UNIVERSITÀ DEGLI STUDI DI UDINE

Dipartimento di Scienze Matematiche, Informatiche e Fisiche

TESI DI LAUREA IN INFORMATICA

Analisi e detection di attacchi ransomware per la protezione di dispositivi IoT in reti domestiche

CANDIDATO
Sebastiano Morson

RELATORE
Prof. Gian Luca Foresti

Anno accademico 2021-2022

Contatti dell'Istituto
Dipartimento di Scienze Matematiche, Informatiche e Fisiche
Università degli Studi di Udine
Via delle Scienze, 206
33100 Udine — Italia
+39 0432 558400
https://www.dmif.uniud.it/

© 2024 Sebastiano Morson
This work is shared under the Creative Commons 4.0 License Attribution-NonCommercial-ShareAlike.

A mia sorella,
le radici del lavoro sono amare, ma il frutto è dolce.

Ringraziamenti

Prima di procedere con la trattazione, vorrei dedicare questa pagina per ringraziare tutte le persone che mi sono state più vicine in questi tre anni di studi universitari.
In primis un ringraziamento va a mio padre e mia madre per avermi dato l'opportunità di intraprendere questo splendido percorso durante il quale ho potuto conoscere tante persone meravigliose ed essere costantemente stimolato a migliorarmi. Sono ancora lontano da ciò che voglio diventare, ma grazie a voi adesso sono un passo più vicino.
Un sentito ringraziamento al mio relatore, Professore Gian Luca Foresti, che mi ha seguito nella scrittura di questa tesi e che si è sempre dimostrato gentile e disponibile.
Un ringraziamento speciale è per mia zia Donatella. Il tuo sostegno e i consigli che hai saputo darmi durante questo percorso sono stati preziosi.
Un grazie di cuore a Gaia. Ci sei stata nella gioia e nello sconforto, ascoltandomi e non smettendo mai di credere in me e nelle mie capacità. Grazie per essere stata la mia polvere di neve nelle giornate più cupe.
Un doveroso ringraziamento va ai miei amici Cristiano, Giulia, Daniel, Othniel e Laura per le risate e i momenti di sfogo. Anche se ora le nostre strade si stanno dividendo non dimenticherò tutti i momenti di spensieratezza passati assieme.
Grazie anche ai miei colleghi Alessandro, Sandro e Davide per tutte le lezioni e le birre dopo gli esami. Siete e sarete sempre per me non soltanto degli amici, ma un punto di riferimento sia sul piano professionale che umano.
Infine ringrazio me stesso, perchè nonostante le difficoltà di questi ultimi mesi ho continuato a perseverare e crederci fino alla fine.

Indice

Elenco delle tabelle

Elenco delle figure

1

Introduzione

La digitalizzazione e Internet hanno permesso di far crescere l'economia globale come nulla prima d'ora nella storia ha mai saputo fare, oltre a rivoluzionare il modo di pensare e di vivere di miliardi di persone. Se però i lati positivi sono stati molti, e per molti, lo sono stati anche per la criminalità, che ha saputo evolversi e adattarsi per riuscire a creare nuove forme di profitto abusando dei moderni strumenti a disposizione. Una di queste sono gli attachi ransomware, che nel 2021 hanno causato un danno economico di 20 miliardi di dollari e che nel 2031 si stima porteranno a una perdita di $265 miliardi[1].

Il fenomeno dei ransomware è in costante evoluzione e minaccia di coinvolgere sempre più marcatamente non solo le aziende, ma anche, direttamente o indirettamente, le persone comuni, che in futuro potrebbero divenire allo stesso tempo vittime e complici a causa della permeazione del mondo IoT nelle realtà domestiche.

Il documento proposto ha l'obiettivo di indagare sulla natura degli attacchi ransomware offrendo infine uno strumento che possa contribuire alla protezione delle reti di dispositivi IoT.

Il contenuto di questa tesi è diviso in due parti: nella prima, che comprende i capitoli 2 e 3, vado a presentare il fenomeno degli attacchi ransomware. In particolare, nel capitolo 2 descrivo l'incidenza che questo tipo di compromissione ha nel mondo aziendale e quali siano le conseguenze dirette e indirette a cui sono esposte le vittime. Nel capitolo 3 illustro le differenze e le similitudini dei ransomware rispetto ad altri malware ed espongo l'anatomia di un moderno attacco ransomware.

Nella seconda parte, composta dai capitoli 4 e 5, descrivo uno strumento di intrusion detection da me ideato e realizzato, per la messa in sicurezza dei dispositivi smart al fine di impedire che i cyber criminali possano sfruttarne le capacità di calcolo e le scarse misure di protezione di questi ultimi.

Nel capitolo 4 vado dunque a delineare il legame tra mondo IoT e minaccia digitale, con particolare enfasi sul modo in cui i cyber criminali possono giovare dalla compromissione dei dispositivi intelligenti.

All'interno del quinto capitolo descrivo lo strumento ISSS, acronimo di IoT Security Subnetting System. Questo tool open-source è pensato per fornire una sottorete sicura a cui poter connettere i dispositivi smart e su cui viene applicato un modello di machine learning per analizzare i flussi di rete al fine di individuare dispositivi infetti e impedire che questi possano continuare a comunicare con l'esterno.

Nel capitolo 6 concludo l'elaborato esponendo gli sviluppi futuri dello strumento descritto nel precedente capitolo, proponendo possibili miglioramenti al fine di ottimizzare la protezione offerta.

I

Presentazione e analisi degli attacchi ransomware

2

Panoramica sugli attacchi ransomware

La categoria dei ransomware, nota già a partire dagli anni '90 con il celebre PC Cyborg[2], dal 2014 ha iniziato ad essere sempre più presente come principale software malevolo impiegato negli attacchi informatici in tutto il mondo, tanto che, dallo studio condotto da Datto per l'anno 2021 sui casi di infezioni da malware nei confronti di aziende medio-piccole, risulta presente nel 64% dei casi, sopra a qualsiasi altra forma di minaccia software.
I ransomware, come altri tipi di malware, ereditano il nome dal loro comportamento, che in questo caso è la cifratura delle risorse contenute all'interno dei sistemi, o il blocco degli stessi, al fine di obbligare la vittima a pagare un riscatto per il ripristino dei dispositivi compromessi.
La natura estorsiva dei ransomware fa sì che qualsiasi tipo di realtà da cui può derivare un ritorno economico, diretto o indiretto, per l'attaccante, possa finire per divenire vittima.
Per quanto la platea di potenziali obiettivi sia estremamente vasta, confrontando i report sugli attacchi ransomware svolti negli ultimi anni da parte di alcune delle maggiori aziende in ambito informatico e di sicurezza, ho potuto evidenziare come le categorie di mercato primariamente attaccate siano state il settore sanitario, governativo, IT e finanziario (vedasi figura2.1).
La ragione dietro a questa stabilizzazione dei target, è legata al ruolo critico di tali infrastrutture, che permette ai criminali di fare leva sui danni economici derivanti dall'interruzione di servizio per richiedere riscatti più elevati e aumentare la probabilità che le vittime paghino. In figura 2.1 sono schematizzate le relazioni tra numero di aziende colpite per ciascun settore e numero di attacchi a loro carico.

2.1 Conseguenze degli attacchi ransomware

Per quanto il riscatto richiesto alle vittime possa essere molto alto, la pericolosità degli attacchi ransomware non risiede solamente nella richiesta di una somma di denaro per il ripristino del sistema, ma bensì in una serie di conseguenze che vedono coinvolta in primo luogo la vittima della violazione e in seconda battuta i partner/clienti ad essa associati.
Descrivo ora gli effetti principali derivanti dagli attacchi ransomware.

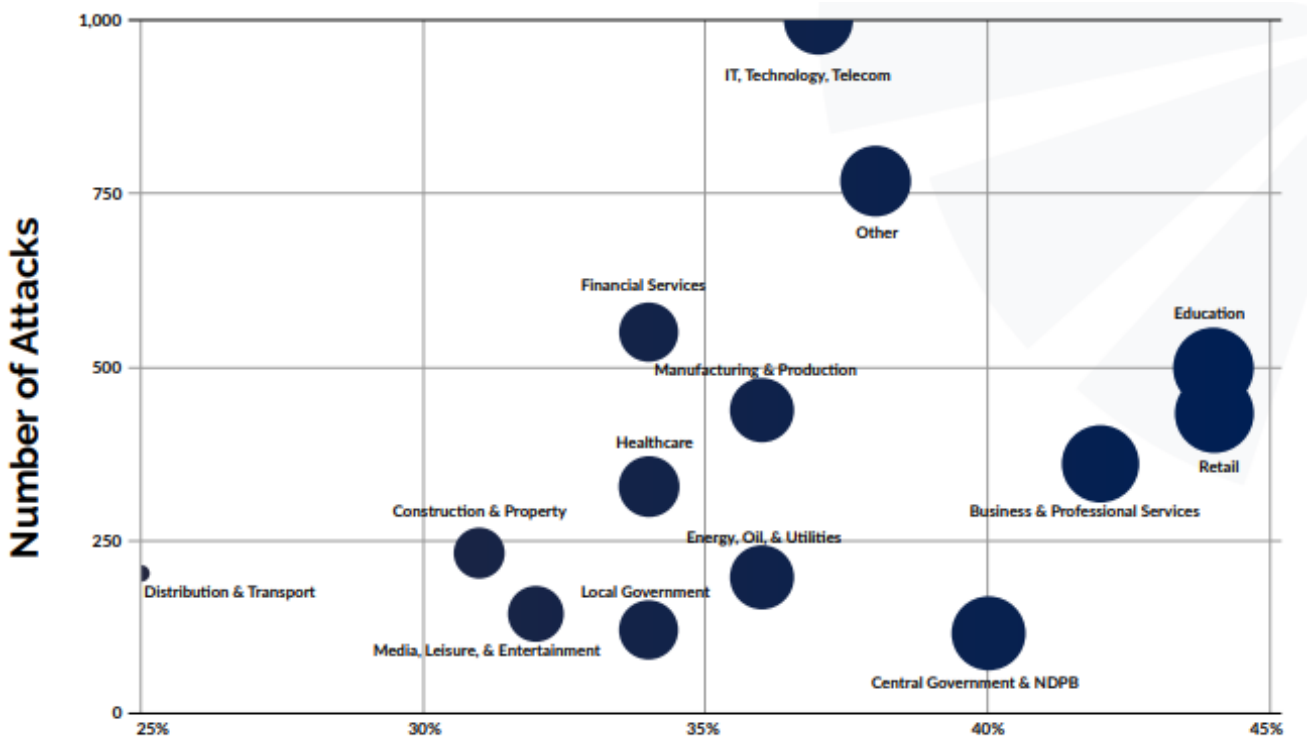

Figura 2.1: Numero di attacchi ransomware per industria [3]

2.1.1 Interruzione dei servizi e riscatto iniziale

Come anticipato, la ripercussione più evidente degli attacchi ransomware riguarda l'interruzione dei servizi e il riscatto iniziale. Il valore richiesto dell'attaccante per interrompere il blocco delle infrastrutture generalmente varia in base al tipo di target e al numero di informazioni che l'attaccante detiene sulla vittima. Banalmente, un attaccante in possesso dei sistemi di una grossa multinazionale chiede a quest'ultima un riscatto nettamente più alto rispetto a un'azienda di importanza minore.
I costi del riscatto vanno poi a sommarsi a quelli di ripristino dell'infrastruttura colpita e a quelli di downtime. L'interruzione dei servizi, che in media dura 22 giorni [4], può venire infatti a costare alle aziende fino a 50 volte la cifra domandata dai cyber criminali [5].

2.1.2 Contraccolpo reputazionale

Le aziende in seguito a un attacco ransomware possono subire conseguenze pesanti sul piano delle relazioni con i clienti e i fornitori. La cattiva luce si può tradurre nel lungo termine in una ingente perdita economica.

2.1.3 Ripercussioni legali

Ai danni economici e reputazionali, si aggiungono infine le conseguenze legali. Gli attacchi ransomware negli ultimi anni sono spesso associati a data leak, fughe di dati che l'European Union Agency for Cybersecurity (Enisa) stima corrispondere mensilmente a 10Terabytes di risorse trafugate [6]. A conferma di questi dati, CrowdStrike, all'interno del suo report "Global Threat Report 2022", ha osservato nel

2021 un incremento dell'82% dei data leak associati ad attacchi ransomware "con 2,686 attacchi effettivi nel Dicembre 2021, contro i 1,474 in 2020"[7].
Queste fughe di dati, che nell'ultimo periodo coinvolgono sempre più spesso dati sensibili, hanno dirette conseguenze sul piano legale. La normativa Europea sui dati sensibili impone infatti, all'interno del GDPR (art. 83, par. 4, GDPR [8]), sanzioni pecuniarie "fino a 10 000 000 EUR, o per le imprese, fino al 2% del fatturato mondiale totale annuo dell'esercizio precedente, se superiore" in caso di violazione delle disposizioni.

2.1.4 Conseguenze indirette

I data leak derivanti dagli attacchi ransomware possono generare un effetto negativo non solo sulle vittime dirette delle compromissioni, ma anche sui soggetti a cui afferiscono le informazioni contenute in essi.
Per spiegare la gravità dei data leak voglio portare come caso di studio l'attacco ransomware che ha visto coinvolto il Comune di Gorizia a settembre di quest'anno [9]. In questa occasione la cyber gang LockBit aveva minacciato di pubblicare, e poi effettivamente pubblicato sul loro data leak site (DLS), 53GB provenienti dai sistemi comunali.
Accedendo al DLS della cyber gang ho potuto constatare come i dati trafugati dai criminali siano stati di varia natura:

- Dati Personali
- Documenti di identità
- Informazioni sugli uffici di Anagrafe, Protocollo, Elettorale, Cimitero
- Bilanci interni
- Informazioni sui redditi e nascite
- Gare di appalto
- Informazioni sul Covid
- Tagli di Budget
- Stipule con le ditte
- Piani ferie interni

Esemplare è stato il caso di una signora, di cui non menzionerò il nome per ovvie ragioni di privacy, che sul proprio account comunale manteneva scansioni dei documenti di identità dei membri familiari, della patente del marito, foto e video degli stessi, referiti medici dei propri figli, oltre a due file word nominati come *passowrd_lavorative.doc* e*password_personali.doc*, contenenti l'associazione in chiaro di password, sito e nome utente per account ad uso aziendale e privato. La pubblicazione online di questo genere di documenti rappresenta ovviamente una minaccia sotto vari punti di vista, e agevola la diffusione di altri crimini informatici.
I rischi principali derivanti dai data leak sono:

Furto di identità

Attraverso i dati pubblicati online i truffatori possono eseguire attacchi di phishing mirato per indurre le vittime a rivelare le credenziali di accesso ai servizi da lui richiesti attraverso email o messaggi provenienti da indirizzi che riteniamo legittimi. Una volta ottenuti i segreti di cui ha bisogno, può spacciarsi per la vittima per compiere attività illecite o acquisti a nostre spese, esponendoci al rischio di importanti conseguenze sul piano legale ed economico.

Data and sex extortion

Un criminale informatico in possesso dei dati sensibili è potenzialmente in grado di violare i sistemi domestici con maggior facilità. Se il dispositivo violato è dotato di webcam, il rischio è che possa riprenderci furtivamente e utilizzare le registrazioni come forma di estorsione. Questo metodo di coercizione potrebbero essere applicato dal malintenzionato al fine di ottenere un ritorno economico oppure per forzare la vittima a compiere azioni atte alla violazione di un altro sistema (ad esempio la vittima potrebbe venir manovrata per lasciare una chiavetta infetta sul luogo di lavoro, finendo poi per diventare cosiddetti “trusted insiders” legalmente coinvolti nel reato).
Il problema non è assolutamente isolato e sotto controllo, tanto che nel luglio del 2021 l’FBI ha comunicato di aver ricevuto più di 16 mila denunce di sex extortion nel 2021, con perdite che superano gli 8 milioni di dollari [10].

