Homework per l'esame di SICUREZZA, corso di Laurea triennale in Informatica

Davide Belcastro, 1962536

June 19, 2023

1 Quesiti

Le tecniche di Machine Learning (ML) e Intelligenza Artificiale (AI) sono oggi adottate nell'ambito dell'Intrusion Detection, ad esempio con lo scopo di aumentare la capacità degli IDS di identificare intrusioni non note.
Si chiede di analizzare la letteratura scientifica e tecnica relativa a soluzioni di intrusion detection al fine di dare risposta alla seguenti domande

- Quali sono le principali tecniche di ML, deep learning, ed AI in generale, che sono state proposte per risolvere il problema dell'intrusion detection? (dare anche una descrizione del loro funzionamento)
- Quali tipologie di intrusioni è possibile identificare con tali tecniche, e quali tipi di intrusioni invece non possono essere identificate?
- Le soluzioni proposte sono di carattere generale, ovvero permettono di identificare più tipi di attacchi, oppure per ogni tipo di attacco va creata una soluzione dedicata?
- Quali sono i dataset a disposizione per sperimentare (training e testing) le tecniche di intrusion detection precedentemente identificate, e quali sono le principali caratteristiche di questi dataset?

Le pubblicazioni considerate devono essere state pubblicate su riviste o conferenze IEEE, ACM, Elsevier, Springer; l'analisi si deve basare su almeno 10 pubblicazioni.

2 Risposta 1

2.1 Support Vector Machine

Una tecnica usata per l'intrusion detection sono le Support Vector Machine (SVM) che sono comunemente utilizzate come algoritmo di classificazione per la distinzione tra traffico di rete normale e anomalo.
Per fare ciò, le SVM vengono addestrate a riconoscere i modelli del traffico di rete normale e ad identificare eventuali comportamenti anomali. Questo processo di addestramento richiede tipicamente l'utilizzo di grandi quantità di dati di traffico di rete pre-classificati, che consentono all'algoritmo di apprendere i modelli di traffico normale.
Una volta addestrate, le SVM possono essere utilizzate per rilevare eventuali anomalie nel traffico in tempo reale, valutando se un determinato flusso di pacchetti di rete è anomalo rispetto al modello di traffico normale precedentemente appreso.
Distinguiamo le SVM di classe uno con le SVM di classe due.
In particolare, le SVM di classe 1 sono progettate per risolvere problemi di classificazione in cui solo una delle due classi è presente nei dati di addestramento. Questo significa che l'obiettivo delle SVM di classe 1 è quello di trovare un iperpiano che separi i dati di addestramento della classe presente in modo tale che sia possibile classificare correttamente i nuovi dati che appartengono a questa classe.
D'altra parte, le SVM di classe 2 sono utilizzate per risolvere problemi di classificazione binaria in cui sono presenti entrambe le classi nei dati di addestramento. In questo caso, l'obiettivo delle SVM di classe 2 è quello di trovare un iperpiano che separi le due classi in modo da riconoscere se un dato è in una delle due classi.
In altre parole, le SVM di classe 1 cercano di trovare un iperpiano che separa una classe dagli altri dati, mentre le

SVM di classe 2 cercano di trovare un iperpiano che separa le due classi tra di loro.
Le SVM annidate (o Nested SVM) sono una tecnica avanzata di apprendimento automatico che utilizza SVM di classe 1 multiple in sequenza per migliorare la precisione della classificazione.
In pratica, le SVM annidate sono una serie di SVM di classe uno, dove i risultati di una SVM vengono utilizzati come input per la successiva SVM nella catena.
L'utilizzo di SVM annidate può contribuire a migliorare la precisione della classificazione rispetto all'utilizzo di una singola SVM. Questo perché l'uso di più SVM in sequenza consente di affrontare problemi più complessi e di scomporre la classificazione dei dati in una serie di decisioni più semplici.

