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. [**?**]

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) [**?**] 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

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

Nel nostro caso ci occupiamo dei Netowrk Intrusion Detection Systems (NIDS), che rilevano intrusioni in una rete analizzando il suo traffico.

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 (SD)
- Anomaly-based (AD)
- Stateful Protocol Analysis (SPA)

2.1.1 *Signature-Based*

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 di una conoscenza pregressa sono anche chiamati Knowledge-based.

2.1.2 *Anomaly-based*

Questo tipo di rilevamento utilizza un modello di riferimento di come la rete normalmente opera. Quindi, se viene rilevato un evento che non è coerente con il modello di riferimento, allora viene segnalato come anomalia. Questo tipo di rilevamento infatti è chiamato anche Behavior-Based.

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.

La principale distinzione da gli AD è che gli SPA non conoscono il comportamento di una specifica rete ma solo quello standard.

2.2 MACHINE LEARNING PER RIDEREVAMENTO INTRUSIONI

2.3 XGBOOST

2.4 DATASET

3

SOFT BROWNIAN OFFSET

4

GENERAZIONE DI ATTACCHI USANDO SOFT BROWNIAN OFFSET

5

RISULTATI

6

CONCLUSIONI

BIBLIOGRAFIA

[1] 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, no. 1, p. e4150, 2021. (Citato a pagina 7.)

[2] 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.)