenza di comportamento nell'analisi del traffico ha comportato un'attività di attenta analisi delle regole esistenti e la conversione delle stesse in policy adeguate alle logiche dei firewall Palo Alto.
- **Integrazione** dei firewall Palo Alto con gli apparati presenti in produzione sulla rete ENEA. L'attività, oltre a richiedere competenze tecniche specifiche ha imposto una meticolosa organizzazione interna al fine di ridurre al minimo i disservizi agli utenti.
- **Migrazione**: la migrazione, cuore dell'attività di installazione e configurazione, ha comportato il coinvolgimento di un nutrito gruppo di esperti di rete al fine di garantire il perseguimento dell'obiettivo di progetto.
- **Ottimizzazione delle configurazioni**: a valle della migrazione sulle soluzioni Palo Alto sono state affinate ed ottimizzate le configurazioni per garantire i servizi esistenti e l'adeguamento ai requisiti di progetto. Per esempio, i test realizzati con attacchi cyber controllati, gestiti utilizzando la piattaforma realizzata in questa Linea di Attività, hanno messo in luce che alcune configurazioni del firewall non erano ottimizzate per il sistema ENEA e le sottoreti utilizzate in questo progetto.

7.3 Installazione e configurazione Server e Storage SuperMicro

L'attività di installazione e configurazione del server di calcolo e dello storage SuperMicro ha richiesto l'implementazione dei seguenti passi:

- Configurazione delle interfacce di management.
- Configurazione delle interfacce iSCI.
- Integrazione nella rete di produzione ENEA.

7.4 Installazione e configurazione di sensori/attuatori

L'attività di installazione e configurazione dei sensori/attuatori ha richiesto:

- La configurazione di indirizzi IP associati ad una VLAN dedicata che permette di isolare logicamente i sensori/attuatori
- l'integrazione dei sensori nella rete di produzione ENEA

7.5 Installazione server Kali Linux

La distribuzione Kali Linux è una fra le migliori tra quelle dedicate agli ethical hacker, dato che contiene numerosi tool adatti allo scopo.

I server Kali Linux sono stati installati sulle sedi di Casaccia e di Portici su Virtual Machines con le seguenti caratteristiche:

- 4 CPU
- 8GB di RAM
- 64GB di spazio disco

Tali server sono stati poi integrati con la rete di produzione di ENEA.

7.6 Generazione attacchi malevoli

L'attività ha previsto la generazione controllata, grazie ai tool disponibili su server Kali Linux, dei seguenti attacchi informatici:

- Attacco DDoS Slowloris versus Sensori attuatori
- Attacco DDoS Slowloris versus WebServer
- Attacco di forza bruta versus WebServer

7.7 Installazione WebServer

Il webserver apache è stato installato sulla sede di Casaccia su una Virtual Machine con le seguenti caratteristiche:

- 4 CPU
- 8GB di RAM
- 64GB di spazio disco

Tale server è stato poi integrato con la rete di produzione di ENEA.

7.8 Configurazione DarkTrace

Darktrace è stato configurato opportunamente per consentirne la comunicazione con il nodo ElasticSearch. I dati sono trasferiti da DarkTrace ad Elasticsearch in modalità push in formato JSON.

7.9 Installazione e Configurazione ElasticSearch

ElasticSearch è stato installato sul server di storage dotato di OS AlmaLinux 9 utilizzando Docker, una piattaforma di containerizzazione che consente di eseguire applicazioni in ambienti isolati e standardizzati chiamati container. Docker è particolarmente utile per installare ElasticSearch perché semplifica il processo di configurazione, eliminando la necessità di installare manualmente tutte le dipendenze. Inoltre, offre flessibilità per eseguire diverse versioni del software senza conflitti, garantendo un'installazione rapida, scalabilità e portabilità. Questa soluzione consente di sfruttare al meglio le risorse del server e di gestire facilmente aggiornamenti e backup.

L'installazione ha coinvolto le seguenti attività:

1. Preparazione dello storage e configurazione della partizione disco
2. Installazione di Docker su AlmaLinux 9

3. **Creazione del container ElasticSearch tramite Docker Compose:** la creazione del container con Docker Compose avviene eseguendo il comando docker compose up, che avvia i servizi definiti nel file docker-compose.yml; durante il processo, i servizi vengono configurati utilizzando le specifiche fornite nel file di composizione, le variabili d'ambiente presenti nel file .env e i file di configurazione .conf necessari per gli altri componenti dell'ecosistema ElasticSearch, come Kibana (per la visualizzazione dei dati), Logstash (per l'elaborazione e il trasporto dei dati).

7.10 Supporto alla definizione del framework privacy

La raccolta dei dati di traffico sulle reti informatiche dei Centri di Casaccia e Portici, e la loro successiva post elaborazione, ha posto un delicato ed importante problema di riservatezza dei dati personali. Per questa ragione sono state sviluppate, in collaborazione con il Data Protection Officer dell'ENEA, le seguenti attività:

- un'analisi attenta e puntuale degli impatti privacy
- la redazione di un documento di Data Protection Impact Assessement e del relativo allegato tecnico
- la redazione del documento di Nomina del Responsabile del Trattamento dei Dati sottoscritto dai partner di progetto.