Scalata sociale

I dati pubblicati online possono servire ad altri cyber criminali per intraprendere violazioni nei confronti di altri obiettivi. La violazione di un sistema informatico di un’azienda prevede l’attacco dei punti deboli della stessa. In molti casi questi ultimi sono dipendenti dell’organizzazione o dipendenti delle società partner che, manipolati attraverso phishing o ingegneria sociale, forniscono un canale di accesso ai malintenzionati. È chiaro dunque che i dati personali di un individuo comune acquistano molto valore agli occhi di un criminale informatico quando si appresta a compromettere un sistema aziendale.

Mancanza di controllo

Nel momento in cui i dati personali di un individuo diventano di dominio pubblico, quest’ultimo perde ogni forma di controllo sui modi e sui tempi in cui tali informazioni circolano e vengono utilizzati da terzi. La mancanza di potere nella gestione dei propri dati espone ai rischio di divenire in futuro complici ignari di nuovi attacchi. Non è inverosimile ipotizzare, infatti, che i criminali informatici potrebbero impiegare i dati sensibili degli utenti all’interno di nuove strategie di attacco volte a eludere i sistemi di sicurezza, esponendoci ancora di più a violazioni e pericoli.

3

Definizione del malware e descrizione di un attacco ransomware

Dopo aver descritto il panorama odierno degli attacchi ransomware, mi concentro ora nella descrizione delle differenze e similitudini con altri malware e sull'anatomia generale di una campagna ransomware.

3.1 Tipologie di malware

Con il termine malware ci si riferisce a un tipo di software o firmware malevolo, che nello specifico "intended to perform an unauthorized process that will have adverse impact on the confidentiality, integrity, or availability of an information system"[11]. Nel mondo i ransomware non sono l'unico tipo di malware, ma ne esistono diverse varietà che si differenziano nelle tecniche e negli obiettivi per cui sono stati realizzati. Di seguito presento alcune delle tipologie di malware più conosciute.

- ADWARE
  Si tratta di un software che mostra a schermo avvisi e banner, spesso contenuti all'interno di pagine web durante la navigazione dal browser. In genere il contenuto visualizzato cerca di apparire legittimo utilizzando nomi e informazioni associati a brand noti, altre volte invece propone soluzioni di sicurezza o fantomatiche vincite. Può trovarsi su ogni dispositivo in quanto utilizza il browser

- SPYWARE
  Lo spyware è un software che osserva e registra le azioni portate avanti dall'utente e le comunica a un soggetto terzo. Nonostante l'utilizzo di uno spyware non sia per forza legato ad attività illecite, le azioni compiute da uno spyware sono spesso non autorizzate dall'utente, per questo è comunemente realizzato per apparire il più possibile trasparente all'interno del sistema così da evitare le rilevazioni.

- VIRUS
  Si tratta di un software malevolo allegato a un altro programma che, una volta eseguito, si auto replica modificando altri programmi e infettandoli con il suo stesso codice.

- WORMS
  Come per i virus, anche i worm possiedono la caratteristica di essere auto replicanti, ma a differenza

dei primi, essi sono in grado di diffondersi all'interno dei sistemi senza che l'utente debba interagire con le risorse compromesse per avviare l'infezione. Attualmente le funzionalità dei worms possono trovarsi all'interno di altri tipi di malware come tecnica per aumentare lo spread della compromissione.

- TROJAN
  Questo tipo di software ha lo scopo di realizzare un canale di comunicazione tra la macchina in cui viene installato e un dispositivo terzo. In genere questo tipo di malware viene distribuito attraverso strumenti apparentemente legittimi, ma contraffatti che inducono al loro utilizzo senza preoccupazioni. Una volta avviato il trojan è in grado di effettuare operazioni di upload e download di file, e in generale offre all'operatore il completo controllo della macchina e delle periferiche a disposizione. Per questo motivo durante un attacco è spesso utilizzato come malware d'appoggio per altri tipi di strumenti.

- ROOTKIT
  Sono una forma di malware che permette all'attaccante di elevare i propri permessi a quelli di root. I rootkit sono software che tipicamente vengono realizzati per rimanere nascosti sia dall'utente che dal sistema operativo e dagli altri software installati.

- KEYLOGGER
  I keylogger svolgono un'azione molto simile a quella degli spyware, ma si distinguono da essi dal momento che limitano le azioni di logging agli eventi generati da tastiera. Attraverso un keylogger un utente terzo è in grado di catturare tutte le digitazioni compiute permettendo quindi di estrapolare credenziali di accesso e informazioni sensibili.

- MALICIOUS CRYPTOMINING o CRYPTOJACKING
  I malware appartenenti a questa categoria permettono a un utente terzo di guadagnare sfruttando le capacità di calcolo del computer compromesso per effettuare il mining di cripto valute come Bitcoin o Monero. In genere è installato tramite trojan.

- EXPLOIT
  Gli exploit sono software che approfittano dei bug e delle vulnerabilità dei sistemi per ottenerne l'accesso. Vengono utilizzati dai malintenzionati come prima forma di attacco attraverso cui potersi infiltrare all'interno dei sistemi per attuare in seguito altre tecniche di compromissione.

- SCAREWARE
  Gli scareware è una forma di malware che utilizza tattiche di social engineering per causare shock, ansia o la percezione di una minaccia per convincere l'utente a comprare il software indesiderato. Si differenziano dagli adware in quanto rispetto ad essi, gli scareware aggiungono il pericolo di un'imminente minaccia per far leva psicologicamente sull'utente.

- WIPER
  I wiper sono una tipologia di software malevolo utilizzata per compromettere in modo irreversibile i dati contenuti all'interno del dispositivo. I wiper tradizionalmente per danneggiare il sistema

applicano uno o più di questi metodi: sovrascrivono i file, cifrano i file, sovrascrivono il MBR. L'utilizzo di questo tipo di malware è generalmente legato a operazioni di insabbiamento delle prove, sabotaggio e cyber war. È interessante notare come nel 2022 è stato individuato un incremento delle varietà di wiper, impiegati principalmente per sferrare pesanti attacchi contro le compagnie ucraine.

3.2 Tipologie di ransomware

Un ransomware come definisce Fortinet, è "uno specifico tipo di malware o software malevolo che tiene in ostaggio dei dati chiedendo un riscatto. La minaccia è di pubblicare, bloccare o corrompere dei dati - o prevenire l'utente da lavorarci o accedere ai suoi computer, a meno che le richieste degli attaccanti non vengano soddisfatte" [12] . Un ransomware è dunque un programma indesiderato che utilizza le risorse contenute all'interno della macchina compromessa per attuare una forma di ricatto contro i legittimi proprietari.
Attualmente le categorie di ransomware più conosciute sono tre e si differenziano essenzialmente nella tipologia di estorsione utilizzata e nella tipologia di risorse compromesse.
Di seguito riporto una breve descrizione per ciascuna delle tre categorie oltre a definire la differenza tra fileless e file-persistent ransomware.

3.2.1 Locker ransomware

I locker ransomware sono malware che vietano l'accesso alle risorse della macchina impedendo alla vittima di poter raggiungere i propri file. Ad essere compromesse sono le funzionalità base del dispositivo come per esempio le schermate di login o i dispositivi di puntamento e di scrittura. I dati in questo caso non vengono alterati in alcun modo e la minaccia si limita al confinare l'utente fuori dai propri sistemi fino a che non venga pagata la cifra richiesta.

3.2.2 Crypto ransomware

I crypto ransomware sono la categoria più avanzata e potenzialmente letale di ransomware. Al contrario dei locker ransomware, i crypto ransomware affliggono i dati personali contenuti nella macchina target. I malware di questo tipo, una volta in esecuzione, ricercano all'interno dell'ambiente i file aventi specifiche estensioni, cifrandoli al termine. Una volta completata questa prima fase il programma si manifesta all'utente attraverso una "ransom note" in cui vengono comunicate le istruzioni per ottenere la chiave di decifratura.

3.2.3 Doxware

La categoria dei doxware è una tipologia di ransomware che non apporta modifiche ai sistemi che affligge. Il funzionamento di questi ransomware si basa sulla fantomatica minaccia di diffusione pubblica di informazioni recuperate all'interno del sistema della vittima. Non si tratta di minacce fondate, bensì di pure intimidazioni che però mettono in allarme l'utente che a volte, preso dal panico, finisce per pagare il riscatto.

3.2.4 Fileless e File Persistent Ransomware

I ransomware possono essere persistenti o fileless. Nel primo caso il ransomware applica passaggi per garantire la stabilità temporale del malware all'interno del sistema, in modo che esso venga eseguito ad ogni riavvio della macchina o venga nuovamente installato in caso di rimozione. I ransomware di questo tipo hanno il vantaggio di poter protrarre nel tempo l'operazione di compromissione, in quanto permettono di eseguire il payload di cifratura più avanti nel tempo rispetto all'esecuzione delle operazioni di persistenza. Questa capacità torna utile nei casi in cui si voglia nascondere all'utente cosa abbia generato l'esecuzione del malware. Nel caso di ransomware fileless, questi non si stabiliscono nel sistema vittima e non richiedono di scaricare tool aggiuntivi, bensì applicano un approccio *live-off-the-land* in cui ad essere utilizzati sono strumenti pre-installati sulla macchina (ad esempio lo strumento Windows PowerShell o i comandi whoami, ping e net). Questi tipi di ransomware sono vulnerabili al reboot del sistema, per questo una volta avviati eseguono direttamente il payload di cifratura. È da evidenziare come gli attacchi basati su tecniche fileless siano notevolmente più efficaci degli attacchi tradizionali, tanto che nel 2018, il 35% degli attacchi totali è stato perpetrato tramite malware fileless e con un tasso di successo 10 volte superiore [13].

3.3 Differenze e similitudini dei ransomware rispetto agli altri malware

Le tipologie di malware illustrate all'interno del paragrafo 3.2 hanno messo in luce come tendenzialmente essi cerchino di celare la loro presenza all'utente del sistema. Questo non vale per i ransomware, programmi che rivelano la loro esistenza in modo inequivocabile nel momento in cui reclamano il riscatto alla vittima. Tale comportamento non è sclusivo dei ransomware, bensì è una caratteristica comune dei malware appartenenti alla categoria dei *programmi malevoli di estorsione digitale*, come adware e scareware. Rispetto a questi ultimi però, i malware appartenenti alla classe dei ransomware si manifestano soltanto dopo che le risorse del dispositivo sono state compromesse, perciò nel momento in cui appare il messaggio di richiesta di riscatto è già troppo tardi.
Un'altra importante distinzione dei ransomware rispetto ad altri malware è che non danneggiano il dispositivo infetto, difatti cercando di preservarne le funzionalità. Tale caratteristica è generalmente motivata dalla necessità di lasciare la possibilità all'utente di effettuare il pagamento del riscatto mediante lo stesso dispositivo. In particolare, rispetto ai wiper i ransomware prevedono di ripristinare completamente l'accesso ai file o alla macchina.

3.4 Anatomia di un attacco ransomware

Come definito all'interno del paragrafo 3.3, i ransomware sono malware che nel loro stadio finale espongono platealmente la loro presenza alla vittima. Per questo motivo il loro impiego avviene principalmente come unico payload ad esempio all'interno di campagne di email malevoli, o come ultimo atto offensivo di un attacco articolato in più fasi.
Nel seguito descriverò gli eventi riscontrabili nel caso di una compromissione basata su un generico crypto ransomware persistente che sfrutta come vettore iniziale un loader, ma la metodologia illustrata

è facilmente generalizzabile al caso di un attacco più complesso. Le tecniche utilizzate e le attività possono variare nel tipo e nei tempi di esecuzione in base al ransomware impiegato e al grado di presenza dell'attaccante.

3.4.1 Accesso iniziale

La prima fase di un attacco ransomware è l'accesso iniziale al sistema, ossia l'attaccante cerca di garantirsi un punto di appoggio all'interno di un dispositivo della vittima da cui poter perpetrare azioni contro di esso. L'obiettivo finale per l'attaccante è riuscire a introdurre il ransomware. Il report ENISA "Threat Landscape for ransomware attacks 2022"[14], in accordo con il MITRE ATT&CK framework, riporta come tecniche di infiltrazione maggiormente riscontrate:

- T1133 External Remote Services
  Questo vettore di attacco si basa sull'idea di compromettere servizi di connessione remota legittimi come le VPN per inserirsi all'interno della rete privata.

- T1566 Phishing
  Gli attacchi di phishing consistono nell'invio di messaggi ingannevoli volti a manipolare la vittima per farle compiere azioni che altrimenti non compirebbe. Il phishing può essere finalizzato verso uno specifico obiettivo (in questo caso si parla di spearphishing) oppure non indirizzato, ossia l'azione è rivolta verso chiunque (ad esempio tramite campagne di email spam).

- T1195 Supply Chain Compromise
  Questa tecnica mira a corrompere furtivamente il meccanismo di distribuzione di un articolo, o l'articolo stesso, in modo tale da compromettere indirettamente chiunque ne faccia uso. L'uso di questa strategia è motivato dal fatto che spesso le aziende fanno affidamento su strumenti di terze parti per sviluppare i loro prodotti. Questo tipo di tecnica di accesso è molto efficace in presenza di policy aziendali che non prevedono che gli strumenti terzi siano periodicamente soggetti a una verifica dell'integrità e della provenienza.

- T1078 Valid Accounts
  La tecnica in questione consiste banalmente nell'abuso di credenziali rubate. Questa tecnica fondamentalmente permette di svolgere un numero di azioni direttamente proporzionale all'autorità detenuta dal profilo che si utilizza. È possibile che l'attaccante sfrutti più di un account per svolgere le sue attività, così da abbassare le possibilità di essere individuato.

Chiaramente queste non sono gli unici vettori iniziali. Un attaccante a seconda dell'importanza e del valore di un attacco, può spingersi ad usare un trusted insider che funga da infiltrato in grado di installare manualmente malware all'interno dell'infrastruttura, come nel celebre caso che ha coinvolto la multinazionale Tesla [15].
È importante sottolineare che l'utilizzo di una specifica tecnica è condizionato alle informazioni che l'attaccante detiene sul target. Se l'azione offensiva è diretta contro un'entità di cui si possiede sufficienti informazioni, è probabile che vengano utilizzati tali dati per l'accesso iniziale. Al contrario, avendo soltanto informazioni generiche è più plausibile vengano utilizzate tecniche di phishing.

Esecuzione del loader/dropper

Durante la fase di accesso ad essere introdotto non è il ransomware, bensì uno script privo di capacità offensive. Quest'ultimo, una volta avviato, installa furtivamente all'interno del sistema tutte le componenti essenziali per proseguire la compromissione e installare correttamente il ransomware. Tale componente può essere di due tipi:

- *dropper*: un programma che ingloba all'interno del suo codice il malware (anche chiamato payload). Una volta eseguito, il dropper estrae il payload e lo salva nella memoria del dispositivo. Non richiede una connessione alla rete.

- *loader (o stub)*: un programma con funzionalità pressoché identiche al dropper, ma che rispetto a questo ha bisogno di contattare un server per ottenere gli strumenti da caricare nella macchina target.

