DOI: 10.26314/GARR-Conf18-proceedings-12

Soluzioni di Deep Learning per la Cyber Security

Nicola Cannistrà, Fabio Cordaro, Francesco La Rosa, Umberto Ruggeri, Riccardo Uccello

Università degli Studi di Messina

Abstract. Un componente fondamentale di un'infrastruttura di cyber security è il sistema di Network Intrusion Detection (NIDS). Un NIDS viene usato per identificare, analizzando il traffico di rete su nodi chiave, attività malevole volte a violare la confidenzialità, l'integrità e la disponibilità dei dati e dei sistemi. Molti tra i NIDS più moderni fanno uso di tecniche di Machine Learning (ML) o Deep Learning (DL) per la loro capacità di adattamento ad attacchi di rete sconosciuti (zero-day). In questo articolo proponiamo un NIDS basato su una deep neural network già adottata con risultati notevoli in ambiti come quelli della Computer Vision e del Natural Language Processing.

Keywords. Cybersecurity, Machine Learning, Deep Learning, AI, network

Introduzione

Creare difese efficaci contro vari tipi di attacchi di rete e garantire la sicurezza delle apparecchiature di rete e delle informazioni è diventato un problema di particolare importanza. I Network Intrusion Detection System identificano attacchi dannosi analizzando il traffico di rete su nodi chiave e sono diventati una parte importante dell'architettura di cyber security.

Esistono tre tipi di analisi del traffico di rete in base alle quali è possibile classificare i NIDS: misuse-based, anomaly-based e hybrid. Nel caso dell'analisi misuse-based il rilevamento dell'abuso è basato sul confronto di eventi registrati con schemi predefiniti di attacco, che, ad oggi, costituisce l'approccio più utilizzato. Questo sistema di rilevamento delle intrusioni è lento nel generare uno schema, rendendo difficile rilevare efficacemente nuovi tipi di attacco emergenti su Internet. Di contro, essi vengono utilizzati per tipi noti di attacchi senza generare un numero elevato di falsi allarmi.

Le soluzioni anomaly-based si basano su un modello del normale comportamento della rete e del sistema e identificano le anomalie come deviazioni da esso. Sono interessanti per la loro capacità di rilevare attacchi zero-day. Il principale svantaggio delle tecniche anomaly-based è il potenziale elevato tasso di falsi allarmi.

La detection ibrida combina l'approccio misuse-based con quello anomaly-based. Diverse sono le soluzioni proposte (Buczak and Guven, 2016) in cui tecniche di ML sono state applicate al problema dell'intrusion detection (Ford and Siraj, 2014). Il ML si concentra principalmente sulla classificazione e la discriminazione in base a caratteristiche note precedentemente apprese per mezzo di dati raccolti per l'addestramento (learning). Il Deep Learning (LeCun et al., 2015) è un nuovo campo di ricerca che ricade nell'ambito del ML.

La sua motivazione sta nella creazione di una rete neurale che simula il comportamento del cervello umano nell'apprendimento analitico. I metodi di DL si possono classificare in metodi di apprendimento supervisionato e apprendimento non supervisionato.

Il vantaggio del DL, rispetto alle soluzioni tradizionali di ML (Meshram and Haas, 2017), è la possibilità di un apprendimento non supervisionato o semi-supervisionato che consente l'estrazione automatica (ed efficiente) di feature (Fiore et al., 2013). Gli algoritmi di DL (Alrawashdeh and Purdy, 2016) comunemente utilizzati includono ad esempio gli Autoencoders, le Belief Networks (DBMs), le Convolutional Neural Networks (CNNs) e le Long Short-Term Memory networks (LSTMs). Di recente sono stati proposti diversi NIDS basati su Convolutional Neural Networks (Wang et al., 2018) che hanno mostrato una precisione nel rilevamento (detection rate) degli attacchi superiore a quella degli approcci precedenti. In questo articolo proponiamo un anomaly-based NIDS che fa uso di una Convolutional Neural Network. I risultati sperimentali mostrano che la nostra soluzione supera approcci analoghi, presenti in letteratura, in termini di accuratezza, Detection Rate (DR) e False Acceptance Rate (FAR).

1. Il Sistema

1.1 Convolutional Neural Networks

Le Convolutional Neural Networks (CNNs) sono un tipo specializzato di rete neurale adatta al processamento di dati sotto forma di matrice, come serie temporali e immagini. La tipica architettura di una CNN (LeCun et al., 2015) consiste di un layer di input ed uno di output completati da diversi layer nascosti.