8 Contributo delle eventuali consulenze alle attività sopra descritte

Non è stata necessaria la partecipazione di alcuna società di consulenza per lo svolgimento delle attività e per il perseguimento delle finalità di progetto.

9 Pubblicazioni scientifiche

Le pubblicazioni scientifiche sono quelle congiunte con i partner di progetto, relative allo sviluppo e alla valutazione di algoritmi di machine learning e intelligenza artificiale eseguiti sul sistema IT integrato. Nello specifico sono in fase di scrittura le seguenti pubblicazioni:

- Leveraging Machine Learning (Supervised and Unsupervised) to Enhance Cybersecurity: Mitigating DoS Attacks in Enea Network Infrastructures. Rivista MAKE di MDPI.
- An unsupervised approach to intrusion detection in IoT networks. Rivista Future Generation Computer Systems

10 Eventi di disseminazione

La partecipazione all'evento conclusivo di presentazione dei risultati del Progetto Integrato Cybersecurity dei sistemi energetici (Workshop-Progetto-Cybersecurity.pdf), presso

l'Auditorium GSE in viale Maresciallo Pilsudski 92, Roma, il 6 Dicembre ha dato l'opportunità di condividere con un'ampia platea di partecipanti i risultati delle attività di progetto.

11 Descrizione dei risultati ottenuti

11.1 Risultati Attesi - Infrastruttura di calcolo HPC

In Figura 1 si riporta lo schema adottato per implementare la infrastruttura di calcolo HPC. L'architettura di rete si compone di due parti afferenti rispettivamente alla rete di Casaccia e a quella di Portici, connesse tra di loro in tunnel su un collegamento geografico fornito dal GARR. Entrambe le reti sono delimitate da Firewall Palo Alto 3260 di ultima generazione, che svolgono la funzione di elementi di sicurezza perimetrale ed al contempo forniscono servizi di sicurezza avanzata. La presenza sia sul campus di Casaccia che su quello di Portici dei nodi DarkTrace garantisce un mirroring dei dati e l'accesso in tempo quasi reale agli stessi per scopi di monitoraggio, di analisi o di post-elaborazione. I nodi DarkTrace svolgono ruoli diversi nell'ambito dell'architettura di rete: il nodo di Portici è configurato in modalità slave e trasmette i dati collezionati al nodo master di Casaccia, il quale è quindi in grado di visualizzare i flussi dati di entrambi i siti. Il nodo ElasticSearch che consente di indicizzare, archiviare e analizzare dati di qualsiasi tipo e dimensione con prestazioni quasi in tempo reale è stato installato sul nodo HPC, descritto nei precedenti paragrafi, ed integrato con il nodo master DarkTrace presente in Casaccia. Il nodo HPC, oltre alla storicizzazione del traffico dati, consente elaborazioni in streaming grazie alla sua capacità di effettuare elaborazioni ad alte prestazioni. Il flusso dati raccolto da ElasticSearch per il tramite di DarkTrace è il traffico in ingresso ed in uscita da e verso i siti di Casaccia e Portici generato sia dagli utenti che usufruiscono dei servizi di rete che dai dispositivi installati presso i campus. Tale traffico è stato segmentato logicamente con la definizione di apposite VLAN che consentono di segregare le reti in maniera efficiente e sicura. Seguendo questa logica anche i sensori/attuatori sono stati configurati su una VLAN dedicata al fine di segregare il traffico da essi generati rispetto al resto della Rete ENEA. La VLAN dedicata ai sensori è accessibile, sulla base di specifiche policy di sicurezza configurate sul firewall, solamente dal nodo SuperMicro che implementa sia gli algoritmi di Machine Learning applicati agli header dei protocolli di comunicazione che gli algoritmi di crittazione del contenuto informativo (payload) dei dati trasmessi dai sensori/attuatori. L'infrastruttura descritta consente inoltre la comunicazione con la SmartGrid di Portici.

11.2 Risultati Attesi – Framework Privacy

La raccolta dei dati di traffico sulle reti informatiche dei Centri di Casaccia e Portici e la loro successiva post elaborazione con algoritmi di Machine Learning pone un delicato ed importante problema privacy, in quanto include dati personali indiretti quali l'indirizzo IP degli utenti che stanno usufruendo dei servizi di rete. Per questa ragione è stata condotta, in collaborazione con il Data Protection Officer dell'ENEA, un'analisi attenta e puntuale degli impatti privacy, formalizzata nella redazione di un documento di Data Protection Impact Assessment (DPIA) e nel relativo allegato tecnico.

11.2.1 Allegato Tecnico DPIA

L'allegato tecnico della DPIA, parte integrante del processo di analisi e di valutazione degli impatti privacy, fornisce i dettagli delle soluzioni tecniche che permettono di perseguire gli obiettivi di progetto garantendo al contempo il rispetto della normativa privacy di riferimento (GDPR).

