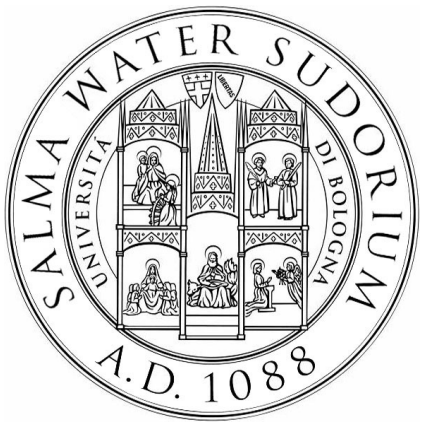

ALMA MATER STUDIORUM
UNIVERSITÁ DI BOLOGNA

Dipartimento di Scienze

Corso di laurea triennale in
Informatica Per Il Management

Ransomware e l'impatto del Machine Learning

Relatore

Prof. Davide Sangiorgi

Laureando

Manuel Castiglia

Appello 25/03/2025

Anno Accademico 2023/2024

Introduzione

L'obiettivo di questa tesi è quello di analizzare il malware più diffuso dell'era odierna, noto come Ransomware, e di vedere come l'avvento dell'Intelligenza Artificiale ha modificato questo fenomeno. In un mondo in cui la tecnologia fa passi da gigante, chi si occupa di sicurezza informatica non può essere da meno, quindi conoscere e capire come un virus si comporta, aiuta non solo a prevenire eventuali infezioni, ma anche a contrastarle nel momento in cui si presentano.

Nel primo capitolo viene descritto cos'è il Ransomware, viene raccontata la sua storia, vengono descritte le varie tipologie che si possono trovare, vengono elencati quelli che sono i vettori di infezione e, infine, viene descritto come avviene un attacco.

Nel secondo capitolo viene affrontato il tema dei mezzi di difesa tradizionali da adottare per proteggersi da questo attacco. Vengono descritti sia metodi generali, come l'eseguire i backup o l'utilizzare approcci diversificati di difesa, sia livelli di protezione direttamente sul sistema operativo, come la cifratura del disco, discutendo degli standard e delle funzionalità principali che si trovano sui vari sistemi operativi. Infine viene anche descritta una soluzione alternativ come l'affidamento dei dati ad aziende esterne, affrontando nello specifico "Cubbit", una startup nata a Bologna che si occupa di cloud storage geo-disribuito.

Nel terzo capitolo, viene introdotta l'Intelligenza Artificiale e il suo impatto sul tema della cybersicurezza. Vengono ripercorse le principali tappe della nascita di questa scienza e si va a spiegare come sia diventata una minaccia da tenere sotto controllo se combinata con questo malware.

Nel quarto capitolo vengono descritte le principali tecniche di difesa che si basano sull'AI, in particolar modo vengono trattati EDR, XDR e MDR, che sono soluzioni avanzate di cybersecurity progettate per rilevare, rispondere e miti-

gare le minacce informatiche in un'azienda o in un'infrastruttura IT, offrendo un approccio più proattivo e automatizzato nella gestione della sicurezza informatica. Infine viene trattato anche il tema della Data Loss Prevenction, che è un'insieme di tecnologie e persone che lavora per evitare la perdita di dati sensibili.

Nel quinto capitolo viene analizzato il ruolo dell'AI e del Machine Learning in un contesto di attacco Ransomware. L'utilizzo di queste tecnologie permette al malware di cambiare ed adattarsi in tempo reale ai sistemi di sicurezza delle vittime, oltre che aumentare le probabilità di infezione creando trappole fatte su misura per ogni vittima. Viene inoltre affrontato il tema delle Generative Adversarial Network, ossia architetture progettate per addestrare un modello generativo che hanno rivoluzionato la creazione di contenuti artificiali.

Nel sesto capitolo invece si tratterà di quelli che sono i limiti del Machine Learning. L'efficacia dei sistemi di difesa basati su ML viene data dalla qualità del dataset con cui viene addestrato il modello e della dipendenza da esso: un dataset troppo sensibile, potrebbe rilevarsi troppo rigido, mentre un dataset troppo preciso potrebbe risultare troppo permissivo. Un'altra minaccia per il ML è il cosiddetto Data Poisoning, un attacco mirato ad inserire informazioni ingannevoli all'interno dell'algoritmo di addestramento del modello.

Nel settimo capitolo viene affrontato il tema del Quantum Computing, una scienza che potrebbe rivoluzionare il mondo della cybersicurezza. I computer quantistici potrebbero compromettere la sicurezza degli algoritmi crittografici tradizionali, accelerando la decifratura delle chiavi e rendendo vulnerabili molti sistemi attuali. D'altro canto però, il loro utilizzo potrebbe rendere molto più efficace e rapido il rilevamento del Ransomware, e migliorare le risposte a un qualsiasi tipo di attacco informatico.

Indice

Capitolo 1

Che cos'è il Ransomware

Ransomware è un tipo di malware (software dannoso progettato per compromettere o ottenere accesso non autorizzato a dispositivi, reti o dati) progettato appositamente per un guadagno finanziario. A differenza dei virus utilizzati negli attacchi di cracking, il Ransomware non è stato progettato per accedere ad un computer e rubare i dati da esso, ma nasce con lo scopo di bloccare l'accesso degli utenti ai file o ai sistemi, finché la vittima non paga un riscatto in cambio di una chiave che consente all'utente di decriptare i propri file. Tecnicamente sono trojan horse[1] crittografici che hanno come unico scopo l'estorsione di denaro, attraverso un sequestro di file, che mediante la cifratura li rende inutilizzabili.

1.1 L'evoluzione del Ransomware

1.1.1 1989: AIDS Trojan

Nel 1989 fece la sua comparsa **AIDS Trojan**, che criptava con chiave simmetrica AES[2] i file delle vittime. Il primo attacco prende il nome di "PC Cyborg"

[1] Trojan horse: file che può infettare qualsiasi dispositivo prendendone possesso.

[2] AES: tipo di crittografia che protegge i dati combinando velocità di utilizzo con una solida sicurezza, rendendoli impenetrabile agli attacchi di brute force.

(in quanto i pagamenti erano diretti alla “PC Cyborg Corporation”) e fu avviato da Joseph Popp, un accademico di Harvard che compiva ricerche sull’AIDS: fece distribuire circa 20.000 floppy intitolati “Informazioni sull’AIDS”, destinati ai delegati di una conferenza dell’OMS sull’AIDS. Il malware bloccava il funzionamento del computer giustificandolo attraverso la presunta “scadenza della licenza di un non meglio specificato software”, richiedendo 189 dollari come riscatto per tornare alla normalità.

Figura 1: AIDS Trojan

1.1.2 2005: GPCoder

L’uso intensivo di internet e della tecnologia da parte di tutte le fasce di utenti ha portato via via alla nascita del phishing, una nuova modalità di attacco legato al social engineering, che è un insieme di tecniche, non necessariamente informatiche, che possono indurre la vittima ad eseguire le azioni desiderate dall’attaccante, solitamente con l’obiettivo di ottenere informazioni, sfruttando le debolezze e i fattori psicologici che fanno normalmente parte della natura umana.

Il ritorno sulla scena dei Ransomware avvenne con il trojan **GPCoder**. Identificato nel 2005, GPCoder ha infettato i sistemi Windows, colpendo file di diverse varietà di estensioni. Una volta trovati i file, venivano copiati in forma crittografata e gli originali venivano eliminati. I nuovi file erano illeggibili e l’uso di una crittografia RSA-1024 assicurava che i tentativi di sbloccarli erano

estremamente improbabili. Veniva in seguito pubblicato un file `.txt` sul desktop della vittima, contenente i dettagli sul pagamento del riscatto e su come riavere i propri dati.

1.1.3 Dal 2009 al 2012: Locker

Nel 2009 si diffusero i primi fake antivirus che, dopo una scansione e la rilevazione di gravi minacce, imponevano alla vittima di pagare una somma per rimuovere i malware. Dal 2011, i fake antivirus hanno subito una mutazione: oltre a rilevare i finti virus, tendevano a bloccare i dispositivi delle vittime, segnando così la nascita dei **locker**. L'esplosione del Ransomware si ha a partire dal 2012 a seguito di una serie di fattori tecnologici, economici e sociali che ne hanno favorito la diffusione.

1. **Adozione delle Criptovalute**: dal 2012, Bitcoin ed altre criptovalute hanno iniziato a guadagnare popolarità: si tratta di una forma di denaro decentralizzato che funziona senza il controllo di una banca centrale o di un'autorità governativa. Funzionano su una rete distribuita chiamata blockchain, gestita da una comunità globale di computer. La loro natura anonima e decentralizzata ha fornito ai criminali un metodo sicuro per richiedere e ricevere riscatti;

2. **Miglioramento degli algoritmi di crittografia**: nel 2012, i Ransomware hanno iniziato ad integrare algoritmi di crittografia asimmetrica più robusti e difficili da decifrare, come RSA[3];

3. **Incremento delle superfici di attacco**: nel 2012 si ha l'aumento dell'uso dei dispositivi digitali e la dipendenza dai dati elettronici, fa-

[3]RSA: algoritmo di crittografia asimmetrica utilizzato per cifrare informazioni basato sulla difficoltà di fattorizzare un numero molto grande in due numeri primi: pur avendo accesso alla chiave pubblica, è molto difficile scoprire la chiave privata necessaria per decodificare il messaggio.

cendo incrementare le potenziali vittime di Ransomware, in quanto non disponevano ancora di adeguate misure di protezione.

1.2 Tipologie di Ransomware

1.2.1 Crypto Ransomware

Il **Crypto Ransomware** è una delle forme più comuni e più pericolose di Ransomware, progettato per trovare e crittografare i file delle vittime, rendendoli inutilizzabili senza una chiave di decrittazione che, nella maggior parte dei casi, viene fornita solo in cambio di un riscatto in criptovaluta.
Una volta infettato un sistema, il malware si attiva eseguendo una serie di operazioni mirate:

1. scansiona il dispositivo e la rete locale alla ricerca di file importanti, come documenti, immagini, database e backup;

2. applica algoritmi di crittografia avanzati, come AES o RSA, per bloccare l'accesso ai dati;

3. cancella o sovrascrive i backup per impedire il recupero dei file senza il pagamento del riscatto;

4. mostra una richiesta di pagamento, minacciando di cancellare o pubblicare i dati se il riscatto non viene versato entro un tempo limite.

1.2.2 Locker Ransomware

Il **Locker Ransomware** rappresenta una variante particolarmente insidiosa che, a differenza del Crypto Ransomware, è progettato per bloccare completa-

mente l'accesso al dispositivo, impedendone l'uso senza necessariamente alterare i dati presenti al suo interno.
Una volta infettato un sistema, il Locker Ransomware segue una strategia ben precisa:

1. blocca l'accesso al sistema operativo, impedendo all'utente di utilizzare il dispositivo;

2. mostra una schermata di blocco, spesso con un messaggio di riscatto che richiede un pagamento per ripristinare l'accesso;

3. simula avvisi legali fasulli, come notifiche da forze dell'ordine che accusano l'utente di attività illegali per spingerlo a pagare;

4. limita le opzioni di intervento, disabilitando il Task Manager, la modalità provvisoria e altre funzioni di ripristino.

1.2.3 Ransomware as a Service (RaaS)

Il **Ransomware as a Service** è un tipo di attacco in cui gli sviluppatori di Ransomware non eseguono direttamente gli attacchi, ma vendono o affittano il loro malware ad altri criminali informatici, permettendo anche a chi non ha competenze avanzate di lanciare campagne su vasta scala.
Gli sviluppatori creano il virus, progettandolo per essere efficace e difficile da rilevare, lo vendono (o affittano) su forum del dark web (spesso con pacchetti personalizzati e supporto tecnico), ottenendo anche una percentuale dei profitti generati dall'attacco.