Sia il dropper che il loader sono in genere programmi piuttosto semplici che mutano spesso. Per questo e per il fatto che spesso possiedono la capacità di rilevare se sono eseguiti su una macchina virtuale, generalmente risultano piuttosto difficili da individuare. Possono essere inoltre persistenti o non persistenti. Nel primo caso, una volta eseguiti per la prima volta, tentano di stabilirsi all'interno del sistema in modo da essere automaticamente eseguiti dopo un riavvio, uno spegnimento o una cancellazione. Nel secondo caso al termine dell'esecuzione entrambi i programmi rimuovono automaticamente dal sistema se stessi oltre ai possibili artefatti da cui è possibile ricondurre alla loro identità.

Comunicazione con il server C&C

Una volta installato, un loader stabilisce una connessione con un server malevolo per ricevere istruzioni sui comportamenti da adottare. Questo genere di server, chiamati server *command-and-control* (spesso abbreviati in C&C o C2), sono fondamentali ai cyber criminali per effettuare movimenti laterali, in quanto permettono loro di inviare e ricevere dati da una rete che hanno compromesso. Gli strumenti da scaricare si trovano all'interno di un server di download, a cui il loader è istruito a connettersi dopo la prima connessione con il C&C server (figura 3.1).

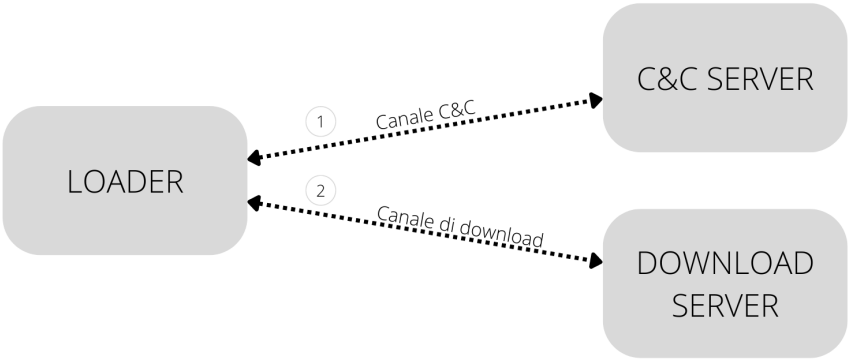

Figura 3.1: Schema di funzionamento dei loader

La comunicazione tra un loader e un C&C, per eludere i moderni sistemi di analisi del traffico di rete, spesso utilizza

- il protocollo TLS [16]
- porte conosciute e generalmente associate ad altri servizi
- comunicazioni DNS
- servizi conosciuti e fidati di terze parti come Telegram e servizi Google

Per la stessa ragione, l'indirizzo del server C&C con cui comunica il loader solitamente cambia dinamicamente nel tempo [17]. Una delle prime informazioni che lo stub richiede al server è perciò l'indicazione su quando e come dovrà contattare il server C&C la volta successiva.
Durante la prima comunicazione, al fine di personalizzare le successive tecniche di lateral movement e privilege escalation, è possibile inoltre che il loader condivida all'attaccante informazioni sull'host quali versione del sistema operativo, servizi in uso e informazioni generali sul sistema, in modo da scaricare i tool compatibili per la specifica macchina [18] oppure interrompere l'attacco se viene rilevato un layout di tastiera corrispondente a un territorio specifico, come nel caso di certi ransomware filo-governativi [19]. In base alla sua complessità e al tipo di attacco, può capitare che il programma ransomware finale venga sostituito al loader fin dalle prime fasi dell'attacco, oppure che venga scaricato un nuovo loader con più funzionalità come BazarLoader o SmokeLoader. Nel primo caso, il ransomware si comporta essenzialmente come uno stub, facendo frequenti controlli al server C&C e auto-update [20]. In questo caso il loader scarica dal download server l'eseguibile del ransomware che viene poi copiato all'interno di una directory locale (spesso la cartella */temp* o *%AppData%/local/temp*) ed infine eseguito (su sistemi Windows a volte si stabilisce all'interno di un processo comune come *svchost.exe*).
Una volta fatto ciò la comunicazione con il C&C viene ripresa dal ransomware.

Garantire la persistenza

Affinchè il ransomware non venga rimosso dal sistema, esso tenta di applicare meccanismi di persistenza, ossia "tecniche che un avversario utilizza per mantenere l'accesso al sistema in seguito a un riavvio, un cambiamento delle credenziali o un'interruzione che può fermare l'accesso" [21].
Le tecniche di persistenza cambiano da malware a malware. Ryuk [22] e LockBit [23] ad esempio aggiungono la chiave di registro di Windows

$$HKEY\_CURRENT\_USER\backslash SOFTWARE\backslash Microsoft\backslash Windows\backslash CurrentVersion\backslash Run$$

con valore il path dell'eseguibile, mentre Maze nella maggior parte dei casi utilizza credenziali utente rubate per creare nuovi account privilegiati.

3.4.2 Ricognizione

La ricognizione del sistema durante un attacco ransomware ha l'obiettivo di reperire sufficienti informazioni per poter disabilitare le difese presenti nella macchina, consolidare la presenza dell'attaccante, scovare la presenza e la posizione dei file utilizzabili in fase di estorsione ed infine permettere una diffusione del ransomware nei sistemi all'interno della stessa rete.

Esplorazione del sistema e privilege escalation

Per effettuare un'esplorazione del sistema il ransomware scarica dal download server diverse componenti aggiuntive, come ad esempio Cobalt Strike, strumento che nel quarto trimestre del 2020 è stato impiegato nel 66% degli attacchi ransomware [24]. Il suo impiego è motivato dal numero di funzionalità di ricognizione del sistema che mette a disposizione, oltre alla capacità di agire come un downloader grazie alla funzionalità di beacon HTTP, HTTPS, o DNS [25].
Altri esempi di tool utilizzati durante la fase di ricognizione sono

- Mimikatz, un tool di post-exploitation utilizzato per ricavare le credenziali di un amministratore di dominio locale e rilevato in diverse campagne come Maze e Sodinokibi
- BloodHound, utilizzato per scansionare il dominio dell'Active Directory e determinare i target più interessanti
- Process Hacker, utilizzato per scoprire e terminare processi e servizi [26][27]

La figura 3.2 mostra come questi strumenti all'interno di una campagna malware sono utilizzati in combinazione tra loro per muoversi all'interno del sistema.

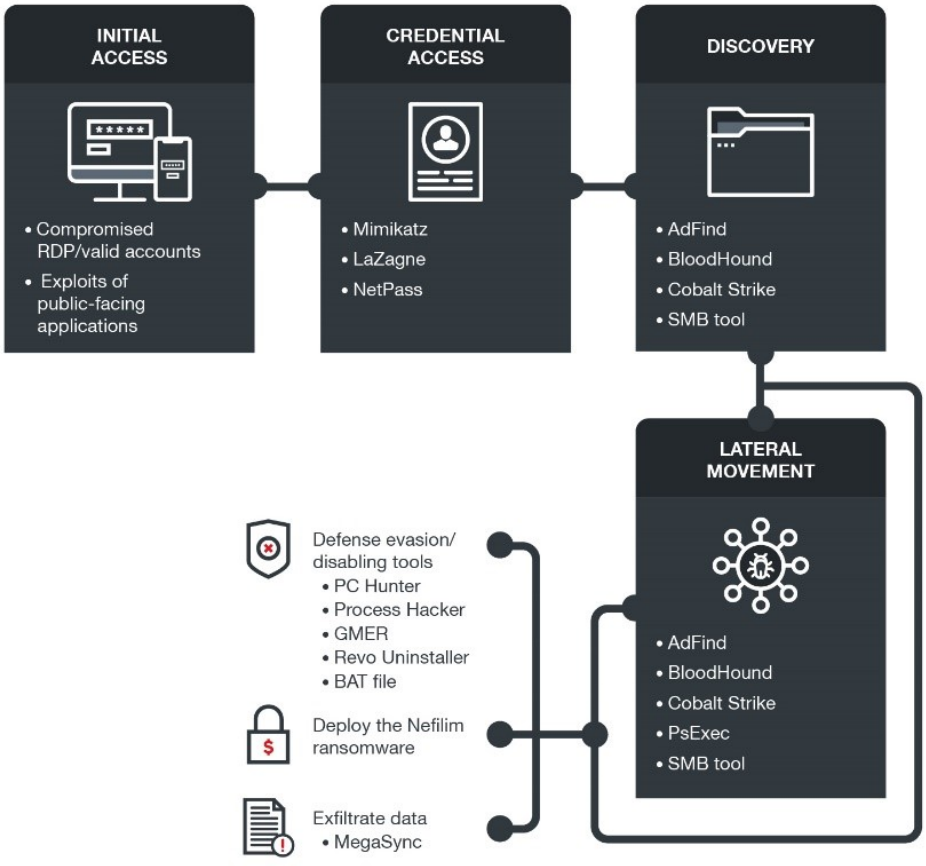

Figura 3.2: Lateral movement[28]

Esfiltrazione dei dati

Nel 70% dei casi, afferma l'analisi condotta da Coveware [29], gli attacchi ransomware prevedono l'esfiltrazione di dati dalla vittima. In genere le motivazioni che spingono l'attaccante ad attuare questa

procedura è per poter applicare strategie di double extortion o per stabilire il giusto prezzo del riscatto. Questa fase dell'attacco richiede particolare cautela da parte dell'attaccante, in quanto trasmettere risorse all'esterno può spingere la vittima a un'analisi approfondita del sistema.
Per esportare i dati fuori dalla rete, l'attaccante può avvalersi di:

- account di posta elettronica istituzionale compromessa
- servizi di messaggistica istantanea e team collaboration già presenti nel sistema (Skype, Slack, Teams, ...)
- servizi di file sharing come FreeFileSync e MEGA malware
- compromissione di protocolli di condivisione di rete
- data pumps
- cloud backup service

La trasmissione dei dati può essere inoltre continuativa o segmentata, e con payload di dimensione variabile. La scelta della strategia dipende dall'obiettivo dell'attaccante e dai tool di sicurezza presenti all'interno del sistema. Questa volubilità chiaramente permette all'attaccante di rendersi invisibile agli occhi delle soluzioni di data loss prevention e di monitoraggio del traffico di rete.

3.4.3 Deployment

Al termine della ricognizione e dell'esfiltrazione dei dati importanti, l'attacco passa alla fase di deployment del payload cifrante/bloccante del ransomware. Questo stadio prevede vari passaggi ad alto rischio eseguiti in un intervallo di tempo ristretto.

Spread del ransomware

I ransomware moderni possiedono features worm-like per cercare di distribuire il ransomware su tutti i dispositivi raggiungibili nella rete senza l'interazione con l'utente. Per farlo utilizza le informazioni ottenute durante la ricognizione del sistema. Alcuni esempi di metodi di spread del ransomware sono:

- sfruttare il Microsoft Group Policy Object per iniettare ed eseguire il ransomware mediante il domain controller
- sfruttare il protocollo SMB.
- abusare della Windows Management Interface (WMI)

[27]
Il famoso ransomware Wannacry [30] faceva uso dell'exploit EternalBlue che affliggeva il protocollo SMB nei sistemi Windows. Una volta eseguito, tramite una richiesta SMB echo, il ransomware riusciva a scoprire le macchine con tale vulnerabilità all'interno della rete. Trovata una macchina, veniva successivamente applicato l'exploit che utilizzava una tecnica conosciuta come heap spray per iniettare una backdoor in grado di installare una copia del ransomware.

Compromissione degli strumenti di recupero dei dati

L'obiettivo primario di un crypto ransomware è impedire l'accesso ai documenti all'interno del sistema. Per questo motivo, una volta in esecuzione, per evitare che il sistema venga ripristinato, il ransomware tenta di cancellare i backup presenti.
Per riconoscere e compromettere i sistemi di backup aziendali, i ransomware sfruttano i nomi assegnati di default alle directory create da questi prodotti. Questi path sono liberamente accessibili sulla documentazione del produttore, permettendo agli attaccanti di creare liste di percorsi da scansionare per rilevare e distruggere i file in esso salvati [2].
Un pattern frequente in molti ransomware è la cancellazione delle shadow copies generate dal Volume Shadow Copy di Windows. Per compromettere il catalogo di copie il metodo più comune è l'abuso dell'utility *vssadmin.exe* per lanciare il comando *vssadmin delete shadows /all /quiet* seguita in certi casi dall'istruzione *wmic shadowcopy delete*, che ha la stessa funzione, ma sfrutta il *Windows Management Instrumentation* [31].

La compromissione delle copie di backup può avvenire prima o dopo la cifratura. Questo perchè, se è vero che la vittima potrebbe recuperare i propri dati, è vero anche che i file salvati potrebbero non essere aggiornati o non venir ripristinati immediatamente, mentre un'azione di manomissione dei salvataggi di backup potrebbe allertare la vittima impedendo che la cifratura avvenga. Esempi di ransomware che distruggono le copie di backup dopo il processo di cifratura sono Wannacry, GandCrap, Ryuk e MegaCortex [27].

Cifratura

Il payload cifrante di un ransomware una volta eseguito avvia un'enumerazione dei file. La scelta dei file da cifrare è comunemente basata su *white-list* o su *black-list*. Le white-list sono liste che definiscono tutte le estensioni e i path che il ransomware non deve intaccare per non compromettere il funzionamento del sistema. Qualsiasi altro file che non presenta nel proprio path alcuna delle stringhe presenti nella white-list, viene cifrato. Un esempio di white-list questo genere di liste è visibile in figura 3.3

RyukReadMe.html	UNIQUE_ID_DO_NOT_REMOVE	PUBLIC	PRIVATE
\\Windows\\	sysvol	netlogon	bin
boot	dev	etc	lib
initrd	sbin	sys	vmlinuz
run	var	AhnLab	Chrome
Mozilla	$Recycle.Bin	WINDOWS	dll
ahrmlog	.aini	.lnk	

Figura 3.3: Estensioni dei file cifrati dal ransomware Ryuk[3]

Al contrario le black-list sono utilizzate per definire il set di stringhe che devono essere presenti all'interno del percorso del file affinché questo sia visto come un target.
Individuato un file importante, il processo di cifratura può avvenire su una copia oppure direttamente sul file originale. Nel primo caso il ransomware crea un doppione cifrato, per poi successivamente rimuovere l'originale. Nella versione in-place la cifratura avviene direttamente sul file originale. In entrambi i casi è frequente che il nome e l'estensione vengano modificati per poter compromettere la relazione tra la nuova versione e quella salvata nel catalogo VSS. Il vantaggio di quest'ultima strategia è che per la vittima risulta impossibile recuperare il file originale attraverso i tool di recovery [27].

Tecniche di cifratura

Prima di poter descrivere in che modo i moderni ransomware cifrano i file, è necessario distinguere e descrivere due categorie di algoritmi crittografici:

Crittografia simmetrica

La crittografia simmetrica è un tipo di crittografia che si basa su una chiave segreta di cifratura che viene utilizzata per cifrare e decifrare un certo contenuto (figura 3.4).

