



# Progetto e sviluppo di un sistema di riconoscimento di malware per dispositivi mobili Android tramite analisi dinamica

Tesi di Laurea Magistrale in Ingegneria Informatica

S. Prestigiacomo

Relatore: Prof. Giuseppe Lo Re

Correlatore: Prof. Alessandra De Paola

# Progetto e sviluppo di un sistema di riconoscimento di malware per dispositivi mobili Android tramite analisi dinamica

*Tesi di Laurea di*
Sergio Prestigiacomo

*Relatore:*
Ch.mo Prof. Giuseppe Lo Re

*Correlatore:*
Prof. Alessandra De Paola

---

## Sommario

Con la proliferazione dei dispositivi mobili, come smartphone e tablet, unito al fatto che molte aziende permettono l'utilizzo di software di terze parti si assiste al conseguente continuo aumento di attacchi alla sicurezza informatica di questi dispositivi. Uno dei sistemi più diffusi su questi dispositivi è quello Android, grazie alla sua natura open source e all'elevato numero di utenti rappresenta un importante mercato per la distribuzione di software, sia benevolo che malevolo. L'obiettivo di questo lavoro di tesi è la progettazione e lo sviluppo di un sistema per il riconoscimento di malware in ambiente Android. Il sistema progettato si basa su un approccio di analisi dinamica, che sfrutta l'osservazione del comportamento delle applicazioni in esecuzione con particolare attenzione al controllo delle operazioni effettuate in presenza di connessione alla rete. Il sistema è progettato in due varianti, utilizzando un singolo classificatore e un sistema di stacking ensemble learning.Il sistema proposto sfrutta algoritmi di intelligenza artificiale per determinare se l'applicazione analizzata può essere considerata malevola o legittima. Per valutare le prestazioni del sistema proposto sono stati confrontati i risultati dell'analisi con e senza connessione di rete. I risultati ottenuti hanno dimostrato le capacità dell'approccio proposto per il riconoscimento dei malware.

# 1 Introduzione

Al giorno d'oggi, l'utilizzo degli smartphone e dei tablet è comune a molte persone, e sono utilizzate in molti ambienti sia pubblici che privati, questo è dovuto all'incremento delle capacità di questi dispositivi [1], però le stesse funzionalità che aiutano gli utenti nella vita di ogni giorno possono essere sfruttate da attaccanti malevoli per attività illecite [13].

Il metodo più comune per estendere queste capacità è attraverso l'utilizzo di applicazioni di terze parti, chiamate "apps". il modo più semplice per ottenere queste applicazioni è tramite l'utilizzo degli "store", ovvero repository che semplificano agli sviluppatori la pubblicazione delle loro applicazioni.

Particolarmente per Android, dove la presenza di diversi store con regole di pubblicazione diverse, ha attirato l'attenzione di sviluppatori applicazioni malevole.

Considerando anche l'alto numero di app sviluppate per questo ambiente, è importante riuscire ad automatizzare il processo di analisi, con risultati buoni in tempi accettabili