1.2.4 Doxware

Il **Doxware** rappresenta un tipo di Ransomware sempre più diffuso che mette a rischio la privacy, poichéesfiltra dati sensibili dal dispositivo infetto e minaccia

di renderli pubblici se la vittima non paga.
Il suo modello di attacco è particolarmente invasivo:

1. infetta il dispositivo tramite uno dei metodi di propagazione tradizionali (vedi paragrafo 1.3);
2. cerca file personali, documenti riservati, credenziali e cronologia di navigazione;
3. carica su un server controllato dagli hacker tutte le informazioni raccolte;
4. minaccia la vittima richiedendo un pagamento per evitare la pubblicazione dei dati su internet o nel dark web.

1.3 Propagazione

I vettori di infezione utilizzati dai Ransomware sono sostanzialmente i medesimi usati per gli altri tipi di attacchi malware.

1.3.1 Phishing

La vittima riceve un'e-mail apparentemente legittima da banche, corrieri o colleghi, al cui interno si trova un link ad un sito dannoso oppure un allegato infetto, che, una volta cliccato, procederà a scaricare e a far eseguire il virus.

1.3.2 Drive-by download

In questo caso, la vittima naviga su siti compromessi dentro i quali sono stati introdotti, da parte di attaccanti che sono riusciti a violare il sito, exploit kit[4]

[4]Exploit kit: consentono agli aggressori di distribuire malware senza avere una conoscenza avanzata del codice utilizzato.

che sfruttano le vulnerabilità del browser e/o Java, presentando banner pubblicitari o pulsanti che, se cliccati, reindirizzano su siti malevoli dove avverrà il download del malware.

1.3.3 Software infetti

Questa è una pratica diventata molto pericolosa oggi giorno, che consiste nell'inserire il virus all'interno di bundle[5] di altri software, per esempio in programmi gratuiti che promettono di "crackare" software costosi per usarli senza pagare.

1.3.4 Dispositivi USB infetti

Questa tecnica si chiama baiting e fa leva sul fattore umano e sulla curiosità delle persone: viene lasciato in un luogo non custodito (ingresso di aziende, parcheggi, mense…) un supporto di memorizzazione (chiavetta USB o hard disk) contenente un malware; la curiosità umana fa sì che in molti casi questa esca funzioni e la persona inserisca il supporto nel proprio computer, facendo eseguire il Ransomware.

1.4 Modalità di esecuzione

Ogni tipologia di malware può quindi utilizzare diverse tecniche di crittografia, cifrare differenti tipi di file, forzare l'interruzione di alcuni processi, e così via, ma l'obbiettivo finale è comune a tutti. Le modalità di esecuzione del virus possono essere definite in cinque passaggi:

[5]Bundle: directory che raggruppa un eseguibile e le relative risorse quali file NIB, immagini, suoni, in una struttura standardizzata.

1. **infezione**: è il vettore di attacco che consente al Ransomware di entrare nel sistema, come abbiamo visto nel paragrafo 1.3;

2. **contatto con i server C&C**: alcuni Ransomware contattano un server di comando e controllo (C&C)[6] per ottenere o archiviare la chiave di crittografia. Questo passaggio può avvenire anche dopo la crittografia dei dati se il Ransomware funziona offline, con il server C&C contattato nell'ultima fase per memorizzare la chiave di decrittazione;

3. **gestione della chiave di crittografia**: la chiave può essere ottenuta dai server C&C o generata localmente. In quest'ultimo caso, il Ransomware crittograferà la chiave generata con una chiave ottenuta dal server C&C e successivamente la memorizzerà sul server;

4. **crittografia dei dati**: in questa fase, il Ransomware crittografa ed elimina i file dell'utente. Solitamente colpisce anche file e volumi montati utilizzando un protocollo di condivisione file in rete, come SMB[7];

5. **estorsione**: in questa ultima fase, il malware richiede il pagamento di un riscatto per decrittare i file, spiega come effettuare il pagamento, imposta una scadenza e minaccia di eliminare alcuni file dell'utente ogni ora.

Figura 2: Schema generale di attacco Ransomware

[6]Server C&C: server che un hacker utilizza come un generale sul campo di battaglia per comandare e controllare altre macchine.

[7]SMB: Server Message Block è un protocollo usato per condividere file, stampanti, porte seriali e comunicazioni di varia natura tra diversi nodi di una rete.

Capitolo 2
Strategie di mitigazione

Quando si affronta il problema del Ransomware, la prevenzione è definita migliore della cura. Questi attacchi sono una minaccia in costante evoluzione e per questo motivo è importante tenersi aggiornati con i nuovi sviluppi, per capire come questi funzionano e si diffondono. Bisogna far in modo che tutti gli utenti siano consapevoli delle tecniche utilizzate dai cybercriminali per diffondere i malware: la consapevolezza di questi attacchi può aiutare gli utenti a riconoscere ed evitare gli attacchi futuri.

2.1 Utilizzo di approcci diversificati di difesa

Gli attacchi Ransomware moderni sono sempre più sofisticati e sfruttano una combinazione di tecniche per compromettere i sistemi informatici. Spesso, un attacco può iniziare con una semplice e-mail di phishing contenente un link a un sito web compromesso o un allegato malevolo. Una volta che l'utente interagisce con il messaggio, il sito dannoso potrebbe sfruttare vulnerabilità del sistema o del browser per scaricare ed eseguire il Ransomware. In altri casi, gli

attaccanti potrebbero utilizzare accessi remoti compromessi, vulnerabilità non patchate o persino attacchi alla catena di fornitura per distribuire il malware. Per mitigare questi rischi, è fondamentale adottare una strategia di difesa multilivello, che affronti ogni fase dell'attacco con misure di protezione specifiche. Tra gli elementi chiave di questa strategia troviamo:

1. **filtraggio avanzato delle e-mail**: soluzioni di sicurezza per la posta elettronica che identificano e bloccano e-mail di phishing o e-mail contenenti allegati e link sospetti;

2. **protezione della navigazione web**: soluzioni come sistemi di filtraggio DNS[8] e firewall[9] possono impedire l'accesso a siti dannosi, bloccando il download di codice malevolo prima che venga eseguito;

3. **aggiornamenti e patch di sicurezza**: mantenere aggiornati sistemi operativi, software e dispositivi aiuta a chiudere le vulnerabilità che gli attaccanti potrebbero sfruttare;

4. **segmentazione della rete e privilegi minimi**: limitare l'accesso ai dati sensibili e segmentare la rete riduce l'impatto di un'eventuale compromissione;

Ogni livello di protezione introduce una barriera aggiuntiva che il Ransomware deve superare, riducendo drasticamente la probabilità di successo dell'attacco.

2.1.1 Negli smartphone

Quando si utilizzano dispositivi mobili, è fondamentale prestare attenzione alla sicurezza delle applicazioni installate, poiché i Ransomware e altre minacce

[8]Sistemi di filtraggio DNS: blocco di siti specifici il cui indirizzo IP viene preidentificato per garantire che l'accesso venga vietato.

[9]Firewall: software o dispositivo hardware che analizza il traffico di rete in entrata e in uscita e, in base a regole predefinite, crea una barriera per bloccare virus e minacce esterne.

possono infiltrarsi attraverso app malevole distribuite tramite store non ufficiali, link inviati via messaggi o siti web che offrono software apparentemente gratuito.

Una prima misura di sicurezza da utilizzare è quella di scaricare applicazioni solo da fonti affidabili, come il Google Play Store o l'Apple App Store, che riduce notevolmente il rischio di infezioni. Prima di installare un'app, è importante inoltre verificare l'elenco delle autorizzazioni richieste ed evitare quelle che richiedono eccessive autorizzazioni. Alcune applicazioni potrebbero tentare di ottenere accesso a funzioni non necessarie, come contatti, fotocamera o dati di localizzazione, anche se non strettamente legate al loro funzionamento. Un'app di torcia, ad esempio, non dovrebbe richiedere l'accesso ai messaggi o alla cronologia delle chiamate.

Un'altra misura essenziale per la protezione dei dispositivi mobili è l'attivazione della funzione di pulizia remota. Questa opzione permette di eseguire un ripristino di fabbrica completo in caso di infezione da ransomware, eliminando tutti i dati e ripristinando il sistema anche se il dispositivo è bloccato. Questa funzionalità è utile anche in caso di smarrimento o furto, impedendo l'accesso ai dati personali da parte di terzi.

2.2 Effettuare backup

Fare dei backup è sempre una buona idea, anche senza la minaccia di un Ransomware. I backup sono anche parti essenziali di un piano di continuità aziendale e di ripristino d'emergenza, che tutte le aziende dovrebbero avere. In ogni caso è consigliabile che tutti gli utenti effettuino almeno i backup dei file che ritengono importanti e che lo facciano regolarmente, a seconda del proprio profilo di rischio. Ci sono vari tipi di backup[1]:

1. **backup completo**: consiste nella copia di tutti i blocchi di cui è composto il file; ogni file è fatto da blocchi di dimensioni predeterminate; la modifica di un file consiste nella modifica di uno o più blocchi. Essendo una copia totale dei file, richiede sempre il massimo dello spazio su disco (pari alla somma delle dimensioni di ciascun file), del tempo necessario per l'esecuzione e delle risorse computazionali;

2. **backup incrementale**: copia solo i blocchi cambiati rispetto all'ultimo backup disponibile. Per eseguire il ripristino occorrono sia il backup completo di riferimento che ciascun backup incrementale fino al giorno scelto. Tra i vantaggi di questo tipo di backup troviamo la velocità di esecuzione e le dimensioni contenute in relazione al backup completo, mentre lo svantaggio principale è che per il ripristino necessita di tutti i backup intermedi;

3. **backup differenziale**: esegue una copia solo dei blocchi cambiati rispetto al backup completo di riferimento. I vantaggi consistono nella rapidità di esecuzione e nelle ridotte richieste di spazio rispetto al backup completo, ma richiede risorse maggiori e piu tempo rispetto al backup incrementale.

2.3 Servizio Copia Shadow (VSS)

Una copia shadow è una feature disponibile nei sistemi operativi Windows (sia Client che Server) che consente salvataggi istantanei cronologici (snapshot) di file e cartelle presenti nei volumi dei nostri computer, anche quando i file sono in uso. È una rete di sicurezza che cattura e conserva lo stato dei file in vari punti, in modo da poter accedere facilmente alle versioni precedenti. Le

copie shadow sono una soluzione pratica per recuperare rapidamente le versioni precedenti dei file senza tempi di inattività elevati.
Una soluzione VSS completa richiede tutti i seguenti componenti di base:

1. **Servizio VSS**: parte del sistema operativo Windows che garantisce che gli altri componenti possano comunicare tra loro correttamente e lavorare insieme;

2. **Richiedente VSS**: software che richiede la creazione effettiva di copie shadow;

3. **VSS Writer**: componente che garantisce che sia disponibile un set di dati coerente per il backup;

4. **Provider VSS**: componente che crea e gestisce le copie shadow.

Per creare una copia shadow, il richiedente, il writer e il provider eseguono le azioni seguenti:

1. il richiedente chiede a VSS di enumerare i writer, raccogliere i metadati del writer e preparare la creazione della copia shadow;

2. ogni writer crea una descrizione XML dei componenti e degli archivi dati di cui è necessario eseguire il backup;

3. VSS notifica a tutti i writer di preparare i dati per la creazione di una copia shadow;

4. VSS indica ai writer di bloccare temporaneamente le richieste di I/O di scrittura dell'applicazione e indica al provider di creare la copia;

5. VSS indica ai writer di sbloccare le richieste di I/O di scrittura;

Esistono varie metodologie per effettuare copie shadow[2]:

1. **copia completa**: crea una copia completa (di sola lettura) dell'intero volume in un determinato momento;

2. **copy-on-write**: effettua una copia differenziale del volume, ossia una copia di tutte le modifiche dopo un determinato momento;

3. **reindirizzamento in scrittura**: esegue una copia differenziale del volume reindirizzando tutte le modifiche a un volume diverso, lasciando invariato quello originale dopo un determinato momento.

Sebbene sia le copie shadow che i backup servano a proteggere i dati, il loro funzionamento è diverso: un backup è un archivio completo dei dati memorizzato in una posizione separata, utilizzato per il ripristino in caso di perdita significativa di dati (rete di sicurezza a lungo termine), mentre le copie shadow consentono di ripristinare rapidamente i file a versioni precedenti senza doverli recuperare da un backup (istantanee a breve termine).

