UNIVERSITÀ
DEGLI STUDI
FIRENZE

Scuola di Scienze Matematiche, Fisiche e Naturali
Corso di Laurea in Informatica

Tesi di Laurea

UTILIZZO DI SOFT-BROWNIAN-OFFSET PER LA GENERAZIONE DI ATTACCHI AI FINI DELL'ADDESTRAMENTO DI RILEVATORI DI INTRUSIONI

APPLICATION OF SOFT-BROWNIAN-OFFSET TO GENERATE CYBER-ATTACKS TO TRAIN INTRUSION DETECTORS

GUGLIELMO BARTELLONI

Relatore: *Relatore*
Correlatore: *Correlatore*

Anno Accademico 2022-2023

INDICE

ELENCO DELLE FIGURE

"Inserire citazione"
— Inserire autore citazione

1

INTRODUZIONE

Un sistema di rilevamento di intrusioni, anche detto intrusion detection system (IDS), è un'applicazione o un dispositivo volta a monitorare continuamente la rete per identificare attività malevole.

In particolare un IDS controlla il traffico di rete oppure i log di sistema, alla ricerca di possibili anomalie che potrebbero indicare la presenza di attacchi.

Per migliorare il rilevamento delle intrusioni, nell'ultimo decennio, si è iniziato ad utilizzare algoritmi di Machine Learning e Deep Learning attingendo informazioni dai Big Data. [2]

Questo però questo ha portato a nuove problematiche, come ad esempio il fatto che i modelli di Machine Learning e Deep Learning sono molto sensibili ai dati di addestramento.

Un dataset contenente pacchetti di rete sarà composto per la maggior parte da pacchetti "normali", cioè pacchetti di traffico abituale e, per la restante parte, da pacchetti di attacchi. I vari dataset sul traffico di rete presenti oggi però, non hanno una quantità sufficiente di attacchi per poter addestrare al meglio i modelli.

Un modello addestrato su un dataset con una percentuale di attacchi troppo bassa, non sarà in grado di rilevare bene gli attacchi.

Un metodo per sopperire a questa mancanza è quello di generare nuovi dati a partire da quelli già esistenti (Data Augmentation). I dati generati devono essere però sufficientemente differenti di modo da rappresentare meglio le anomalie che si hanno nella rete durante un attacco.

In questa tesi esploreremo un algoritmo di Data Augmentation chiamato Soft-Brownian-Offset (SBO) [3] e lo utilizzeremo per generare nuovi dati per addestrare un IDS.

In particolare Soft-Brownian-Offset permette di generare campioni così detti out-of-distribution (OOD), cioè campioni che non fanno parte della distribuzione dei dati di partenza. Questo è coerente col fatto che di solito i pacchetti di attacchi sono fuori dalla distribuzione dei pacchetti comuni.

2

FONDAMENTI

2.1 INTRUSION DETECTION SYSTEM

Un intrusione può essere definita come un evento che causa danni ad un sistema informatico [1].

Gli Intrusion Detection System (IDS) sono delle soluzioni hardware o software che, posti all'interno di una rete o di un sistema, rilevano eventuali intrusioni.

Questi strumenti sono essenziali per tenere al sicuro le persone da attacchi informatici [1].

[4] Le principali funzioni degli IDS sono:

- Monitorare ed analizzare sia le attività utente che di sistema
- Tracciare le violazioni delle policy utente
- Analizzare le configurazioni e le vulnerabilità del sistema
- Rilevare tipici attacchi di rete
- Analizzare di attività anomale

[5] Solitamente gli IDS vengono classificati in base al tipo di analisi che effettuano e come questi rilevano le minacce. Ne esistono di tre tipi principali:

- Signature-Based Detection (SD)
- Anomaly-based Detection (AD)
- Stateful Protocol Analysis (SPA)

2.1.1 *Signature-Based Detection*

Questo tipo di rilevamento utilizza la firma di un attacco per poterlo rilevare. Quindi conoscendo questa firma, gli IDS la comparano agli eventi catturati della rete. Dato che questi attacchi hanno bisogno di una conoscenza pregressa sono anche chiamati Knowledge-based.

2.1.2 *Anomaly-based Detection*

Gli AIDS sono stati introdotti per sopperire alle mancanze del Signature-Based Detection. Questo tipo di rilevamento utilizza un modello che rappresenta il normale comportamento della rete. Quindi, se viene rilevato un evento che non è coerente con il modello di riferimento, allora viene segnalata un'anomalia. Questo tipo di rilevamento è chiamato anche Behavior-Based.