Il documento descrive logicamente gli apparati di rete che sono coinvolti nelle varie fasi dell'elaborazione dei dati e le attività realizzate dal Titolare del Trattamento dei Dati (ENEA) e dai Responsabili del Trattamento dei Dati. Infatti, il trattamento e l'elaborazione dei dati prevede il coinvolgimento, oltre che dell'ENEA in quanto Titolare del Trattamento dei dati, anche dell'Università di Bari e dell'Università di Roma 3 come Responsabili del Trattamento dei dati, che hanno effettuato il training degli algoritmi di Machine Learning.

Come schematizzato in Figura 5, il flusso dei dati, sia nella fase di training che di elaborazione in tempo reale, nel dominio ENEA coinvolge le seguenti risorse:

- **Firewall**: apparato che permette di monitorare le informazioni in ingresso ed in uscita dalla rete ENEA, di garantire la sicurezza della stessa, filtrando i flussi di traffico non ritenuti affidabili.
- **DarkTrace**: apparato che colleziona i dati di traffico generati dai dispositivi sia human che IoT e li presenta in un formato facilmente intelligibile.
- **ElasticSearch**: si tratta di un nodo che svolge molteplici funzioni all'interno del flusso di elaborazione dei dati:
    - **Storage**: I raw data collezionati dal DarkTrace Threat Analyzer sono trasferiti e salvati su questo nodo su un DB apposito
    - **Filtering**: i campi non necessari, ai fini dello scopo di elaborazione, dei record raw data sono filtrati e rimossi per rispettare il principio di minimizzazione
    - **Pseudo-anonimizzazione**: ai dati, precedentemente filtrati, è applicato un processo di pseudo-anonimizzazione, i cui dettagli sono forniti nei successivi paragrafi.

Tutti gli apparati hardware sono installati presso il CR ENEA di Casaccia e/o di Portici, con eventuale utilizzo di infrastruttura virtuale distribuita sui vari centri ENEA.

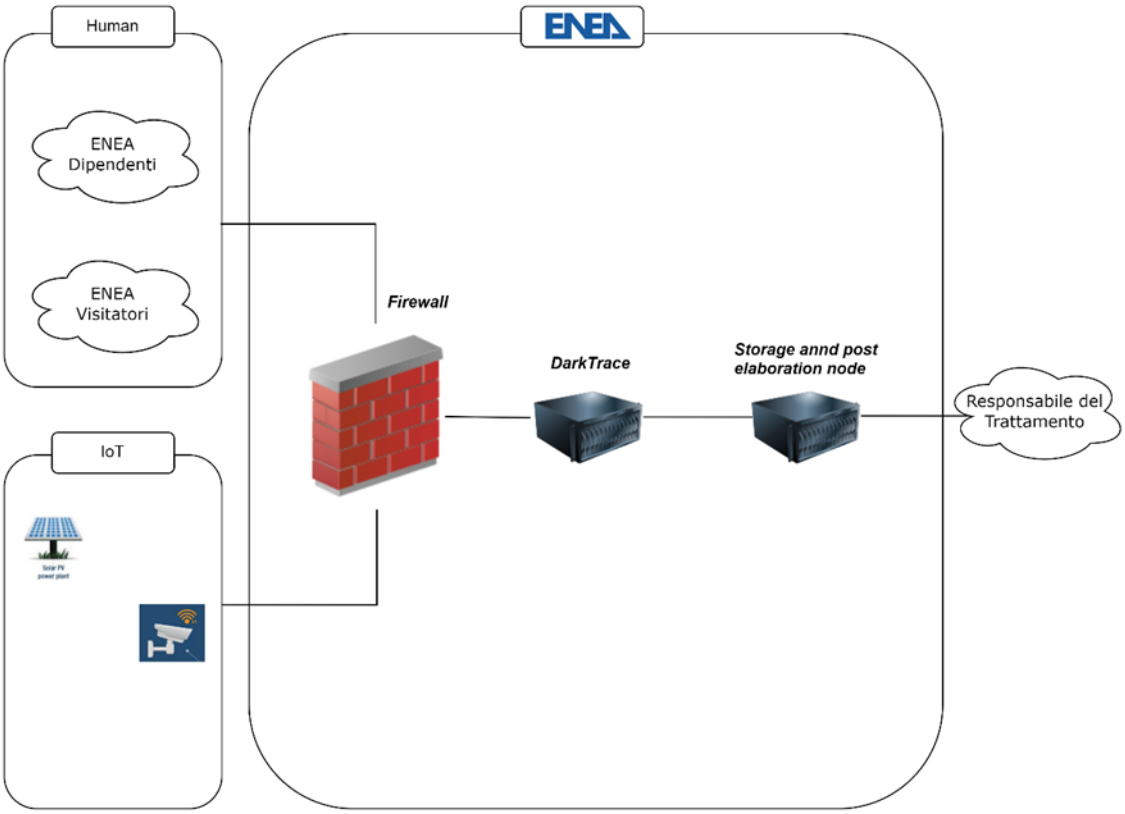

Figura 5: Flusso dei dati per assicurare la privacy

Il dataset a cui è stato applicato il processo di pseudo – anonimizzazione descritto nel paragrafo seguente è stato poi trasferito, tramite una comunicazione sicura basata su VPN, all'Università degli Studi di Bari Aldo Moro ed all'Università degli Studi Roma Tre (Responsabili del Trattamento dei Dati) che hanno implementato il training degli algoritmi di Machine Learning.