2.4 Crittografia del disco

Uno dei modi più efficaci per salvaguardare i dati sensibili è tramite la crittografia del disco rigido o dell'intero disco. Questa misura di sicurezza completa crittografa tutti i dati sul disco rigido di un dispositivo, rendendolo inaccessibile agli utenti non autorizzati. La codifica dei dati viene eseguita utilizzando un algoritmo specifico, o cifratura, per convertire un'unità fisica in un formato codificato accessibile solo con una chiave o una password utilizzata per la crittografia, impedendo l'accesso ai dati a entità non autorizzate. Ne esistono due tipi[3]:

1. **crittografia dell'intero disco (FDE)**: protegge l'intera unità e tutti i suoi file da accessi non autorizzati. Funziona crittografando tutti i dati

su un'unità disco, assicurando che rimangano codificati e sicuri senza la chiave di decrittazione corretta. Quando i dati entrano nell'unità, vengono divisi in blocchi di dimensione fissa, in genere 128 bit o 256 bit. Ogni blocco viene quindi crittografato utilizzando una chiave di crittografia di lunghezza specificata a seconda dell'algoritmo di crittografia utilizzato;

2. **crittografia a livello di file (FLE)**: opera a livello di file system, consentendo la crittografia di singoli file e directory. I processi coinvolti nella crittografia a livello di file sono invisibili all'utente; lui deve essere autenticato nell'ambiente per decrittografare i file e il sistema crittografa automaticamente i file quando vengono memorizzati sull'unità di storage locale. La maggior parte dei servizi utilizza AES, che utilizza la stessa chiave per crittografare e decrittografare i dati, che viene spesso conservata in una posizione sicura sul sistema e crittografata da una chiave privata disponibile solo per l'amministratore.

2.4.1 BitLocker (Windows)

BitLocker[4] è una potente funzionalità di crittografia integrata nei sistemi operativi Windows. Questa tecnologia è stata progettata per proteggere i dati sensibili e informazioni confidenziali presenti sul disco rigido o su specifici dispositivi di archiviazione come unità USB o unità esterne. L'abilitazione del Bitlocker, rende virtualmente inaccessibili a persone non autorizzate tutti i dati presenti in un determinato ambiente.
Il processo di crittografia completa del disco di Windows inizia con una fase di crittografia tramite l'algoritmo AES, che potrebbe richiedere del tempo a seconda delle specifiche dell'unità. Questa fase garantisce che tutti i dati rimangano protetti anche quando il computer è spento o bloccato.
Prima che il sistema operativo venga caricato, è necessario inserire una pas-

sword, utilizzare una chiave USB o utilizzare altre opzioni di autenticazione per sbloccare il disco, che possono essere:

1. **password**: metodo basato sulla conoscenza di una password tipicamente complessa e difficile da indovinare;

2. **PIN**: metodo più conveniente del precedente, basato sula conoscenza di un PIN numerico abbastanza lungo e complesso;

3. **smart card**: soluzione che richiede un hardware aggiuntivo;

4. **chiavetta USB**: soluzione fisica che richiede l'inserimento di una chiavetta USB.

2.4.2 FileVault (macOs)

FileVault[5] è un sistema integrato di crittografia del disco progettato per sistemi operativi macOS. Il suo scopo è quello di crittografare i dati presenti sul disco rigido del Mac in background, in modo tale da renderli accessibili solo al proprietario.
Impiega un sistema di crittografia AES a 128 bit, con una chiave a 56 bit. Una volta abilitato, questo sistema agisce proteggendo i file presenti sul disco: dal momento della sua attivazione, dunque, tutti i dati verranno bloccati in automatico, e protetti da una password specifica. Frammenta i dati e i file suddividendoli in piccoli pacchetti, rendendo di fatto impossibile l'accesso agli utenti non in possesso della chiave apposita.

2.4.3 LUKS (Linux)

LUKS[6], acronimo di "Linux Unified Key Setup" si propone come lo standard per la crittografia dei dischi in sistemi Linux, è indipendente dalla piattaforma e può essere utilizzato per cifrare dischi in un'ampia varietà di strumenti, assicurando piena compatibilità e interoperabilità tra software diversi e garantendo

che la gestione delle password avvenga in una modalità sicura e documentata. Il suo funzionamento si basa su un sistema a due livelli: da un lato, i dati vengono cifrati con una chiave principale (Master Key), e dall'altro, questa chiave principale è protetta da una o più password utente memorizzate in appositi Key Slots. Quando si crea un volume cifrato con LUKS, viene generata una chiave principale casuale, che poi viene cifrata con la password dell'utente e salvata all'interno del LUKS Header, situato all'inizio del disco e fondamentale da proteggere perché una sua corruzione renderebbe i dati completamente irrecuperabili. Oltre alla possibilità di gestire più chiavi di accesso (quindi multi-utenza), LUKS supporta diversi algoritmi di cifratura, come AES-XTS, Serpent e Twofish, e varie modalità operative, tra cui XTS, comunemente usata per cifrare i dischi.
Con LUKS2, l'evoluzione più recente del sistema, sono stati introdotti miglioramenti significativi, come il supporto a Argon2, un algoritmo di derivazione della chiave più resistente agli attacchi brute-force, e un sistema più flessibile per la gestione delle chiavi.

2.5 Affidamento a terze parti: caso Cubbit

Cubbit è il primo provider europeo di cloud storage geo-distribuito. Nata a Bologna, la startup fornisce una soluzione per lo storage dei dati sicura, scalabile, con un sistema di crittografia e frammentazione dei dati, e che consente di risparmiare sui costi di archiviazione e backup dei dati, riducendo l'impatto ambientale. I dati salvati su Cubbit sono cifrati, frammentati e replicati su più sedi, al sicuro da minacce informatiche.

La rete P2P di Cubbit, chiamata Swarm, è composta da più nodi che lavorano insieme per fornire agli utenti un cloud storage sicuro e scalabile.

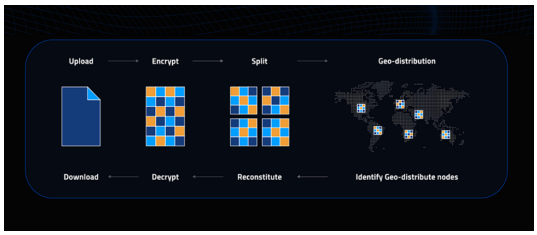

Figura 3: Workflow di Cubbit

Un nodo è una Cubbit Cell, un dispositivo autonomo che funge sia da unità di archiviazione dei dati che da nodo di relay, consentendo di archiviare e condividere i dati in modo sicuro sulla rete. Se una cella va offline, i suoi frammenti vengono automaticamente ridistribuiti e il nodo viene prontamente riparato dalla rete Cubbit.

2.5.1 Resistenza al Ransomware

Per garantire la massima sicurezza, i dati degli utenti vengono criptati con AES-256 e suddivisi in pezzi, che vengono poi elaborati in più frammenti ridondanti tramite codici a correzione d'errore Reed-Solomon e diffusi in modo sicuro sulla rete attraverso canali peer-to-peer criptati end-to-end. Nessuna Cubbit Cell archivia file o oggetti nella sua interezza, ma archivia frammenti criptati dei file di più persone. Con il cloud storage S3 compatibile di Cubbit, è possibile evitare attacchi di tipo Ransomware: permette di scegliere quali dati sono immutabili e una data di scadenza entro la quale, i dati, non possono essere crittografati, alterati o cancellati da niente e nessuno. L'S3 Object Locking è fondamentale per adottare una strategia WORM (Write Once Read Many) che permette di proteggere i dati critici da Ransomware ed eliminazioni accidentali rimanendo al contempo conforme ai requisiti di revisione e conservazione. Si basa su cinque pilastri fondamentali per garantire l'immutabilità dei dati e la loro protezione:

1. **metadati immutabili**: il sistema di storage associa a ogni oggetto (file o dato) un set di metadati immutabili che includono il flag di locking e

il periodo di retention;

2. **protezione a livello di file system o database**: l'immutabilità è applicata mediante il controllo degli accessi sul file system o sul database sottostante;

3. **crittografia e hash per garantire l'integrità**: la protezione del dato è ulteriormente rinforzata da meccanismi crittografici (crittografia end to end, checksum[10] e hash crittografici);

4. **controllo del consenso in sistemi distribuiti**: in sistemi decentralizzati come Cubbit, l'integrità dei dati è protetta anche da protocolli di consenso (la modifica di un frammento deve essere acconsentita da tutti i nodi, cosa che però non possibile se il lock è attivo);

5. **politiche di sicurezza a livello applicativo**: l'applicazione stessa gestisce l'object locking (API RESTful o interfaccia utente) e implementa ulteriori protezioni come il rifiuto automatico delle operazioni non valide e gli audit di log.

[10]Checksum: serve a controllare l'integrità di un dato o di un messaggio che potrebbe subire delle modifiche durante la sua trasmissione.

Capitolo 3

Ransomware e AI: come cambia

Purtroppo, le soluzioni anti-Ransomware e anti-malware tradizionali non sono in grado di contrastare le moderne minacce informatiche in maniera efficace: faticano a riconoscere le minacce che fanno uso di esfiltrazione[11] e crittografia per acquisire informazioni critiche. I criminali informatici sfruttano l'intelligenza artificiale per rendere i Ransomware sempre più evoluti ed effettuare attacchi più efficaci e "produttivi", andando a creare phishing, vishing (phishing telefonico) e smishing (phishing basato su SMS). È altresì possibile effettuare attacchi alla rete, alle applicazioni, nascondere l'esfiltrazione dei dati nel traffico normale, nonché gestire la negoziazione del riscatto da pagare a seguito di un attacco Ransomware.

[11]Esfiltrazione: trasferimento intenzionale, non autorizzato e occulto di dati da un computer o altro dispositivo.

3.1 La nascita dell'Intelligenza Artificiale

Le prime tracce di Intelligenza Artificiale come disciplina scientifica risalgono agli anni Cinquanta, un periodo di grande fermento scientifico sullo studio del calcolatore e il suo utilizzo per sistemi intelligenti. Nel 1956, al Darmouth College, nel New Hampshire, si tenne un convegno al quale presero parte i maggiori esponenti dell'informatica, durante il quale ebbe un ruolo fondamentale il lavoro di Alan Turing, considerato uno dei padri dell'informatica moderna, che prende il nome di "Test di Turing": secondo il test una macchina poteva essere considerata intelligente se il suo comportamento, osservato da un essere umano, fosse considerato indistinguibile da quello di una persona. Il tema dell'Intelligenza Artificiale ricevette una forte attenzione da parte della comunità scientifica e nacquero diversi approcci. I principali furono la logica matematica, per la dimostrazione di teoremi e l'inferenza di nuova conoscenza, e le reti neurali.

3.2 AI Forte e AI Debole

Le aspettative sulle applicazioni dell'Intelligenza Artificiale, col tempo, iniziarono a crescere. Tuttavia, poiché i macchinari dell'epoca non disponevano di una capacità computazionale adeguata, questa e altre aspettative non furono mantenute e ciò portò alla frammentazione dell'Intelligenza Artificiale in distinte aree basate su teorie diverse, facendo emergere due paradigmi principali: Intelligenza Artificiale Forte e Intelligenza Artificiale Debole.

3.2.1 AI Forte

La teoria dell'Intelligenza Artificiale Forte sostiene che le macchine sono in grado di sviluppare una coscienza di sé. Questo paradigma è supportato dal campo di ricerca nominato Intelligenza Artificiale Generale (AGI), che studia sistemi in grado di replicare l'intelligenza umana. Quest'area di ricerca ha ricevuto però poco interesse da buona parte della comunità scientifica che ritiene l'intelligenza umana troppo complessa per essere replicata.

3.2.2 AI Debole

Il paradigma dell'Intelligenza Artificiale Debole, in opposizione al primo, ritiene possibile sviluppare macchine in grado di risolvere problemi specifici senza avere coscienza delle attività svolte. In altre parole, l'obiettivo dell'IA Debole non è realizzare macchine dotate di un'intelligenza umana, ma di avere sistemi in grado di svolgere una o più funzioni umane complesse.

