Attività di R&D in ambito malware detection PDF per PagoPA S.p.A.

Analisi sul 7% dei documenti malevoli senza JS

[Confidential]

Data: 09/05/2023
Progetto: Attività di R&D in ambito malware detection PDF per PagoPA S.p.A.
Versione: 1.0

Tabella dei contenuti

1 Introduzione

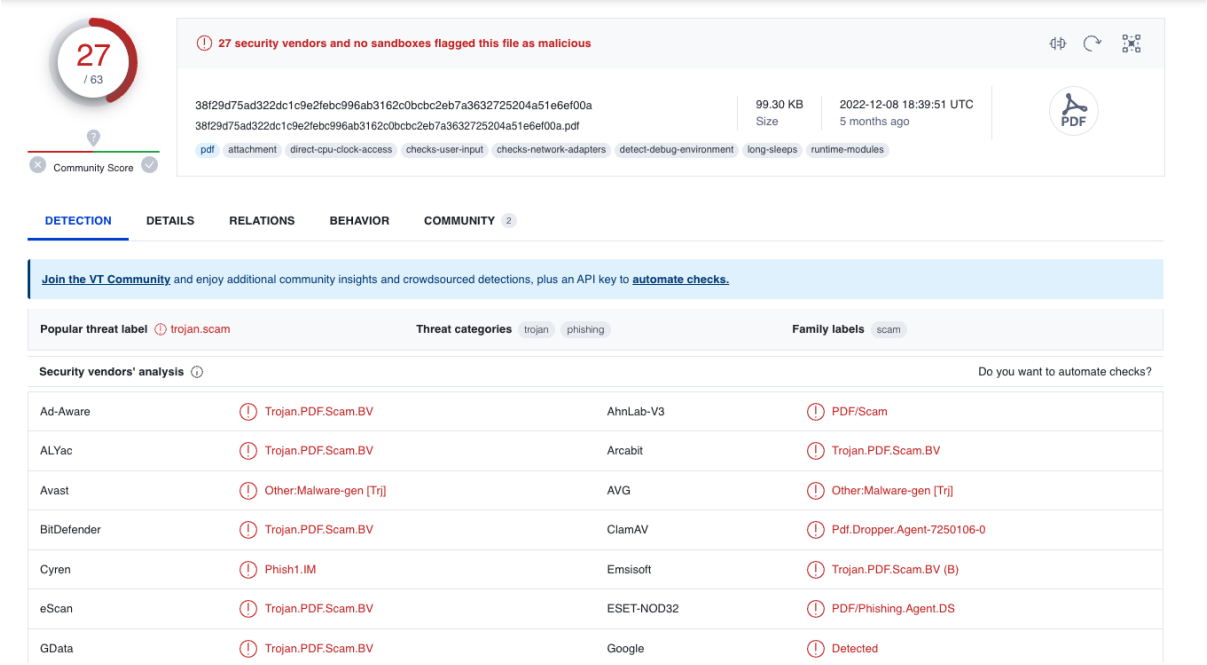
Utilizzando **peepdf** sono risultati tra i PDF malevoli alcuni che non contengono codice JavaScript, pari circa al 7% del totale dei malevoli. Il numero di documenti in questione è **894**. In questa appendice verrà presentata un'analisi utilizzando VirusTotal dei suddetti documenti per identificare gli attacchi da essi sfruttati.

2 Analisi tramite VirusTotal

Per questa analisi è stato **creato uno script Python che effettua query su VirusTotal via API**, usando gli hash dei PDF risultati inizialmente senza JS. Da tale analisi si nota come ci siano comunque dei documenti che riportano **"Trojan JS"** e risultati simili, facendo supporre che **peepdf non sia riuscito a rilevare il codice JS** che in realtà era presente. Ovviamente ciò non vuol dire che se non c'è riferimento a JS nelle firme dei vari antivirus utilizzati da VirusTotal allora i documenti non presentano codice JS. **Non è quindi possibile avere un risultato preciso** in termini di quanti documenti PDF, tra quelli identificati come malevoli non contenenti codice JavaScript, siano davvero esenti da codice JavaScript. Tuttavia, i documenti PDF che non presentavano alcun risultato facente riferimento a JavaScript presentavano invece riferimenti a firme relative a **Phishing** o **Scam**, facendo quindi supporre che tali attacchi si basassero su tecniche di *social engineering* e **non su tecniche di exploit automatico**.