11.2.2 Pseudo-anonimizzazione

I dati relativi al traffico di rete, siano essi generati da dispositivi IoT che da dispositivi human, collezionati da DarkTrace e trasferiti automaticamente su un server di preprocessing, subiscono un processo di pseudo-anonimizzazione che persegue l'obiettivo di bilanciare il legittimo interesse della tutela della sicurezza della rete con la tutela della privacy dei dipendenti ENEA. Il processo di pseudo-anonimizzazione adottato è una delle tecniche suggerite nell'Opinion 216 "Parere 05/2014 sulle tecniche di anonimizzazione" del Working Group 29.

L'hash function è basata sull'algoritmo di hashing SHA256, notoriamente molto robusto, applicato all'indirizzo IP dei dispositivi che si connettono alla rete ENEA. Come espressamente indicato nell'Opinion 216 "Parere 05/2014 sulle tecniche di anonimizzazione" del Working Group 29, il risultato di questa elaborazione, è realizzato tramite funzioni di hashing, che sono per definizione matematicamente irreversibili. È importante sottolineare che il processo di pseudo-anonimizzazione descritto, come tutti i processi di pseudo-anonimizzazione, permette di ridurre la collegabilità (linkability) con i dati personali indiretti dei soggetti interessati. La robustezza del processo proposto permette di ridurre quanto più possibile

questo rischio. D'altra parte, la pseudo-anonimizzazione non elimina la biunivocità che volutamente viene mantenuta sia per garantire la correlazione delle informazioni relative agli stessi dispositivi per scopi di rilevamento di anomalie di sicurezza che per determinare problemi puntuali generati da singoli device.

I dati di traffico raccolti da DarkTrace vengono trasferiti al server storage Supermicro su cui è installato l'ambiente ElasticSearch, per l'archiviazione e le successive analisi. Per garantire la tutela della privacy, i dati vengono pseudo-anonimizzati tramite una ingest pipeline configurata in ElasticSearch. Questa pipeline utilizza uno script che applica un algoritmo di hash (SHA-256) agli indirizzi IP sorgente e destinazione, trasformandoli in valori crittografati non riconoscibili. Questo processo consente di mantenere l'integrità e l'utilità dei dati per analisi e monitoraggio, rispettando al contempo i requisiti di riservatezza. A valle della procedura di pseudo-anonimizzazione, i dati vengono indicizzati quotidianamente, permettendo un accesso rapido e organizzato alle informazioni. Il sistema è configurato per conservare i dati per un periodo massimo di due anni, in conformità con le disposizioni del Data Protection Impact Assessment (DPIA). Questa soluzione garantisce il rispetto delle normative sulla protezione dei dati, mantenendo elevate prestazioni e sicurezza nell'elaborazione delle informazioni.

11.3 Risultati non previsti nel capitolato – Dataset

L'implementazione dell'infrastruttura di calcolo a basso consumo ha consentito ad ENEA non solo il raggiungimento degli obiettivi di progetto ma anche l'opportunità di realizzare un sistema che consenta di generare un dataset completo, rappresentativo di molteplici pattern di traffico, pseudo-anonimizzato in conformità alle normative vigenti e con adeguati volumi di traffico ai fini della rilevazione di eventuali intrusioni.

11.3.1 Dataset in assenza di attacchi informatici

Il primo passo per il conseguimento di tale obiettivo è la scelta di un dataset che sia rappresentativo delle caratteristiche del traffico di telecomunicazioni in assenza e con un attacco informatico in corso. Per quel che riguarda il traffico in assenza di attacchi informatici, si è deciso di analizzare i raw data collezionati dal DarkTrace e poi esportati ad ElasticSearch, i quali forniscono uno spaccato delle abitudini di traffico di tutti i soggetti che accedono alle risorse presenti sul sito di Casaccia e di Portici. Vista e considerata la numerosità dei soggetti coinvolti, la diversa titolarità delle funzioni da essi svolte, la multidisciplinarità degli argomenti trattati e la gamma di servizi forniti sia all'interno che all'esterno del perimetro dell'ente, il dataset preso in considerazione si dimostra essere estremamente interessante ai fini del training degli algoritmi di Intrusion Detection e per la successiva analisi in Stream Analytics.

In molti articoli presenti in letteratura è stato evidenziato come la mancanza di un dataset adeguato ha limitato in maniera significativa lo sviluppo, l'analisi e la valutazione di algoritmi di Machine Learning che consentano di individuare possibili attacchi informatici.

In passato, dataset quali DARPA98, KDD99, ISC2012 e ADFA13 sono stati utilizzati dai ricercatori per valutare le prestazioni di algoritmi di Intrusion Detection. Purtroppo, molti di questi non erano aggiornati, alcuni erano caratterizzati da una mancanza di diversità e di adeguati volumi di traffico ed altri ancora per ragioni privacy non consentivano un'efficace analisi dei metadati.