Il principale vantaggio di questo tipo di IDS è la possibilità di rilevare gli attacchi zero-day [6], in quanto questi sistemi, non si basano sulla firma dei dati o su regole rigide. Inoltre un altro vantaggio è che risulta essere difficile, per un eventuale criminale, capire quale sia il comportamento normale di un utente senza produrre un segnale da parte del sistema [1].

[1] Gli AIDS possono essere classificati in base al metodo utilizzato per la loro implementazione:

- Basati sulla Statistica (Statistical-Based)
- Basati sulla Conoscenza (Knowledge-Based)
- Basati sull'apprendimento automatico (Machine Learning-Based)

Gli AIDS Statistical-Based dopo aver registrato i dati di una porzione di elementi, derivano il modello statistico di un utente nella rete.

I Knowledge-based invece, utilizzano regole predefinite per generare il modello di riferimento. Dall'altra parte troviamo gli AIDS Machine Learning-Based, che utilizzano un algoritmo di apprendimento automatico per generare il modello di riferimento.

La figura 1 mostra più in dettaglio questo tipo di classificazione e i metodi di implementazione associati.

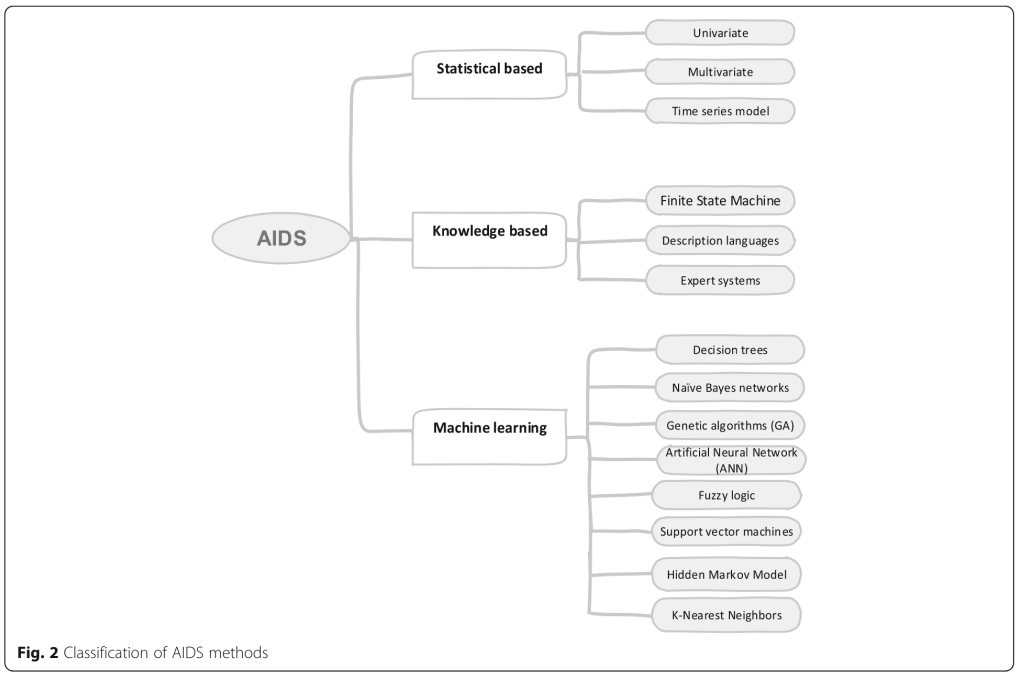

Fig. 2 Classification of AIDS methods

Figura 1: Schema che riassume i vari metodi di implementazione degli AIDS, da [1]. I modelli sono divisi nelle tre categorie descritte precedentemente, ma sono presenti anche i vari tipi di algoritmi utilizzati.

2.1.3 *Stateful-Protocol-Analysis*

In questo caso gli IDS conoscono lo stato e le specifiche del protocollo utilizzato. Vengono quindi rilevati degli eventi che non rispettano gli standard del protocollo, generalmente quelli da specifica e.g. IEEE.

Potrebbe sembrare che gli AD e gli SPA siano simili, in realtà i primi, conoscono il comportamento di una specifica rete, mentre i secondi, conoscono solo gli standard dei protocolli.

2.2 MACHINE LEARNING PER RILEVAMENTO DI INTRUSIONI

2.2.1 *Machine Learning*

Il Machine Learning (ML) è un sottoinsieme dell'intelligenza artificiale che permette ai sistemi di imparare da dati e di migliorare le loro prestazioni nel tempo senza essere esplicitamente programmati. Nel caso degli Intrusion Detection Systems, gli algoritmi di ML permettono di rilevare in maniera precisa e rapida attacchi per grandi quantità di dati in poco tempo [7].