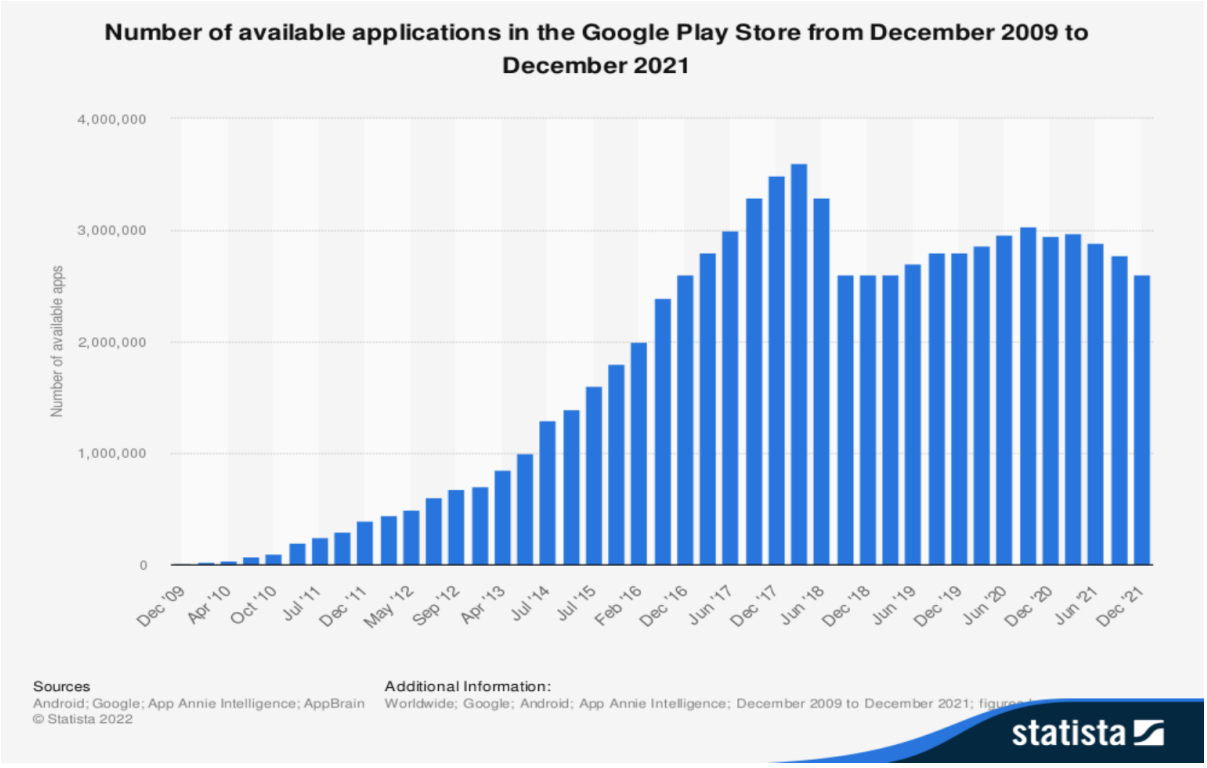


*Figura 1 numero di applicazioni presenti nel Google Store[46]*

Il mercato dei sistemi operativi mobili è praticamente un duopolio [47], a cui fanno capo iOS della Apple e Android di Google, entrambe hanno diversi approcci alla gestione della sicurezza dei loro dispositivi [2], e iOS, pur non essendo esente da problemi di sicurezza [3], è considerato più sicuro rispetto alla sua controparte.

Android invece è la piattaforma mobile preferita dai cracker[1], questo per una serie di motivi come:

- La sua natura opensource, la cui molta libertà permette agli sviluppatori di modificare il funzionamento di alcune parti del sistema, ad esempio, il *launcher* che gestisce l'interfaccia utente, le app che gestiscono telefonate e SMS e anche la tastiera di sistema[4];
- La libertà degli utenti di scaricare applicazioni da store di terze parti o da parti esterne, semplicemente e senza invalidare la garanzia del dispositivo;
- la procedura per la pubblicazione e l'aggiornamento delle app non prevede controlli approfonditi [5];
- il problema della frammentazione, cioè l'elevato numero di varianti di Android create da produttori e operatori che spesso non hanno sufficienti risorse per la manutenzione di un sistema operativo, fa sì che la maggior parte dei dispositivi Android utilizzati non siano aggiornati con le ultime patch di sicurezza [6].
- La crescente diffusione dei dispositivi mobili con sistema operativo Android [7], che ha portato ad un aumento del numero e delle tipologie di malware specifici per questa piattaforma.