L'articolo "Nested One-Class Support Vector Machines for Network Intrusion Detection" del 2018 pubblicato nella rivista scientifica "IEEE Transactions on Information Forensics and Security", si concentra sull'utilizzo di SVM di classe uno annidate per migliorare la capacità di rilevamento dei sistemi di sicurezza di rete.
L'articolo sottolinea che l'utilizzo delle SVM annidate può migliorare la capacità del sistema di rilevamento delle intrusioni di rete rispetto all'utilizzo delle SVM standard. Gli autori mostrano che l'utilizzo di SVM annidate offre un miglioramento significativo nella capacità di individuare correttamente gli attacchi informatici, riducendo al contempo il tasso di falsi positivi.
Le SVM possono essere utilizzate in combinazione cone altre tecniche come ad esempio con K-Mean Clustering.

L'articolo "A K-Means clustering and SVM based hybrid concept drift detection technique for network anomaly detection" del 2022, pubblicato nella rivista Elsevier Expert Systems with Applications, si concentra sull'elaborazione dei dati per la rilevazione delle anomalie nella rete utilizzando una tecnica ibrida che incorpora K-Means Clustering e SVM.
Nell'articolo viene proposto un sistema di rilevamento delle anomalie che è in grado di rilevare i cambiamenti nel comportamento del traffico sulla rete, utilizzando una tecnica ibrida basata su K-Means Clustering e SVM. In particolare, l'obiettivo di questo sistema è identificare i cambiamenti che possono rappresentare una minaccia per la sicurezza, rilevando gli attacchi della rete.
L'algortimo K-Mean Clustering viene utilizzato per suddividere un gruppo di dati in un certo numero di cluster, in modo tale che i dati all'interno di ogni cluster siano più simili tra loro rispetto ai dati presenti in altri cluster. In questo caso, l'algoritmo K-Means viene utilizzato per creare cluster di dati in modo da identificare pattern e correlazioni tra di essi.
La tecnica proposta nell'articolo utilizza K-Means per suddividere i dati in cluster, al fine di identificare i pattern di comportamento della rete. Successivamente, SVM viene utilizzato per classificare il comportamento di ogni cluster, in modo da identificare anomalie all'interno di ogni cluster.
In questo modo, SVM si integra con l'output dell'algoritmo di clustering K-Means nella fase di classificazione, combinando le informazioni dei pattern del cluster e delle anomalie per avere un rilevamento più accurato delle anomalie nella rete.
L'efficacia del sistema proposto viene dimostrata sperimentalmente utilizzando dati di traffico reale. In particolare, i risultati ottenuti confermano che questa tecnica ibrida con K-Means Clustering e SVM migliora l'accuratezza della rilevazione delle anomalie rispetto alle tecniche utilizzate singolarmente.

2.2 Random Forest

Random Forest è un algoritmo di apprendimento automatico utilizzato anche per l'intrusion detection nelle reti informatiche. Esso può essere utilizzato per effettuare la classificazione dei dati, come ad esempio quelli relativi al traffico di rete, e quindi identificare eventuali intrusioni o comportamenti anomali.
In particolare, Random Forest è una tecnica di classificazione supervisionata che si basa sull'aggregazione di più alberi decisionali. Esso combina la predizione di più alberi decisionali, ognuno dei quali esegue una classificazione sui dati di input sulla base di una serie di regole di decisione. La combinazione dei risultati di più alberi decisionali permette di ottenere una classificazione più accurata dei dati.
Per quanto riguarda l'intrusion detection, Random Forest può essere utilizzato per la classificazione dei dati di traffico di rete. Gli alberi decisionali all'interno dell'insieme di classificatori possono individuare le caratteristiche degli eventi di rete associati ad attacchi o ai comportamenti anomali.
Una volta che l'insieme di classificatori è addestrato, Random Forest può essere utilizzato per analizzare in tempo reale il traffico di rete e confrontare i dati di ingresso con le caratteristiche "imparate" e quindi classificare l'evento

in una categoria specifica.

Primo Articolo: "An Efficient Network Intrusion Detection Model based on Random Forest", pubblicato nel 2016 sulla rivista Elsevier Procedia Computer Science.
L'approccio proposto nell'articolo utilizza l'algoritmo Random Forest che viene utilizzato per l'analisi dei dati di traffico di rete e la classificazione degli eventi in modo da identificare eventuali attacchi o comportamenti anomali. In particolare, vengono analizzate le feature di traffico di rete come la durata della connessione, la quantità di byte scambiati, il protocollo utilizzato, l'entropia dei dati, ecc.