3.3 L'Intelligenza Artificiale oggi

Negli ultimi decenni, l'Intelligenza Artificiale Debole ha avuto un'evoluzione straordinaria, passando dalle prime applicazioni in ambito industriale a una tecnologia onnipresente in numerosi settori. Se negli anni Ottanta l'AI era utilizzata principalmente per sistemi esperti e automazione di processi industriali, oggi è al centro di innovazioni che stanno trasformando il modo in cui interagiamo con la tecnologia e affrontiamo problemi complessi.
L'Intelligenza Artificiale moderna si è sviluppata grazie alla crescita esponenziale della potenza di calcolo, della disponibilità di dati e degli algoritmi di apprendimento automatico. Tra le principali aree di sviluppo troviamo:

1. **Machine Learning**: un insieme di tecniche che permette ai sistemi di apprendere dai dati ed effettuare previsioni senza essere esplicitamente programmati;

2. **Deep Learning**: algoritmi come le reti neurali profonde hanno portato a progressi significativi in campi come il riconoscimento delle immagini, la diagnosi medica e la sicurezza informatica;

3. **Elaborazione del Linguaggio Naturale (NLP)**: l'AI è ora in grado di comprendere e generare testo in linguaggio umano, come dimostrano strumenti avanzati come ChatGPT, Google Bard e gli assistenti vocali Siri e Alexa;

4. **AI Generativa**: un settore che utilizza reti neurali per creare nuovi contenuti, come testo, immagini, video e musica;

5. **Robotica e Automazione**: l'integrazione dell'AI nella robotica ha portato alla creazione di robot autonomi in grado di svolgere compiti complessi.

3.4 Ransomware e AI

Il Ransomware esiste da tempo, ma la sua efficacia è aumentata in modo significativo da quando gli aggressori sfruttano la tecnologia AI per identificare i sistemi vulnerabili e lanciare attacchi altamente mirati contro di essi [7]. Utilizzando algoritmi avanzati, gli aggressori possono scansionare rapidamente le reti alla ricerca di punti deboli che possono essere sfruttati e quindi adattare i loro payload[12] di conseguenza, rendendo molto più difficile per le misure di si-

[12]Payload: routine presente in un virus informatico che ne estende le funzioni oltre l'infezione del sistema.

curezza tradizionali, come i firewall o il software antivirus, rilevarli o prevenirli prima che il danno sia fatto, oppure possono creare malware e Ransomware dotati di intelligenza artificiale, consapevoli della situazione e altamente evasivi, in grado di analizzare i meccanismi di difesa del sistema bersaglio e di imparare rapidamente a imitare le normali comunicazioni del sistema, in modo da poter rimanere inosservate fino a quando non hanno causato danni irreparabili. Le vittime non hanno altra scelta se non quella di pagare per sperare di recuperare i dati perduti o di accedere nuovamente ai loro sistemi senza subire perdite finanziarie significative solo per i costi di inattività.

3.4.1 Machine Learning

I ricercatori hanno scoperto che gli attori delle minacce informatiche possono utilizzare modelli di apprendimento automatico Machine Learning (ML) che alimentano l'intelligenza artificiale per distribuire malware e muoversi lateralmente attraverso le reti aziendali. Il Machine Learning si occupa di creare sistemi che apprendono o migliorano le performance in base ai dati che utilizzano. Il suo compito è addestrare i computer a imparare dai dati e a migliorare con l'esperienza, anziché essere appositamente programmato per riuscirci.
Nel machine learning, gli algoritmi vengono addestrati a far emergere schemi e correlazioni da grandi set di dati e a formulare le migliori decisioni e previsioni sulla base di tali analisi. Le applicazioni di machine learning migliorano con l'uso e diventano più accurate man mano che aumentano i dati a cui hanno accesso. Esistono tre principali approcci di ML:

1. **Supervised Learning**: l'algoritmo viene addestrato su un set di dati etichettato. Un esempio di utilizzo di questo algoritmo nella cybersecurity è l'identificazione di malware, dove il sistema impara a riconoscere comportamenti dannosi confrontandoli con esempi precedenti;

2. **Unsupervised Learning**: qui l'algoritmo esplora dati non etichettati per scoprire pattern nascosti. E' utile, ad esempio, nel rilevamento di anomalie all'interno di reti aziendali;

3. **Reinforcement Learning**: il sistema impara attraverso feedback continui, migliorando progressivamente le proprie prestazioni, come avviene nei sistemi di difesa automatizzati che si adattano dinamicamente a nuovi attacchi.

3.5 Ransomware adattivo

Il termine Ransomware adattivo si riferisce a una categoria di Ransomware che, grazie all'integrazione di tecniche di ML, è in grado di adattarsi dinamicamente alle specifiche caratteristiche del sistema bersagliato, migliorando l'efficacia dell'attacco e rendendo più difficile la rilevazione da parte dei sistemi di sicurezza. Un Ransomware adattivo che sfrutta il ML è più sofisticato e riesce a modificare il proprio comportamento in base a diversi fattori, come:

1. il **comportamento dell'utente**: il Ransomware può analizzare come l'utente interagisce con il sistema (quali file aprono, su quali applicazioni lavorano, ecc.) e decidere quali file criptare per massimizzare il danno;

2. le **caratteristiche del sistema**: se il Ransomware rileva che il sistema bersaglio ha determinate difese, potrebbe scegliere di modificare la propria strategia di attacco (ad esempio, evitando determinati file di sistema, se rileva una macchina virtuale[13] o una sandbox[14]);

[13]Macchina virtuale: software che crea un ambiente virtuale che emula tipicamente il comportamento di una macchina fisica.

[14]Sandbox: pratica di sicurezza in cui si utilizza un ambiente isolato all'interno del quale si esegue il codice e lo si analizza senza influire sull'applicazione, sul sistema o sulla piattaforma.

3. il **comportamento delle difese**: analizzando i comportamenti dei sistemi di rilevamento antivirus o di prevenzione delle intrusioni, il Ransomware può scegliere di eseguire operazioni meno evidenti per evitare di essere individuato.

3.5.1 ML come contributo all'adattività del Ransomware

Il ML permette al Ransomware di apprendere e di ottimizzare le sue azioni, utilizzando modelli predittivi che gli consentono di:

1. il **ottimizzare i tempi di crittografia**: il Ransomware alimentato dal ML può analizzare i tempi di risposta del sistema, ottimizzando la velocità con cui agisce e cercando di evitare rallentamenti che potrebbero rivelarlo;

2. il **rilevare e bypassare i meccanismi di difesa**: attraverso tecniche di apprendimento automatico, il Ransomware può analizzare i file di sistema e riconoscere segnali tipici di un ambiente di sicurezza (come un antivirus o una sandbox). Ad esempio, se rileva la presenza di un antivirus o di un sistema di rilevamento, può decidere di non crittografare alcun file, attivando invece una strategia di evasione, come il monitoraggio invisibile dei dati o l'esfiltrazione;

3. il **evoluzione dinamica del comportamento**: grazie agli algoritmi di apprendimento, il Ransomware può anche adattare la propria strategia in base al successo o al fallimento dell'attacco. Se un certo metodo di crittografia è facilmente individuato da un sistema di sicurezza, il Ransomware può imparare a modificarlo nel tempo, rendendosi più difficile da rilevare e da neutralizzare.

Capitolo 4
Tecniche di difesa basate su AI

Il Machine Learning e l'Intelligenza Artificiale sono componenti cruciali della protezione moderna dal Ransomware, poiché consentono di rilevare i modelli di comportamento sospetti anziché limitarsi a confrontare la firma di un'istanza di malware con quelle di un database di minacce note. Senza questo approccio comportamentale, nessuna misura anti-malware è in grado di identificare correttamente le migliaia di istanze di malware zero-day generate quotidianamente dagli hacker. Il Machine Learning consente di rilevare informazioni sugli exploit non noti rifacendosi ai dati delle interazioni precedenti e attuali con il sistema come riferimento per il comportamento sicuro. Dal momento in cui le organizzazioni raccolgono sempre più dati, questo approccio consente di rilevare le minacce zero-day[15] in maniera più affidabile. Adottando soluzioni anti-Ransomware con AI e ML integrati, le aziende traggono vantaggio dalla capacità della tecnologia di imparare e adattarsi.

Con il passare del tempo, i sistemi sono in grado di sviluppare una base di riferimento per il comportamento previsto e di metterla a confronto con dati e comportamenti nuovi e cambiati.

[15]Minacce zero-day: i malintenzionati ottengono l'accesso a un sistema sfruttando una vulnerabilità della sicurezza in un programma software di cui il produttore non ne è a conoscenza.

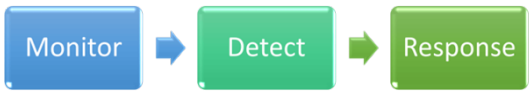

Figura 4: Schema di difesa basato sul ML

4.1 Endpoint Detection and Response

Endpoint Detection and Response (EDR)[8] è una soluzione integrata per la sicurezza degli endpoint che si basa sull'analisi dei dati degli endpoint, sul monitoraggio continuo in tempo reale e sulla risposta automatizzata basata su regole per proteggere un sistema da minacce persistenti avanzate e potenziali incidenti di sicurezza. Le soluzioni di sicurezza EDR sono in grado di rilevare comportamenti sospetti del sistema su host ed endpoint, raccogliere dati sugli endpoint e analizzare singoli eventi, andando quindi ad indagare sulla causa principale del comportamento dannoso per avvisare il team di sicurezza e aiutarlo a correggere le minacce prima che i file dannosi possano influire sull'ambiente. Le soluzioni di sicurezza EDR sono specializzate in diverse funzioni primarie.

4.1.1 Rilevamento automatizzato delle minacce

EDR implementa una visibilità completa su tutti gli endpoint per rilevare vari indicatori di attacco (IOA) e analizza miliardi di eventi in tempo reale per identificare automaticamente le attività sospette verso la rete protetta. Le solide soluzioni di sicurezza EDR si sforzano di comprendere un singolo evento come parte di una sequenza più significativa per applicare la logica di sicurezza. Se una sequenza di eventi punta a un IOA noto, la soluzione EDR lo identificherà come dannoso ed emetterà automaticamente un avviso di rilevamento.

4.1.2 Integrazione dell'intelligence sulle minacce

Le soluzioni integrate combinano il monitoraggio delle minacce e della rete con l'intelligence sulle minacce per rilevare più rapidamente i comportamenti dannosi. Se lo strumento EDR rileva tattiche, tecniche e procedure (TTP) sospette, fornirà dettagli completi sul potenziale incidente di sicurezza, prima che si verifichino violazioni dei dati (possibili aggressori, superficie di attacco più vulnerabile, modalità di distribuzione del malware e altre informazioni già note sull'attacco).

4.1.3 Monitoraggio in tempo reale e visibilità storica

EDR utilizza l'aggregazione attiva dei dati degli endpoint per rilevare gli incidenti di sicurezza subdoli. Agli utenti viene fornita una visibilità completa su tutte le attività sugli endpoint aziendali dal punto di vista della sicurezza informatica. Una soluzione dedicata può tenere traccia di una miriade di eventi relativi alla sicurezza, tra cui la creazione di processi, le modifiche del registro, il caricamento dei driver, l'utilizzo della memoria e del disco, l'accesso al database centrale, le connessioni di rete e altro ancora.

4.1.4 Rapida indagine sulle minacce

Le soluzioni per la sicurezza degli endpoint possono analizzare rapidamente le minacce e accelerare la correzione. È possibile considerarli come analisti della sicurezza, che raccolgono i dati da ogni evento endpoint e li archiviano in un enorme database centralizzato che fornisce dettagli e contesto completi per consentire indagini rapide sia per i dati in tempo reale che per quelli storici.

4.1.5 Integrazione EDR

L'integrazione di un EDR è una delle soluzioni più efficaci per proteggere i dispositivi da minacce avanzate ma, tuttavia, la sua implementazione può va-

riare in difficoltà. Per un utente singolo o una piccola azienda, l'installazione di un EDR è relativamente semplice in quanto molte soluzioni moderne, come Microsoft Defender for Endpoint, offrono interfacce intuitive e installazioni automatizzate, non richiedendo competenze avanzate. La situazione cambia per le aziende più grandi o per ambienti con infrastrutture complesse, dato che l'implementazione di un EDR richiede esperti di cybersecurity, amministratori di rete e specialisti in threat hunting, stabilendo protocolli di risposta alle minacce.