Come esempio, è stato preso il risultato dell'analisi di VirusTotal per il PDF con il seguente sha256: **38f29d75ad322dc1c9e2febc996ab3162c0bcbc2eb7a3632725204a51e6ef00a.**

In esso non è presente **alcun riferimento a JS**, come possiamo vedere in maniera più compatta **su VirusTotal** da browser:



27 / 63

Community Score

27 security vendors and no sandboxes flagged this file as malicious

38f29d75ad322dc1c9e2febc996ab3162c0bcbc2eb7a3632725204a51e6ef00a
38f29d75ad322dc1c9e2febc996ab3162c0bcbc2eb7a3632725204a51e6ef00a.pdf

99.30 KB Size

2022-12-08 18:39:51 UTC
5 months ago

PDF

pdf attachment direct-cpu-clock-access checks-user-input checks-network-adapters detect-debug-environment long-sleeps runtime-modules

DETECTION DETAILS RELATIONS BEHAVIOR COMMUNITY 2

Join the VT Community and enjoy additional community insights and crowdsourced detections, plus an API key to automate checks.

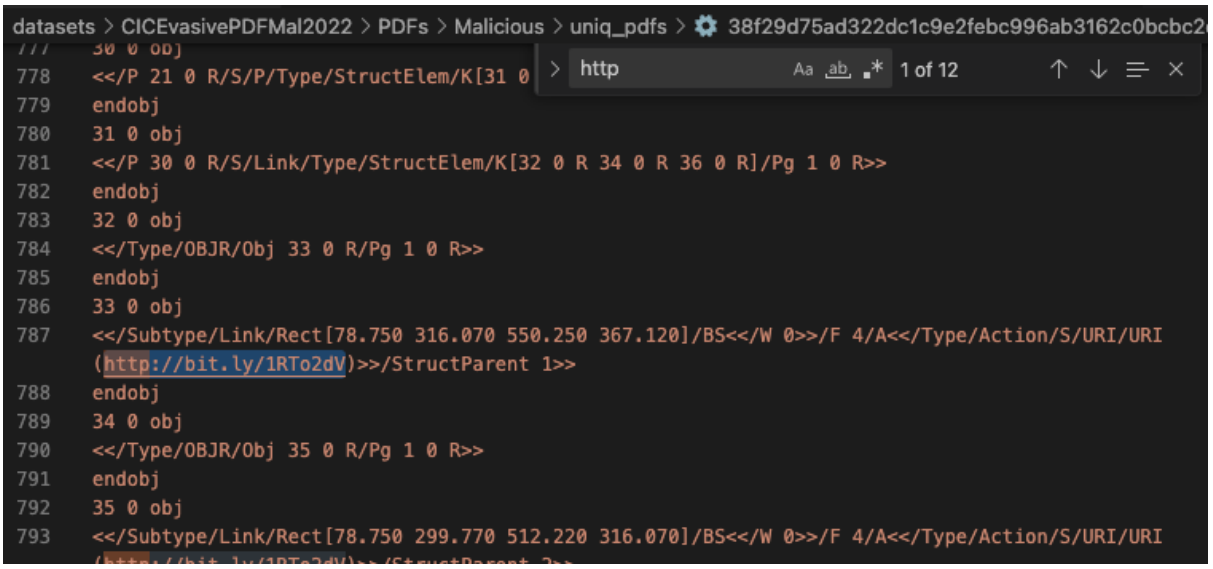
Popular threat label trojan.scam | Threat categories trojan phishing | Family labels scam