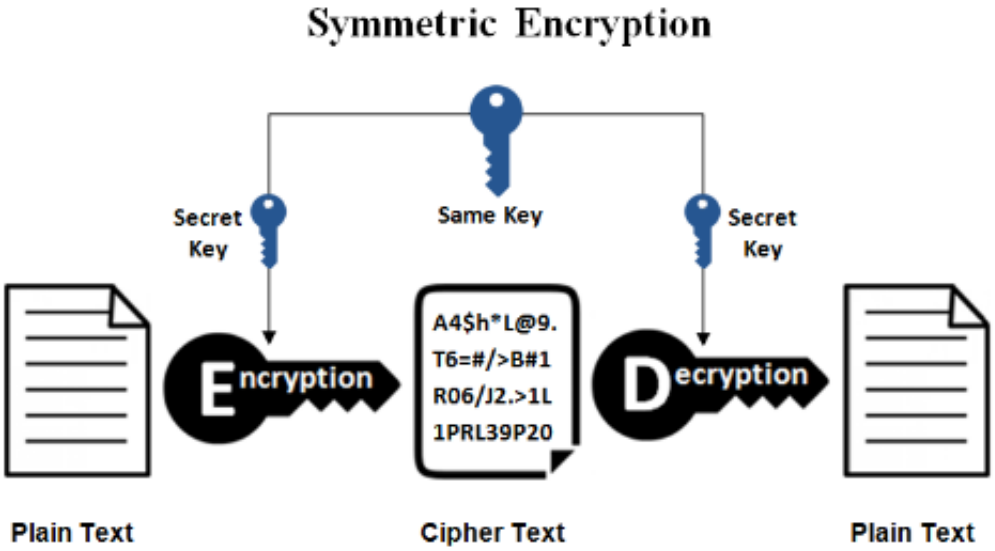

Figura 3.4: Funzionamento della cifratura simmetrica [32]

L'utilizzo da parte di un ransomware di un algoritmo a chiave simmetrica garantisce la capacità di cifrare rapidamente il contenuto dei file. Lo svantaggio è che per poter evitare attacchi *pay-once, decrypt-many*, i ransomware devono utilizzare chiavi sempre diverse. Questa necessità costringe quindi i ransomware con cifratura simmetrica a personalizzare gli eseguibili o a richiedere, in fase di esecuzione, la copia della chiave a un server C&C, esponendola al rischio di intercettamento.

Crittografia Asimmetrica

La crittografia asimmetrica (vedi figura 3.5) è basata sulla generazione di una coppia di chiavi pubblica e privata tali per cui:

- derivare il valore della chiave privata è computazionalmente irrealizzabile conoscendo solamente il

valore di quella privata.

- il contenuto cifrato attraverso la chiave pubblica è decifrabile unicamente attraverso la chiave privata

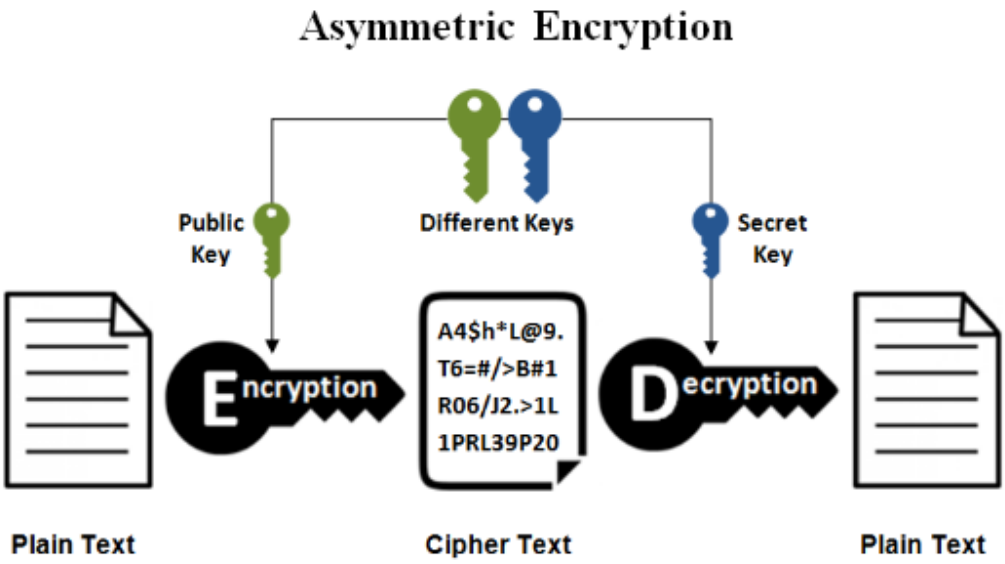

Figura 3.5: Funzionamento della cifratura asimmetrica [32]

Per un ransomware, il vantaggio di utilizzare un algoritmo a chiave pubblica-privata rispetto ad uno simmetrico risiede nel fatto che la chiave pubblica non necessita la riservatezza e può quindi essere scambiata in chiaro. Di contro questo tipo di algoritmi sono estremamente più lenti rispetto alla controparte simmetrica [33].
Se alcuni ransomware in passato utilizzavano approcci puramente simmetrici (JigSaw[34]) o asimmetrici (Unlock92), ora i moderni ransomware come CryptoLocker[35] utilizzano un approccio di cifratura misto, che permette loro di ottenere la velocità di cifratura dell'approccio basato su chiave simmetrica, ma con il livello di segretezza garantito dall'approccio basato su chiave asimmetrica.
Facendo riferimento alla figura 3.6:

1. Inizialmente l'eseguibile del ransomware contiene al proprio interno una chiave pubblica (in rosso) di cui l'attaccante detiene la controparte privata.

2. Il primo passaggio che il malware svolge è richiedere al server C&C una nuova chiave pubblica attraverso una richiesta cifrata con la chiave pubblica iniziale . In questo modo l'attaccante e il ransomware si autenticano vicendevolmente.

3. Successivamente il server risponde con la nuova chiave pubblica.

4. Il ransomware genera quindi una chiave simmetrica che utilizza per cifrare i file

5. Terminata la cifratura dei file, la chiave simmetrica viene cifrata con la chiave pubblica e viene rimossa ogni sua traccia dal dispositivo. A questo punto il processo termina con la garanzia da parte dell'attaccante di essere l'unico in grado di poter decifrare la chiave simmetrica per poter ripristinare i documenti.

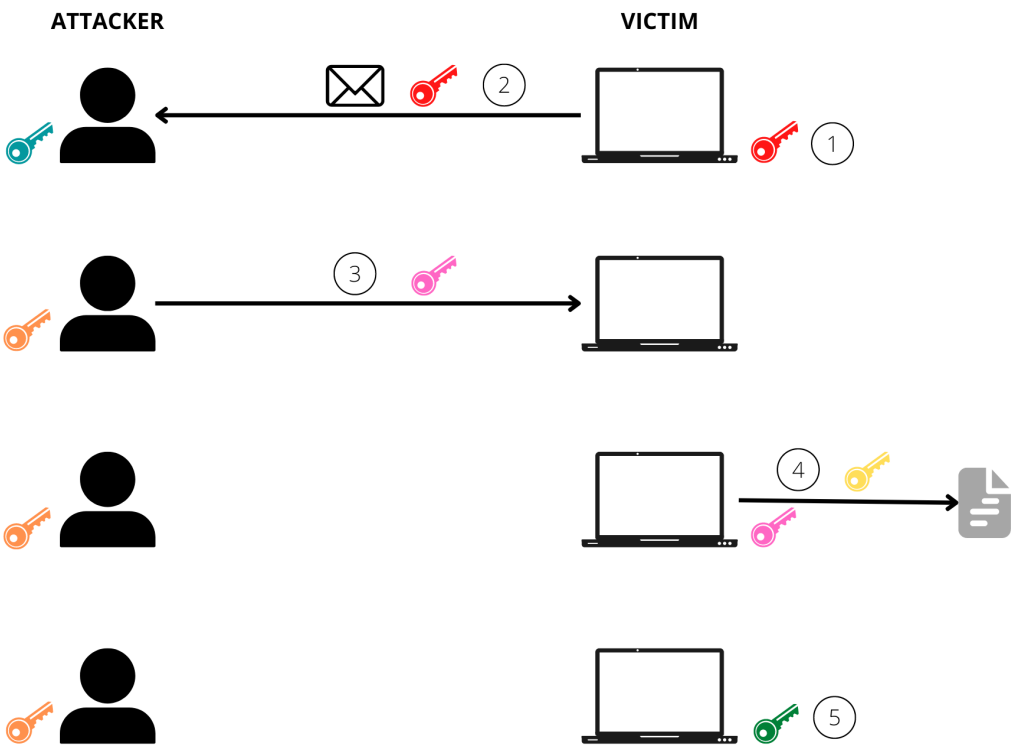

Figura 3.6: Scambio delle chiavi di cifratura dei moderni ransomware

Gli algoritmi maggiormente utilizzati sono RSA e HCC per quanto riguarda la generazione delle chiavi asimmetriche, mentre AES, RC4 e Salsa20 per la creazione delle chiavi simmetriche.

3.4.4 Estorsione

Ultimata la fase di cifratura e inserite le ransom note all'interno delle directory che contengono file cifrati dal ransomware l'attacco è ultimato e il cyber criminale può applicare le varie tecniche di estorsione. Un trend in rapida crescita e previsto nell'83% degli attacchi ransomware che hanno successo, è quella della triple extortion [36][37], ossia una pratica che consiste nel:

1. minacciare inizialmente la vittima di persistere nel mantenere bloccati o distruggere i dati contenuti nei sistemi colpiti
2. far leva sui dati sensibili esfiltrati, spostando la minaccia dalla distruzione/blocco alla diffusione pubblica o alla vendita online degli stessi, nel caso la vittima si rifiuti di pagare
3. aggiungere l'ulteriore minaccia di attacchi DoS durante la fase di ripresa dell'azienda in caso di mancato pagamento

II

Proposta di un tool per la protezione dei dispositivi IoT all'interno delle reti domestiche

4

Relazione tra mondo IoT e minaccia ransomware

4.1 Il problema dell'IoT nelle reti domestiche

L'International Telecommunication Union (ITU) definisce l'IoT come "A global infrastructure for the information society enabling advanced services by interconnecting (physical and virtual) things based on, existing and evolving, interoperable information and communication technologies"[38].
Negli ultimi anni il numero dei dispositivi appartenenti al mondo dell'IoT è stato caratterizzato da un costante trend positivo grazie al fatto che il loro utilizzo ha saputo permeare settori importanti come quello sanitario, industriale e domestico, tanto che Insider stima che nel 2027 gli oggetti connessi saranno 41 miliardi [39].
Per quanto la diffusione globale di questo genere di dispositivi semplificherà certamente la vita a milioni di persone, allo stesso tempo essi rappresentano un'importante minaccia alla sicurezza di imprese e cittadini. Gli oggetti smart venduti negli store di tutto il mondo sono infatti offerti ad una platea di persone solitamente prive della sensibilità sul rischio informatico che necessiterebbe l'utilizzo di tali apparecchi. Questo aspetto, unito al fatto che molto spesso le tecnologie IoT sono sviluppate prive di misure di sicurezza native, offre il terreno ideale per la diffusione di malware e minaccia di alimentare il numero e l'efficacia degli incidenti informatici di natura dolosa. Gli eventi dell'ottobre 2016 che hanno visto la botnet Mirai impiegata per sferrare un massiccio attacco DoS nei confronti del provider DNS Dyn sono un'evidente dimostrazione di ciò che è stato detto.

I ransomware, in particolare, possono giovare dall'insicurezza del mondo IoT per varie ragioni:

- gli apparecchi smart possono venir sfruttati come server C&C [40]
- i dispositivi IoT generalmente vengono utilizzati da individui non consapevoli dell'importanza di aggiornare i propri sistemi, perciò le vulnerabilità conosciute, correggibili attraverso un periodico aggiornamento degli strumenti, potrebbero essere usate come vettori di accesso
- in futuro i dispositivi IoT saranno largamente diffusi, questo implicherà per i cyber criminali la possibilità, prendendo come target le persone comuni invece che le aziende, di guadagnare cifre significative pur chiedendo un riscatto anche modesto ai singoli individui

AZIONE	PERCENTUALE DI RILEVAMENTO
payload dei ransomware	63%
utilizzo di tool come PSExec e Cobalt Strike	53%
esfiltrazione dei dati	49%
accesso iniziale	42%
movimento laterale	31%

Tabella 4.1: Percentuale di rilevamento delle soluzioni di sicurezza business

- Più gli oggetti smart rivestiranno posizioni critiche all'interno delle abitazioni (termostati, serrature, sistemi di illuminazione, ...), più probabile sarà che la vittima paghi piuttosto di attendere che i suoi sistemi tornino operativi, soprattutto se il riscatto richiesto è meno costoso del ripristino del sistema

4.2 I limiti delle soluzioni EDR nelle realtà domestiche e la sfida per la sicurezza dei dispositivi IoT

La sicurezza del mondo IoT rappresenta un tassello essenziale per poter progredire nella digitalizzazione, ma costituiscono una sfida che i tradizionali sistemi di protezione degli endpoint non sono in grado di garantire per diversi motivi, primo tra tutti la limitata capacità di rilevamento. TrendMicro riporta infatti come nel 2021 le statistiche di rilevamento di azioni malevoli per 2958 IT Decision Makers su 26 nazioni siano state preoccupantemente basse[41], vedi tabella 4.1.
Questi dati sono estrapolati da contesti aziendali, ma permettono di intuire il livello di efficacia di rilevamento delle attuali soluzioni domestiche nei confronti dei ransomware e dei tool presenti nella caratteristica catena di attacco del malware. Le soluzioni EDR aziendali, d'altro canto, sono pensate per realtà business, non adatte dunque all'applicazione all'interno di sistemi domestici, sia per i costi che per le conoscenze tecniche richieste in fase di installazione e di manutenzione dei prodotti. Un importante aspetto da tenere in considerazione è inoltre la ristretta capacità di calcolo dei prodotti smart: i tradizionali strumenti EDR come AV, strumenti di web control e web protection divengono inadatti se applicati al mondo IoT.
Questi ed altri fattori pongono dunque una sfida importante per la ricerca di nuovi metodi di protezione. Alla luce delle considerazioni esposte, illustro ora uno strumento open-source da me proposto per la mitigazione della minaccia IoT, basato sulla network segmentation e l'applicazione di un modello di machine learning per il controllo del traffico di rete ed il rilevamento delle intrusioni. Lo scopo ultimo è la messa in sicurezza delle infrastrutture domestiche smart, impedendo da un lato che i ransomware possano impiegati per estorcere denaro alle vittime, e dall'altro che tali dispositivi possano venir sfruttati per ospitare server C&C o per compiere altri tipi di attacchi.
La speranza è che l'impiego di questa soluzione possa contribuire a contrastare le capacità di abuso dei sistemi IoT domestici da parte dei cyber criminali per avanzare attacchi ransomware nei confronti delle vittime stesse o di terzi.