Secondo Articolo: "An improved binary manta ray foraging optimization algorithm based feature selection and random forest classifier for network intrusion detection" pubblicato nel 2022, nella rivista Elsevier Journal of Ambient Intelligence and Humanized Computing, si concentra sull'applicazione dell'algoritmo di selezione delle caratteristiche basato sulla tecnica di ottimizzazione di tipo binary manta ray foraging optimization (BMFO) e sulla tecnica di classificazione Random Forest per la rilevazione delle intrusioni nella rete.
La tecnica di selezione delle caratteristiche basata su BMFO migliora la capacità del modello di rilevare le intrusioni,in quanto è finalizzata alla selezione delle feature più importanti per la classificazione dei dati di traffico di rete.
Dopo aver selezionato le feature più rilevanti attraverso l'algoritmo di selezione delle feature, l'algoritmo Random Forest è utilizzato per la creazione del modello di classificazione e la classificazione dei dati di traffico di rete. Il modello proposto viene sperimentato utilizzando il dataset NSL-KDD per la rilevazione delle intrusioni. I risultati sperimentali dimostrano che il modello proposto supera gli altri modelli di selezione delle caratteristiche basati su altre tecniche di apprendimento automatico in termini di accuratezza della rilevazione delle intrusioni.
I risultati ottenuti dimostrano che l'implementazione dell'algoritmo Random Forest in combinazione con la selezione delle feature basata sull'algoritmo "binary manta ray foraging optimization" produce prestazioni migliori in termini di accuratezza nella rilevazione di attacchi di rete.

2.3 Convolutional Neural Network

In generale, le CNN sono delle reti neurali che vengono utilizzate per l'analisi di immagini(soprattutto in ambito medico).
Recentemente però le CNN hanno attirato l'attenzione nella rilevazione delle intrusioni, in quanto questo tipo di reti sono in grado di analizzare i dati di traffico di rete e identificare eventuali pattern anomali che possano indicare un attacco.

L'articolo "Network Intrusion detection approach based on convolutional neural network" del 2022 presente nel dataset IEEE descrive più nel dettaglio il loro funzionamento.
Secondo quanto scritto nell'articolo,il modello di CNN è stato configurato per l'analisi dei dati di traffico di rete, ovvero la scansione dei pacchetti di dati per identificare eventuali anomalie o attacchi.
In particolare, gli autori utilizzano un dataset di traffico di rete, denominato NSL-KDD, per l'addestramento della CNN. Dopo aver addestrato la CNN attraverso questo dataset, il modello è stato in grado di classificare in modo corretto i pacchetti di dati di traffico di rete come "normale" o "anomalo".
I risultati ottenuti dimostrano che la CNN utilizzata come classificatore per la rilevazione delle intrusioni di rete è in grado di ottenere una buona accuratezza nella rilevazione degli attacchi di rete, superando in alcuni casi altri metodi di rilevazione delle intrusioni come la rilevazione basata sulla firma e la rilevazione basata sul comportamento.

2.4 Elliptic Envelope

Il metodo Elliptic Envelope è un algoritmo di apprendimento automatico utilizzato nella rilevazione delle intrusioni informatiche, che si basa sulla definizione di una "envelope" (o involucro), uno spazio di dati multidimensionale per isolare eventuali anomalie presenti nel dataset.
Nella rilevazione delle intrusioni, l'algoritmo Elliptic Envelope può essere utilizzato per analizzare i dati di traffico di rete e identificare eventuali attività anomale che possano indicare un attacco non autorizzato. L'algoritmo costruisce una "envelope" ideale dei dati, che descrive un'ipotesi di normalità sulla distribuzione dei dati sulla base di quelli che sono considerati come dati normali. In questo modo, gli elementi che si trovano al di fuori dell'involucro possono

essere considerati come potenziali anomalie.
L'approccio di Elliptic Envelope può essere utile nella rilevazione di attacchi zero-day e altre attività anomale, poiché è in grado di rilevare questo tipo di attività senza la necessità di conoscenze predefinite degli attacchi. Tuttavia, l'efficacia dell'algoritmo dipende dalla qualità dei dati di input e dalla corretta definizione della "envelope" durante la fase di training.
In sintesi, l'algoritmo di Elliptic Envelope è un metodo di rilevazione delle intrusioni che utilizza un approccio non parametrico basato su un modello probabilistico per individuare eventuali attività anomale nei dati di traffico di rete.

