Rilevamento delle Intrusioni mediante Ensemble Machine Learning

Tesi di Laurea Magistrale in Ingegneria Informatica

S. P. Romeo

Relatore: Prof. Giuseppe Lo Re

Relatore: Prof. Alessandra De Paola

Correlatore: Ing. Antonio Bordonaro

Rilevamento delle Intrusioni mediante Ensemble Machine Learning

Tesi di Laurea di
Simon Pietro Romeo

Relatore:
Prof. Alessandra De Paola
Prof. Giuseppe Lo Re

Correlatore:
Ing. Antonio Bordonaro

Sommario

Uno dei problemi più rilevanti nell'ambito della sicurezza informatica è la gestione degli accessi non autorizzati a un sistema informativo da parte di un attaccante.
I sistemi di rilevamento delle intrusioni basati su tecniche di machine learning consentono di rilevare questi accessi ricavando in maniera automatica un profilo di base del sistema e classificando come possibile intrusione tutto quel traffico che non rientra nei limiti dei parametri appresi.
Tuttavia questo tipo di approccio soffre di alcuni problemi tra cui l'indispo- nibilità di dataset aggiornati e la difficoltà nell'individuare un'architettura di machine learning che massimizzi i risultati ottenuti.
Lo scopo del presente lavoro è stato la ricerca della migliore tecnica di de- cision fusion basata sull'ensemble machine learning nell'ambito di un sistema di rilevamento delle intrusioni tramite classificazione multilivello.
Tra le varie tecniche di ensemble presenti in letteratura si è presa in considerazione la Stacked Generalization. Per verificarne l'efficacia sono stati analizzati più insiemi di meta-caratteristiche differenti proposti in letteratura e da questo lavoro. Inoltre si sono confrontati diverse tecniche di classificazione tra cui alberi decisionali e reti Bayesiane.
Il sistema multilivello risultante consente di superare le limitazioni degli approcci esistenti, come dimostrato dalla valutazione sperimentale condotta..

Indice

Capitolo 1

Introduzione

Nell'ambito della sicurezza informatica, i sistemi di rilevamento delle intrusioni, o IDS (Intrusion Detection System), sono dei componenti software o hardware il cui scopo è quello di stabilire, mediante varie tecniche, se all'interno della rete, nella quale stanno operando, si è verificata una violazione del perimetro di sicurezza virtuale.

Un'intrusione è definita come un'attività o un insieme di attività non autorizzate atte a recare un qualche tipo di disservizio o danno ad un sistema informativo.
Le intrusioni hanno l'obiettivo principale di far venir meno almeno uno degli obiettivi della sicurezza informatica ovvero [13]:

- **Confidenzialità**: L'accesso ai dati è consentito soltanto alle persone che ne hanno l'autorizzazione e non possono essere letti da nessun altro attore esterno.
- **Integrità**: I dati devono essere protetti da modifiche accidentali o non autorizzate. In caso di modifica i dati devono essere contrassegnati come non autentici.
- **Disponibilità**: L'accesso in lettura ai dati deve essere garantito per un tempo stabilito, che può essere anche potenzialmente infinito, e in modo continuativo.

Dal momento che l'informazione è un bene fondamentale per le aziende, e che ormai la maggior parte delle informazioni sono custodite all'interno di

sistemi informativi, ogni organizzazione deve essere in grado di garantire la sicurezza dei propri dati dato il continuo aumento dei fenomeni di attacchi virtuali.
La recente tendenza dei cyber-criminali di sferrare attacchi che mirano a "tenere in ostaggio" i dati di importanti aziende e infrastrutture pubbliche ha portato all'attenzione dell'opinione pubblica il problema legato agli attacchi informatici.

Una delle soluzioni più avanzate per mitigare le attività malevole sono i sistemi di rilevamento delle intrusioni.
Lo scopo di questi sistemi è quello di identificare diversi tipi di traffico malevolo e di uso scorretto della rete non individuabili da dispositivi più semplici come i Firewall.
Gli IDS possono operare nel contesto delle reti locali, proteggendole dagli attacchi provenienti dall'esterno e dalle infrazioni commesse dagli utenti interni alla rete, o nell'ambito di una singola macchina, monitorando le azioni intraprese dall'utente e dal sistema operativo, e stabilendo se queste costituiscano una minaccia all'integrità del sistema e dei suoi dati o se rappresentino un uso ammissibile del sistema.