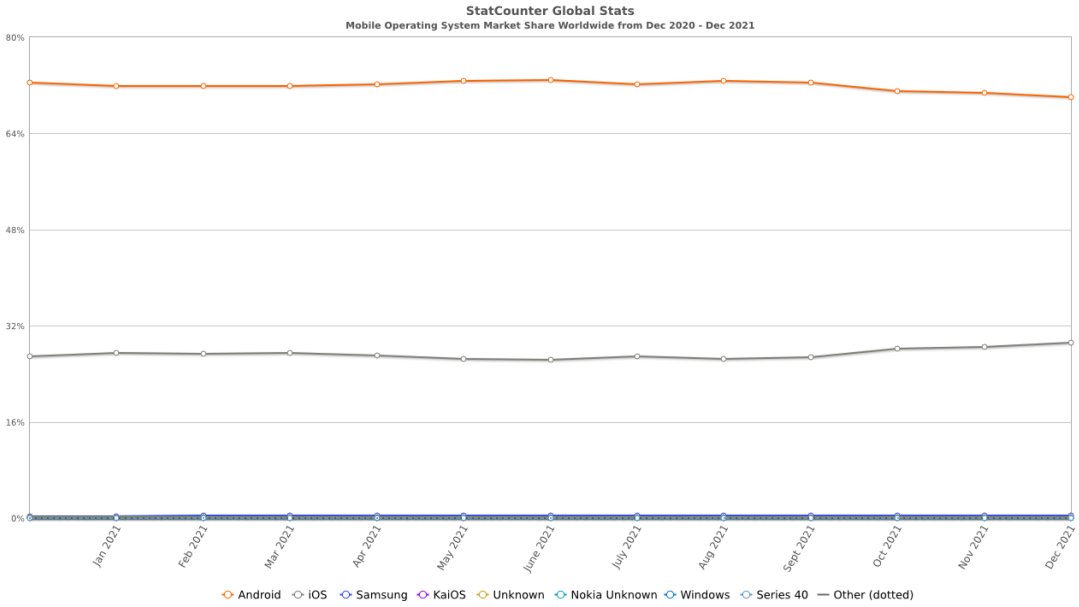


*Figura 2 market share Android[7]*

[1] Con cracker si intende un esperto di sicurezza informatica che sfrutta le sue conoscenze per fini malevoli, viene definito semplicemente hacker sebbene questo ultimo termine indica solamente un esperto di cyber security.

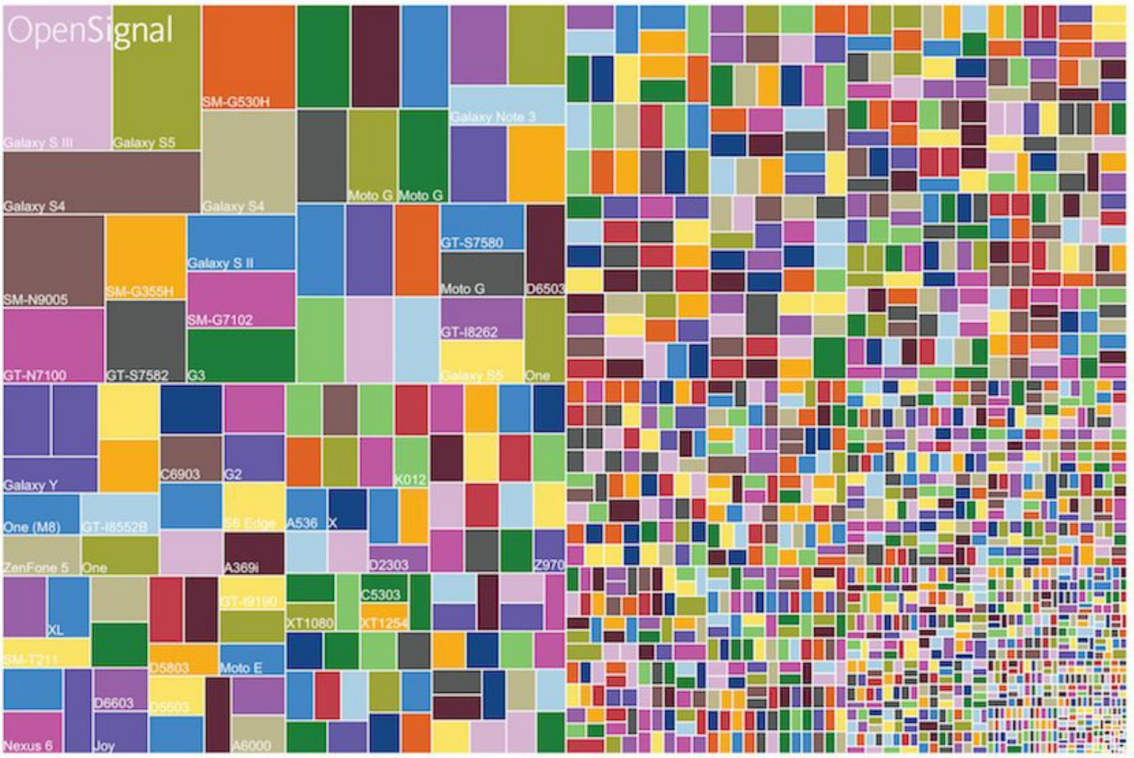


*Figura 3 frammentazione dispositivi Android[8]*

Queste caratteristiche hanno portato ad un aumento del numero e delle tipologie di malware specifici per questa piattaforma.

