UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO

DIPARTIMENTO DI INFORMATICA

CORSO DI LAUREA IN INFORMATICA

TESI DI LAUREA
IN
Vedi cosa fa Giacomo

APCMAM:
Inserisci un titolo Giacomo

RELATORE:
Prof.ssa Annalisa Appice

LAUREANDO:
Giacomo Gaudio

CORRELATORI:

Dott.ssa Giuseppina Andresini

ANNO ACCADEMICO 2024 2025

"Ad Oleg, a cui devo il 90% degli esami"
— Julius Robert Oppenheimer

Dedico questo lavoro ai miei cari

Indice

Disclaimer

Tutti i marchi, nomi commerciali, prodotti e loghi menzionati in questa tesi sono di proprietà dei rispettivi titolari. L'autore non rivendica alcun diritto di proprietà su tali marchi o nomi commerciali e li utilizza solo a scopo informativo e descrittivo, senza alcun intento di violazione.

Sommario

Questo è un abstract.

Capitolo 1

Introduzione

Con l'avvento dell'era digitale, sempre più attività e servizi sono migrati verso il mondo online: dall' e-commerce alla pubblica amministrazione, fino ai servizi finanziari e sanitari. In italia, secondo i dati ISTAT,l'86,2% delle famiglie italiane dispone di un accesso ad internet a testimonianza di un uso sempre più popolare[1].

Tuttavia, l'aumento della connettività comporta inevitabilmente anche una maggiore esposizione alle minacce informatiche. Nel 2024, l'Italia ha registrato il 10% degli attacchi informatici globali; secondo il rapporto Clusit, circa un terzo di tali attacchi è stato veicolato tramite l'utilizzo di malware[2].

Per malware si intende un istanza di un programma il cui intento sia malevole come carpire informazioni personali e/o arrecare danni ai dispositivi[3].

Per poter identificare i malware esistono due tipi di analisi: statica e dinamica. L'analisi statica prevede l'analisi del codice binario, analizzando ogni ramo di esecuzione possibile alla ricerca di codice malevolo. L'analisi dinamica invece prevede l'esecuzione del programma cercando di identificare un insieme di comportamenti noti tra i malware[4].

Per comportamenti intendiamo l'insieme (ordinato) delle chiamate Api al Kernel per permettere l'esecuzione dell'applicativo. Un singolo "comportamento" è effettivamente la singola chiamata Api prendendo il nome di "Api Call"; mentre la lista (ordinata) prende il nome "API call sequence"[5].

L'obbiettivo di questo caso di studio è di poter classificare un programma in base ai suoi comportamenti. Per raggiungere tale scopo sono stati raccolti diversi dataset di sequenza di chiamate api sul sistema operativo Windows ed è stato progettato e idealizzato un applicativo python (seguendo il paradigma OOP) che attraverso algoritmi di machine-learning riesca a classificare un programma in base alla sua sequenza di chiamate api.

1.1 Formato Windows PE malware

Per poter eseguire un programma in `Microsoft Windows`, la logica applicativa deve essere interpretabile dal sistema operativo. Il formato che mette a disposizione il sistema operativo `Microsoft Windows` è `PE` (Portable Executable), denominato cosi poiche si riferisce che il formato non è specifico dell'architettura[6].

Un file in formato **PE** ha nel seguente ordine le seguenti intestazioni:

- **Stub MS-DOS**: area iniziale del file PE, compatibile con il sottosistema `MS-DOS`. In assenza di istruzioni specifiche da parte dello sviluppatore, contiene un programma di default che visualizza il messaggio: "Impossibile eseguire questo programma in modalità DOS".

- **Firma PE**: campo che identifica il file come appartenente al formato **Portable Executable (PE)** e ne consente il corretto riconoscimento da parte del loader di Windows.

- **Intestazione COFF**: fornisce informazioni fondamentali sul file, come il tipo di macchina su cui è destinato a essere eseguito, il numero di sezioni e la data di compilazione.

- **Intestazione facoltativa (solo immagine)**: sezione che specifica le informazioni necessarie al caricamento ed esecuzione del programma. È composta da tre sottosezioni principali:

  - **Campi standard intestazione facoltativi** : includono l'indirizzo di entry point del programma, la dimensione del codice e dei dati, e altri parametri di base.
  - **Campi specifici dell'intestazione facoltativa di Windows**: contengono informazioni più dettagliate, come la versione minima del sistema operativo richiesto, le dimensioni massime di heap e stack, e i valori di allineamento.
  - **Directory dati intestazione facoltative**: riferimenti alle tabelle e alla loro dimensione delle risorse esterne come le **ddl**.