Nell'ambito degli IDS esiste una fondamentale distinzione che li differenzia in base al loro principio di funzionamento. Ogni sistema appartiene ad una delle seguenti categorie:

- **Signature based IDS** ovvero IDS basati sulle Signature
- **Anomaly based IDS** ovvero IDS basati sulle Anomalie

Un Sistema di Rilevamento delle Intrusioni basato sulle signature utilizza tecniche di *pattern matching* per riconoscere schemi di attacco già conosciuti per cui è stata prodotta una firma [14].
L'allarme viene lanciato quando la firma del traffico in ingresso coincide con una delle firme presente all'interno del database del sistema; questi sistemi sono particolarmente efficaci per tutte quelle tipologie di attacchi conosciuti per cui è stata prodotta una firma.
Tuttavia lo svantaggio principale risiede nella difficoltà nell'individuare attacchi di tipo *zero-day* ovvero gli attacchi che sfruttano una vulnerabilità

ancora non nota agli esperti e per il quale non è stata prodotta ancora una firma.

Un altro svantaggio legato a questi sistemi è quello di non riuscire ad individuare attacchi che sfruttano la segmentazione dei pacchetti per eluderne il controllo. Infatti, un IDS basato sulle Signature analizza i singoli pacchetti e li confronta individualmente con le firme degli attacchi conosciuti. Tutti quei malware che utilizzano più pacchetti per sferrare un attacco difficilmente vengono individuati.
Per risolvere questo problema sono state sviluppate soluzioni che tengono conto dello stato del sistema mediante un diagramma a stati finiti [35].

Il continuo aumento di malware basati su vulnerabilità *zero-day* e la natura polimorfica di alcune tipologie di malware hanno reso inadeguato questo tipo di IDS tradizionali.
Una soluzione è quella di utilizzare gli IDS basati sulle anomalie che, invece di individuare gli attacchi conosciuti in sé, profilano i comportamenti accettabili all'interno di un sistema informativo e tutto quel traffico che non è classificato come accettabile è segnalato come potenziale attacco.

Nei sistemi di rilevamento delle intrusioni basati sulle anomalie, come già anticipato, viene creato un modello comportamentale "normale" di un sistema informativo utilizzando tecniche di apprendimento automatico basate sulla statistica o sulla conoscenza.
Ogni deviazione significativa del comportamento osservato dal modello costruito viene vista come un'anomalia e quindi, potenzialmente, un'intrusione.

In un sistema di questo tipo si distinguono due fasi:

- **Fase di Addestramento**, in cui viene analizzato il traffico durante una situazione di normalità per apprendere un modello di comportamento "normale" del sistema.
- **Fase di Test**, in cui viene utilizzato un insieme di record di attacchi sconosciuti all'IDS per stabilirne la capacità di generalizzare il rilevamento delle intrusioni.

Il vantaggio principale degli IDS basati sulle anomalie è l'abilità di riuscire a identificare gli attacchi zero-day [2].
Tra gli altri benefici troviamo anche la possibilità di individuare attività malevole condotte dagli utenti interni alla rete.

Inoltre, questa tipologia di IDS sono configurati sulla base degli schemi comportamentali degli utilizzatori abituali dei sistemi in cui operano. Questa costituisce una sfida per i cyber criminali in quanto è complesso individuare quale possa essere un comportamento normale di un utente del sistema senza produrre un allarme.

Nonostante i grossi benefici di questa nuova classe di IDS, essi sono relativamente complessi da tarare e tenere aggiornati.
I principali svantaggi sono dettati dalla scarsa disponibilità e completezza di dataset utilizzati per la fase di addestramento e dai tempi di apprendimento del modello del sistema. Infatti, per mantenere efficaci questi sistemi, l'aggiornamento del modello comportamentale degli utenti del sistema informativo deve essere eseguito con costanza. Il mancato aggiornamento del modello può portare il sistema a segnalare un numero elevato di *falsi positivi*, ovvero casi in cui il sistema informativo è utilizzato in modo corretto ma che vengono segnalati come potenziali attacchi dall'IDS.
Dato che gli allarmi prodotti dall'IDS vengono analizzati dagli amministratori per verificare la veridicità dell'attacco e bloccarlo in caso affermativo, un alto numero di falsi allarmi potrebbe portare ad un lavoro extra per il personale incaricato alla sicurezza aumentando il tasso di errore nella gestione degli attacchi.
Come analizzato successivamente, alto è l'interesse verso questa tipologia e molte sono le tecniche utilizzate in letteratura per l'implementazione di IDS basati sulle anomalie.