Il dataset proposto rappresentativo del traffico benigno ha le seguenti caratteristiche:

- **Anonymity:** i dati di traffico sono stati oggetto di un processo di pseudo-anonimizzazione i cui dettagli sono descritti nei paragrafi precedenti.
- **Available Protocols:** i dati collezionati su DarkTrace ed in seguito esportati su ElasticSearch sono relativi a tutti i protocolli di comunicazione anche se l'elaborazione degli algoritmi di Machine Learning di Detection è stata poi effettuata solo sui protocolli HTTP, HTTPS e SSH.
- **Complete Network Configuration:** il traffico da analizzare, sia per scopi di training degli algoritmi di machine learning sia per la successiva fase di elaborazione in streaming, è stato spillato dalla rete di produzione di ENEA ed in quanto tale rappresentativo di una topologia di rete completa, la cui composizione è stata descritta nei paragrafi precedenti.
- **Feature Set**: le features dei raw data, trasmessi da DarkTrace ad ElasticSearch e funzionali al progetto, sono abbastanza numerose da evitare che l'addestramento degli algoritmi di Machine Learning possa essere affetto da overfitting.
- **Metadata**: la struttura del dataset e la relativa descrizione è assicurata dalla formattazione delle informazioni in Json.
- **Volume**: i volumi giornalieri analizzati sono di circa 35 GB/giorno.

11.3.2 Dataset con attacchi informatici

Il dataset rappresentativo di attacchi informatici è stato generato artificiosamente con il supporto di tool che consentono attacchi DDoS e di forza bruta. Anche in questo caso, l'insieme del traffico, risultato dell'integrazione del traffico benigno descritto nella precedente sezione e di quello generato con tool open source è collezionato dal DarkTrace ed esportato su Elastic Search per successive elaborazioni.

11.3.2.1 Attacco Slowloris

L'attacco DDoS scelto per il test-case è di tipo Slowloris (Figura 6). È considerato molto efficace ed alto impatto per le seguenti ragioni:

- **Nascosto**: è difficile da rilevare poiché utilizza una larghezza di banda ridotta **(LOW)** e consuma lentamente le risorse del server nel tempo **(SLOW).**
- **Impatto diffuso**: ha un impatto non solo sul sito Web o sul server preso di mira, ma anche sui suoi utenti e clienti, che potrebbero riscontrare rallentamenti, errori o completa indisponibilità.
- **Utilizzo intensivo di risorse**: l'attacco richiede grandi quantità di risorse del server per difendersi, portando potenzialmente a una riduzione delle prestazioni o al completo fallimento del bersaglio.
- **Persistenti**: questi attacchi possono durare per un periodo prolungato, rendendoli più difficili da mitigare rispetto ad altri tipi di attacchi DDoS.

Quando un server Web riceve una richiesta HTTP da un browser, elabora la richiesta e invia una risposta HTTP al browser. Il server web alloca inoltre una certa quantità di risorse (come memoria, tempo di CPU e larghezza di banda della rete) per gestire ciascuna richiesta. Il server web può gestire solo un numero limitato di richieste alla volta, a seconda della sua configurazione e capacità. Un attacco Slowloris sfrutta questo limite inviando più richieste HTTP parziali al server web.

In questo modo l'aggressore può occupare tutte le connessioni disponibili sul server web e impedire ad altri utenti legittimi l'accesso al sito web.

Figura 6: Modalità di funzionamento di un web server, senza e con attacco Slowloris

Un esempio di attacco Slowloris potrebbe essere determinato da una richiesta di http GET incompleta come quella riportata nell'esempio sotto:

GET /index.html HTTP/1.1

Host: www.example.com

User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:89.0) Gecko/20100101 Firefox/89.0

Accept: text/html, application/xhtml+xml,application/xml;q=0.9,image/webp,/*;q=0.8*

Accept-Language: en-US,en;q=0.5

Accept-Encoding: gzip, deflate

Connection: keep-alive

Upgrade-Insecure-Requests: 1

***X-a**:*

Da notare che nella parte di http Header non è presente alcun carattere di fine riga (\r\n) dopo l'ultimo campo di intestazione (X-a). Non è presente alcuna riga vuota (\r\n\r\n) dopo la sezione dell'intestazione. Ciò significa che questa richiesta è incompleta e non valida secondo il protocollo HTTP. Tuttavia, alcuni server Web accetteranno comunque questa richiesta e attenderanno più dati dal client prima di inviare una risposta o chiudere la connessione.

11.3.2.2 Testbed Attacco Slowloris vs Sensori IoT

Il test implementato per la generazione del dataset, che riproduce in maniera controllata un attacco informatico indirizzato ai sensori IoT, si compone di:

- un server Kali Linux (attaccante) installato presso la sede di Portici e su cui è abilitato un tool che simula un attacco Slowloris
- un server Kali Linux (attaccante) installato presso la sede di Casaccia e su cui è abilitato un tool che simula un attacco Slowloris
- sensori IoT, le cui caratteristiche sono state descritte nel dettaglio nei precedenti paragrafi

Implementazione attacco Slowloris vs Sensori IoT: l'attacco è stato sviluppato con il supporto del tool SlowHttpTest installato sui server Kali Linux (attaccanti), attivando progressivamente più connessioni in parallelo ognuno dei 2 sensori IoT.