I tipi di malware sviluppati sono molteplici, ognuno con metodi di attacco e scopi diversi [9], ma in generale vengono raggruppati in diverse categorie, Google classifica questi in varie categorie [10]

| | | |
|---|---|---|
| Click fraud | Trojan | Spyware |
| Hostile downloader | Denial of service | Backdoor |
| Stalkerware | Privilege escalation | Phishing |
| Ransomware | Rooting | Spamware |

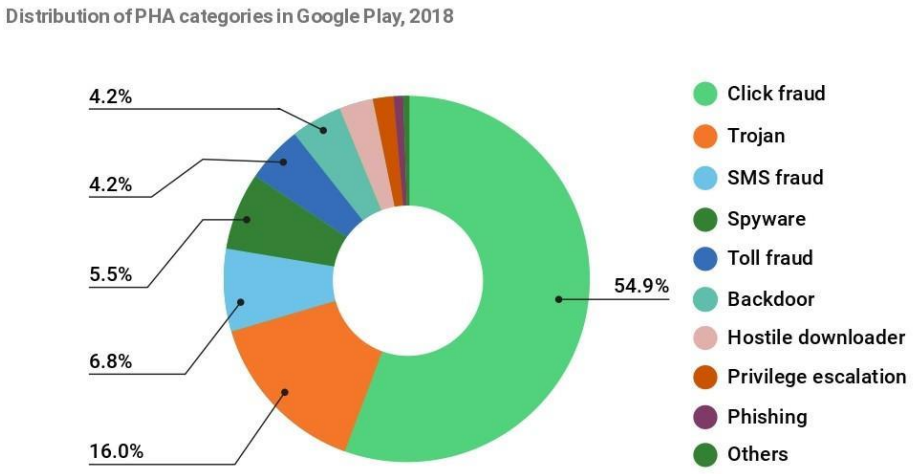


*Figura 4 frammentazione dispositivi Android[9]*

**Click fraud:** malware che simula click su banner pubblicitari inesistenti al fine ottenere illeciti guadagni truffando le società che gestiscono tali banner. Nonostante tale frode colpisca principalmente le società che gestiscono gli spazi pubblicitari online, si possono verificare anche disagi per l'utente che esegue tali apps. Per esempio, spesso si verificano diminuzioni delle prestazioni dello smartphone e una diminuzione dell'autonomia. Inoltre, l'indirizzo IP del dispositivo può venire contrassegnato come sospetto provocando un rallentamento o l'impossibilità di accedere a determinati servizi.

**Trojan:** I trojan sono dei software che appaiono benigni, come giochi o app di utilità, ma in realtà compiono azioni indesiderate a discapito dell'utente, in generale i trojan sono utilizzati in combinazione con altri PHA[2], come i click fraud.

**SMS fraud:** consistono in applicazioni che inviano messaggi di testo a numeri a tariffazione maggiorata controllati dagli attaccanti in modo da rubare il credito della linea

[2] Con PHA si definiscono tutte le applicazioni potenzialmente pericolose (Potentially Harmful Applications)

di telefonia mobile.

**Spyware:** si tratta di una tipologia di malware che ottiene illecitamente i dati personali presenti nei dispositivi degli utenti, come la lista dei contatti informazioni delle e-mail. Tali dati possono in seguito essere monetizzati dagli attaccanti tramite la vendita o l'estorsione.

**Toll fraud:** malware che implementano una truffa analoga a quella impiegata in SMS fraud, utilizzando ad esempio chiamate telefoniche a sovraprezzo o, più comunemente, connessioni ad access point a pagamento(WAP fraud).

**Backdoor:** le backdoor sono delle vulnerabilità nel sistema che vengono sfruttate dagli attaccanti per prendere il controllo completo o parziale del dispositivo di un utente inconsapevole, queste ultime possono essere presenti anche in app benevole.

**Hostile downloader:** si tratta di applicazioni che di per sé non sono malevole ma che scaricano sul dispositivo altre applicazioni che invece lo sono. Non rientrano in questa categoria i browser e i software di file sharing a condizione che scarichino altre app solo su richiesta esplicita dell'utente.

**Privilege escalation:** questa categoria comprende tutte le applicazioni che consentono di ottenere il livello massimo di permessi (root) nei dispositivi Android. Tale pratica minando alla base il modello della sicurezza Android costituisce sempre un pericolo, ma in alcuni casi questo comportamento è richiesto consapevolmente dall'utente che desidera personalizzare il sistema oltre ciò che Android consente per impostazione predefinita.

**Rooting:** Codici che ottengono il livello massimo di privilegio nei dispositivi. Bisogna sottolineare il fatto che il rooting di per sé non è malevolo, ma l'utilizzo di questi codici da parte di applicazioni malevole è comune.

**Other:** in questa categoria rientrano tutti i malware che non fanno parte di nessuna delle categorie elencate precedentemente.