2.5 Naive Bayes

Il classificatore di Naive Bayes è un algoritmo di apprendimento automatico basato sulla conoscenza probabilistica delle caratteristiche del problema. Nella rilevazione delle intrusioni, l'algoritmo Naive Bayes può essere utilizzato per addestrare un classificatore che prende in input diverse feature di rete per identificare le attività anomale dei traffici di rete.

L'articolo "A New Two-Phase Intrusion Detection System with Naïve Bayes Machine Learning for Data Classification and Elliptic Envelop Method for Anomaly Detection" del 2021 pubblicato nel dataset Elsevier, si concentra sull'utilizzo dell'algoritmo Naive Bayes e sul metodo Elliptic Envelope nella rilevazione delle intrusioni informatiche.
L'articolo propone un sistema di rilevamento delle intrusioni a due fasi, dove la prima fase usa l'algoritmo di Naive Bayes per classificare i dati di traffico di rete (traffico normale o anomalo). In caso di attività sospette, la seconda fase scansiona il traffico con il metodo Elliptic Envelope per individuare eventuali anomalie.
In particolare, il modello Naive Bayes viene allenato su un dataset di traffico di rete contenente sia dati di traffico di rete "normale" che dati di traffico di rete "anomalo". Una volta che il classificatore viene allenato, il modello può essere utilizzato per l'analisi del traffico in tempo reale.
Nella fase di rilevamento degli attacchi, il metodo dell'Elliptic Envelope viene usato per identificare eventuali campioni anomali all'interno del dataset classificato.
L'articolo presenta anche i risultati degli esperimenti che dimostrano come il sistema proposto abbia un'elevata efficienza per la rilevazione di tutte le categorie di attacchi, migliorando la precisione e riducendo il tasso di falsi positivi.

2.6 VirusTotal

Fin'ora sono stati visti algoritmi che utilizzano tecniche di AI,ML e DL per rilevare intrusioni analizzando, principalmente, il traffico in rete.
Esistono delle tecniche che non analizzano il traffico in rete ma file presenti sul computer.
La rilevazione di intrusione basata sui file sul computer può essere relativamente più complessa rispetto alla rilevazione delle intrusioni basata sul traffico di rete,in quanto i malware all'interno dei file possono usare tecniche molto sofisticate per nascondersi,esistono infatti malware mefamorfici che, ad ogni copia all'interno del computer infetto,cambiano completamente il loro aspetto e il loro comportamento,in questo modo risulta difficile trovare numerosi file con pattern uguali che possono considerarsi malware.
VirusTotal è un servizio online gratuito che consente di analizzare file, URL e indirizzi IP per individuare la presenza di malware o altre minacce informatiche. Il servizio utilizza diverse tecniche di analisi, tra cui la Signature Detection e l'Anomaly Detection, per rilevare le minacce.
Il funzionamento di VirusTotal è abbastanza semplice. Gli utenti possono caricare un file o inserire un URL o un indirizzo IP da analizzare. Il servizio esegue quindi l'analisi del file o dell'URL utilizzando diversi motori di scansione antivirus e altre tecnologie di sicurezza, restituendo un rapporto dettagliato sulla presenza di eventuali minacce.
Per quanto riguarda la Signature Detection, VirusTotal utilizza un'ampia gamma di motori antivirus per rilevare la presenza di malware all'interno del file. Ogni motore antivirus utilizza una specifica definizione di regole per individuare le minacce informatiche, in modo da poter coprire il maggior numero possibile di minacce.
Per quanto riguarda l'Anomaly Detection, VirusTotal utilizza diverse tecniche di machine learning e intelligenza artificiale per individuare eventuali comportamenti anomali nel file o nell'URL analizzato. Queste tecniche si basano sulla creazione di modelli di comportamento normale, in modo da poter individuare eventuali variazioni che potreb-

bero indicare la presenza di una minaccia.