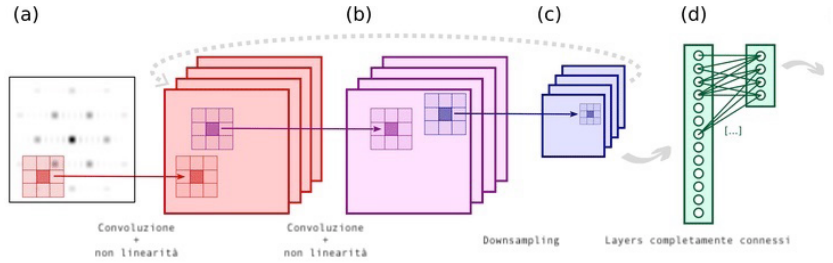

Fig. 1
L'architettura di una Convolutional Neural Network

I layer nascosti sono sia dei convolutional layer che dei pooling layer o dei layer completamenti connessi. Un'architettura tipica delle CNNs è mostrata in Fig. 1. Il convolutional layer costituisce il blocco principale di una CNN e i suoi parametri coincidono con i coefficienti dei filtri (o kernel) che verranno applicati sull'immagine d'ingresso (o matrice). Un altro elemento tipico delle CNNs è il pooling, che è una forma di down-sampling non lineare. Lo strato di pooling serve a ridurre le dimensioni spaziali della rappresentazione, a ridurre il numero di parametri e quindi a limitarne l'overfitting. Strati completamente connessi, infine, connettono ogni neurone di uno strato a ciascun neurone nello strato successivo, come in una rete neurale multi-strato tradizionale (MLP).

1.2 L'architettura proposta

Il classificatore, che ha come core la CNN, è stato implementato usando Tensorflow (Abadi et al., 2016). La soluzione proposta è caratterizzata da 24 canali generati dall'applicazione di altrettanti kernel di dimensione 3*3 e con un passo pari a 1. La matrice, da cui si ottengono i canali, contiene i dati raw acquisiti dalla rete ed ha una dimensione fissa di 28*28 elementi (784 bytes). La CNN è composta da due convolutional layer, 2 layer di pooling (con un 2*2 max-pooling) ed una rete completamente connessa con 3 strati nascosti e 5 uscite. Per il training si è usato l'Adam optimizer su batch di 60 esempi con una funzione di costo cross-entropy. Si sono adottati, inoltre, un learning rate ed un training time rispettivamente pari a 0.002 e 50 epoche.

2. Misure Sperimentali

Per favorire il confronto tra la soluzione proposta e altre presenti in letteratura, si è realizzato un benchmark basato sul dataset DARPA98 (Graf et al., 1998). Nel dataset DARPA98, al traffico normale, sono sovrapposti attacchi di rete appartenenti a 4 famiglie diverse: DoS, Probe, R2L e U2R (vedi tab.1). Per valutare le prestazioni della soluzione proposta, abbiamo adottato tre metriche, accuratezza, DR e FAR, comunemente utilizzate nel campo dell'intrusion detection. L'accuratezza è utilizzata per ottenere una stima complessiva delle prestazioni del NIDS. La DR è usata per stimare le prestazioni del sistema rispetto all'identificazione degli attacchi. Il FAR, invece, permette di quantificare gli errori di classificazione in caso di traffico normale. La definizione di ciascuna delle metriche adottate è riportata in eq. 1. Con riferimento alla eq. 1, dato X come evento attacco e non-X come evento traffico normale, potremo indicare con TP (true positive) il numero di istanze correttamente classificate come X, con TN (true negative) il numero di istanze classificate correttamente come non-X, con FP (false positive) il numero di istanze classificate erroneamente come X e con FN (false negative) il numero di istanze classificate erroneamente come non-X.

$$ACC = \frac{TP+TN}{TP+FP+FN+TN} \quad DR = \frac{TP}{TP+FN} \quad FAR = \frac{FP}{FP+TN} \quad (1)$$

Categoria	Pattern di attacco
Dos	back, land, neptune, pod, smurf, teardrop
R2L	ftp-write, guess-passwd, imap, multihop, phf, spy, warezclient, warezmaster
U2R	buffer-overflow, loadmodeule, perl, rootkit
Probe	ipsweep, nmap, portsweep, satan

Tab. 1
Categorie
degli attacchi

Dataset	Acc%	DR%	FAR
Dos	99.62	99.23	0.03
Probe	99.30	83.43	0.02
R2L	99.75	75.1	0.03
U2R	99.98	67.2	0.03
Totale	99.72	97.82	0.08

Tab. 2
Prestazioni
del NIDS

In tab.2 sono riportati accuratezza, DR e FAR riscontrati per ciascuna tipologia di attacco. Dai risultati ottenuti è emersa una "buona" classificazione per quasi tutte le categorie di attacchi oggetto di questa sperimentazione. Da un'analisi più approfondita dei risultati ottenuti è emerso che una parte degli attacchi di DDoS sono stati "confusi" con del traffico normale. La ragione di tale risultato sta nel fatto che alcuni flussi di rete relativi a DDoS sono molto simili a quelli del traffico normale.