Dal punto di vista delle risorse hardware, gli EDR più moderni sono ottimizzati per funzionare su dispositivi standard senza impattare troppo sulle prestazioni, ma in ambienti aziendali con migliaia di endpoint, il monitoraggio continuo può richiedere risorse di calcolo elevate, soprattutto quando vengono analizzati grandi volumi di dati o si effettuano analisi avanzate basate su intelligenza artificiale. Un altro aspetto da considerare è il costo: alcune soluzioni EDR sono integrate nei sistemi operativi, mentre altre richiedono licenze a pagamento che possono diventare costose, soprattutto se si vogliono funzionalità avanzate come il roll-back[16] automatico dei file in caso di attacco Ransomware.

Figura 5: EDR

[16]Roll-back: operazione che permette di riportare i dati a uno stato precedente.

4.2 eXtended Detection and Response

XDR (eXtended Detection and Response)[9] raccoglie e correla automaticamente i dati tra più livelli di sicurezza: e-mail, endpoint, server, workload in cloud e rete. Ciò permette di rilevare più velocemente le minacce e di migliorare i tempi di indagine e di risposta attraverso l'analisi della sicurezza.

4.2.1 Come funziona XDR

L'architettura XDR combina più funzionalità di sicurezza e fonti di dati in un'unica piattaforma centralizzata.

1. **Raccolta di dati**: XDR raccoglie dati in tempo reale da un'ampia gamma di fonti nell'infrastruttura IT, tra cui:
   (a) Endpoint (computer, server, dispositivi mobili);
   (b) Rete (firewall, router, switch, traffico di rete);
   (c) Applicazioni (e-mail, software aziendali, sistemi SaaS[17]);
   (d) Cloud (infrastrutture cloud pubbliche e private);
   (e) Identità e accessi (sistemi IAM[18], Active Directory).
2. **Normalizzazione e centralizzazione**: i dati raccolti vengono normalizzati e aggregati in un data lake centralizzato in modo da rendere compatibili tra loro i dati provenienti da fonti diverse e fornire un contesto unificato per l'analisi successiva;
3. **Analisi avanzata con AI**: una delle caratteristiche distintive di XDR è l'uso di tecniche avanzate di analisi:

[17] Sistema SaaS: consente agli utenti di connettersi ad app basate sul cloud tramite Internet e usare tali app.
[18] Sistemi IAM: controllo fondamentale della sicurezza del cloud in quanto autentica gli utenti e regola l'accesso a sistemi, reti e dati.

(a) i modelli di ML analizzano grandi volumi di dati per identificare anomalie e comportamenti sospetti;

(b) si fa il confronto del comportamento normale di utenti, dispositivi e applicazioni con attività anomale, come accessi da località insolite, modifiche a file non autorizzate e incremento anomalo del traffico di rete;

(c) XDR utilizza feed di informazioni sulle minacce per rilevare schemi di attacco noti.

4. **Correlazione degli eventi**: XDR correla i dati provenienti da più fonti per identificare minacce complesse e multistadio.

5. **Risposta automatizzata alle minacce**: una volta identificata una minaccia, XDR può attivare risposte come isolamento degli endpoint compromessi, blocco del traffico di rete sospetto, blocco o revoca degli accessi a risorse critiche, creazione automatica di regole di sicurezza per future minacce simili. Le risposte possono essere di due tipologie:

   (a) **automatiche** se vengono eseguite immediatamente per bloccare la minaccia in tempo reale;

   (b) **guidate** se forniscono ai team di sicurezza dei suggerimenti per la gestione manuale degli incidenti;

6. **Dashboard centralizzata e visione olistica**: XDR offre una dashboard unica per monitorare e gestire la sicurezza su tutti i livelli dell'infrastruttura IT.

4.2.2 Integrazione XDR

L'adozione di un XDR può essere più semplice per aziende che dispongono già di una solida infrastruttura IT, ma per le PMI potrebbe risultare più complessa, specialmente se non hanno un team dedicato alla cybersecurity: se da

un lato l'XDR automatizza molti processi di rilevamento, riducendo il numero di falsi positivi e migliorando il tempo di risposta, dall'altro è fondamentale configurarlo correttamente per evitare problemi di compatibilità o inefficienza operativa. Essendo basato in gran parte su cloud, l'XDR non richiede investimenti significativi in server o infrastrutture locali, rendendolo scalabile e accessibile anche alle PMI. Tuttavia, l'analisi costante dei dati provenienti da molteplici fonti può comportare un aumento del traffico di rete (soprattutto in ambienti con numerosi dispositivi connessi) e un carico di elaborazione significativo.

Figura 6: XDR

4.3 Managed Detection and Response

MDR (Managed Detection and Response)[10] è un servizio di sicurezza informatica che consente di proteggere in modo proattivo le organizzazioni dalle minacce informatiche usando il rilevamento avanzato e la rapida risposta agli eventi imprevisti. I servizi MDR includono una combinazione di tecnologia e competenze umane per eseguire la ricerca, il monitoraggio e la risposta alle minacce informatiche. A differenza di strumenti tradizionali che richiedono gestione e monitoraggio interni (come EDR), con l'MDR, una terza parte esperta in sicurezza (il provider MDR) si occupa del monitoraggio, della rilevazione e della risposta alle minacce.

4.3.1 Come funziona MDR

Il processo di rilevamento e risposta gestito include in genere cinque passaggi:

1. **scala di priorità**: è estremamente dispendioso in termini di tempo per i team di sicurezza esaminare gli infiniti avvisi di cybersecurity che ricevono ogni giorno. Questo è il motivo per cui molti partner MDR offrono ciò che è noto come definizione di priorità gestita. Usando una combinazione di automazione e analisi umana, MDR ordina l'enorme volume di avvisi dell'organizzazione e separa i falsi positivi dalle minacce informatiche significative, presentando quindi un flusso di avvisi di alta qualità al team di sicurezza;

2. **ricerca**: MDR offre funzionalità proattive e complete per la ricerca di minacce informatiche 24 ore su 24. Le piattaforme di Intelligence sulle minacce informatiche raccolgono dati critici sui potenziali rischi e queste informazioni vengono quindi passate agli analisti. Questi esperti umani hanno competenze e conoscenze estese per identificare e rispondere a minacce informatiche furtive che a volte non sono presenti nelle soluzioni tecniche automatizzate;

3. **analisi**: gli analisti MDR esaminano le minacce informatiche per offrire all'organizzazione una chiara comprensione della portata e del significato delle minacce informatiche, fornendo informazioni dettagliate come il tipo di attacco informatico, il momento in cui si è verificato, chi è interessato e la gravità dell'attacco informatico. Usando queste informazioni preziose, riescono a tracciare una risposta efficace e ad identificare i passaggi successivi;

4. **correzione**: è il processo di interruzione dell'attacco informatico per impedirne la diffusione. Ciò può comportare la rimozione di malware, l'isolamento di reti o sistemi interessati, l'eliminazione di intrusioni, la

pulizia del registro di sistema e l'eliminazione dei meccanismi di persistenza del malware. Una correzione efficace garantisce che la rete venga ripristinata allo stato precedente all'attacco informatico;

5. **neutralizzazione**: dopo che l'attacco informatico è stato arrestato e la rete è stata ripristinata allo stato precedente, gli analisti eseguono un'analisi della causa radice. In questo modo possono eliminare completamente il cyberattacker e impedire che si verifichino occorrenze future dello stesso tipo di minacce informatiche.

4.3.2 Integrazione MDR

A differenza di un EDR, che viene installato e gestito internamente, un MDR è un servizio completamente gestito da un provider di sicurezza, il quale monitora costantemente l'infrastruttura IT, identifica anomalie e risponde alle minacce in tempo reale, rendendolo particolarmente adatto per piccole e medie imprese, che spesso non dispongono di un team IT dedicato alla cybersecurity. Per le grandi aziende, invece, l'integrazione di un MDR può risultare più complessa: se da un lato offre il vantaggio di un monitoraggio H24 senza la necessità di gestire un Security Operations Center (SOC) interno, dall'altro potrebbe essere necessario integrarlo con strumenti di sicurezza esistenti, come per esempio firewall aziendali, e, inoltre, il team IT interno deve comunque interfacciarsi con il provider per definire policy di sicurezza e gestire eventuali azioni di risposta agli incidenti.
Dal punto di vista delle risorse hardware, un MDR ha un impatto minimo sulle infrastrutture aziendali, poiché sfrutta principalmente il cloud per le operazioni di monitoraggio e analisi. In alcuni casi però, potrebbe essere richiesta l'installazione di agenti sugli endpoint per migliorare la capacità di rilevamento delle minacce, il che potrebbe avere un leggero impatto sulle prestazioni dei dispositivi.

Figura 7: MDR

4.4 Data Loss Prevention

La prevenzione della perdita dei dati (DLP) è una combinazione di persone, processi e tecnologia che lavora per rilevare e impedire la perdita di dati sensibili. Una soluzione DLP utilizza risorse come software antivirus, IA e apprendimento automatico per rilevare attività sospette confrontando i contenuti con i criteri DLP dell'organizzazione o dell'individuo, che definiscono come i dati vengono condivisi e protetti, senza esporli a utenti non autorizzati[11].
Le soluzioni DLP automatizzate sono molto più sicure del tracciamento manuale dei dati tra ambienti e dispositivi, in quanto ispezionano dati in tempo reale, verificano la presenza di parole chiave ed elementi che segnalano informazioni sensibili, andando quindi, a crittografare o limitare l'accesso in base ai criteri scelti dall'utente. Lavorano principalmente in tre aree principali:

1. **dati in uso**, ossia i dati con cui si sta lavorando;
2. **dati in movimento**, ossia i dati condivisi tra le reti;
3. **dati in tempo reale**, ossia i dati memorizzati (locale e cloud).

4.4.1 Principali caratteristiche di DLP

Queste soluzioni sono in grado di offrire varie funzionalità, tra cui[12]:

1. **rilevamento dei dati**: identificazione e localizzazione dei dati sensibili all'interno della rete, dei sistemi e delle repository di archiviazione di un'organizzazione o di un individuo;

2. **crittografia e mascheramento dei dati**: permette alla soluzione DLP di proteggere i dati, crittografando o mascherando le informazioni sensibili evitando così, che siano esposte ad accessi non autorizzati;

3. **risposta e segnalazione degli incidenti**: con questa funzione, la soluzione DLP è in grado di fornire avvisi e report sugli incidenti di sicurezza dei dati;

4. **meccanismi di correzione e quarantena degli utenti**: la soluzione di DLP è in grado di isolare o di bloccare automaticamente l'accesso ai dati sensibili quando vengono rilevate minacce;

5. **protezione degli endpoint e dei dati sul cloud**: la soluzione DLP è in grado di salvaguardare i dati sensibili archiviati o elaborati in ambienti cloud, oltre che i dati sui dispositivi degli utenti.

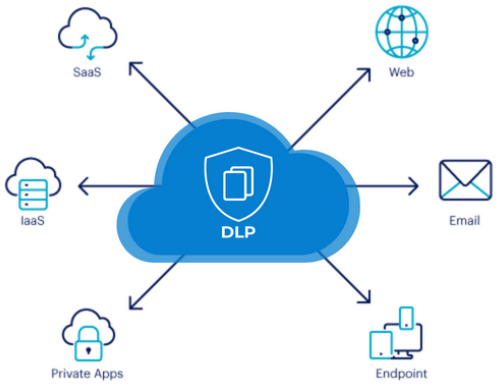

Figura 8: DLP

Capitolo 5
ML come arma per i cybercriminali

L'uso del Machine Learning a favore dei Ransomware rappresenta una preoccupante evoluzione nel campo della sicurezza informatica. Invece di limitarsi a tentativi di attacco più rudimentali, i criminali informatici possono sfruttare il ML per rendere gli attacchi più sofisticati, adattivi e difficili da rilevare, massimizzando il danno e aumentando le probabilità di successo.