5

IoT Subnetting Security System

5.1 Architettura della soluzione proposta

Al giorno d'oggi un sistema domestico nella maggior parte dei casi rispecchia il modello rappresentato in figura 5.1, ossia un'architettura che presenta un modem centrale che funge da router e access point

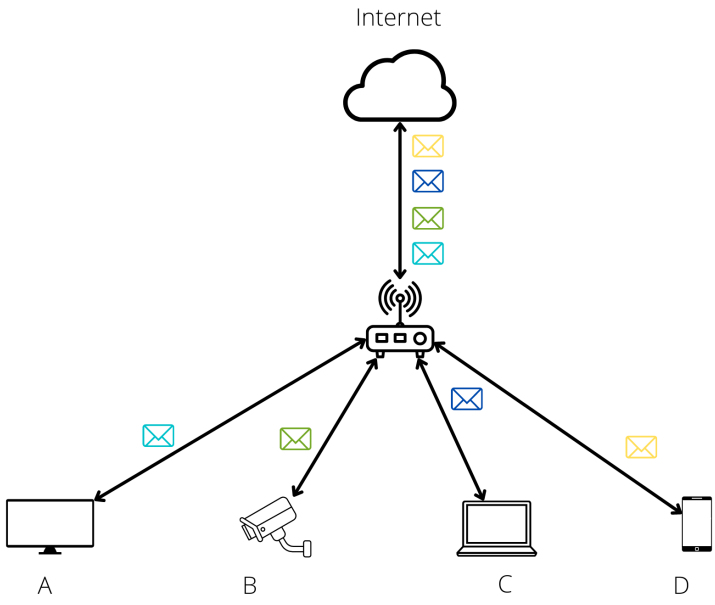

Figura 5.1: Il modem funge da router per tutti i dispositivi domestici

a cui tutti i dispositivi dotati di interfacce di rete vengono connessi. Questa architettura pone un problema di sicurezza importante in quanto un malintenzionato introdotto all'interno della rete locale finisce per essere potenzialmente in grado di comunicare con tutti i dispositivi connessi, potendo quindi effettuare una enumerazione ad ampio spettro dei servizi e delle porte vulnerabili sui vari dispositivi presenti nella stessa sottorete, o compiere altre attività illecite come la comunicazione con server C&C. Il sistema che vado ad avanzare si basa sul concetto di segmentazione della rete, ossia la divisione di una rete di calcolatori in sottoreti di dimensioni minori con lo scopo di migliorare le performance e/o la sicurezza. L'idea è separare il normale traffico dei device privati come laptop, smartphone, desktop o server nas rispetto a quello dei dispositivi smart, offrendo ai dispositivi IoT una sottorete controllata

a cui connettersi e da cui non si conosca l'esistenza del resto dei dispositivi. La soluzione che si vuole porre in essere fa dunque affidamento su un calcolatore terzo[1] che funge da routed wireless access point, il quale si pone come nodo intermedio tra il modem di casa e i dispositivi smart [figura 5.2]

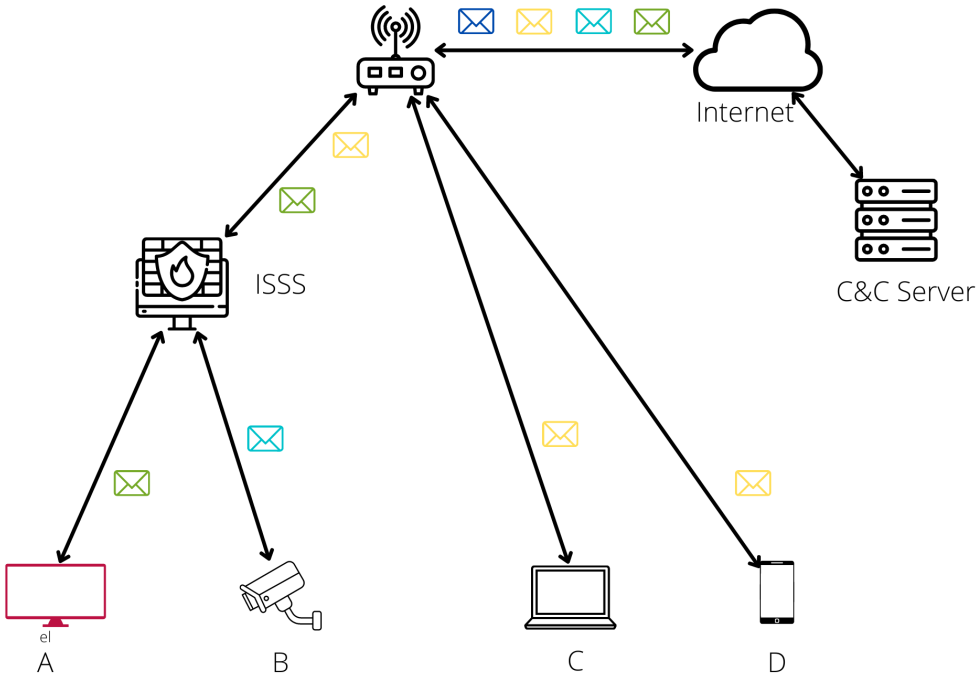

Figura 5.2: I dispositivi IoT si connettono all'access point offerto dall'ISSS

L'ISSS è l'unico dispositivo connesso alla rete locale, di conseguenza, ad esempio, il traffico proveniente dall'endpoint A e destinato ad Internet sarà visto dai dispositivi C e D come proveniente dal dispositivo ISSS, i quali ignoreranno l'esistenza dei dispositivi A e B.
La segmentazione di rete è una tecnica adoperata già da parecchi anni all'interno dei sistemi IT aziendali per questioni di performance e sicurezza, ma meno diffusamente presente nelle realtà domestiche a causa delle conoscenze tecniche necessarie all'applicazione della stessa, nonostante sia fondamentale quando si parla di sicurezza del mondo IoT. Offrire un dispositivo che automatizzi la configurazione di un sistema di rete segmentato può quindi aiutare concretamente nella mitigazione del rischio.
La network segmentation permette di limitare lo spread di malware tra i dispositivi connessi in rete all'interno della stessa infrastruttura, ma chiaramente non permette di individuare le minacce esistenti. Per questa ragione la soluzione suggerita prevede che l'ISSS possieda le capacità analizzare il traffico di rete e, mediante un modello di intrusion detection, isolare i dispositivi ritenuti potenzialmente compromessi tramite la modifica delle regole di instradamento. L'architettura del programma che funge da ISSS è schematizzata in figura 5.3. Come si può vedere le componenti essenziali sono:

- Initializer_Component: si occupa di effettuare la prima configurazione delle interfacce di rete del device che ospita l'ISSS in modo che il dispositivo funga da router per la sottorete protetta.
- Run-time_Component: implementa le funzionalità di detection e response ed è costituito a sua volta da:

[1] da qui in avanti farò riferimento a tale dispositivo con l'acronimo di ISSS (IoT Subnetting Security System)

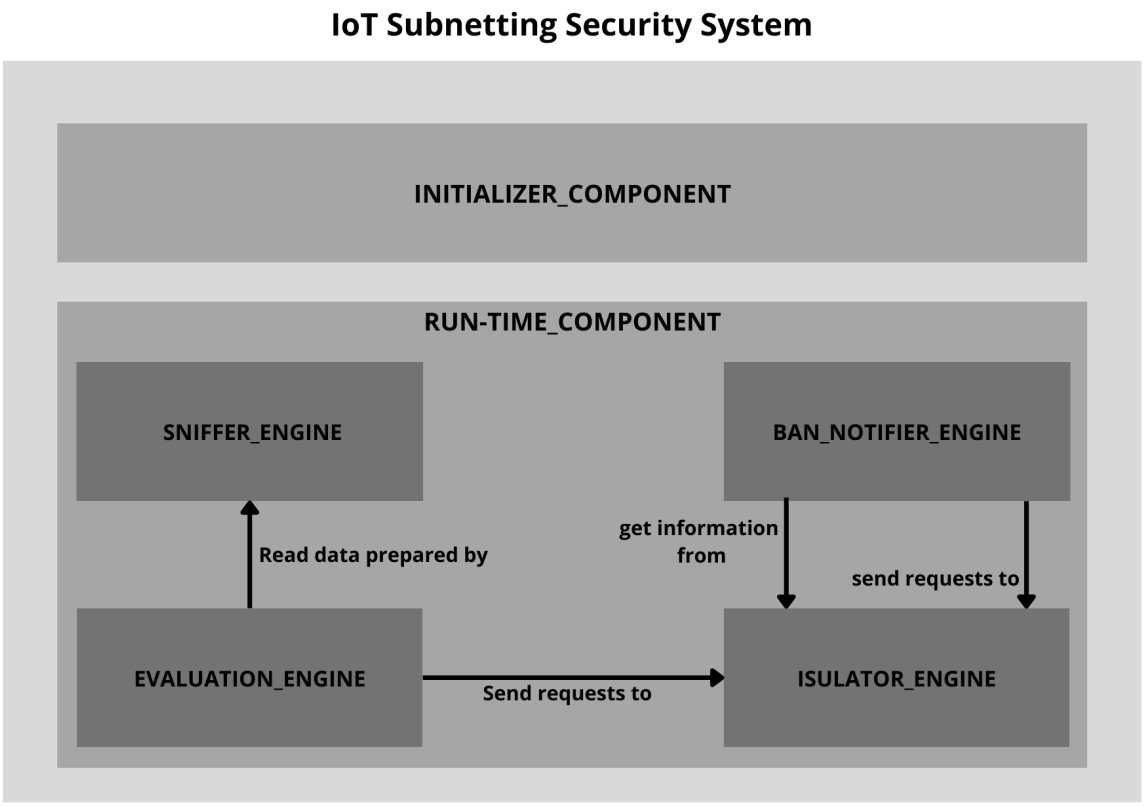

Figura 5.3: Architettura interna dell'ISSS

- Sniffer_Engine: cattura i pacchetti che circolano sulla sottorete e formatta le statistiche sui flussi in modo da esser manipolabili dal Evaluation_Engine. Per flusso, in accordo con il RFC 7011, si definisce "a set of packets or frames passing an Observation Point in the network during a certain time interval. All packets belonging to a particular Flow have a set of common properties."[42]
- Evaluation_Engine: analizza le statistiche sui flussi catturate dallo Sniffer_Engine e stabilisce se si tratta di traffico lecito o illecito
- Isulator_Engine: in base alle valutazioni dell'Evaluation_Engine e del Ban_Notifier_Engine modifica le regole di instradamento del dispositivo configurato come ISSS
- Ban_Notifier_Engine: componente che offre varie funzionalità all'utente, tra le quali la possibilità di ottenere la lista dei dispositivi attualmente in isolamento e di autorizzare il traffico proveniente da specifici indirizzi

Questa architettura permette di demandare le operazioni computazionalmente più costose al dispositivo configurato come ISSS, consentendo di non aggiungere alcun tipo di carico sui dispositivi IoT che generalmente godono di scarse risorse di calcolo.
Essendo l'ISSS un tool installato su un dispositivo indipendente, questo gli consente:

- una maggior modularità hardware

- l'inserimento di tale componente all'interno di un'infrastruttura preesistente senza la necessità di dover modificare o sostituire la strumentazione già in opera
- di non demandare nessuna operazione di calcolo in real time al modem di casa, evitando quindi di inficiare sul throughput della rete dei dispositivi di lavoro
- la messa in sicurezza anche di dispositivi legacy o che non godono di aspetti di sicurezza nativa

5.2 Descrizione di un caso d'uso

Descrivo ora un esempio generale sul comportamento atteso dal sistema durante l'esecuzione.
Assumo una rete domestica in cui sono presenti: 2 device IoT di varia natura (A, B); 2 device privati (C,D); un dispositivo che funga da ISCS; un modem domestico che offre capacità di routing verso Internet; un server C&C in qualche parte del mondo e raggiungibile attraverso la rete.

Normalmente il traffico di rete procede regolarmente. I pacchetti inoltrati dai dispositivi C e D transitano verso il modem, mentre i pacchetti inviati dai dispositivi smart inizialmente transitano verso l'ISSS il quale gli analizza e da cui poi avviene il forwarding al modem che a sua volta gli indirizza in Internet.
Ipotiziamo a questo punto che il dispositivo A venga compromesso da un cyber criminale e istanzi una comunicazione con un server C&C (vedi figura 5.4).

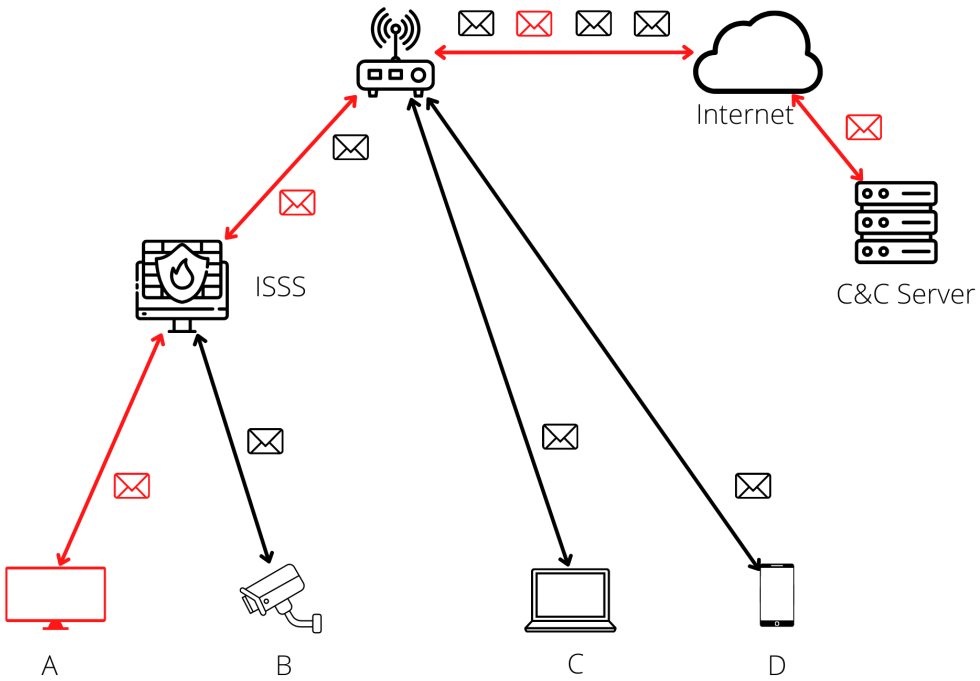

Figura 5.4: N il dispositivo A viene compromesso e tenta di comunicare con un server C&C.
I pacchetti inviati dall'host vengono inoltrati all'ISSS

Una volta che la comunicazione tra il server C&C e il dispositivo A viene ritenuta sospetta dall'ISSS, esso aggiorna le proprie regole di forwarding etichettando i due device coinvolti come bannati (figura 5.5: ogni pacchetto transitante e avente come mittente o destinatario uno di essi viene automaticamente

scartato.
Fintanto che non viene garantita la bontà del dispositivo A e del dispositivo C&C, esso rimane in quarantena e dunque privo della capacità di interagire con altri dispositivi impedendo quindi lo spread dell'infezione.

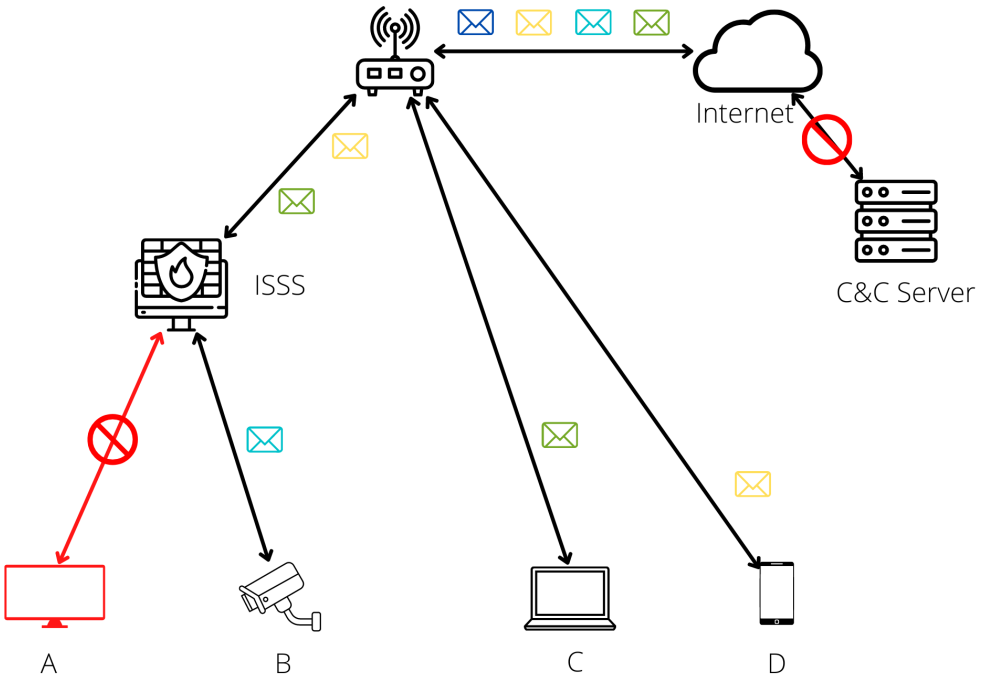

Figura 5.5: L'ISSS impedisce al dispositivo A infetto di comunicare

5.3 Implementazione dello strumento

Ai fini sperimentali come hardware fisico per ospitare l'ISSS ho deciso di impiegare un Raspberry Pi 3b+ provvisto del sistema operativo Raspbian alla versione Bullseye.