Il lavoro di tesi si concentra su tecniche di ensemble machine learning [8], ovvero tecniche che sfruttano più algoritmi di machine learning per estrarre un'unica classificazione.
Nello specifico, il lavoro svolto, si è concentrato nell'individuare una tecnica di ensemble basata sullo stacking di classificatori e sull'utilizzo di algoritmi di intelligenza artificiale per ottenere la combinazione migliore delle predizioni effettuate dai singoli classificatori.

Il presente lavoro è così strutturato:

- Il capitolo 2 contiene una panoramica sugli IDS basati sulle anomalie e i diversi metodi usati per implementarli.
- Il capitolo 3 contiene i cenni sulla teoria delle tecniche utilizzate nel presente lavoro.
- Il capitolo 4 contiene la descrizione del sistema proposto, l'insieme delle meta-caratteristiche utilizzato e le tecniche di classificazione studiate.
- Il capitolo 5 contiene i risultati ottenuti dalla fase sperimentale.
- Il capitolo 6 contiene l'analisi delle prestazioni del sistema completo.
- Il capitolo 7 contiene le conclusioni e i possibili sviluppi futuri.

Bibliografia

[1] *A Standard for the Transmission of IP Datagrams over Ethernet Networks*. RFC 894. Apr. 1984. DOI: 10.17487/RFC0894.

[2] Ammar Alazab et al. «Using feature selection for intrusion detection system». In: *2012 International Symposium on Communications and Information Technologies (ISCIT)*. 2012, pp. 296–301. DOI: 10.1109/ISCIT.2012.6380910.

[3] Leyla Bilge et al. «DISCLOSURE: Detecting botnet command and control servers through large-scale netflow analysis». In: *ACSAC 2012, 28th Annual Computer Security Applications Conference, December 3-7, 2012, Orlando, Florida, USA*. 2012.

[4] Kevin W. Bowyer et al. «SMOTE: Synthetic Minority Over-sampling Technique». In: *CoRR* abs/1106.1813 (2011). arXiv: 1106.1813.

[5] L. Breiman et al. *Classification and Regression Trees*. Taylor & Francis, 1984. ISBN: 9780412048418.

[6] Leo Breiman. «Bagging predictors». In: *Machine learning* 24.2 (1996), pp. 123–140.

[7] Leo Breiman. «Random Forests». In: *Machine Learning* 45.1 (ott. 2001), pp. 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324.

[8] Thomas G Dietterich. «Ensemble methods in machine learning». In: *International workshop on multiple classifier systems*. Springer. 2000, pp. 1–15.

[9] Harris Drucker et al. «Boosting and Other Ensemble Methods». In: *Neural Computation* 6.6 (nov. 1994), pp. 1289–1301.

[10] Saso Džeroski e Bernard Ženko. «Is Combining Classifiers with Stacking Better than Selecting the Best One?» In: *Machine Learning* 54.3 (mar. 2004), pp. 255–273. ISSN: 1573-0565. DOI: 10.1023/B:MACH.0000015881.36452.6e.

[11] Eibe Frank et al. «Using Model Trees for Classification». In: *Machine Learning* 32.1 (lug. 1998), pp. 63–76. ISSN: 1573-0565. DOI: 10.1023/A:1007421302149.

[12] Farnaz Gharibian e Ali A. Ghorbani. «Comparative Study of Supervised Machine Learning Techniques for Intrusion Detection». In: *Fifth Annual Conference on Communication Networks and Services Research (CNSR '07)*. 2007, pp. 350–358. DOI: 10.1109/CNSR.2007.22.

[13] R. Heady et al. «The architecture of a network level intrusion detection system». In: (gen. 1990). DOI: 10.2172/425295.

[14] Hannes Holm. «Signature Based Intrusion Detection for Zero-Day Attacks: (Not) A Closed Chapter?» In: *2014 47th Hawaii International Conference on System Sciences*. 2014, pp. 4895–4904. DOI: 10.1109/HICSS.2014.600.

[15] M.A. Jabbar, Rajanikanth Aluvalu e Sai Satyanarayana Reddy S. «RFAODE: A Novel Ensemble Intrusion Detection System». In: *Procedia Computer Science* 115 (2017). 7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-24 August 2017, Cochin, India, pp. 226–234. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2017.09.129.

[16] M.A. Jabbar, Rajanikanth Aluvalu e S. Sai Satyanarayana Reddy. «Intrusion Detection System Using Bayesian Network and Feature Subset Selection». In: *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*. 2017, pp. 1–5. DOI: 10.1109/ICCIC.2017.8524381.

[17] Youssef Laarouchi et al. «A language-based intrusion detection approach for automotive embedded networks». In: *International Journal of Embedded Systems* 10 (gen. 2018), p. 1. DOI: 10.1504/IJES.2018.10010488.