5.1 Ottimizzazione degli attacchi

Il Machine Learning può essere utilizzato dai criminali per ottimizzare l'esecuzione degli attacchi, migliorandone l'efficacia. Alcuni approcci prevedono:

1. **apprendimento dalle vittime precedenti**: i criminali informatici possono usare algoritmi di ML per raccogliere informazioni sulle vittime precedenti e sulle caratteristiche comuni per selezionare i target più redditizi;

2. **automazione della selezione dei target**: utilizzando l'analisi predittiva e il clustering[19], il Ransomware potrebbe scegliere in modo intelligente i sistemi da colpire, identificando le vittime con la configurazione più vulnerabile o quelle con i dati più preziosi;

3. **ottimizzazione del metodo di crittografia**: il Ransomware alimentato da ML può adattarsi e scegliere automaticamente il metodo di crittografia più efficace, modificandolo in base alle risorse disponibili.

5.2 Attacchi di ingegneria sociale più convincenti

Le tecniche di apprendimento automatico sono molto potenti nell'ingegneria sociale. Con l'utilizzo del ML è possibile sferrare attacchi più convincenti grazie a:

1. **generazione automatica delle e-mail di phishing**: i criminali possono creare messaggi di phishing altamente personalizzati, grazie alle informazioni raccolte precedentemente sulle vittime, che sono più difficili da riconoscere come malevoli;

2. **deepfake**: con l'uso del Deep Learning, è possibile creare video o audio falsi (deepfake) per impersonare un individuo e convincere la vittima a eseguire azioni che compromettono la sicurezza;

3. **analisi dei comportamenti dell'utente**: i modelli di ML possono analizzare i comportamenti dell'utente per personalizzare l'attacco.

[19]Clustering: insieme di tecniche di analisi multivariata dei dati volte alla selezione e raggruppamento di elementi omogenei in un insieme di dati.

5.3 Evasione e mimetizzazione dell'attacco

Il Machine Learning consente ai Ransomware di essere più furtivi ed evitare il proprio rilevamento, migliorando la loro capacità di mimetizzarsi all'interno dei sistemi compromessi:

1. **rilevamento e bypass dei sandbox**: i sistemi di sicurezza usano spesso ambienti sandbox per analizzare i malware. Con l'aiuto del ML, il Ransomware può riconoscere quando si trova in un ambiente sandbox e modificare il proprio comportamento per non attivarsi;

2. **evasione dalle difese tradizionali**: un Ransomware alimentato da ML può imparare a distinguere tra "sistemi di difesa attivi" e "sistemi normali", scegliendo di non eseguire azioni sospette quando rileva software di sicurezza o modelli di difesa;

3. **modifica dinamica del comportamento**: se il Ransomware rileva che una certa sequenza di operazioni è stata già identificata come sospetta, può cambiare automaticamente il suo comportamento.

5.4 Ottimizzazione del pagamento del riscatto

Negli attacchi Ransomware tradizionali, i dati venivano cifrati e veniva richiesto come riscatto una somma standard espressa in criptovalute, senza tener conto del potere economico delle vittime. Con l'integrazione del ML, si è passati a richiedere il pagamento di somme personalizzate in base alle disponibilità della vittima. Prima di procedere alla crittografia dei dati, o all'invio della metodologia di riscatto, il ML raccoglie informazioni sulla vittima (dati finanziari,

dati personali, dati comportamentali e/o informazioni sull'organizzazione bersaglio) e in base a questi, formula importo e modalità di pagamento ad hoc. Sfruttare il ML consente ai cybercriminali di aumentare l'efficienza operativa (utilizzando la AI, si ha una riduzione del tempo di attacco), aumentare il tasso di pagamento (somme personalizzate diminuiscono il rischio di rifiuto) e aumentare i profitti.

5.5 Generative Adversarial Networks

Il ML, in particolare attraverso le Generative Adversarial Networks (GANs)[13], può essere utilizzato per generare varianti del codice del Ransomware che cambiano continuamente per evitare il rilevamento da parte di software di sicurezza basati su firma. Le GAN sono un'architettura per addestrare un modello generativo di AI, che ha rivoluzionato il modo in cui si creano dei contenuti artificiali.

5.5.1 Struttura GAN

Questo tipo di rete è stato introdotto per la prima volta nel 2014 da Ian Goodfellow, all'epoca ricercatore presso l'Università di Montreal. Le GAN sono costituite da due reti neurali: un **generatore** e un **discriminatore**. La prima ha il compito di creare nuovi dati che possano ingannare il discriminatore, mentre la seconda ha la mansione di distinguere tra i dati creati dal generatore e quelli reali; funge da classificatore, ossia una rete da addestrare a distinguere i dati reali e quelli "fake" generati dal generatore, estraendone le caratteristiche. Il generatore parte da un dato completamente inventato, detto "latent vector", e lo passa al discriminatore che si addestra a riconoscere un determinato tipo di dato.

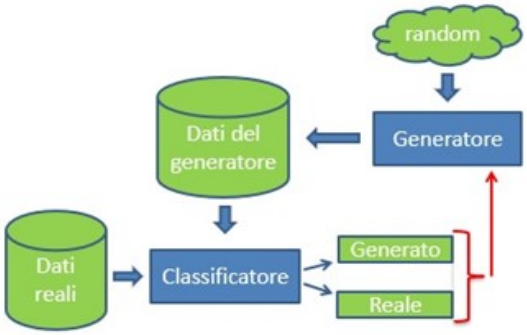

Figura 9: Struttura Generative Adversarial Network

5.5.2 Creazione di varianti uniche di malware

Gli attaccanti possono utilizzare le GAN per generare varianti di malware che sfuggono ai tradizionali sistemi di rilevamento basati su firme o su Machine Learning. Il processo prevede l'addestramento del generatore a creare malware modificato che il discriminatore, simulando un sistema di rilevamento, non riesce a identificare come malevolo. Questo ciclo continuo migliora la capacità del malware di passare inosservato.

L'integrazione di GAN nei RaaS, per esempio, porta a un boom degli attacchi, con minacce sempre più sofisticate e difficili da prevenire. Alcuni possibili scenari futuri prevedono anche attacchi mirati su larga scala completamente automatizzati con Ransomware adattato automaticamente a ciascun target, combinando il malware con strategie di ricatto e disinformazione.

5.5.3 Deepfake e phishing avanzato

Grazie alla loro capacità di generare contenuti realistici, hanno avuto un impatto significativo nelle tecniche di social engineering e phishing avanzato. Gli attaccanti sfruttano questa tecnologia per creare deepfake sofisticati per ingannare utenti e sistemi di sicurezza: il generatore crea contenuti falsi, il discriminatore valuta se i contenuti sembrano realistici e il sistema si allena in modo iterativo, migliorando continuamente la qualità dei contenuti fino a

renderli indistinguibili dalla realtà.
Le GANs possono essere usate anche per creare e-mail di phishing avanzate, rendendo gli attacchi di social engineer più difficili da scovare: analizzando i dati delle vittime, possono generare messaggi con un tono e un aspetto più realistico, creare pagine di login più realistiche e generare firme elettroniche aziendali più credibili.

5.5.4 GAN per il superamento di controlli di sicurezza

I CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) sono strumenti progettati per distinguere un utente umano da un bot o un programma automatizzato. Le GAN rappresentano una minaccia significativa in questo campo, poiché sono in grado di apprendere e superare le sfide progettate per bloccare i bot. Attraverso un processo iterativo, il generatore tenta di creare risposte corrette, mentre il discriminatore agisce come giudice per valutarne l'accuratezza, fino a creare output indistinguibili da quelli reali.
Superare i CAPTCHA è un passo fondamentale per gli attaccanti Ransomware che cercano di automatizzare e ampliare la portata delle loro operazioni, consentendo loro di creare account falsi per diffondere link malevoli, accedere a infrastrutture di comando e controllo, mascherare operazioni o creare botnet[20] per la diffusione del malware.

5.5.5 Integrazione delle GAN

L'integrazione delle GAN nei Ransomware non è immediata e presenta diverse sfide, legate principalmente a tre fattori:

1. **competenze informatiche**: richiede conoscenze avanzate in machine learning, reti neurali e cybersecurity offensiva. È necessario comprendere e sviluppare reti neurali generative per creare ransomware polimorfi,

[20]Botnet: insieme di dispositivi controllati da cybercriminali per attaccare un bersaglio.

avere competenze in attacchi adversariali e padroneggiare le tecniche di evasione. Queste capacità non sono alla portata di un hacker improvvisato, ma piuttosto di gruppi ben finanziati, come gli APT (Advanced Persistent Threats[21]);

2. **potenza di calcolo**: per addestrare una GAN efficace si ha la necessità di GPU[22] di fascia alta o infrastrutture cloud avanzate (es. Google Cloud), dataset di riferimento per apprendere schemi utili per evadere i sistemi di rilevamento e tempo, in quanto l'ottimizzazione di una GAN non è immediata. Una volta addestrata però, permette al ransomware di evolversi direttamente sui dispositivi infetti, senza bisogno di calcoli complessi in tempo reale;

3. **complessità di implementazione**: sviluppare un Ransomware che utilizzi le GAN in modo efficace non è semplice. Generare malware poliformi in modo che evadano i sistemi di detection AI per effettuare attacchi mirati, richiede test, adattamenti continui e una forte conoscenza della cybersecurity difensiva per essere efficaci.

[21] APT: attacco informatico sofisticato e sistematico che continua per un periodo di tempo prolungato

[22] GPU: componente hardware in grado di eseguire calcoli matematici ad alta velocità.

Capitolo 6
I limiti del Machine Learning

Nonostante il Machine Learning rappresenti uno strumento potente per affrontare e diffondere le minacce Ransomware, ci sono diversi limiti e sfide che ne riducono l'efficacia in questo ambito. Questi limiti sono legati sia alla natura delle tecnologie ML, sia all'evoluzione delle tecniche di attacco dei criminali informatici.

6.1 Evasione attraverso attacchi adversaliari

I modelli di ML possono essere facilmente ingannati da attacchi adversariali, che consistono nella creazione di input appositamente manipolati per indurre il modello a prendere decisioni sbagliate.

6.1.1 Come funziona un attacco adversaliare

Un attacco adversariale consiste nell'apportare piccole modifiche intenzionali ai dati di input, in modo che siano percepite dal modello come un comportamento legittimo. Un Ransomware può essere progettato per modificare leggermente

il suo comportamento per assomigliare a un processo legittimo, oppure un file binario malevolo può essere alterato in modo che le sue caratteristiche non corrispondano ai pattern identificati dal modello di ML come minacce. Queste modifiche sono spesso impercettibili agli esseri umani, ma ingannano i modelli di ML, facendo passare il Ransomware inosservato.

6.1.2 Tipologie di attacchi adversaliari

Gli attacchi adversariali nel contesto Ransomware possono essere classificati in base all'accesso che l'attaccante ha al sistema di ML:

1. **White-box attack**: l'attaccante ha pieno accesso al modello di ML, incluse le sue architetture e i parametri. Utilizzando queste informazioni, può progettare attacchi altamente specifici per ingannare il sistema. Un esempio di questo attacco è l'attacco C&W, usato per ingannare un antivirus o un sistema di rilevamento malware basato su ML che va a modificare un file binario in modo impercettibile affinché venga classificato come benigno;

2. **Black-box attack**: l'attaccante non ha accesso diretto al modello, ma può osservare le risposte del sistema (ad esempio, se una minaccia viene rilevata o meno), e con queste informazioni, può iterare su diversi input per trovare varianti di Ransomware che sfuggano al rilevamento. Un esempio di attacco di questo tipo è DeepLocker, un malware che utilizza l'AI per attivarsi solo in presenza di bersagli specifici, rendendo il rilevamento da parte degli antivirus quasi impossibile.