5.3.1 Initializer_Component

Questa componente, realizzata tramite uno script bash, ha il compito di:

- configurare il dispositivo come access point con forwarding dei pacchetti all'interfaccia di rete connessa al modem di casa.
- installare dei tool necessari al Run-Time_Component
- istanziare un servizio di sistema che all'avvio del dispositivo esegue il payload del RunTime-Component.

5.3.2 RunTime_Component

Il Run-Time_Component è un programma scritto in Python che, una volta eseguito, avvia i sottoprogrammi che lo compongono, ossia lo Sniffer_Engine, l'Evaluation_Engine ed infine l'Isulator_Engine.

Sniffer_Engine & Evaluation_Engine

Sebbene queste due componenti svolgano funzionalità distinte e nell'architettura generale del sistema siano considerate come parti separate, a livello di impletamentazione sono realizzate all'interno dello stesso script Python.

Lo Sniffer_Engine è realizzato sfruttando il framework open-source, NFStream. Questo strumento, come riporta il sito ufficiale, è uno tool multipiattaforma che permette di analizzare i flussi di rete, estrapolando per ciascuno di essi fino a 88 features. NFStream è in grado di estrapolare i flussi sia in modalità offline, attraverso file di cattura .pcap, sia in real-time specificando un'interfaccia di rete.

L'Evaluation_Engine è un programma Python che analizza singolarmente ogni entry catturata dallo Sniffer_Engine. Per ciascun flusso, tramite l'algoritmo di machine learning, determina se è di natura malevola o benigna. Per evitare singoli casi falsi negativi (dove per negativo intendo un rilevamento malevolo), l'Evaluation_Engine è provvisto di una tabella sql, realizzata tramite l'ausilio del motore sqlite3, che memorizza il numero di anomalie riscontrate per ciascun mac address e l'ultimo aggiornamento del valore del contatore. Nel caso in cui il contatore debba essere incrementato in seguito a un rilevamento, l'Evaluator_Engine provvede a determinare quanto tempo è trascorso dall'ultima segnalazione. Se questo tempo è minore un valore prestabilito e se il contatore è superiore a una soglia anch'essa predeterminata, le trasmissioni vengono bannate definitivamente come malevoli. Queste soglie sono fissate arbitrariamente, e rispettivamente, a 120 secondi e 3 rilevamenti successivi massimi consentiti prima del banning della connessione. Nel caso in cui venga rilevato una sospetta intrusione, i passi che avvengono sono quindi:

1. Ricavare il timestamp attuale, il mac address e l'ip address degli attori coinvolti nel flusso
2. Verificare quanto tempo è passato dalla precedente rilevazione associata all'ip e al mac address di sorgente e destinatario
3. Se necessario incrementare o azzerare il contatore delle intrusioni
4. Se le informazioni sul flusso evidenziano un comportamento malevolo viene richiesto l'isolamento del dispositivo all'Isulator_Engine

Per comunicare la necessità di isolare un dispositivo, l'Evaluation_Engine apre in lettura un file csv denominato *lock-unlock-devices.csv*, e condiviso tra Isulator_Engine e Ban_Notifier_Engine, su cui effettua l'append della stringa così composta:

<ACTION>, <MAC ADDRESS>, <AUTHOR>, <TIMESTAMP>

Dove:

- ACTION individua l'azione da voler compiere. Può essere di due tipi: LOCK e UNLOCK
- MAC ADDRESS è il mac address del dispositivo su cui si vuole compiere l'azione
- AUTHOR è l'autore della richiesta. Può essere USER o EVALUATOR a seconda che l'azione sia inoltrata tramite l'interfaccia offerta dal Ban_Notifier_Engine o l'Evaluator_Engine
- TIMESTAMP è l'indicazione temporale associata alla richiesta

In figura 5.6 è rappresentata l'architettura appena descritta.

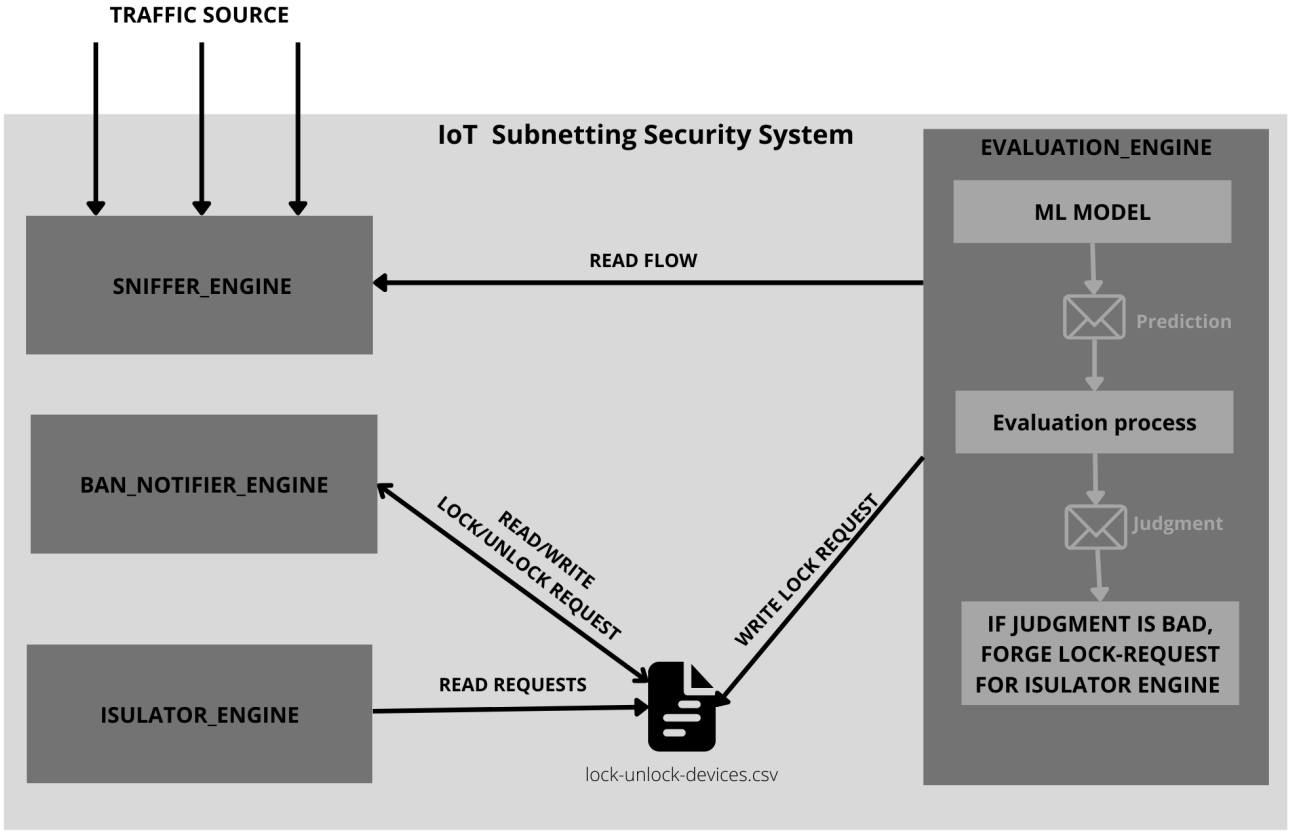

Figura 5.6: Schema di funzionamento dell'Evaluation Engine

Nel seguito riporto nello specifico i dettagli del modello di machine learning impiegato e le scelte prese durante la realizzazione dello stesso.

Descrizione dell'algoritmo Random Forest

Per discriminare i flussi di tipo maligno da quelli benigni ho deciso di usare l'algoritmo Random Forest. Le Random Forest sono un modello di classificazione supervisionato, basato su enseamble learning, ideato inizialmente da Tin Kam Ho nel 1995 [43] e poi evoluto successivamente da Leo Breiman nel 2001[44]. La scelta di questo algoritmo è stata basata sull'analisi di vari studi comparativi tra i metodi di classificazione tradizionale, i quali hanno evidenziato come questo metodo, applicato all'intrusion detection, ottenga nella maggior parte dei casi ottime performance rispetto a modelli di apprendimento supervisionato quali KNN, SVM, Decision Tree, MultinomialNB, DNN, in particolare nel numero di falsi positivi generati [45] [46] [47].

Per comprendere il funzionamento delle Random Forest è necessario definire primariamente il concetto di Decision Tree (vedi figura 5.7).

I Decision Tree sono un metodo usato per effettuare sia studi di regressione che di classificazione.

L'idea alla base dell'algoritmo dei Decision Tree è creare, a partire da un dataset iniziale, un albero in

cui:

- ciascun nodo rappresenta un test sull'attributo
- ogni ramo uscente da un nodo corrisponde ad uno dei possibili valori che l'attributo può assumere
- ogni foglia corrisponde all'assegnamento di una classe.

La classe individuata per una istanza di input corrisponde quindi al valore della foglia raggiunta valutando, a partire dalla radice, il valore attribuito dall'istanza all'attributo associato a ciascun nodo presente nel path radice-foglia.
La fase di addestramento di un Decision Tree serve per costruire l'albero ed in genere per farlo viene applicato un approccio top-down. Un metodo utilizzato per scegliere la variabile associata a ciascun nodo consiste nel scegliere la feature che massimizza il guadagno entropico derivato dal partizionamento dei dati rispetto a tale attributo, dove per guadagno entropico si intende il valore G(S,A) tale che:

$$G(S, A) = E(S) - \sum_{v \in Elementi(A)} \frac{|S_{A=v}|}{|S|} \cdot E(S_{A=v})$$

In cui:

- S è il l'insieme degli elementi
- A è l'insieme dei valori associati all'attributo su cui si intende partizionare l'insieme
- E è la funzione per il calcolo dell'entropia associata a un insieme S di elementi e corrisponde a

$$E(S) = -\sum_{i=1}^{n} P_i \cdot \log_2 P_i$$

  con P_i la percentuale di elementi dell'insieme S, associati alla classe i e n il numero di classi di S. Un valore entropico prossimo a 1 corrisponde a un forte sbilanciamento positivo del numero di elementi legati ad una classe, mentre un valore di 0.5 indica che il numero di elementi per ciascuna classe è bilanciato.

Il problema degli alberi di decisione evidenziato da Tin Kam Ho, è che più la profondità dell'albero diventa grande, più si tende ad avere overfitting dei dati. Per risolvere questo limite, Ho propose dunque l'algoritmo Random Forest, il quale attraverso il metodo di enseable *random subspace method* [48] combina assieme i risultati ottenuti da un certo numero di Decision Tree, detti *stimatori*, per far sì che la predizione finale non derivi da un unico albero di decisione, ma bensì dalla maggioranza dei valori ottenuti da diversi alberi di decisione (vedi figura 5.8). In particolare, nel caso del modello di Ho, ciascun albero di decisione effettua una valutazione soltanto su un subset delle features totali del dataset di input, così da generalizzare meglio il problema e ridurre la correlazione tra gli stimatori.

La versione del 2001 di Breiman evolve il modello proposto da Ho, integrando la tecnica del bootstrap aggregation, in modo che ciascun albero di decisione venga allenato su un subset delle features totali e su un subset dei dati di training.

È da sottolineare che le Random Forest non solo permettono di ottenere ottime prestazioni in termini di classificazione, ma inoltre al crescere del numero di Decision Trees utilizzati, permettono di aumentare

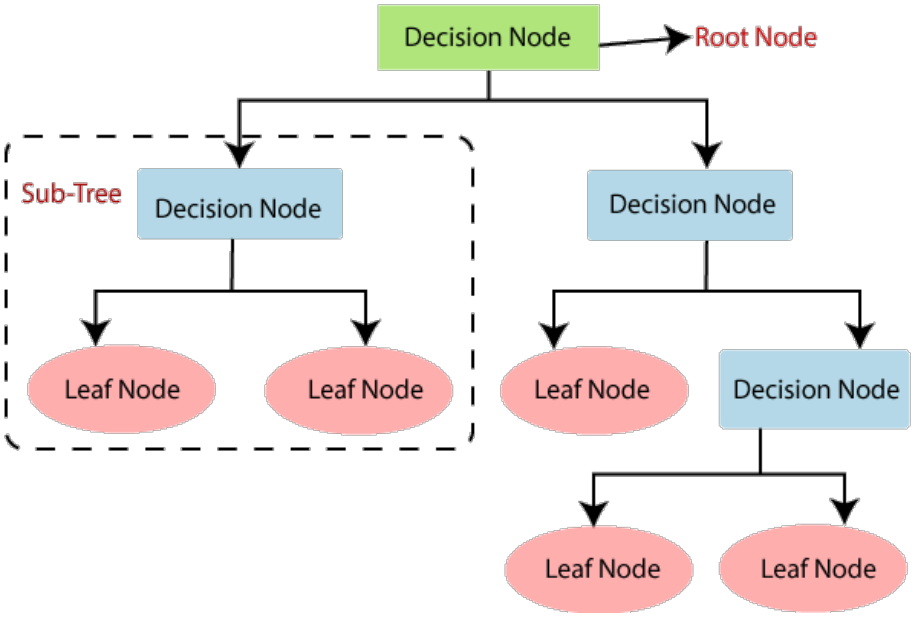

Figura 5.7: Esempio di Decision Tree

l'accuratezza senza generare per questo overfitting grazie al fatto che, come evidenziato da Breiman, per la legge dei grandi numeri è improbabile che un numero elevato di stimatori producano in larga misura un marcato errore complessivo. [44].

Descrizione del dataset

Il dataset utilizzato per allenare e testare il modello di machine learning è stato realizzato da zero partendo dai file pcap pubblicati online da 5 istituti di ricerca differenti. La scelta di optare su un dataset personalizzato è motivata dal fatto che le collezioni già presenti in rete come il famoso KDD-99-Cup, o la versione NSL-KDD del Canadian Institute for Cybersecurity, derivano da catture su calcolatori di diversa natura e non per forza di dispositivi smart. Il traffico generato dai device IoT si suppone sia in genere più ripetitivo, meno soggette a variazioni nei protocolli usati. Essendo l'obiettivo l'individuazione delle intrusioni nel traffico generato da dispositivi IoT, un dataset personalizzato su tale categoria ho pensato potesse portare a un livello di precisione maggiore.
Fatte le dovute premesse, passo quindi a spiegare come ho realizzato il dataset.
Come detto, la caratteristica primaria che volevo ottenere era partire da catture di traffici generati da dispositivi IoT. Di seguito riporto le fonti di file pcap individuate, con una breve descrizione sulla provenienza delle catture:

- HCRL Iot Network Intrusion Dataset [49]
  I 42 file .pcap contenuti in questo dataset sono stati generati a partire dalla cattura del traffico proveniente da dall'home assistant SKT NUGU (NU 100), dalla Wi-Fi camera EZVIZ (C2C Mini O Plus 1080P) e da altri laptop e smartphone connessi nella stessa rete.