L'articolo "VirusTotal: A Cloud-based Multi-engine Malware Detection and Analysis Service" di Hisao Kameda, pubblicato su IEEE nel 2018, descrive il funzionamento e le caratteristiche di VirusTotal.
L'articolo descrive in dettaglio l'architettura di VirusTotal,inoltre discute come VirusTotal utilizza una serie di motori di scansione di malware per eseguire l'analisi di un file, combinando l'output di questi motori per fornire una valutazione complessiva del rischio di un file.
L'articolo conclude discutendo alcune delle limitazioni di VirusTotal, come il rischio che utenti malitenzionati possano entrare nel sistima e/o la possibilità di falsificare i risultati delle analisi inviando file modificati.
Nell'articolo, viene spiegato che uno dei possibili modi per falsificare i risultati delle analisi su VirusTotal è quello di modificare il file che si vuole analizzare, ad esempio tramite l'uso di un packer che maschera il vero contenuto del file. In questo modo, il risultato dell'analisi potrebbe essere diverso da quello che dovrebbe essere in realtà,ciò può comportare la non rilevazione di malware in un file in cui il malware in relatà è presente.
Il pericolo di ciò è che VirusTotal è utilizzato da molte organizzazioni e professionisti della sicurezza informatica come fonte di informazioni sulla presenza di malware. Se gli utenti falsificano i risultati delle analisi, questo potrebbe portare a conclusioni errate e ad azioni di sicurezza inadeguate o addirittura dannose. Inoltre, l'assenza di controlli di autenticità degli utenti potrebbe consentire a utenti malintenzionati di manipolare i risultati delle analisi per propagare il malware.

3 Risposta 2

3.1 SVM

SVM può essere utilizzato per rilevare vari tipi di intrusioni nella rete, come ad esempio

- Port Scanning: SVM analizza i dati di traffico di rete in cerca di un elevato numero di tentativi di connessione alle porte del sistema, che sono tipici di un attacco Scanning. Se il modello di traffico anomalo viene rilevato, l'SVM può classificare l'evento come un attacco di tipo Scanning.
- Attacchi DDoS: SVM può rilevare i pacchetti di dati malformati e l'elevato traffico in entrata sul sistema, che possono essere indicativi di un attacco DoS. Inoltre, l'SVM può anche analizzare le metriche di performance del sistema, come l'utilizzo della CPU e la latenza, per rilevare un aumento significativo rispetto ai valori normali.
- SQL injection: SVM può rilevare il modello di traffico anomalo generato dall'inserimento di query SQL malevole in una richiesta web. Ad esempio, l'SVM può cercare i caratteri di escape come il segno di apertura parentesi "(" o il simbolo ";" seguito da una query SQL, che sono tipici di un attacco di SQL injection.

Tuttavia, SVM potrebbe avere difficoltà nella rilevazione di alcune tipologie di intrusioni, come ad esempio

- Attacchi di tipo zero-day: Gli attacchi zero-day sono una forma di attacco informatico in cui l'attaccante sfrutta una vulnerabilità del sistema sconosciuta, prima che il produttore del software o il team di sicurezza siano in grado di identificare e correggere la vulnerabilità.
  L'SVM ha difficoltà con gli attacchi di tipo zero-day perché questi attacchi sfruttano vulnerabilità sconosciute nei sistemi informatici, per le quali non esiste ancora un patch di sicurezza disponibile. Di conseguenza, l'SVM non ha mai visto prima questi tipi di dati anomali e non ha una conoscenza sufficiente per identificare in modo affidabile gli attacchi zero-day.
- Attacchi che utilizzano tecniche di mascheramento: SVM ha difficoltà con le tecniche che utilizzano mascheramento perché questi attacchi sono progettati per nascondere il loro vero scopo e mascherarsi come traffico di rete normale. Poiché l'SVM è basato su modelli di apprendimento supervisionato che utilizzano un insieme di dati di addestramento per identificare i pattern di traffico anomalo, può avere difficoltà a rilevare gli attacchi che utilizzano tecniche di mascheramento.