3. Conclusioni

In questo articolo abbiamo proposto un Network Intrusion Detection System basato su una Convolutional Neural Network. I risultati sperimentali evidenziano le prestazioni del sistema in termini di accuratezza, DR e FAR. Due sono i problemi che richiedono un ulteriore approfondimento. Il primo problema consiste nel migliorare le prestazioni della rete su dataset sbilanciati (il traffico dovuto ad attacchi è piccolo rispetto al traffico normale). Il secondo consiste nella possibilità di migliorare le prestazione del NIDS proposto combinando i dati (raw) raccolti con feature tradizionali (Wang et al., 2018).

Riferimenti bibliografici

Alrawashdeh K., Purdy C. (2016), Toward an online anomaly intrusion detection system based on deep learning, 15th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 195–200.

Buczak A., Guven E. (2016), A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys Tutorials, (18), pp. 1153–1176.

Abadi M. et al. (2016), Tensorflow: Large-scale machine learning on heterogeneous distributed systems.

Fiore U., Palmieri F., Castiglione A., De Santis A. (2013), Network anomaly detection with the restricted boltzmann machine, Neurocomput., (122), pp. 13–23.

Ford V., Siraj A. (2014), Applications of machine learning in cyber security, (10).

Graf I., Lippmann R., Cunningham R., Fried D., Kendall K., Webster S., Zissman, M., (1998). Results of DARPA 1998 offline intrusion detection evaluation. https://www.ll.mit.edu/r-d/datasets/1998-darpa-intrusion-detection-evaluation-data-set.

LeCun Y., Bengio Y., Hinton G. (2015), Deep learning, (521), pp. 436–444.

Meshram A., Haas C. (2017), Anomaly detection in industrial networks using machine learning: A roadmap. Machine Learning for Cyber Physical Systems, pp. 65–72.

Wang W., Sheng Y., Wang J., Zeng X., Ye X., Huang Y., Zhu M. (2018), Hast-ids: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection, IEEE Access, (6), pp. 1792–1806.

Autori

Nicola Cannistrà - nicola.cannistra@unime.it
APM-GARR per l'Università di Messina. Responsabile dell'U.Op. "Infrastrutture ICT" dell'Università degli Studi di Messina, ha sviluppato competenze in ambito sistemistico, progettazione e gestione reti in ambito MAN, sistemi di WiFi centralizzato, sistemi VoiP, sicurezza e sistemi di autenticazione.

Fabio Cordaro - fabio.cordaro@unime.it
Vice responsabile dell'Unità Operativa "Infrastrutture ICT" presso l'Università degli Studi di Messina, Technical Contact per il dominio della stessa università, ha sviluppato competenze nell'amministrazione di sistemi Unix-like, progettazione e gestione reti, amministrazione di servizi di rete, implementazione di applicativi web.

Francesco La Rosa - francesco.larosa@unime.it
Ha conseguito un dottorato di ricerca in Computer Science c/o l'Università degli Studi di Messina. E' coautore di decine di articoli pubblicati su atti di convegno e riviste internazionali. Negli ultimi anni ha ricoperto ruoli di responsabilità c/o il CIAM (Centro Informatico Ateneo di Messina), Università degli Studi di Messina.

Umberto Ruggeri - umberto.ruggeri@unime.it
Laureato in Ingegneria Elettronica presso l'Universita degli Studi di Messina. Responsabile dell'Unità Organizzativa Sistemi ed Infrastrutture ICT presso l'Universita degli Studi di Messina. Ha sviluppato competenze in ambito sistemistico, virtualizzazione di sistemi, progettazione di reti, sicurezza e sistemi di autenticazione.

Riccardo Uccello - riccardo.uccello@unime.it
Laureato in Fisica presso l'Universita degli Studi di Messina. Ha ricoperto negli ultimi anni ruoli di responsabilità nel settore dell'ICT, prima presso l'Universita Mediterranea ultimamente presso il Centro Informatico dell'Università degli Studi di Messina. Ricopre il ruolo di APA-GARR per l'Ateneo messinese.