Nello specifico è stata applicata la seguente configurazione:

- Client-side: è stato lanciato il seguente comando slowhttptest -H -c 1000 -r 10 -g -u http://192.168.118.151/ -l 604800 -i 3 -v 4" che corrisponde all'attivazione delle sessioni HTTP sulla base dei seguenti parametri:
  - Numero di connessioni target: 1000
  - Rate di connessioni: 10 connessioni al secondo
  - Url: Web Server Apache ospitato sul server 192.168.118.151
- Server-side: sono stati utilizzati i parametri di configurazione standard del web server apache.

11.3.2.3 Effetti Attacco Slowloris vs Sensori IoT

L'assenza di meccanismi di protezione per la mitigazione da attacchi DoS sulla componente webserver dei sensori IoT, che, come tanti apparati IoT dispongono di una ridotta gamma di funzionalità attinenti alla sicurezza informatica, ha comportato un blocco del servizio erogato dagli stessi. Alla conclusione dell'attacco, il servizio web erogato dai sensori IoT non è stato in grado di ripristinare in autonomia le sue normali funzionalità web e per tale ragione è stato necessario un riavvio dei sensori.

Considerata questa forte limitazione dei sensori in termini di capacità di mitigazione di attacchi DDoS e la conseguente riduzione del perimetro di ricerca del progetto, è stato implementato un'ulteriore testbed in cui come target dell'attacco è stato considerato un webserver apache che include nativamente meccanismi di protezione da attacchi DoS. In questo modo, è possibile simulare gli effetti di un attacco DoS su sensori IoT che presentano un più elevato grado di robustezza nei confronti di attacchi DoS.

11.3.2.4 Test Bed Attacco Slowloris vs Apache Server

Il test bed implementato per la generazione del dataset con attacchi informatici consiste in:

- un server Kali Linux (attaccante) installato presso la sede di Portici e su cui è abilitato un tool che simula un attacco Slowloris
- un server Kali Linux (attaccante) installato presso la sede di Casaccia e su cui è abilitato un tool che simula un attacco Slowloris
- un apache server (vittima) installato presso la sede di Casaccia

11.3.2.5 Implementazione attacco Slowloris vs WebServer Apache

L'attacco è stato sviluppato con il supporto del tool SlowHttpTest presente sui server Kali Linux (attaccanti), attivando progressivamente più connessioni in parallelo verso un web server apache.

Nello specifico è stata applicata la seguente configurazione:

- **Client-side**: è stato lanciato il seguente comando slowhttptest -H -c 1000 -r 10 -g -u http://<IP_WebServer_Apache>/ -l 604800 -i 3 -v 4" che corrisponde all'attivazione delle sessioni HTTP sulla base dei seguenti parametri:
  - Numero di connessioni target: 1000

    - Rate di connessioni: 10 connessioni al secondo
    - Url: Web Server Apache ospitato sul server <IP_WebServer_Apache>
- **Server-side**: sono stati utilizzati i parametri di configurazione standard del web server apache

11.3.2.6 Meccanismi di protezione Apache WebServer

Il web server apache, nella sua configurazione standard, prevede dei meccanismi di mitigazione nel caso di eventuali attacchi informatici.

Request Timeout

Nel dettaglio il file /etc/apache2/mods-enabled/reqtimeout.conf implementa una logica di difesa contro attacchi DoS sulla base di un timeout che specifica che le richieste HTTP header devono essere correttamente inviate entro un periodo variabile tra i 20 ed i 40 secondi

mod_reqtimeout limits the time waiting on the client to prevent an attacker from causing a denial of service by opening many connections but not sending requests. This file tries to give a sensible default configuration, but it may be necessary to tune the timeout values to the actual situation.

Wait max 20 seconds for the first byte of the request line+headers. From then, require a minimum data rate of 500 bytes/s, but don't wait longer than 40 seconds in total.

RequestReadTimeout header=20-40, minrate=500

Quando un client non completa la sua richiesta entro il tempo definito dal timeout, il server risponde con un messaggio "408 Request Timeout".

Http.maxhdr

Un altro parametro di configurazione che protegge il server da attacchi DoS è http.maxhdr. Si tratta del numero massimo di header in una richiesta. Il valore di default è 101 richieste header per sessione. Quando un client invia un numero eccessivo di richieste, il server risponde con un messaggio "400 Bad Request".

11.3.2.7 Effetti Attacco Slowloris vs Apache Server

I meccanismi di protezione del webserver apache mitigano l'effetto dell'attacco DDoS, come evidenziato dall'immagine sotto, determinando un effetto ciclico che può essere riassunto dai seguenti passi:

- Gli attaccanti richiedono l'apertura di più connessioni in parallelo al bersaglio della loro attività malevola impegnando progressivamente le risorse del web server e rendendo quindi il servizio non disponibile ad utenti legittimi.
- Il web server rimane in attesa per un tempo pari a quello definito nel mod_reqtimeout prima di considerare la richiesta del client (in questo caso l'attaccante) non completa, di chiudere la sessione e di rispondere con un messaggio di "408 Request Timeout".
- Progressivamente le risorse del web server vengono liberate, dopo essere state rilasciate per incompletezza della richiesta, ed il servizio torna ad essere disponibile anche per gli utenti legittimi.