Questi malware sono diffusi sia tramite i canali ufficiali (ad esempio Google Play Store ed Amazon App Store) ma soprattutto attraverso canali alternativi come i marketplace di terze parti (ad esempio Aptoide e Uptodown).

## 1.1 Struttura della tesi

L'elaborato di tesi è strutturato come segue: nel capitolo 2 vi è una breve descrizione delle tipologie di analisi dei malware, mostrando le varie differenze che le caratterizzano; nel capitolo 3 vengono mostrate e descritte le tecniche di apprendimento automatico utilizzate per la fase di classificazione; nel capitolo 4 vengono illustrate le scelte progettuali a livello di architettura e configurazione; nel capitolo 5 viene mostrato l'ambiente di sviluppo, illustrando le varie componenti utilizzate nel sistema dinamico; nel capitolo 6 vengono mostrati i risultati sperimentali ottenuti e, infine, nel capitolo 7 sono presenti alcune conclusioni sul lavoro svolto.

BIBLIOGRAFIA

1.He, D. &. (2015). Mobile application security: Malware threats and defenses. Wireless Communications, IEEE.
2.Kataria, T. Anjali and R. Venkat, "Quantifying smartphone vulnerabilities," 2014 International Conference on Signal Processing and Integrated Networks (SPIN), 2014, pp. 645-649, doi: 10.1109/SPIN.2014.6777033.
3.I. Mohamed and D. Patel, "Android vs iOS Security: A Comparative Study," 2015 12th International Conference on Information Technology - New Generations, 2015, pp. 725-730, doi: 10.1109/ITNG.2015.123.
4.Kharisma, Awal. "What Is Android?" http://developer. android. com/guide … (2011): n. pag. Print.
5.Jacob Leon Kröger, Jens Lindemann, and Dominik Herrmann. 2020. How do app vendors respond to subject access requests? A longitudinal privacy study on iOS and Android Apps. In Proceedings of the 15th International Conference on Availability, Reliability and Security (ARES '20).
Association for Computing Machinery, New York, NY, USA, Article 10, 1–10. DOI:https://doi.org/10.1145/3407023.3407057
6 Mehran Mahmoudi and Sarah Nadi. 2018. The Android update problem: an empirical study. In Proceedings of the 15th International Conference on Mining Software Repositories (MSR '18). Association for Computing Machinery, New York, NY, USA, 220–230. DOI:https://doi.org/10.1145/3196398.3196434
7.L. Caviglione et al., "Tight Arms Race: Overview of Current Malware Threats and Trends in Their Detection," in IEEE Access, vol. 9, pp. 5371-5396, 2021, doi: 10.1109/ACCESS.2020.3048319.
8.OpenSignal. «AndroidFragmentationVisualized» . Disponibile a http://opensignal.com/reports/fragmentation-2013/. Lug. 2013
9. Kiss, Nicolas, et al. "Kharon dataset: Android malware under a microscope." The {LASER} Workshop: learning from authoritative security experiment results ({LASER} 2016). 2016.
10. Tratto da: https://developers.google.com/android/play-protect/phacategories
11. Radzikowski, Przemek Shem. "Analyzing the Power Consumption of Mobile Antivirus Software on Android Devices." (2015).
12 Mohamed, Manar, Babins Shrestha, and Nitesh Saxena. "Smashed: Sniffing and manipulating android sensor data for offensive purposes." IEEE Transactions on Information Forensics and Security 12.4 (2016): 901-913.
13 Nobles, Calvin. "Botching human factors in cybersecurity in business organizations." HOLISTICA–Journal of Business and Public Administration 9.3 (2018): 71-88.
14 Sartea, R. &. (2016). Active Android malware analysis: an approach based on stochastic games.
15 A. De Paola, S. G. (2018). A hybrid system for malware detection on big data. IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS).
16 William, S. (2016). Cryptography and network security : principles and practice (Seventh ed)
17 YARA. (2020, 04 06). Tratto da https://virustotal.github.io/yara/
18 Fabrizio Biondi, T. G.-W. (2018). Tutorial: anOverview of Malware Detection and Evasion Techniques. ISoLA 2018 - 8th International SymposiumOn Leveraging Applications of Formal Methods, Verification and Validation,. Limassol,.
19. B. Kang, S. Y. (2016). N-opcode analysis for android malware classification and categorization. 2016 International Conference On Cyber Security And Protection Of Digital Services (Cyber Security). London.
20. W. Park, K. L. (2014). Analyzing and detecting method of Android malware via disassembling and visualization. International Conference on Information and Communication Technology Convergence (ICTC). Busan.
21. Marastoni, N. &. (2017). GroupDroid: Automatically Grouping Mobile Malware by Extracting Code Similarities. SSPREW-7 , December 4–5, 2017. San Juan, PR, USA.