Una **DDL** è una libreria che contiene codice e dati utilizzabili da più di un programma contemporaneamente. Ogni programma può utilizzare le `DDL` del sistema operativo windows per ottenere memoria, accesso a risorse, far apparire elementi a schermo ed accedere alla rete[7].

I malware utilizzano queste librerie per poter eseguire comportamenti malevoli.

1.2 Categorie Malware

I malware, in base al risultato delle loro azioni, assumono un nome più specifico. Nel nostro caso di studio sono state ritrovate le seguenti specializzazioni:

- **Unknown**: Applicazione di cui non si conosce se il suo intento sia malevolo o no.
- **Goodware**: Applicazione benevola; non ha nessun obbiettivo di danneggiare l'utente.
- **Malware**: Appellativo generico per malware in caso in cui non si riesca effettivamente a discriminare l'esito delle sue azioni.
- **Backdoor**: Applicazione che ha come scopo quello di fornire accesso alla macchina attacatta all'attaccante[8].
- **Trojan**: Software il cui scopo è quello di modificare, bloccare o cancellare dati esistenti[9].
- **Virus**: Applicazione disegnata per diffondersi su più host; ha come scopo quello di interrompere il normale funzionamento della macchina host[9].
- **Worm**: Si comportano come i virus ma non hanno bisogno di essere eseguiti manualmente sulla macchina obbiettivo[9].
- **Dropper**: Applicazione che ha come sccopo quella di far scaricare altri software malevoli[10].
- **Spyware**: Applicazione che ha come compito di carpire dati sensibili sulla macchina host e riportare tali dati sulla macchina disegnata[9].
- **Adware**: Applicazione che ha come scopo capire le abitudini dell'utente per propinare pubblicità mirata; possono anche deviare la navigazione web su determinati siti esca[9].
- **Packed**: Malware che è stato compresso almeno una volta per eludere i controlli antivirus[11].

1.3 Chiamata Api Windows

1.4 Contributi

Bibliografia

[1] L. Zannella, M. Zanella, and M. Rizzo, "Report Cittadini E ICT," Istat, Report/Study, 2025. [Online]. Available: https://www.istat.it/wp-content/uploads/2025/04/REPORT_CITTADINI-E-ICT_2024.pdf

[2] C. , "© Clusit - Rapporto 2025 sulla Cybersecurity in Italia e nel mondo," Clusit, Tech. Rep., Mar. 2025. [Online]. Available: https://clusit.it/wp-content/uploads/download/Rapporto_Clusit_03-2025_web.pdf

[3] N. Idika and A. P. Mathur, "A survey of malware detection techniques," *Purdue University*, vol. 48, no. 2, pp. 32–46, 2007. [Online]. Available: https://profsandhu.com/cs5323_s17/im_2007.pdf

[4] Y. Ki, E. Kim, and H. K. Kim, "A novel approach to detect malware based on api call sequence analysis," *International Journal of Distributed Sensor Networks*, vol. 11, no. 6, p. 659101, 2015. [Online]. Available: https://doi.org/10.1155/2015/659101

[5] Y. Qiao, Y. Yang, L. Ji, and J. He, "Analyzing malware by abstracting the frequent itemsets in api call sequences," in *2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, July 2013, pp. 265–270. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6680850

[6] Microsoft, "Formato pe - win32 apps." [Online]. Available: https://learn.microsoft.com/it-it/windows/win32/debug/pe-format

[7] ——, "Libreria di collegamento dinamico." [Online]. Available: https://learn.microsoft.com/it-it/troubleshoot/windows-client/setup-upgrade-and-drivers/dynamic-link-library

[8] Kaspersky, "Trojan droppers." [Online]. Available: https://encyclopedia.kaspersky.com/glossary/backdoor/

[9] Cisco, "What is a malware." [Online]. Available: https://www.cisco.com/site/us/en/learn/topics/security/what-is-malware.html

[10] Kaspersky, "Trojan droppers." [Online]. Available: https://encyclopedia.kaspersky.com/glossary/trojan-droppers/

[11] ——, "Multipacked." [Online]. Available: https://encyclopedia.kaspersky.com/knowledge/multipacked/

Ringraziamenti