[18] Kingsly Leung e Christopher Leckie. «Unsupervised Anomaly Detection in Network Intrusion Detection Using Clusters.» In: gen. 2005, pp. 333–342.

[19] Alexandru Niculescu-Mizil e Rich Caruana. «Predicting good probabilities with supervised learning». In: *Proceedings of the 22nd international conference on Machine learning*. 2005, pp. 625–632.

[20] Nils J Nilsson. *Intelligenza artificiale*. Apogeo Editore, 2002.

[21] Ranjit Panigrahi e Samarjeet Borah. «A detailed analysis of CICIDS2017 dataset for designing Intrusion Detection Systems». In: 7 (gen. 2018), pp. 479–482.

[22] F. Pedregosa et al. «Scikit-learn: Machine Learning in Python». In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[23] Moacir P. Ponti Jr. «Combining Classifiers: From the Creation of Ensembles to the Decision Fusion». In: *2011 24th SIBGRAPI Conference on Graphics, Patterns, and Images Tutorials*. 2011, pp. 1–10. DOI: 10.1109/SIBGRAPI-T.2011.9.

[24] Swarnalatha Purushotham e B. K. Tripathy. «Evaluation of Classifier Models Using Stratified Tenfold Cross Validation Techniques». In: *Global Trends in Information Systems and Software Applications*. A cura di P. Venkata Krishna, M. Rajasekhara Babu e Ezendu Ariwa. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 680–690. ISBN: 978-3-642-29216-3.

[25] Smitha Rajagopal, Poornima Panduranga Kundapur e Katiganere Siddaramappa Hareesha. «A Stacking Ensemble for Network Intrusion Detection Using Heterogeneous Datasets». In: *Security and Communication Networks* 2020 (gen. 2020), p. 4586875. ISSN: 1939-0114. DOI: 10.1155/2020/4586875.

[26] Smitha Rajagopal, Poornima Panduranga Kundapur e Hareesha K. S. «Towards Effective Network Intrusion Detection: From Concept to Creation on Azure Cloud». In: *IEEE Access* 9 (2021), pp. 19723–19742. DOI: 10.1109/ACCESS.2021.3054688.

[27] Ludovico Scalavino. *Rilevamento delle intrusioni tramite classificazione multilivello*. Tesi di Laurea Magistrale in Ingegneria Informatica, Università degli studi di Palermo. Mar. 2021.

[28] Robert E. Schapire. «Explaining AdaBoost». In: *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*. A cura di Bernhard Schölkopf, Zhiyuan Luo e Vladimir Vovk. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 37–52. ISBN: 978-3-642-41136-6. DOI: 10.1007/978-3-642-41136-6_5.

[29] Alexander Seewald. «How to Make Stacking Better and Faster While Also Taking Care of an Unknown Weakness.» In: gen. 2002, pp. 554–561.

[30] Bayu Adhi Tama et al. «An Enhanced Anomaly Detection in Web Traffic Using a Stack of Classifier Ensemble». In: *IEEE Access* 8 (2020), pp. 24120–24134. DOI: 10.1109/ACCESS.2020.2969428.

[31] K. M. Ting e I. H. Witten. «Issues in Stacked Generalization». In: *Journal of Artificial Intelligence Research* 10 (mag. 1999), 271–289. ISSN: 1076-9757. DOI: 10.1613/jair.594.

[32] Kai Ting e Ian Witten. «Stacked Generalization: when does it work?» In: (nov. 1997).

[33] Ljupčo Todorovski e Sašo Džeroski. «Combining Classifiers with Meta Decision Trees». In: *Machine Learning* 50.3 (mar. 2003), pp. 223–249. ISSN: 1573-0565. DOI: 10.1023/A:1021709817809.

[34] Jouni Viinikka et al. «Processing intrusion detection alert aggregates with time series modeling». In: *Information Fusion* 10.4 (2009), pp. 312–324.

[35] Neil Walkinshaw, Ramsay Taylor e John Derrick. «Inferring Extended Finite State Machine models from software executions». In: *2013 20th Working Conference on Reverse Engineering (WCRE)*. 2013, pp. 301–310. DOI: 10.1109/WCRE.2013.6671305.

[36] David Wolpert. «Stacked Generalization». In: *Neural Networks* 5 (dic. 1992), pp. 241–259. DOI: 10.1016/S0893-6080(05)80023-1.