Un articolo del 2017 che analizza i tipi di attacco che può rilevare SVM è il seguente: "An investigation of SVM and ANN for intrusion detection system. Journal of Network and Computer Applications". Questo articolo confronta l'efficacia di SVM come metodo di rilevamento delle intrusioni e descrive come SVM sia in grado di rilevare diversi

tipi di intrusioni, tra cui attacchi DoS, attacchi di port scanning e attacchi di SQL injection. Tuttavia, come descritto in precedenza, SVM potrebbe avere difficoltà a rilevare alcune intrusioni che utilizzano tecniche di mascheramento. Questo articolo è pubblicato nella rivista "Journal of Network and Computer Applications", che è pubblicata da Elsevier.

3.2 Random Forest

Questo algoritmo è in grado di rilevare una vasta gamma di tipologie di attacchi, tra cui

- Scanning: Per riconoscere attacchi di tipo scanning con l'algoritmo di Random Forest, si utilizzano tecniche di machine learning supervisionato, in cui l'algoritmo viene addestrato su un set di dati di input, in cui sono presenti informazioni su esempi di attacchi di scanning e su esempi di traffico di rete normale.
- Attacchi di tipo DoS
- attacchi zero-day: Come già visto,questi tipi di attacchi sfruttano tecniche che non sono state ancora scoperte,tuttavia RF può essere in grado di rilevare alcuni tipi di attacchi zero-day grazie alla sua capacità di adattarsi ai nuovi schemi di traffico anomalo. Questo significa che se un attacco zero-day genera un comportamento di rete anomalo che non è stato visto prima, Random forest potrebbe essere in grado di rilevarlo.

Potrebbe non essere in grado di rilevare alcune tipologie di attacchi, come ad esempio

- Attacchi di mascheramento: Se l'attaccante riesce a mascherare il proprio traffico in modo tale da assomigliare al traffico normale, può essere difficile per Random forest rilevare l'attacco.
- Attacchi di tipo phishing o spear phishing: Questi attacchi mirano a rubare le informazioni personali degli utenti attraverso la creazione di siti web e/o e-mail fraudolente. Random Forest potrebbe non essere in grado di rilevare tali attacchi basandosi sui soli dati di input, poiché la rilevazione potrebbe richiedere una maggiore comprensione del contenuto dei siti web.
- Attacchi di tipo Social Engineering: Questi attacchi cercano di ottenere l'accesso a informazioni riservate dall'utente, sfruttando le debolezze umane, come la mancanza di consapevolezza della sicurezza informatica o la fiducia eccessiva. Questo tipo di attacco potrebbe essere difficile da individuare utilizzando un algoritmo di rilevazione delle intrusioni.

E' importante precisare che RF può essere utilizzato per rilevare alcuni tipi di attacchi di phishing e social engineering, ma non tutti.
Ad esempio, Random forest può essere in grado di rilevare phishing tramite analisi del contenuto dei messaggi di posta elettronica, ad esempio, cercando di rilevare messaggi che richiedono all'utente di inserire le proprie credenziali di accesso o che contengono link sospetti. Tuttavia, gli attaccanti di phishing possono utilizzare tecniche sofisticate per mascherare il contenuto del messaggio in modo da evitare la rilevazione da parte dell'algoritmo di Random forest.
Per quanto riguarda gli attacchi di social engineering, Random forest potrebbe essere in grado di rilevare alcune tecniche comuni, come l'inganno o la manipolazione dell'utente, analizzando il comportamento dell'utente sul sito o sulle applicazioni web. Tuttavia, se l'attaccante è in grado di ingannare l'utente in modo tale da farlo agire normalmente, Random forest potrebbe non essere in grado di rilevare l'attacco.
Un articolo del 2020 che spiega nel dettaglio le performance di RF è il seguente:
"Performance analysis of Random Forest Classifier for intrusion detection system. Procedia Computer Science".
Questo articolo esamina l'efficacia di Random Forest come metodo di rilevamento delle intrusioni e descrive come il classificatore sia in grado di rilevare diverse categorie di intrusioni, inclusi attacchi di tipo DoS e attacchi di port scanning. Tuttavia, viene anche rilevato che RF potrebbe avere difficoltà a rilevare alcune intrusioni che utilizzano tecniche di mascheramento, come la manipolazione del traffico di rete per nascondere l'attività sospetta.
L'articolo è pubblicato sulla rivista "Procedia Computer Science", che è pubblicata da Elsevier.