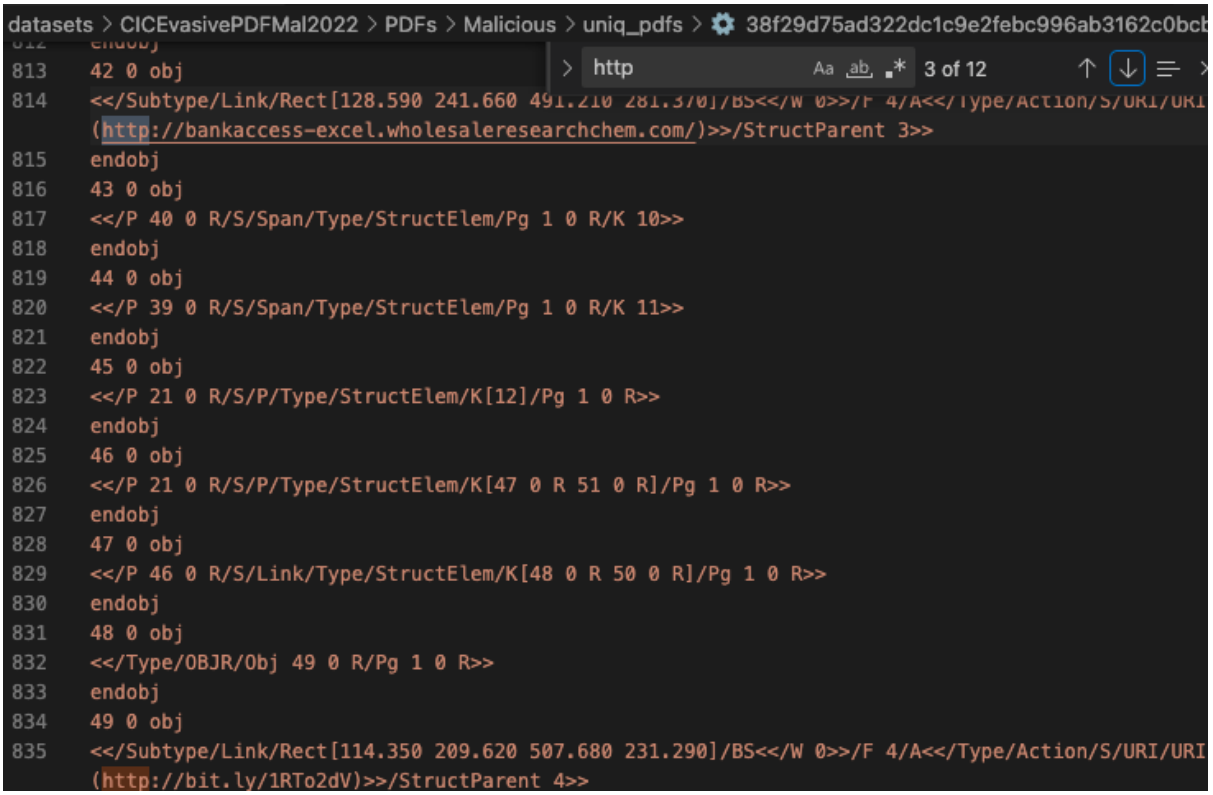
Security vendors' analysis — Do you want to automate checks?

Ad-Aware	Trojan.PDF.Scam.BV	AhnLab-V3	PDF/Scam
ALYac	Trojan.PDF.Scam.BV	Arcabit	Trojan.PDF.Scam.BV
Avast	Other:Malware-gen [Trj]	AVG	Other:Malware-gen [Trj]
BitDefender	Trojan.PDF.Scam.BV	ClamAV	Pdf.Dropper.Agent-7250106-0
Cyren	Phish1.IM	Emsisoft	Trojan.PDF.Scam.BV (B)
eScan	Trojan.PDF.Scam.BV	ESET-NOD32	PDF/Phishing.Agent.DS
GData	Trojan.PDF.Scam.BV	Google	Detected