- Aalto IoT devices captures [50]
  Il dataset è composto dalle catture effettuate durante il setup di 31 dispositivi IoT di 27 categorie differenti. Non sono presenti dunque catture di tipo malevolo.

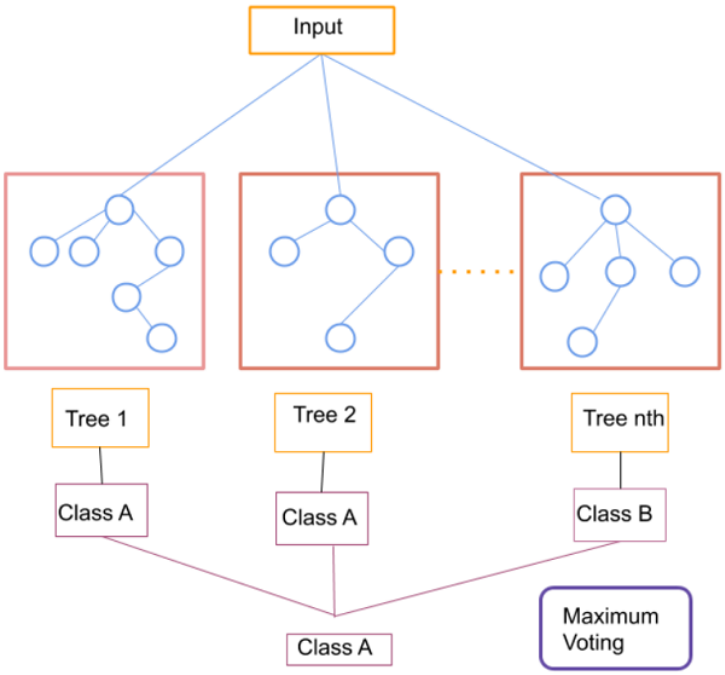

Figura 5.8: Soluzione proposta da Ho

- ETF IoT Botnet Dataset [51]
  L'ETF IoT Botnet Dataset è composto da catture di tipo benigno e maligno ed è stato realizzato eseguendo i sample di diversi malware collezionati dall'Università di Belgrado nel periodo 2019-2021, su un dispostivo RaspberryPi.

- IoT-23 [52]
  IoT-23 è un dataset composto da 20 catture generate a partire da un dispostivo RaspberryPi compromesso da malware differenti e 3 originati da dispositivi non compromessi (una Philips HUE smart LED lamp, l'assistente Amazon Echo home ed infine la Somfy smart doorlock).

- MQTT-IOT-IDS2020 [53]
  Questa collezione di file pcap contiene le catture di 5 scenari differenti provenienti da un'architettura di rete MQTT formata da 12 sensori, un broker, una camera simulata e un attaccnte. Gli scenari simulati sono: un traffico normale, uno scan della rete, uno scan UDP, un attacco bruteforce alle porte SSH e un attacco bruteforce al broker mqtt.

Per semplicità da qui in avanti mi riferirò ai dataset appena presentati, rispettivamente con le abbreviazioni: HCRL, AIDC, ETF, IOT23 ed MIDS2020.

Analizzando ciascun dataset mi sono potuto rendere conto che, eccetto per ETF, le trasmissioni di tipo malevolo collezionate in IOT-23, HCRL e MIDS2020 sono contaminate da traffici benigni. Per questa ragione ai fini del training e del testing del modello ho deciso di utilizzare come flussi malevoli solamente quelli estratti a partire dal dataset ETF e di sfruttare i traffici maligni offerti dagli altri tre dataset per validare in un secondo momento l'efficacia dello strumento una volta completato.
Il passo successivo è stato quindi estrapolare dai file pcap contenuti in AIDC ed ETF le caratteristiche sui flussi presenti in ciascuno di essi. Per farlo ho sfruttato il tool già citato NFStream.
Ai singoli file csv ho poi aggiunto una feature "label" a cui ho associato la stringa "benign" o "malicious"

DATASET	NUMBER OF TOTAL SAMPLES	NUMBER OF BENIGN SAMPLES	NUMBER OF MALICIOUS SAMPLES
MQTT	88468	88468	0
AALTO	16124	16124	0
MENDELEY	39320	9860	29460
IOT23	2879	2879	0
HCRL	401	401	0

Tabella 5.1: Distribuzione dei sample di partenza nel dataset finale

a seconda che il csv fosse stato generato da una cattura priva di infezione o in cui era presente una forma compromissione.
L'ultimo passaggio è stato effettuare il merging dei vari file csv da cui, come si può vedere in figura 5.9 ho ottenuto un dataset fortemente sbilanciato, composto per il 78.5% da dati benigni e dal restante 21.5% da dati malevoli.

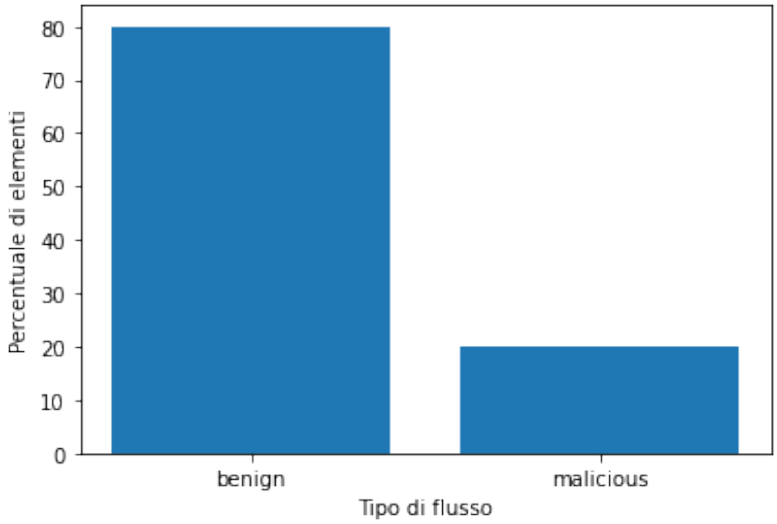

Figura 5.9: Distribuzione dei campioni benigni e maligni nel dataset

Il numero di entry per ciascun dataset di partenza è raffigurato in tabella 5.1.

Preprocessing del dataset

Una volta generato il dataset sono passato alla fase di preprocessing in cui ho studiato le caratteristiche dei dati con lo scopo di ottimiazzare le capacità di previsione del modello.
Inizialmente il numero di features che componeva il mio dataset erano 89, 88 generate dal framework NFStream più la label da me assegnata. Per prima cosa ho rimosso le variabili che ai fini dell'addestramento del modello possedevano un basso contenuto informativo, ossia indirizzi ip, mac address e i vari timestamp. In questo modo ho ridotto il numero di features a 65. In seguito ho rimosso le variabili che presentavano un coefficiente di correlazione di Pearson, in valore assoluto, superiore al 95%, riducendo il numero di colonne di ulteriori 12 unità e ottenendo globalmente 53 features. L'ultimo passaggio è stato discretizzare i valori qualitativi.

Scomposizione del dataset

Per addestrare il modello e per testarlo ho scomposto il dataset iniziale in due parti: la prima, composta dal 75% degli elementi del dataset l'ho utilizzata per allenare il modello, mentre la seconda, a cui appartengono il restante 25% è a sua volta ripartita in 3 sottoinsiemi di dati da 12219, 10000 e 10000 elementi, che ho sfruttato per testare la Random Forest su diversi scenari. Il primo e il secondo scenario sono composti da dati fortemente sbilanciati verso i casi benigni, mentre il terzo presenta un bilanciamento del 50% tra le due classi.

Metriche di valutazione impiegate

Per valutare il modello addestrato le metriche che ho analizzato sono state:

- **False positive rate (FPR)**:
  Questa metrica misura la probabilità che il modello generi un falso positivo. In questo caso il suo valore corrisponde alla probabilità che un flusso venga etichettato come benigno pur essendo maligno. Il FPR è calcolato come il rapporto tra il numero di falsi positivi generati dal modello e la somma dei casi di falso positivo e di vero negativo.

$$FPR = \frac{FP}{FP + TN}$$

- **False negative rate (FNR)**:
  Al contrario del FPR, il False Negative Rate misura la probabilità che il modello generi un falso negativo. In questo caso il suo valore corrisponde alla probabilità che un flusso venga etichettato come maligno pur essendo benigno. Il FNR è calcolato come il rapporto tra il numero di falsi negativi generati dal modello e la somma dei casi di falso negativo e di vero positivo.

$$FPR = \frac{FN}{FN + TP}$$

- **F1-Score**:
  L'F1-score è un valore che rappresenta la media armonica tra precisione e recall, due metriche che indicano rispettivamente la capacità di un modello di predirre ed individuare le classi positive. Un valore di F1 prossimo ad 1 corrisponde a un alto livello di precision e recall. Il calcolo dell'F1-score permette di ottenere una stima più accurata della bontà del modello nel caso in cui il numero di elementi per ciascuna classe non è bilanciato.

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

- **F2-Score**:
  Rispetto all'F1-Score questa metrica pone maggior importanza sul valore della recall. Lo score F2 permette di determinare se il modello genera molti o pochi falsi positivi. Nel nostro caso, ottenere un F2 score particolarmente alto indica che il modello riesce ad ottenere un basso tasso di falsi positivi, ossia è più difficile che riconosca come benigno un traffico che in realtà è maligno.

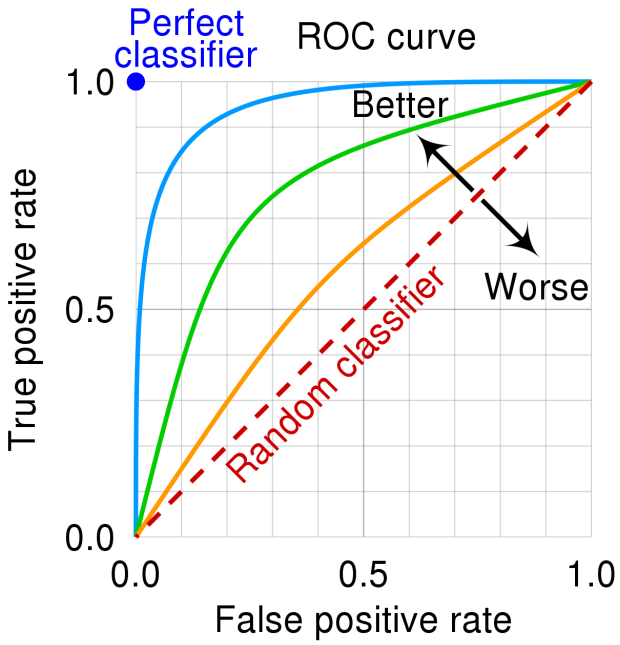

Figura 5.10: ROC Curve

$$F2 = 3 \cdot \frac{Precision \cdot Recall}{4 \cdot Precision + Recall}$$

- **Area Under the Curve (AUC)**:
  La metrica AUC si basa sulla definizione di curva ROC (Receiving Operating Characteristic Curve), ossia una rappresentazione grafica dei valori di recall e FPR. Un esempio di curva ROC è visibile in figura 5.10. Il valore di AUC individua l'area sottesa alla curva di ROC, consentendo di derivare il suo grado di curvatura. Maggiore è il valore di AUC, maggiore è la curvatura verso l'alto della ROC curve, che si traduce in un minor numero di falsi positivi in favore di un alto tasso di true positive. Ottenere da un modello un AUC score prossimo ad 1 corrisponde perciò ad una maggior propensione a risultati positivi piuttosto che ai falsi positivi.

- **Matthews correlation coefficient (MCC)**:
  Il coefficiente di correlazione Matthews è una metrica che può essere impiegata per valutare risultati prodotti a partire da un dataset sbilanciato. Un coefficiente di 1 indica una stima perfetta, un valore di 0 una stima casuale, mentre MCC= 1 evidenzia una stima inversa. La metrica MCC riesce a racchiudere il contenuto informativo della confusion matrix [54].

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}$$

Testing del modello RandomForest con dati di training sbilanciati

Addestrato il modello su un set di training fortemente sbilanciato (18.5% di elementi maligni e 81.5% di dati benigni), i risultati ottenuti sono stati quelli riportati in tabella 5.2.

Test	#Sample	F1	F2	AUC	MCC	FNR	FPR
Sbilanciato	12219	99.06%	99.10%	99.45%	98.84%	0.23%	0.87%
Sbilanciato 2	10000	99.18%	98.92%	99.33%	99.00%	0.08%	1.25%
Bilanciato	10000	99.41%	99.15%	99.41%	98.82%	0.16%	1.02%

Tabella 5.2: Risultati ottenuti dal modello allenato su un training set fortemente sbilanciato

Testing del modello RandomForest con dati di training bilanciati usando SMOTE

Per cercare di incrementare le prestazioni ottenute con il precedente modello ho cercato di bilanciare il numero di training samples maligni. Per far ciò ho applicato la tecnica di data augmentation Synthetic Minority Oversampling Technique (SMOTE) per bilanciare la classe minoritaria nei dati di training. Questa tecnica consiste nell'individuare gli elementi della classe minoritaria e per ciascuno di essi determinare i vettori tra il punto considerato e ciascuno dei k punti vicini ad esso. I vettori vengono successivamente moltiplicati per un numero casuale nell'intervallo [0,1] e sommati al punto iniziale, permettendo di generare un nuovo valore (vedi figura 5.11) Utilizzare la strategia SMOTE ha il

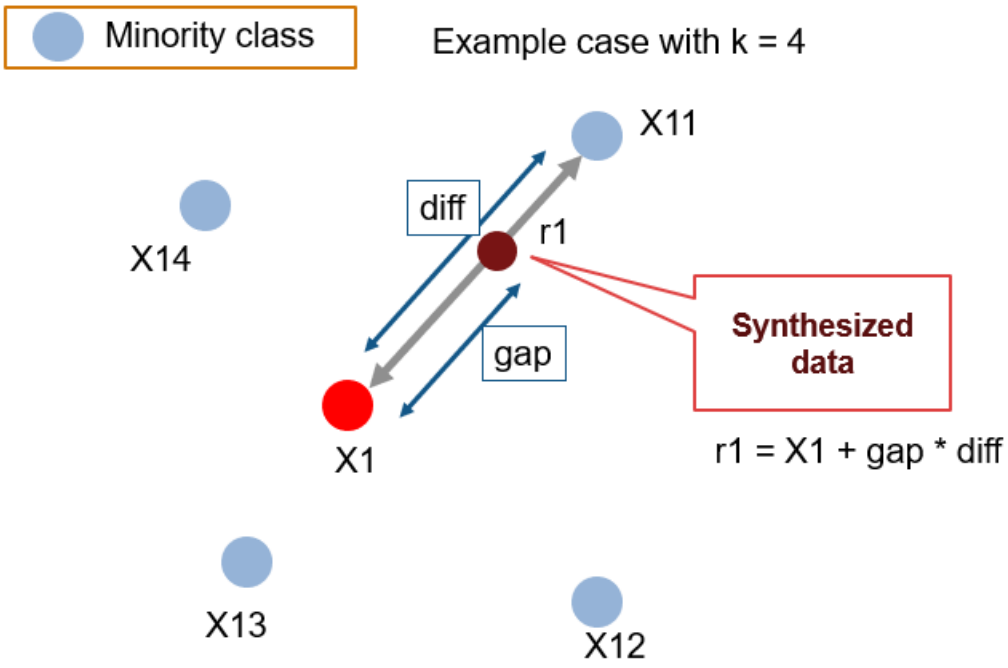

Figura 5.11: Esempio di generazione di un punto con la tecnica SMOTE [55]

vantaggio di permettere di generare punti nuovi e non duplicati come avviene nelle tradizionali tecniche di oversample, mantenendo allo stesso tempo una somiglianza con i dati della classe iniziale.