3.3 Naive Bayes

Il classificatore Naive Bayes è un altro algoritmo di apprendimento automatico ampiamente utilizzato per la rilevazione delle intrusioni nella rete. Come per gli altri algoritmi di rilevazione delle intrusioni, anche il classificatore

Naive Bayes può rilevare alcune tipologie di attacchi, ma potrebbe non essere in grado di rilevare altre tipologie. Ad esempio, il classificatore Naive Bayes può rilevare

- Attacchi di scanning
- Attacchi di Denial of Service
- Attacchi di tipo phishing
- Attacchi di tipo spam: può classificare e-mail in spam e non spam. Il suo approccio probabilistico e la sua capacità di elaborare grandi quantità di dati lo rendono particolarmente efficace nell'identificazione di modelli di spamming. In particolare, il modello di classificazione di Naive Bayes utilizza la probabilità di un certo termine o combinazione di termini in un messaggio per calcolare la probabilità che il messaggio sia spam o non spam. Questo lo rende particolarmente utile per l'identificazione di modelli di spamming e per l'elaborazione di grandi volumi di messaggi.

Tuttavia, il classificatore Naive Bayes potrebbe avere difficoltà nella rilevazione di alcune tipologie di attacchi avanzati e sofisticati, come ad esempio

- Attacchi basati sull'ingegneria sociale
- Attacchi basati sul mascheramento
- Attacchi basati sull'evasione o elusione, come ad esempio attacchi che modificano in modo dinamico il loro comportamento per eludere i sistemi di rilevamento delle intrusioni.

In sintesi, il classificatore Naive Bayes può rilevare diverse tipologie di attacchi comuni, ma potrebbe non essere in grado di rilevare alcune tipologie di attacchi avanzati o sofisticati che richiedono una maggiore conoscenza e comprensione del contesto e del contenuto del traffico di rete.
Un articolo del 2017 molto interessante che tratta in maniera dettagliata i pro e i contro di Naive Bayes è il seguente: "An efficient intrusion detection system using Naive Bayes classifier. Procedia Computer Science".
Questo articolo è pubblicato sulla rivista "Procedia Computer Science", che è pubblicata da Elsevier.

In generale possiamo notare come gli attacchi, ad oggi, ancora diffili da prevenire anche utilizzando tecniche di intelligenza artificiale sono quelli basati sull'ingegneria sociale dove l'obiettivo è ingannare l'utente a compiere azioni che non dovrebbe compiere facendole sembrare azioni legittime, oppure attacchi sofisticati "ad hoc" per il sistema da attaccare sfruttando tecniche di intrusione non ancora scoperte e/o vulnerabilità specifiche.