- Al raggiungimento del numero massimo delle richieste di connessioni (parametro di configurazione del tool Slowhttptest) indirizzate al web server, gli attaccanti interrompono temporaneamente la loro azione per poi attaccare nuovamente dopo un tempo variabile randomicamente.

11.3.2.8 Attacco Forza Bruta

L'attacco di forza bruta scelto per il test-case è un attacco a dizionario. Nella sua forma più semplice, un attacco a dizionario è un tipo di attacco di forza bruta in cui gli hacker tentano di indovinare la password degli account di un utente scorrendo rapidamente un elenco di parole, frasi e combinazioni numeriche comunemente utilizzate. Una volta decifrata una password mediante un attacco a dizionario, l'hacker può utilizzarla per accedere al profilo utente.

11.3.2.9 Testbed Attacco hydra vs Apache Server

Il test implementato per la generazione del dataset, che riproduce in maniera controllata un attacco di forza bruta, si compone di:

- un server Kali Linux (attaccante) installato presso la sede di Portici e su cui è abilitato un tool che simula un attacco di forza bruta
- un apache server (vittima) installato presso la sede di Casaccia

Implementazione attacco di forza bruta vs Apache Server: l'attacco è stato sviluppato con il supporto del tool Hydra installato sul server Kali Linux (attaccante), effettuando molteplici tentativi di accesso SSH al web server Apache con una tecnica di attacco a dizionario.

Nello specifico è stata applicata la seguente configurazione:

hydra -L userlist.txt -P rockyou.txt <IP_WebServer_Apache> ssh -t 4 -V -o ./hydra_results.txt

11.3.3 PostElaborazione del dataset

Il traffico dati è collezionato da DarkTrace e poi esportato in modalità push verso ElasticSearch in formato Json, grazie ad un'integrazione nativa frutto della collaborazione tra le due aziende.

11.3.3.1 Architettura Elastic Search

Come schematizzato in Figura 7, gli elementi chiave dell'architettura ElasticSearch sono:

- **Document:** ElasticSearch serializza e archivia i dati sotto forma di documenti JSON. Un documento, identificato da un ID univoco, è un insieme di campi con associazione coppie chiave-valore che permettono la memorizzazione dei dati.
- **Index**: L'indice è l'unità di archiviazione fondamentale in ElasticSearch, uno spazio dei nomi logico per l'archiviazione di dati che condividono caratteristiche simili. Un indice è una raccolta di documenti identificati in modo univoco da un nome o da un alias. Questo nome univoco è importante perché viene usato per indirizzare l'indice nelle query di ricerca e in altre operazioni.
- **Shard**: L'elemento chiave dietro la tolleranza ai guasti e la scalabilità di ElastiSearch è lo sharding. Gli indici vengono suddivisi in partizioni, una o molte. Uno shard è

essenzialmente una parte di un indice. In realtà è un indice Lucene che è la tecnologia alla base del modo in cui ElasticSearch indicizza i documenti. In questo modo è possibile prendere gli indici e dividerli orizzontalmente nel cluster. Pertanto, le operazioni di indicizzazione e ricerca possono essere eseguite su più sistemi per lo stesso tempo per lo stesso utente o thread. Inoltre, suddividendo gli indici in shard, si possono gestire le repliche in modo efficiente. È quindi possibile gestire shard primari e di replica per ogni indice. Ad esempio, se si dispone di un indice con 3 shard e 1 replica per ogni shard (come si può vedere nell'immagine sotto), si hanno in totale 6 shard. Affinché questa opzione sia fault tolerant, le repliche non possono mai essere allocate nello stesso nodo della partizione primaria replicata.

- **Node**: Un nodo è un'istanza indipendente di ElasticSearch che gestisce gli shards che appartengono a uno o più indici. I nodi possono avere ruoli diversi, ad esempio nodo di dati, nodo master e nodo di inserimento. Nella configurazione del "Progetto Integrato Cybersecurity dei sistemi energetici" è stato installato un solo nodo ElasticSearch.
- **Cluster**: Un cluster ElasticSearch è una raccolta di nodi interconnessi. Tutti i nodi di un cluster possono gestire le richieste dei client e comunicare tra loro. Ogni nodo di un cluster è proprietario di un subset delle partizioni che appartengono a un indice. Nella configurazione del "Progetto Integrato Cybersecurity dei sistemi energetici" il cluster è composto da un solo nodo.