I risultati dell'impiego di questo algoritmo sono riportati in tabella 5.3. Come si può notare c'è stato un lieve e diffuso miglioramento negli score registrati a partire dagli stessi dati di test. La scelta di utilizzare SMOTE al posto di una più semplice tecnica di undersampling è stata per cercare evitare la perdita di informazioni utile sui dati benigni. Allo stesso modo, ho evitato di utilizzare strategie di oversampling per mantenere il più basso possibile il numero di duplicati nei dati di training.

Test	#Sample	F1	F2	AUC	MCC	FNR	FPR
Sbilanciato	12219	99.69%	99.82%	99.90%	99.62%	0.12%	0.08%
Sbilanciato 2	10000	99.13%	98.93%	99.34%	98.93%	0.12%	1.19%
Bilanciato	10000	99.49%	99.29%	99.49%	98.98%	0.18%	0.84%

Tabella 5.3: Risultati ottenuti dal modello allenato su un training set bilanciato

Isulator_Engine

L'Isulator_Engine è un programma, anch'esso scritto in Python, che durante l'esecuzione rimane in ascolto del file lock-unlock-devices.csv. Le stringhe che si aspetta di ricevere sono strutturate come definito all'interno del paragrafo in cui descrivo l'Evaluation_Engine. Non appena l'Isulator_Engine legge una nuova riga nel file csv i passi che effettua sono:

1. Controllare il tipo di azione da intraprendere
2. Se il campo ACTION è impostato su LOCK allora verifica all'interno della tabella sql *granted_devices* se il dispositivo con il mac address specificato è presente. Se lo è non intraprende alcuna azione. Se non lo invece:
   - accede alla tabella *locking_devices* e aggiunge una entry per tale dispositivo in modo che l'utente attraverso il Ban_Notifier_Engine possa venire a conoscenza dell'azione intrapresa
   - richiama il firewall iptables nativo nelle distribuzioni Linux, modificando le regole applicate attraverso il comando
     sudo iptables A INPUT m mac mac-source <MAC> j DROP
     affinchè i pacchetti con tale mac address vengano scartati.
3. Se il campo ACTION è impostato su UNLOCK allora verifica all'interno della tabella sql *locking_devices* se il dispositivo con il mac address specificato nel campo MAC della riga, è presente in una delle entry. Se è presente allora lo elimina e richiama il comando
   sudo iptables D INPUT m mac macsource <MAC> j DROP
   affinchè i pacchetti con tale mac address possano riprendere a circolare.

Ban_Notifier_Engine

Il Ban_Notifier_Engine è un programma Python che integra le funzionalità di interfacciamento tra l'utente e l'ISSS. Lo script, che sfrutta le tabelle già citate *locking_devices*, *granted_devices* e *allerts_table*, offre le opzioni indicate in tabella 5.4. Attualmente il Ban_Notifier_Engine si basa su linea di comando, e l'utente per poterne sfruttare le capacità è necessario acceda al dispositivo da remoto o collegandolo un monitor, ma in futuro non è escluso possa evolvere in un'interfaccia web accessibile dalla rete locale, previa autenticazione.

OPZIONE	OUTPUT
-l, –list	mostra il mac address dei dispositivi attualmente in isolamento
-s, –show-logs	mostra lo storico delle richieste inviate all'Isulator_Engine
-a, –show-allerts	mostra il contenuto della tabella dei rilevamenti dell'Evaluation_Engine
–show-granted-devices	mostra la lista dei dispositivi autorizzati dall'utente
–lock-device <MAC ADDRESS >	isola il dispositivo specificato
–unlock-device <MAC ADDRESS >	rimuove dall'isolamento il dispositivo specificato
–grant-device <MAC ADDRESS >	autorizza il dispositivo specificato Impedisce l'isolamento del dispositivo indicato
–remove-grant <MAC ADDRESS >	rimuove l'autorizzazione al il dispositivo specificato

Tabella 5.4: Opzioni offerte dal Ban_Notifier_Engine

6

Conclusioni

6.1 Contesti di applicazione e motivazioni per l'utilizzo del tool

Il range di applicazione della soluzione proposta è piuttosto vasto. Può infatti essere utilizzato per monitorare la propria smart home infrastructure, ma anche piccoli studi e aziende private (fotografi, creatori di contenuti sul web, professionisti freelance, ...) che potrebbero risentire fortemente di un attacco con impiego di ransomware. Adottare soluzioni di detection e response come quella illustrata è necessario per varie ragioni. Innanzitutto un maggior numero di utilizzatori significa più feedback, che si riflettono in prodotti che evolvono più rapidamente. Dopodichè, è da considerare il fatto che prima che le soluzioni di sicurezza inizino ad essere adottate da un grande numero di persone è necessario del tempo, tempo che in futuro farsi trovare impreparati davanti alle minacce del mondo IoT significherà lasciare tempo ai cyber criminali di guadagnare ulteriormente a nostre spese.

6.2 Sviluppi futuri

Lo strumento presentato chiaramente non ha la pretesa di essere una panacea per gli attacchi ransomware, ma bensì si pone come punto di partenza per lo sviluppo di un efficace sistema di protezione domestica. Difendere le reti di dispositivi presenti nelle case delle persone comuni è essenziale per il contrasto agli attacchi informatici del futuro, specialmente quelli con impiego di ransomware.
Uno sviluppo futuro del progetto potrebbero essere la sostituzione dell'algoritmo Random Forest utilizzato, con modelli di anomaly detection e di apprendimento continuo per incrementare le capacità di mitigazione delle minacce 0-day.

6.3 Considerazioni finali

Per quanto criminalità informatica e digitalizzazione globale siano e saranno sempre un connubio indivisibile, l'analisi proposta mette in luce come sia possibile cercare di mitigare tale minaccia attraverso strumenti nuovi di individuazione e risposta che sfruttino le tecniche di machine learning conosciute.
La speranza è che in futuro l'impiego di tool di sicurezza diventi la norma e non un optional come ad oggi sembrerebbe essere.

Bibliografia

[1] Di Freeze. Global ransomware damage costs predicted to exceed $265 billion by 2031, Jun 2022.

[2] KnowBe4. Aids trojan: Pc cyborg: Original ransomware. https://www.knowbe4.com/aids-trojan.

[3] CyberSaint. State of ransomware attacks report 2022. *The Computer Journal*, 27(2):2, 2022.

[4] Published by Statista Research Department and Jul 7. Average length of downtime after a ransomware attack 2021, Jul 2022. https://www.statista.com/statistics/1275029/length-of-downtime-after-ransomware-attack.

[5] Datto. Datto's global state of the channel ransomware report. Technical report, Datto, 2020. https://www.datto.com/resource-downloads/Datto-State-of-the-Channel-Ransomware-Report-v2-1.pdf.

[6] European Union Agency for Cybersecurity. Enisa threat landscape 2021, October 2021. https://www.enisa.europa.eu/publications/enisa-threat-landscape-2021.

[7] CrowdStrike. 2022 global threat report. Technical report, CrowdStrike, 2022. https://go.crowdstrike.com/rs/281-OBQ-266/images/Report2022GTR.pdf.

[8] Art. 83 gdpr – general conditions for imposing administrative fines, Mar 2018.

[9] Stefano Bizzi. L'attacco hacker al comune di gorizia: Pubblicati i documenti rubati, Sep 2022.

[10] Federal Bureau of Investigation. Fbi warns about an increase in sextortion complaints, alert number i-090221-psa,, September 2021. https://www.ic3.gov/Media/Y2021/PSA210902.

[11] CSRC Content Editor. Malware - glossary: Csrc. https://csrc.nist.gov/glossary/term/malware.

[12] What is ransomware? definition, types & how it works. https://www.fortinet.com/resources/cyberglossary/ransomware.

[13] Ransomware fileless, criptano i pc senza lasciare tracce: Come difendersi, Feb 2019. https://www.cybersecurity360.it/nuove-minacce/ransomware/ransomware-fileless-criptano-i-pc-senza-lasciare-tracce-come-difendersi/.

[14] European Union Agency for Cybersecurity. Enisa threat landscape for ransomware attacks. November 2022. https://www.enisa.europa.eu/publications/enisa-threat-landscape-2022.

[15] Andy Greenberg. A tesla employee thwarted an alleged ransomware plot, Aug 2020. https://www.wired.com/story/tesla-ransomware-insider-hack-attempt/.

[16] Written by Luca Nagy. Nearly a quarter of malware now communicates using tls, Aug 2021. https://news.sophos.com/en-us/2020/02/18/nearly-a-quarter-of-malware-now-communicates-using-tls/.

[17] Roger A. Grimes. *Ransomware protection playbook*. Wiley, 2022.

[18] Christian Rossow, Christian Dietrich, and Herbert Bos. Large-scale analysis of malware downloaders. In Ulrich Flegel, Evangelos Markatos, and William Robertson, editors, *Detection of Intrusions and Malware, and Vulnerability Assessment*, Lecture Notes in Computer Science, pages 42–61. Springer LNCS, 2013. 9th GI International Conference on Detection of Intrusions and Malware and Vulnerability Assessment, DIMVA 2012 ; Conference date: 26-07-2012 Through 27-07-2012.

[19] KrebsonSecurity. Try this one weird trick russian hackers hate, May 2021. https://krebsonsecurity.com/2021/05/try-this-one-weird-trick-russian-hackers-hate/.

[20] Roger A. Grimes. *Ransomware protection playbook*. Wiley, 2022.

[21] Persistence.

[22] What is ryuk ransomware? the complete breakdown, Mar 2022. https://www.crowdstrike.com/blog/big-game-hunting-with-ryuk-another-lucrative-targeted-ransomware/.

[23] Unit 42 by Palo Alto Network. Ransomware's new trend: Exfiltration and extortion. Technical report, Unit 42 by Palo Alto Network, December 2020. https://www.paloaltonetworks.com/resources/whitepapers/ransomwares-new-trend-exfiltration-and-extortion.

[24] Jonathan Munshaw. Quarterly report: Incident response trends in summer 2020, Aug 2022. https://blog.talosintelligence.com/2020/09/CTIR-quarterly-trends-Q4-2020.html.

[25] CobaltStrike. Post exploitation. https://hstechdocs.helpsystems.com/manuals/cobaltstrike/current/userguide/content/topics/post-exploitation_main.htm.

[26] TrendMicro. Locked, loaded, and in the wrong hands: Legitimate tools weaponized for ransomware in 2021, Apr 2021. https://www.trendmicro.com/vinfo/au/security/news/cybercrime-and-digital-threats/locked-loaded-and-in-the-wrong-hands-legitimate-tools-weaponized-for-ransomware-in-2021.

[27] SophosLabs. How ransomware attacks whitepaper. Technical report, https://www.sophos.com, 2019. https://www.sophos.com/en-us/medialibrary/PDFs/technical-papers/sophoslabs-ransomware-behavior-report.pdf.

[28] Janus Agcaoili and Earle Earnshaw. Locked, loaded, and in the wrong hands: Legitimate tools weaponized for ransomware in 2021, Apr 2021. https://www.trendmicro.com/vinfo/au/security/news/cybercrime-and-digital-threats/locked-loaded-and-in-the-wrong-hands-legitimate-tools-weaponized-for-ransomware-in-2021.

[29] Bill Siegel. Ransomware payments decline in q4 2020, Nov 2021. https://www.coveware.com/blog/ransomware-marketplace-report-q4-2020.

[30] Adam McNeil. How did the wannacry ransomworm spread?: Malwarebytes labs, Mar 2019. https://www.malwarebytes.com/blog/news/2017/05/how-did-wannacry-ransomworm-spread.

[31] Inhibit system recovery. https://attack.mitre.org/techniques/T1490/.

[32] Harshit. What is cryptography?, Sep 2020. https://www.technoarchsoftwares.com/blog/what-is-cryptography/.

[33] Kazi Alam, Saifuddin Mahmud, and Mohammad Khan. A comparison between traceable and untraceable blind signature schemes through simulation. pages 1–4, 05 2013.

[34] Stephen Cooper. What is jigsaw ransomware & how to protect against it, Nov 2022. https://www.comparitech.com/net-admin/jigsaw-ransomware/.

[35] Elise. Cryptolocker - a new ransomware variant, Dec 2021. https://blog.emsisoft.com/en/1615/cryptolocker-a-new-ransomware-variant/.

[36] Stu Sjouwerman. 83% of all successful ransomware attacks featured double and triple extortion, May 2022. https://blog.knowbe4.com/83-of-all-successful-ransomware-attacks-featured-double-and-triple-extortion.

[37] Albert Zsigovits. Lockbit ransomware borrows tricks to keep up with revil and maze, Apr 2020. https://news.sophos.com/en-us/2020/04/24/lockbit-ransomware-borrows-tricks-to-keep-up-with-revil-and-maze/.

[38] C Zavazava. Itu work on internet of things, 2015. ictp workshop, 2015.

[39] Peter Newman. The internet of things 2020: Here's what over 400 iot decision-makers say about the future of enterprise connectivity and how iot companies can use it to grow revenue, Mar 2020. https://www.businessinsider.com/internet-of-things-report.

[40] Mark Manahan Marco Dela Vega, Jeanne Jocson. Emotet adds new evasion technique, Apr 2019. https://www.trendmicro.com/en_us/research/19/d/emotet-addsnewevasiontechnique-andusesconnecteddevicesasproxyccservers.html.

[41] TrendMicro. Everything is connected: Uncovering the ransomware threat from global supply chains a global study. Technical report, https://www.trendmicro.com, 2022. https://www.trendmicro.com/explore/glrans.

[42] Paul Aitken, Benoît Claise, and Brian Trammell. Rfc 7011: Specification of the ip flow information export (ipfix) protocol for the exchange of flow information, Sep 2013.

[43] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282 vol.1, 1995.

[44] Leo Breiman. Random forests. volume 45, october 2001.

[45] Paulo Angelo Alves Resende and André Costa Drummond. A survey of random forest based methods for intrusion detection systems. *ACM Comput. Surv.*, 51(3), may 2018.

[46] Shyla, Kapil Kumar, and Vishal Bhatnagar. Machine learning algorithms performance evaluation for intrusion detection. *Journal of Information Technology Management*, 13(1):42–61, 2021.

[47] Yunpeng Zhang, Yash Gandhi, Zhixia Li, and Zhiwen Xiao. Improving the classification effectiveness of network intrusion detection using ensemble machine learning techniques and deep neural networks. In *2022 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, pages 117–123, 2022.

[48] Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998.

[49] Hyunjae Kang, Dong Hyun Ahn, Gyung Min Lee, Jeong Do Yoo, Kyung Ho Park, and Huy Kang Kim. Iot network intrusion dataset. 2019.

[50] Samuel Marchal. Iot devices captures. april 2017.

[51] Pavle Jovanović, Đorđe; Vuletić. Etf iot botnet datasetl. 2021.

[52] Sebastian Garcia, Agustin Parmisano, and Maria Jose Erquiaga. Iot-23: A labeled dataset with malicious and benign iot network traffic, Jan 2020. More details here https://www.stratosphereips.org/datasets-iot23.

[53] Hanan Hindy, Christos Tachtatzis, Robert Atkinson, Ethan Bayne, and Xavier Bellekens. Mqtt-iot-ids2020: Mqtt internet of things intrusion detection dataset, 2020.

[54] Chauvin Y Andersen C Nielsen H Baldi P, Brunak S. Assessing the accuracy of prediction algorithms for classification: an overview. bioinformatics. page 412–424, 2000.

[55] Swastik Satpathy. Smote: Overcoming class imbalance problem using smote, Jan 2021.