6.1.3 Difficoltà degli attacchi

L'efficacia e la realizzabilità di un attacco adversariale dipendono da diversi fattori, tra cui:

1. **competenze informatiche**: la realizzazione di un attacco adversariale sofisticato richiede una solida conoscenza di machine learning, reti neurali e tecniche di ottimizzazione. È fondamentale comprendere il funzionamento dei modelli di deep learning, il modo in cui elaborano i dati di input, come vengono appresi i pattern utilizzati per prendere decisioni e, inoltre, giocano un ruolo cruciale nella loro generazione, concetti come il calcolo del gradiente, l'ottimizzazione iterativa e le funzioni di perdita;

2. **potenza di calcolo**: la potenza di calcolo necessaria dipende dalla complessità del modello bersaglio e dalla tecnica utilizzata. Gli attacchi più semplici possono essere eseguiti con risorse hardware limitate, rendendoli accessibili anche a utenti con computer di fascia media, mentre attacchi più avanzati, come quelli basati sul calcolo del gradiente, possono richiedere una notevole potenza computazionale. Se il sistema target è protetto da difese avanzate, come il training adversariale o meccanismi di rilevamento delle anomalie, il costo computazionale dell'attacco può aumentare ulteriormente, ma, nonostante queste difficoltà, l'accesso a risorse cloud e GPU avanzate, hanno reso più facile per attori meno sofisticati condurre attacchi adversariali efficaci.

Gli attacchi più semplici, come quelli Black-Box, sono estremamente frequenti perché non richiedono conoscenze avanzate e possono essere eseguiti con un comune PC. Grazie a strumenti open-source come Foolbox[23], anche cybercriminali con competenze limitate possono generare input adversariali e testarne l'efficacia. Man mano che la complessità aumenta, diminuisce la frequenza di questi attacchi: quelli che sfruttano i gradienti del modello per creare input adversariali, richiedono una conoscenza più avanzata di machine learning e reti neurali e, sebbene questi attacchi possano essere eseguiti su GPU di fascia media, la loro implementazione richiede esperienza in deep learning e accesso a

[23]Foolbox: quantifica le perturbazioni avversarie utilizzando metriche di distanza (L0, L2..) per valutare la sottigliezza e l'efficacia degli attacchi.

risorse di calcolo più potenti. Gli attacchi più avanzati, spesso di tipo White-Box, sono rari perché richiedono una profonda conoscenza delle reti neurali, delle tecniche di ottimizzazione e una significativa potenza di calcolo: per eseguire tali operazioni, sono necessarie GPU di fascia alta o infrastrutture cloud avanzate, rendendo questi attacchi accessibili solo a governi, organizzazioni criminali ben finanziate o team di ricerca specializzati in AI adversariale.

6.2 Dipendenza dai dati di addestramento

Il successo di un sistema di ML dipende fortemente dalla qualità e dalla quantità dei dati usati per addestrarlo, mettendolo davanti a diverse sfide:

1. i dataset potrebbero non contenere abbastanza esempi di Ransomware nuovi o varianti sconosciute, portando a modelli meno efficaci;

2. i Ransomware si evolvono rapidamente, e i dati utilizzati per addestrare i modelli possono diventare obsoleti in breve tempo, riducendo l'efficacia del modello contro nuove minacce;

3. è difficile raccogliere campioni di Ransomware in modo sicuro e creare dataset rappresentativi senza rischi per l'ambiente di ricerca.

6.2.1 Il problema dell'overfitting

L'overfitting[14] è uno dei principali problemi in Machine Learning, in cui un modello diventa troppo specifico rispetto al dataset di addestramento. In pratica, il modello apprende non solo i pattern generali utili a fare previsioni, ma anche dettagli irrilevanti o rumore, riducendo la sua capacità di generalizzare su nuovi dati.

I Ransomware sono in costante evoluzione con nuove varianti che introducono tecniche innovative per evitare il rilevamento. Se un sistema di ML soffre di overfitting, sarà in grado di rilevare con precisione solo i Ransomware già presenti nel dataset di addestramento e fallirà nel riconoscere Ransomware polimorfici o generati automaticamente con caratteristiche diverse dal dataset di addestramento.

6.2.2 Sovraccarico di falsi positivi e falsi negativi

Quando si utilizzano sistemi basati su Machine Learning per rilevare Ransomware, il fenomeno dei falsi positivi e falsi negativi rappresenta una delle principali sfide operative e tecniche. Questi errori possono influenzare significativamente l'efficacia dei sistemi di sicurezza, causando conseguenze che vanno dall'intrusione di Ransomware alla perdita di fiducia nelle soluzioni di rilevamento.
Un falso positivo si verifica quando un sistema di rilevamento segnala un file, un processo o un'attività legittima come malevola, anche se non lo è. Un falso negativo, al contrario, si verifica quando il sistema non riesce a rilevare una minaccia effettiva, classificandola come innocua.
Nella progettazione di un sistema di rilevamento Ransomware, esiste un compromesso intrinseco tra la riduzione dei falsi positivi e quella dei falsi negativi. Questo compromesso si basa su due metriche fondamentali:

1. **sensibilità**, ossia la capacità del sistema di rilevare tutte le minacce, anche a costo di segnalare qualche falso positivo;

2. **precisione**, ossia la capacità di segnalare solo minacce reali, riducendo i falsi allarmi.

Aumentare la sensibilità spesso significa un aumento dei falsi positivi, perché il sistema diventa più "prudente" e segnala tutto ciò che somiglia ad una minac-

cia. Al contrario, un sistema troppo preciso rischia di ignorare alcune minacce per ridurre i falsi positivi, aumentando il rischio di falsi negativi.

6.3 Data Poisoning

Il data poisoning[15] è una forma di attacco informatico diretto contro i modelli di AI e ML che si verifica quando un attaccante introduce deliberatamente nei dati di addestramento informazioni false, ingannevoli o malevole con l'intento di alterare il processo addestrativo dell'algoritmo, andando a compromettere la validità delle sue risposte, dei suoi processi decisionali, della sua affidabilità, e causare di conseguenza errori nelle previsioni del modello, evadere sistemi di sicurezza, indurre bias[24] nelle decisioni dell'algoritmo e permettere agli aggressori di aggiungere una backdoor che consente di indurre i modelli a fare ciò che essi desiderano.

Ricerche sugli attacchi di data poisoning nei processi di Machine Learning hanno identificato due tipi di avvelenamento dei dati, definiti split-view poisoning e front-running poisoning. Nel caso dello **split-view poisoning** un aggressore potrebbe agire ottenendo il controllo di una risorsa web (come un dominio scaduto) indicizzata da un particolare set di dati, potendo quindi avvelenare i dati raccolti, rendendoli imprecisi, con il potenziale di influenzare negativamente l'intero algoritmo. Nel caso del **front-running poisoning**, invece, l'aggressore non ha il pieno controllo dello specifico set di dati, ma ha la possibilità di prevedere con precisione quando una risorsa sarà accessibile per essere inclusa nel set di dati, avvelenandolo appena prima che le informazioni vengano raccolte.

[24]Bias: distorsioni nelle valutazioni di fatti e avvenimenti

6.3.1 Come difendersi

Gli attacchi più comuni di Data Poisoning avvengono quando i dataset di addestramento sono pubblicamente accessibili andandoli ad avvelenare per indurre a classificare codice malevolo come sicuro (per esempio, la manipolazione di fonti come Wikipedia). Poiché possono essere modificati facilmente, gli attacchi richiedono competenze informatiche moderate e possono essere eseguiti senza un grande dispendio di risorse computazionali rendendoli di conseguenza piuttosto frequenti. Quando il dataset è parzialmente protetto, la frequenza diminuisce, rimanendo però un rischio concreto. Un esempio tipico è la manipolazione dei modelli di cybersecurity, che va ad alterare i dati di addestramento per ridurre la capacità di rilevare minacce emergenti. La frequenza di questi attacchi è più bassa rispetto ai precedenti, ma richiedono competenze avanzate in machine learning, hacking e accesso indiretto ai dati. Gli attacchi più rari, infine, colpiscono modelli privati e ben difesi, come quelli usati in sanità, difesa e infrastrutture critiche. In questi casi, alterare i dati richiede un accesso diretto ai server o la collaborazione di un insider. La loro complessità e le barriere di accesso ai dati li rendono rari, ma quando avvengono, possono avere conseguenze estremamente gravi.

L'intervento umano nei cicli operativi dell'AI assicura una supervisione continua sulla qualità dei dati e sulle decisioni prese dai modelli, consentendo un affinamento delle decisioni prese dai modelli, assicurando che riflettano valori, etica e considerazioni sociali. Includere l'umanità nel processo di apprendimento dell'AI permette ai modelli di beneficiare direttamente dall'esperienza e dall'intuizione umane che possono fornire feedback immediato al sistema, aiutandolo a adattarsi più efficacemente a nuovi scenari o a correggere percorsi decisionali errati. Questo controllo di qualità è indispensabile quando si tratta di prevenire o mitigare gli effetti del data poisoning, permettendo agli operatori umani di correggere i dati corrotti prima che influenzino negativamente il modello.

Capitolo 7
Prospettive future: Quantum Computing

Il quantum computing rappresenta una delle frontiere più promettenti e al contempo più inquietanti della tecnologia moderna. Con la sua capacità di elaborare informazioni a velocità esponenzialmente superiori rispetto ai computer tradizionali, il quantum computing potrebbe rivoluzionare la cybersecurity, offrendo nuove opportunità ma anche creando nuove minacce.

7.1 Panoramica sul Quantum Computing

L'informatica quantistica rappresenta un salto significativo rispetto all'informatica classica. A differenza dei computer classici, che elaborano i dati in bit binari (0 e 1), i computer quantistici utilizzano bit quantistici, o qubit. Questi qubit sfruttano i principi della meccanica quantistica consentendo loro di eseguire calcoli complessi a velocità irraggiungibili dalle loro controparti classiche, in particolare la sovrapposizione e l'entanglement.

1. **sovrapposizione**: mentre un bit classico può essere solo 0 o 1, un qubit può trovarsi in più stati contemporaneamente. Questo permette ai computer quantistici di esplorare simultaneamente molteplici soluzioni a un problema, aumentando enormemente la velocità di calcolo.

2. **entanglement**: è fenomeno per cui due qubit, anche se distanti tra loro, diventano correlati in modo istantaneo. Ciò significa che manipolare un qubit influisce immediatamente sull'altro, indipendentemente dalla distanza, permettendo ai computer quantistici di eseguire operazioni con un livello di coordinazione e parallelismo che sarebbe impossibile nei sistemi classici.

Grazie a queste proprietà, i computer quantistici non sono semplicemente più veloci, ma affrontano i problemi in modo completamente, diverso valutando simultaneamente tutte le possibili soluzioni, riducendo drasticamente i tempi di calcolo per problemi complessi[16].

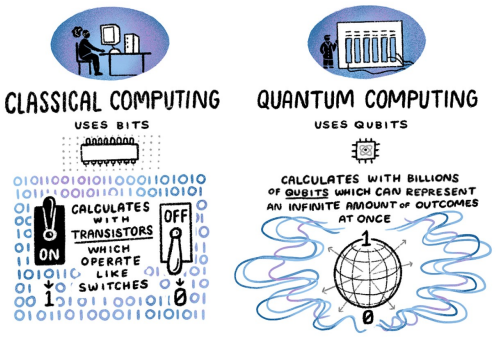

Figura 10: Classical vs Quantum Computing

7.2 Opportunità nella cybersicurezza

Il quantum computing offre numerose opportunità per migliorare la cybersecurity, rivoluzionando vari aspetti della protezione dei dati e della sicurezza delle reti.

1. **Crittografia avanzata**: la crittografia quantistica offre un livello di sicurezza teoricamente inviolabile grazie alla legge della meccanica quantistica: qualsiasi tentativo di intercettazione di una chiave quantistica altera lo stato dei qubit, rendendo immediatamente rilevabile l'intrusione.

2. **Rilevamento delle minacce**: grazie alla sua potenza di calcolo, il quantum computing può analizzare rapidamente grandi quantità di dati per identificare modelli e anomalie, migliorando il rilevamento delle minacce in tempo reale, permettendo di identificare e neutralizzare minacce avanzate molto più velocemente di quanto sia possibile oggi.

3. **Ottimizzazione della sicurezza**: gli algoritmi quantistici possono ottimizzare le strategie di sicurezza per massimizzare l'efficacia delle difese: potrebbero, per esempio, essere utilizzati per simulare e prevedere i percorsi di attacco più probabili, consentendo di rafforzare le difese nei punti più vulnerabili.

4. **Generazione di numeri casuali**: questo aspetto è fondamentale per la crittografia e la sicurezza informatica, in quanto i computer quantistici possono generare numeri veramente casuali (cosa che i computer classici non possono fare con la stessa precisione) rendendo le chiavi crittografiche generate molto più sicure.