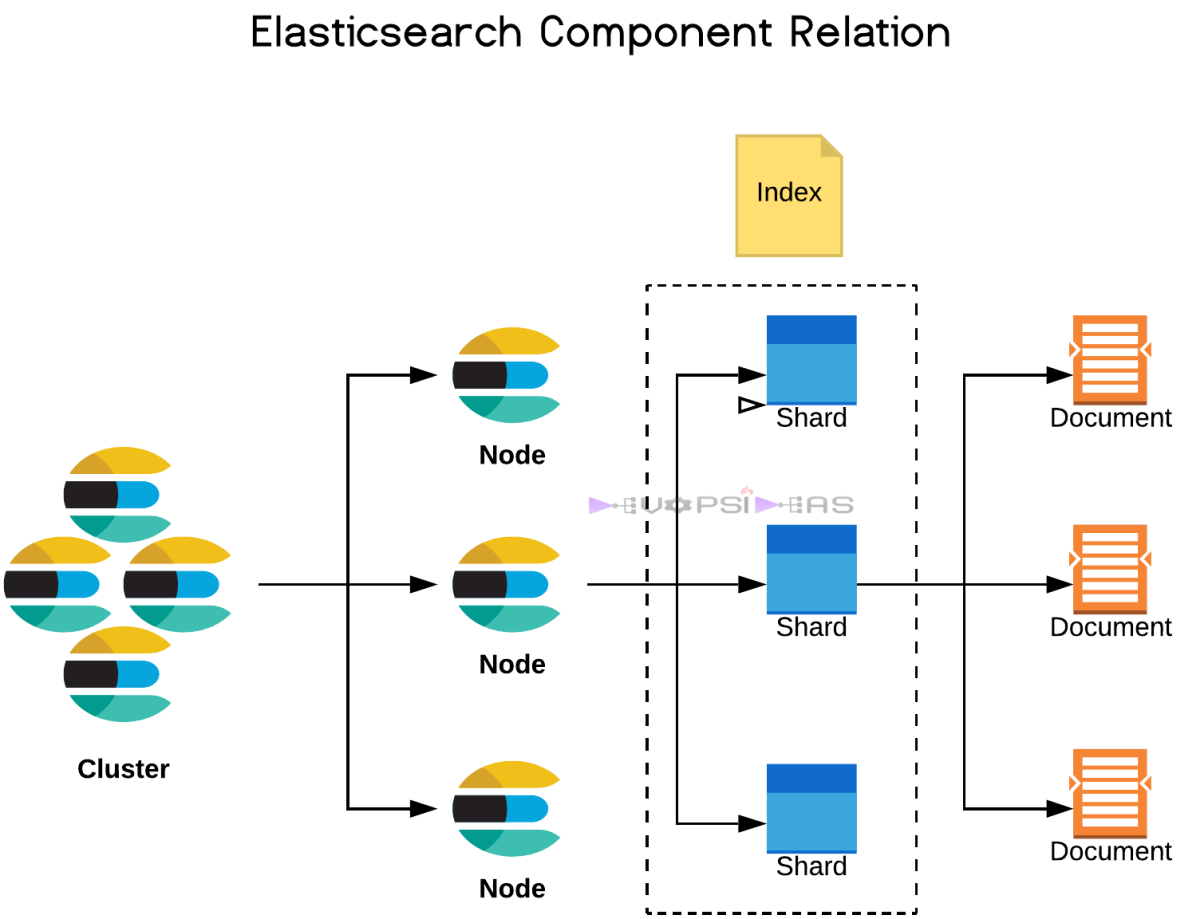

Figura 7: Elementi chiave dell'architettura ElasticSearch

11.3.4 Esempio di documento

Ogni documento è identificato univocamente dalla coppia di un ID documento ("_id"), univoco all'interno dello stesso indice, e di un indice ("_index"). Dall'esempio, è inoltre evidente come sia stato applicato un processo di pseudo-anonimizzazione (i cui dettagli sono descritti nei paragrafi precedenti) dell'indirizzo IP sia di origine che di destinazione al file di garantire un adeguato livello di privacy.


```
{
  "_id": "pYMzuZMB2NkcYgPlvh8R",
  "_index": "darktrace-dt-28154-01-2024.12.12",
  "_score": 3.6932797,
  "_source": {
    "@host": "dt-28154-01",
    "@timestamp": "2024-12-12T04:49:45",
    "@type": "conn",
    "conn_state": "SF",
    "conn_state_full": "SYN/FIN completion",
    "dest_ip": "1edb4d62e7099c6178d5297d4f65007f58ab099cf9f9ef5f0a6e7d070c486988",
    "dest_port": 9000,
    "duration": 0.0040149688720703125,
    "epochdate": 1733978985.796905,
    "history": "ShADadfF",
    "local_orig": true,
    "local_resp": true,
    "missed_bytes_orig": 0,
    "missed_bytes_resp": 0,
    "orig_bytes": 225,
    "orig_ip_bytes": 493,
    "orig_pkts": 5,
    "orig_ttl": 63,
    "proto": "tcp",
    "resp_bytes": 573,
    "resp_cc": "IT",
    "resp_ip_bytes": 841,
    "resp_pkts": 5,
    "resp_ttl": 62,
   "service": "http",
    "source_ip": "cca5d72e4cd19d47753ebbd685675716b3f8f3d1cc63645c4d15bd586c3ce5c8",
    "source_port": 58676,
    "start_ts": 1733978985.796905,
    "uid": "CbuabK1B38wWvZPX0b00",
    "vlan": "20"
  }
}
```