4 Risposta 3

Le soluzioni proposte permettono di identificare più tipi di attacchi. Come abbiamo visto nella risposta precedente, ogni tecnica è in grado di rilevare alcuni categorie di attacchi mentre non è in grado di rilevare altre categorie.
Ogni soluzione di conseguenza ha i suoi pro e i suoi contro. E' possibile ottenere soluzioni ibride unendo più tecniche, come citato nella risposta alla domanda 1 menzionando alcuni articoli.
In conclusione, le tecniche proposte non sono "ad hoc" per un determinato tipo di attacco, ma sono di carattere generale.
Gli attacchi difficili da prevenire in ogni tecnica proposta riguardano attacchi che implementano strategie nuove e/o di mascheramento, questo perchè sono attacchi molto specifici e una soluzione di carattere generale può riscontrare difficoltà.
Creare una soluzione dedicata proprio per attacchi specifici come gli attacchi che sfruttano il mascheramento è possibile.
Tuttavia, poiché gli attacchi di mascheramento cercano di eludere le soluzioni di rilevamento delle intrusioni, la creazione di una soluzione di rilevamento efficace può essere molto difficile.
In generale, la creazione di una soluzione di rilevamento delle intrusioni ad hoc richiede una conoscenza approfondita degli attacchi specifici che si desidera rilevare.
Soluzioni di intrusion detection ad hoc per il social engineering dovrebbero includere l'analisi del comportamento dell'utente, la verifica dell'autenticità delle richieste di informazioni e la verifica della provenienza delle richieste. Ci

sono anche tecniche di phishing detection che utilizzano tecniche di machine learning e analisi dei dati per rilevare tentativi di phishing, che è una forma comune di attacco di social engineering.
Tuttavia, è importante notare che la prevenzione degli attacchi di social engineering richiede anche la consapevolezza degli utenti, in modo che gli utenti possano riconoscere e resistere a questi tipi di attacchi.
In un sistema è più efficiente avere una/due soluzioni di carattere generale o tante soluzioni ad hoc?
Dipende,ognuna ha i suoi pro e i suoi contro, avere una soluzione di carattere generale può essere più conveniente in termini di costo e complessità rispetto ad avere molte soluzioni ad hoc, ognuna delle quali deve essere progettata e implementata separatamente. Inoltre, una soluzione generale può essere più flessibile e adattarsi a una vasta gamma di minacce. Tuttavia, come abbiamo visto, può essere necessario avere soluzioni ad hoc specifiche per alcune minacce particolari, in cui le soluzioni generiche potrebbero non essere efficaci.

5 Risposta 4

Di seguito vengono citati alcuni dei dataset di rilevazione delle intrusioni più comunemente associati alle tecniche di apprendimento automatico SVM, RF, Naive Bayes.

NSL-KDD: questo è un dataset di reti di computer che contiene connessioni di rete etichettate come "normale" o "anomala". È una versione migliorata del dataset KDDCup99, che presenta alcune limitazioni e difetti.
NSL-KDD è stato creato per risolvere alcune limitazioni del dataset KDDCup99, tra cui la presenza di un elevato numero di record duplicati e l'utilizzo di tecniche di pre-processing obsolete che rendevano il dataset poco realistico rispetto alle attuali minacce informatiche. Inoltre, NSL-KDD contiene un set di dati di test separato(Training and Testing) per valutare le prestazioni dei modelli di rilevamento delle intrusioni, cosa che mancava nel dataset KDDCup99.
Avere un set di dati separato è importante poichè in questo modo, si può avere una valutazione più realistica delle prestazioni del modello su dati che non sono stati utilizzati per l'addestramento del modello stesso.
Questi dataset vengono usati principalmente pervalutare l'efficacia di algoritmi come SVM e Naive Bayes.

UNSW-NB15: questo dataset contiene dati di rete reali e artificiali(generati attraverso modelli), tra cui dati di traffico TCP, UDP e ICMP. Il dataset contiene pacchetti che rappresentano anche diverse categorie di attacchi, come attacchi di DoS, esecuzione remota di codice,port stanning e molti altri.
Questo dataset è stato utilizzato per valutare l'efficacia di algoritmi come Random Forest, Naive Bayes, SVM.

Infine,ci sono diversi dataset che contengono file con virus che possono essere utilizzati per l'addestramento di software che analizzano i file in un computer per l'individuazione di malware. Uno di questi dataset è il Microsoft Malware Classification Challenge Dataset, che contiene una collezione di esempi di file PE(formato di file per eseguibili che contengono intestazioni di file, tabelle di sezioni e codice eseguibile) infetti da virus e una serie di metadati associati. Tuttavia, è importante notare che l'utilizzo di dataset contenenti malware deve essere effettuato con cautela e in un ambiente controllato, per evitare la diffusione di eventuali infezioni.