7.3 Minaccia per la cybersicurezza

Nonostante le opportunità, il quantum computing presenta anche significative minacce per la cybersecurity:

1. **decifratura delle chiavi crittografiche**: una delle minacce più gravi è la possibilità che i computer quantistici possano decifrare rapidamente le chiavi crittografiche utilizzate oggi, rendendo obsoleti molti dei sistemi di

sicurezza attuali. Gli algoritmi come RSA che attualmente proteggono la maggior parte delle comunicazioni sicure su Internet, potrebbero essere facilmente infranti da un computer quantistico sufficientemente potente;

2. **attacchi più veloci e sofisticati**: gli hacker potrebbero utilizzare il quantum computing per sviluppare attacchi più veloci e sofisticati, capaci di superare le difese tradizionali, come attacchi di brute force accelerati o l'uso di algoritmi quantistici per trovare vulnerabilità nel codice di sicurezza esistente;

3. **necessità di nuovi standard di sicurezza**: con l'avvento del quantum computing, è necessario sviluppare nuovi standard e protocolli di sicurezza per proteggere le informazioni sensibili.

7.4 Un'arma contro il Ransomware

Il quantum computing può anche diventare un prezioso alleato nella lotta contro il Ransomware. Per contrastare i rischi posti da questa tecnologia emergente, esperti di cybersecurity e istituzioni stanno sviluppando nuove soluzioni avanzate.

7.4.1 Crittografia Post-Quantistica

Organizzazioni come il NIST[25], stanno lavorando su nuovi algoritmi crittografici resistenti ai computer quantistici, basati su problemi matematici che neanche un quantum computer potrebbe risolvere facilmente e, quando questi standard saranno adottati, le difese contro gli attacchi Ransomware quantistici diventeranno più robuste.

[25]NIST: è un'agenzia che promuove l'innovazione attraverso il progresso della scienza, degli standard e della tecnologia delle misurazioni.

7.4.2 Quantum Key Distribution

La distribuzione quantistica delle chiavi (QKD) sfrutta le leggi della meccanica quantistica per creare chiavi crittografiche impossibili da intercettare senza essere rilevati. Se implementata su larga scala, questa tecnologia potrebbe garantire comunicazioni ultra-sicure, rendendo inefficaci i tentativi di intrusione da parte degli hacker.

7.4.3 Quantum Machine Learning per il rilevamento delle minacce

I computer quantistici potrebbero essere usati per analizzare enormi quantità di dati e individuare pattern anomali, migliorando i sistemi di rilevamento delle minacce e permettendo di bloccare i Ransomware prima che possano causare danni.

7.5 Perché non si utilizza ancora?

Questa tecnologia ha il potenziale di rivoluzionare completamente il modo di calcolare e risolvere problemi, ma ci sono ancora diverse sfide che devono essere affrontate prima che possa essere utilizzata su larga scala. In particolare, ci sono tre aspetti principali che stanno frenando l'utilizzo del quantum computing, che sono la complessità tecnologica, la mancanza di algoritmi pratici e software compatibili, e i costi elevatissimi che questa tecnologia comporta.

7.5.1 Complessità tecnologica

Il quantum computing non è semplicemente una versione più potente dei computer tradizionali, ma si basa su un paradigma completamente diverso. I qubit, sono estremamente sensibili e facilmente influenzabili da qualsiasi perturbazione esterna, e quindi mantenere la loro stabilità richiede condizioni di funzionamento estremamente rigorose, come temperature vicine allo zero assoluto. Inoltre, per funzionare correttamente, i qubit devono essere protetti

da interferenze esterne come il rumore e le vibrazioni, e questo significa che i computer quantistici devono essere ospitati in ambienti altamente controllati e protetti da interferenze ambientali. Questi ostacoli tecnologici rendono il quantum computing ancora una sfida tecnica da superare, limitando la sua applicazione pratica al momento.

7.5.2 Mancanza di algoritmi e software pronti all'uso

Un altro punto cruciale riguarda la mancanza di algoritmi e software compatibili. Per usare davvero un computer quantistico, non basta costruirlo, ma bisogna anche sviluppare software in grado di sfruttarne appieno le potenzialità. Mentre nel mondo dell'informatica classica esistono migliaia di applicazioni pronte all'uso, nel mondo del quantum computing siamo ancora lontani da una situazione simile: la programmazione quantistica richiede un approccio completamente nuovo, basato su modelli matematici che non hanno alcuna corrispondenza diretta con quelli che usiamo oggi.
Sebbene siano stati fatti dei progressi, il quantum advantage, cioè il punto in cui un computer quantistico supera un supercomputer tradizionale in un compito pratico, è ancora un obiettivo da raggiungere. Fino a quando non si disporrà di algoritmi adatti per risolvere problemi concreti, il quantum computing rimane una tecnologia più teorica che applicabile a livello quotidiano.

7.5.3 Costi elevatissimi

La costruzione e la manutenzione di un computer quantistico richiedono infrastrutture estremamente sofisticate e costose e, oltre ai costi per i materiali altamente specializzati necessari per costruire i qubit, bisogna considerare le tecnologie di raffreddamento criogenico, che sono fondamentali per mantenere il sistema a temperature estremamente basse. I laboratori che ospitano questi computer richiedono ambienti altamente protetti per evitare interferenze esterne, il che comporta una spesa significativa. Non è solo una questione

di hardware: la formazione di esperti in quantum computing è un altro costo, i professionisti in questo campo sono ancora pochi e richiedono anni di formazione.

Conclusioni

Il Ransomware continua a rappresentare una delle minacce informatiche più pericolose e in continua evoluzione. Nel corso di questa analisi, abbiamo esaminato la nascita e lo sviluppo di questo malware, le principali strategie di difesa e l'impatto delle nuove tecnologie, in particolare dell'Intelligenza Artificiale e del Machine Learning, sulla sicurezza informatica.

Se da un lato le tecniche di difesa basate sul ML consentono di rilevare e prevenire attacchi in modo più efficace, dall'altro lato i criminali informatici sfruttano la stessa tecnologia per rendere i loro attacchi più sofisticati e mirati. Questo scenario crea un continuo equilibrio tra attaccanti e difensori, in cui ogni innovazione nel campo della sicurezza viene rapidamente seguita da nuove tattiche di attacco.

Abbiamo analizzato come il ML possa essere utilizzato come strumento di difesa, ad esempio attraverso algoritmi capaci di identificare anomalie e prevenire intrusioni prima che si verifichino danni significativi. Tuttavia, il ML stesso presenta limiti e vulnerabilità, tra cui l'esposizione a tecniche di evasione come gli attacchi adversariali, che manipolano i modelli per eludere il rilevamento, la dipendenza del dataset di apprendimento, oppure l'avvelenamento dei dati nell'algoritmo di addestramento del modello.

Alla luce di queste considerazioni, appare evidente che la lotta contro il Ransomware non possa essere affrontata con un'unica soluzione definitiva: è ne-

cessario un approccio combinato e adattivo, che includa tecnologie avanzate basate su AI e ML per il rilevamento e la risposta agli attacchi, strategie di difesa proattive (segmentazione della rete, backup continuo ed educazione degli utenti) e collaborazione tra enti di cybersicurezza.

Bibliografia

[1] CoreTech. *Backup: Completo, Differenziale, Incrementale*. 2018. URL: https://1backup.me/it/why/backup-cloud-completo-differenziale-incrementale.php (visitato il giorno 29/01/2025).

[2] Microsoft. *Servizio Copia Shadow del volume (VSS)*. 2024. URL: https://learn.microsoft.com/it-it/windows-server/storage/file-server/volume-shadow-copy-service#how-vss-works (visitato il giorno 09/02/2025).

[3] Tanishq Mohite. *Che cosa è la crittografia completa del disco? Importanza e best practice*. 2024. URL: https://blog.scalefusion.com/it/full-disk-encryption/ (visitato il giorno 14/01/2025).

[4] Ilaria Della Queva. *BitLocker: funzionalità e vantaggi*. 2023. URL: https://cyberment.it/sicurezza-informatica/bitlocker-funzionalita-e-vantaggi/ (visitato il giorno 16/01/2025).

[5] Recovery Data. *Filevault: cos'è e come funziona la tecnologia di cifratura MAC*. 2019. URL: https://www.recovery-data.it/filevault/#Come_funziona_FileVault (visitato il giorno 16/01/2025).

[6] Simone Piccardi. *Gestire filesystem cifrati usando LUKS*. 2024. URL: https://www.truelite.it/gestire-filesystem-cifrati-usando-luks/ (visitato il giorno 16/01/2025).

[7] Acronis. *Il ruolo dell'intelligenza artificiale e del machine learning nella protezione dal ransomware*. 2024. URL: https://www.acronis.com/it-it/blog/posts/role-of-ai-and-ml-in-ransomware-protection/ (visitato il giorno 20/01/2025).

[8] Acronis. *Che cos'è l'Endpoint Detection and Response (EDR)?* 2024. URL: https://www.acronis.com/it-it/blog/posts/what-is-endpoint-detection-and-response-edr/ (visitato il giorno 22/01/2025).

[9] Tred Micro. *Cos'è XDR?* 2024. URL: https://www.trendmicro.com/it_it/what-is/xdr.html (visitato il giorno 22/01/2025).

[10] Microsoft. *Cos'è MDR?* 2024. URL: https://www.microsoft.com/it-it/security/business/security-101/what-is-mdr-managed-detection-response (visitato il giorno 22/01/2025).

[11] Microsoft. *Cos'è la prevenzione della perdita dei dati (DLP)?* 2024. URL: https://www.microsoft.com/it-it/security/business/security-101/what-is-data-loss-prevention-dlp (visitato il giorno 03/02/2025).

[12] Federica Maria Rita Livelli. *Data Loss Prevention: le normative, le best practice, le soluzioni.* 2025. URL: https://www.zerounoweb.it/techtarget/searchsecurity/data-loss-prevention-le-normative-le-best-practice-le-soluzioni/ (visitato il giorno 03/02/2025).

[13] Nanni Bassetti. *GAN (Generative Adversarial Networks): cosa sono, applicazioni e vantaggi.* 2023. URL: https://www.agendadigitale.eu/cultura-digitale/gan-generative-adversarial-networks-cosa-sono-applicazioni-e-vantaggi/ (visitato il giorno 24/01/2025).

[14] IBM. *Che cos'è l'overfitting?* 2023. URL: https://www.ibm.com/it-it/topics/overfitting (visitato il giorno 26/01/2025).

[15] Luisa Di Giacomo. *Intelligenza artificiale e data poisoning: come "avvelenare" i dati.* 2024. URL: https://www.diritto.it/intelligenza-artificiale-data-poisoning-avvelenare/#block-000baf0b-4c92-46ff-8319-eb5dadf091b5 (visitato il giorno 30/01/2025).

[16] Onova. *Quantum Computing e Cyber Security: Opportunità e Minacce.* 2024. URL: https://www.onova.it/news/quantum-computing-e-cyber-security-opportunita-e-minacce/ (visitato il giorno 13/02/2025).

Ringraziamenti

Un grazie di cuore a tutti.
Grazie alla mia famiglia per avermi accompagnato in questo percorso, per avermi appoggiato e spronato a dare sempre il massimo. Nonostante la distanza, ho sempre potuto fare affidamento su di voi per qualsiasi cosa e grazie ai vostri consigli, ho potuto affrontare questa avventura con una marcia in più.
Grazie al mio relatore, il Prof. Davide Sangiorgi, per avermi affiancato nella stesura della tesi.
Grazie a tutti coloro che hanno fatto parte di questo capitolo della mia vita, sia nei momenti migliori che nei momenti peggiori. A quelle persone che ci sono state sin dall'inizio e a quelle conosciute in corso d'opera.. Avete reso unico e indimenticabile questo viaggio.
Infine, un grazie speciale a chi ha saputo darmi le sveglie nei momenti di blackout, sopportandomi nei momenti più bui e più complicati, risollevandomi, e facendomi capire che ce l'avrei fatta a superare qualsiasi ostacolo.
Se oggi sono arrivato qui, è merito vostro.. E ve ne sarò per sempre grato.