[37] Qingtao Wu e Zhiqing Shao. «Network Anomaly Detection Using Time Series Analysis». In: *Joint International Conference on Autonomic and Autonomous Systems and International Conference on Networking and Services - (icas-isns'05)*. 2005, pp. 42–42. DOI: 10.1109/ICAS-ICNS.2005.69.

[38] N. Ye et al. «Multivariate statistical analysis of audit trails for host- based intrusion detection». In: IEEE Transactions on Computers 51.7 (2002), pp. 810–820. doi: 10.1109/TC.2002.1017701.

[39] N. Ye et al. «Multivariate statistical analysis of audit trails for host- based intrusion detection». In: IEEE Transactions on Computers 51.7 (2002), pp. 810–820. doi: 10.1109/TC.2002.1017701.

[40] Jiong Zhang, Mohammad Zulkernine e Anwar Haque. «Random-Forests- Based Network Intrusion Detection Systems». In: IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 38.5 (2008), pp. 649–659. doi: 10.1109/TSMCC.2008.923876.

[41] Agate V., De Paola A., Ferraro P., Lo Re G., Morana M., SecureBallot: A secure open source e-Voting system, (2021) Journal of Network and Computer Applications, 191, art. no. 103165, DOI: 10.1016/j.jnca.2021.103165

[42] Agate V., De Paola A., Lo Re G., Morana M.A Simulation Software for the Evaluation of Vulnerabilities in Reputation Management Systems (2021) ACM Transactions on Computer Systems, 37 (1-4), art. no. 3458510 DOI: 10.1145/3458510

[43] Concone F., Lo Re G., Morana M., SMCP: a Secure Mobile Crowdsensing Protocol for fog-based applications, (2020) Human-centric Computing and Information Sciences, 10 (1), art. no. 28, DOI: 10.1186/s13673-020-00232-y

[44] Bordonaro A., De Paola A., Lo Re G., Morana M., Smart Auctions for Autonomic Ambient Intelligence Systems, (2020) Proceedings - 2020 IEEE International Conference on Smart Computing, SMARTCOMP 2020, art. no. 9239687, pp. 180 – 187, DOI: 10.1109/SMARTCOMP50058.2020.00043

[45] Agate V., De Paola A., Lo Re G., Morana M. DRESS: A distributed RMS evaluation simulation software, (2020) International Journal of Intelligent Information Technologies, 16 (3), DOI: 10.4018/IJIIT.2020070101

[46] Agate V., Curaba M., Ferraro P., Lo Re G., Morana M., Secure e-voting in smart communities, (2020) CEUR Workshop Proceedings, 2597, pp. 1 – 11

[47] Agate V., De Paola A., Lo Re G., Morana M. A Platform for the Evaluation of Distributed Reputation Algorithms, (2019) Proceedings of the 2018 IEEE/ACM 22nd International Symposium on Distributed Simulation and Real Time Applications, DS-RT 2018, art. no. 8601020, pp. 182 – 189, DOI: 10.1109/DISTRA.2018.8601020

[48] De Paola A., Gaglio S., Lo Re G., Morana M. A hybrid system for malware detection on big data, (2018) INFOCOM 2018 - IEEE Conference on Computer Communications Workshops, pp. 45 – 50, DOI: 10.1109/INFCOMW.2018.8406963

[49] De Paola A., Favaloro S., Gaglio S., Lo Re G., Morana M., Malware detection through low-level features and stacked denoising autoencoders, (2018) CEUR Workshop Proceedings, 2058

[50] Concone F., De Paola A., Lo Re G., Morana M., Twitter analysis for real-Time malware discovery, (2017) 2017 AEIT International Annual Conference: Infrastructures for Energy and ICT: Opportunities for Fostering Innovation, AEIT 2017, 2017-January, pp. 1 – 6, DOI: 10.23919/AEIT.2017.8240551

[51] Agate V., De Paola A., Lo Re G., Morana M., Vulnerability evaluation of distributed reputation management systems, (2017) ValueTools 2016 - 10th EAI International Conference on Performance Evaluation Methodologies and Tools, pp. 235 – 242, DOI: 10.4108/eai.25-10-2016.2266868

[52] Agate V., De Paola A., Gaglio S., Lo Re G., Morana M., A framework for parallel assessment of reputation management systems, (2016) ACM International Conference Proceeding Series, 1164, pp. 121 – 128, DOI: 10.1145/2983468.2983474

[53] Agate V., de Paola A., Lo Re G., Morana M., A simulation framework for evaluating distributed reputation management systems, (2016) Advances in Intelligent Systems and Computing, 474, pp. 247 – 254, DOI: 10.1007/978-3-319-40162-1_27