22. Leonid Batyuk, M. H.-D. (2011). Using Static Analysis for Automatic Assessment and Mitigation of Unwanted and Malicious Activities Within Android Applications, In Proceedings of the International Conference on Malicious and Unwanted Software. Tratto il giorno 04 06, 2020 da http://lilicoding.github.io/SA3Repo/papers/2011_batyuk2011using.pdf
23 D. Ö. Şahın, O. E. (2018). New Results on Permission Based Static Analysis for Android Malware. 6th International Symposium on Digital Forensic and Security (ISDFS). Antalya
24 Arini Balakrishnan, C. S. (2012). Code Obfuscation Literature Survey. Tratto da http://pages.cs.wisc.edu/~arinib/writeup.pdf
25. B. Amos, H. T. (2013). Applying machine learning classifiers to dynamic Android malware detection at scale. 2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC). Sardinia.
26. V. G. Shankar, G. S. (2017). AndroTaint: An efficient android malware detection framework using dynamic taint analysis. 2017 ISEA Asia Security and Privacy (ISEASP).
27 Williamson, S. &. (2012). Active Malware Analysis using Stochastic Games. Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AA-MAS 2012).
28 Hasan, Hayyan, Behrouz Tork Ladani, and Bahman Zamani. "Enhancing Monkey to trigger malicious payloads in Android malware." 2020 17th International ISC Conference on Information Security and Cryptology (ISCISC). IEEE, 2020.
29. N. Varol, A. F. Aydogan and A. Varol, "Cyber attacks targeting Android cellphones," 2017 5th International Symposium on Digital Forensic and Security (ISDFS), 2017, pp. 1-5, doi: 10.1109/ISDFS.2017.7916511.
30. Rashid, W. (2017). Automatic Android Malware Analysis. Kiel University of Applied Sciences,, Computer Science and Electrical Engineering. Tratto il giorno 03 13, 2020 da https://github.com/waqarrashid33/internship_report/blob/master/main.pdf
31. Anwar, S. Z. (2018). Android botnets: A serious threat to android devices. Pertanika Journal of Science & Technology 26.1.
32. Crussell, J. R. (2014). "Madfraud: Investigating ad fraud in android applications.". Proceedings of the 12th annual international conference on Mobile systems, applications, and services.
33. Gorecki, Christian, et al. "Trumanbox: Improving dynamic malware analysis by emulating the internet." Symposium on Self-Stabilizing Systems. Springer, Berlin, Heidelberg, 2011.
34. Rashid, W. (2017). Automatic Android Malware Analysis. Kiel University of Applied Sciences,, Computer Science and Electrical Engineering. Tratto il giorno 03 13, 2020 da https://github.com/waqarrashid33/internship_report/blob/master/main.pdf
35. Shannon, c. E. (1948). A Mathematica lTheory of Communication. The Bell SystemTechnical Journal.
36. Mitchell, T. M. (1997). Machine Learning. McGraw-Hill Science.
37. Ho, T. K. (1995). Random decision forests. Proceedings of 3rd International Conference on Document Analysis and Recognition. Montreal, Quebec, Canada.
38. Breiman, L. (1994). Bagging Predictors. Bagging PredictorsByLeo Breiman*Technical Report No. 421September 1994*PartiaDepartment of StatisticsUniversity of CaliforniaBerkeley, California 94720.
39. Lombardo, A. (2016). Probabilità e statistica per ingegneri
40. Dempster, M. L., Laird, N. M., & Rubin, D. B. (1997). Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological), Vol.
41. Tratto da virusshare: https://virusshare.com/
42. Tratto da Android Malware Genome Project : http://www.malgenomeproject.org/
43. Tratto da Drebin: https://www.sec.cs.tu-bs.de/~danarp/drebin/
44. Tratto da Androzoo: https://androzoo.uni.lu/
45. Hossin, M. a. (2015). A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process.
46. Tratto da Statista: https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/

47. Steinberg, M. (2019). The Platform Economy. How Japan Transformed the Consumer Internet. University of Minnesota Press.
48. 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
49. 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
50. 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
51. 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
52. 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
53. 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
54. 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
55. 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
56. 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
57. 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
58. 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
59. 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
60. 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