Solitamente gli questi algoritmi sono divisi in:

- Supervised
- Unsupervised
- Semi-supervised

Inoltre gli algoritmi di ML possono essere usati per la classificazione o per la predizione.

La classificazione è un problema di apprendimento, in cui il modello viene addestrato su un dataset con classi conosciute e cerca di trovare una relazione tra le caratteristiche dei dati e le classi stesse. Un classificatore mappa una funzione:

$$f(x_1, \dots x_S) = \hat{y}$$

che assegna uno scalare, facente parte di un insieme C di elementi disgiunti (le classi), ad un insieme di S elementi vettoriali (le caratteristiche). [8]

Nella regressione invece, si cerca di approssimare la funzione:

$$f(x_1, \dots x_S) = y$$

a partire dalle caratteristiche. La funzione f è approssimata utilizzando varie tecniche di interpolazione, estrapolazione, analisi di regressione e curve fitting. [8]

Supervised Machine Learning

Il Supervised Machine Learning utilizza un dataset con classi completamente etichettate cercando di trovare una relazione tra gli elementi del dataset e le classi di questi dati. La classificazione è composta da due parti: l'addestramento e il test. L'addestramento utilizza una variabile di controllo, mentre nel test, si fa predire parte del dataset al modello controllando poi i risultati. Ci sono vari algoritmi di questo tipo e.g. Support Vector Machine (SVM), Naïve Bayes, Reti Neurali, Alberi di Decisione, Random Fores.

Unsupervised Machine Learning

In questo caso, gli algoritmi di ML non hanno un dataset con classi etichettate. Questi algoritmi cercano di trovare delle relazioni tra i dati, cercando di raggrupparli in base a delle caratteristiche comuni.

L'unsupervised machine learning mostra però basse prestazioni per quanto riguarda il rilevamento quindi non possono essere utilizzati come principale strunento per gli IDS. Il motivo di questo scarso rendimento è dovuto al fatto che è molto probabile generare dei Falsi Positivi (vengono rilevati attacchi quando in realtà non cene sono) oppure generare dei Falsi Negativi (non venongo rilevati attacchi quando in realtà ce ne sono) questo va a degradare le performance generali del modello. [6]

D'altra parte però questi modelli si sono dimostrati opinabilmente migliori per rilveare gli attacchi zero-day. [6]

2.3 XGBOOST

2.4 DATASET

3

SOFT BROWNIAN OFFSET

4

GENERAZIONE DI ATTACCHI USANDO SOFT BROWNIAN OFFSET

5

RISULTATI

6

CONCLUSIONI

BIBLIOGRAFIA

[1] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: Techniques, datasets and challenges," *Cybersecurity*, vol. 2, p. 20, July 2019. (Citato nelle pagine 3, 9, 10, and 11.)

[2] Z. Ahmad, A. Shahid Khan, C. Wai Shiang, J. Abdullah, and F. Ahmad, "Network intrusion detection system: A systematic study of machine learning and deep learning approaches," *Transactions on Emerging Telecommunications Technologies*, vol. 32, Jan. 2021. (Citato a pagina 7.)

[3] F. Moller, D. Botache, D. Huseljic, F. Heidecker, M. Bieshaar, and B. Sick, "Out-of-distribution Detection and Generation using Soft Brownian Offset Sampling and Autoencoders," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, (Nashville, TN, USA), pp. 46–55, IEEE, June 2021. (Citato a pagina 7.)

[4] A. S. Ashoor and S. Gore, "Importance of Intrusion Detection System," vol. 2, no. 1, 2010. (Citato a pagina 9.)

[5] H.-J. Liao, C.-H. Richard Lin, Y.-C. Lin, and K.-Y. Tung, "Intrusion detection system: A comprehensive review," *Journal of Network and Computer Applications*, vol. 36, pp. 16–24, Jan. 2013. (Citato a pagina 9.)

[6] T. Zoppi, A. Ceccarelli, and A. Bondavalli, "Unsupervised Algorithms to Detect Zero-Day Attacks: Strategy and Application," *IEEE Access*, vol. 9, pp. 90603–90615, 2021. (Citato a pagina 10.)

[7] T. Saranya, S. Sridevi, C. Deisy, T. D. Chung, and M. Khan, "Performance Analysis of Machine Learning Algorithms in Intrusion Detection System: A Review," *Procedia Computer Science*, vol. 171, pp. 1251–1260, 2020. (Citato a pagina 12.)

[8] F. Hoffmann, T. Bertram, R. Mikut, M. Reischl, and O. Nelles, "Benchmarking in classification and regression," *WIREs Data Mining and Knowledge Discovery*, vol. 9, Sept. 2019. (Citato nelle pagine 12 and 13.)