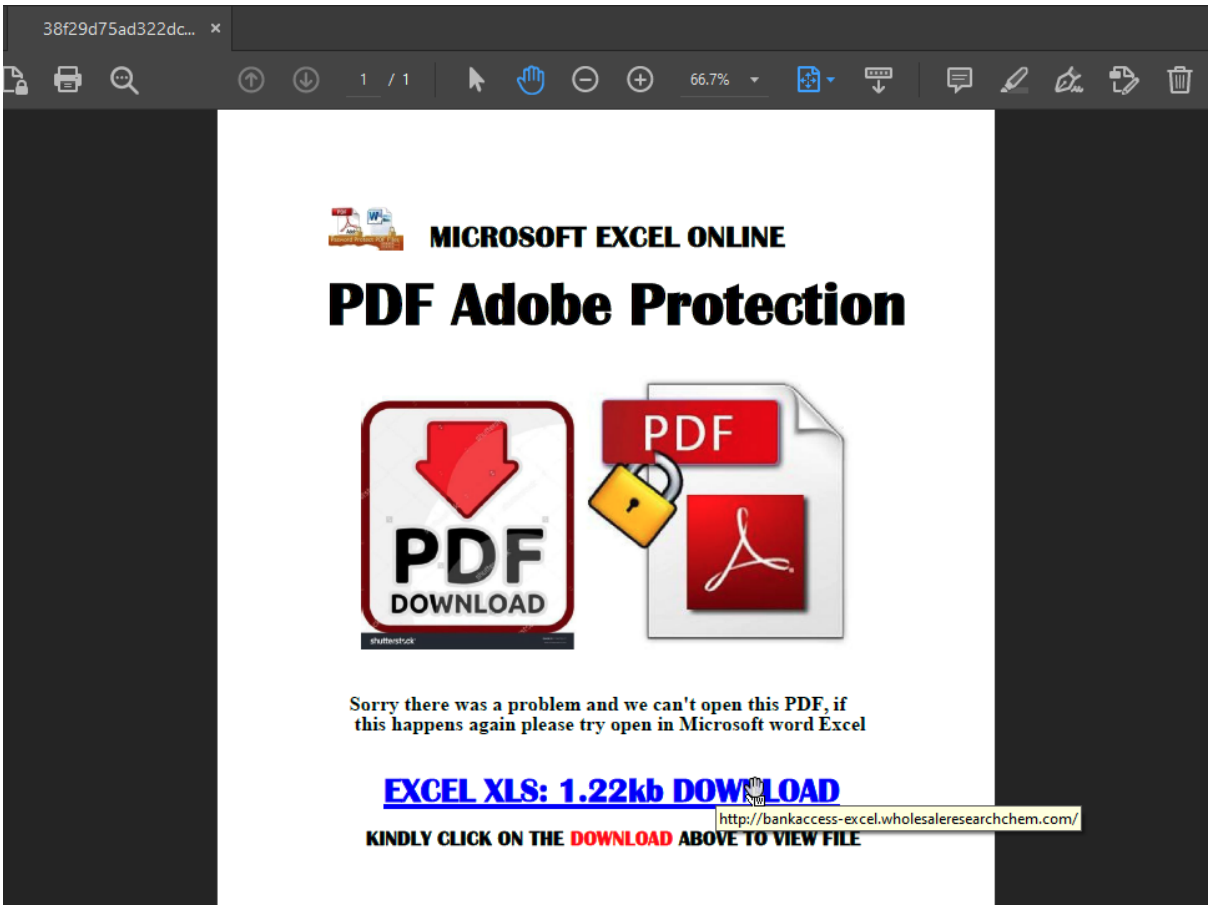
Da qui si vede che sembra essere un PDF di **phishing/scam**, quindi potrebbe contenere qualche **URL malevolo**.

Guardando il contenuto testuale del file PDF è possibile notare la presenza di diversi URL al suo interno:



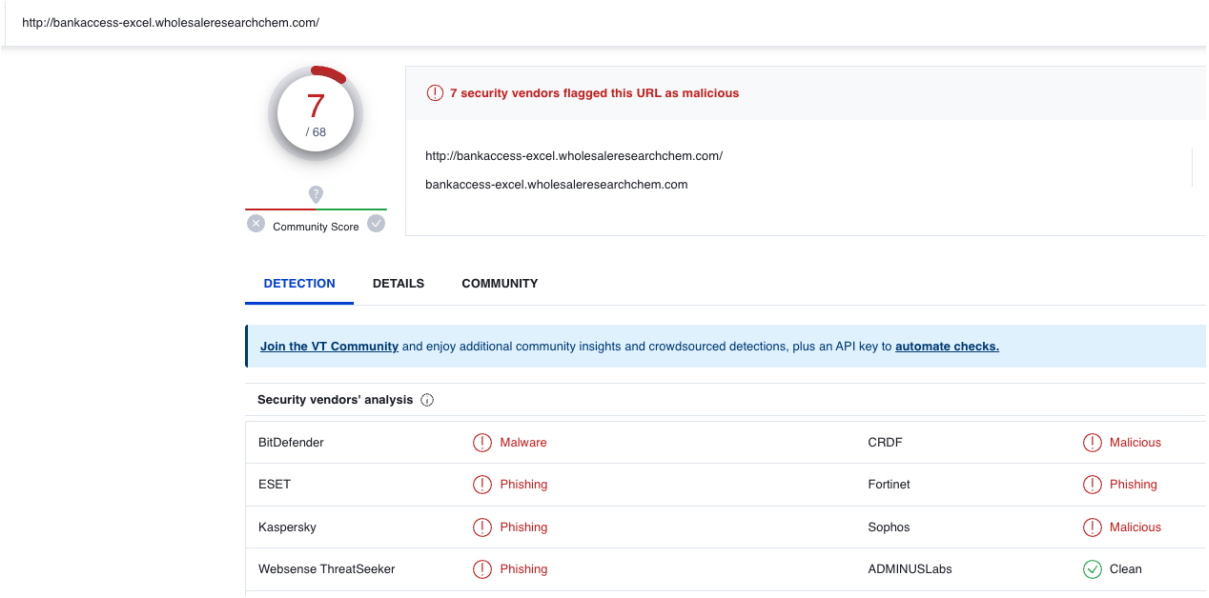


Utilizzando una **VM windows opportunamente isolata dall'host e da Internet**, si è proceduto ad **aprire il file in questione**:



Come è possibile notare, è un **PDF di phishing**, quindi la vittima dovrebbe cadere nella trappola dell'attaccante e **cliccare sul link in maniera sprovveduta**.

Tale link risulta essere un **URL malevolo**, secondo **VirusTotal**:



2 Conclusioni

Non è possibile conoscere con certezza gli attacchi effettuati dai PDF appartenenti al 7%, in quanto i documenti in questione andrebbero effettivamente **analizzati uno per uno con tecniche di malware analysis specifiche**.

In generale, si è notato come alcuni **strumenti avanzati come VirusTotal** rilevano apparentemente la presenza di codice JS all'interno di molti dei documenti inizialmente contrassegnati come non contenenti codice JS. Questo si evince dal fatto che alcuni motori antivirus li rilevano con firme del tipo **"Trojan.JS"** e simili.

Una possibile causa di tale risultato è che gli **strumenti da riga di comando** utilizzati durante l'analisi, **seppur molto efficaci in generale** anche contro documenti offuscati, **potrebbero fallire in caso di documenti offuscati in maniera eccessivamente complessa**, ottenendo quindi dei **falsi negativi**.

Un'altra possibile causa è che i **sistemi di rilevamento automatici** come VirusTotal diano dei **falsi positivi**, contrassegnando dei malware come contenenti codice JS quando in realtà non è così.

Tuttavia, molti dei documenti analizzati risultano comunque contrassegnati con firme del tipo **"Scam"** o **"Phishing"**, il che significa che tali documenti rientrano nella categoria numero 4 precedentemente esposta, ossia quella relativa allo sfruttamento di *social engineering*.

Si può concludere pertanto che:

- **più del 93% dei documenti malevoli del dataset contiene codice JS**;
- **i documenti rimanenti** risultano essere relativi ad **attacchi che necessitano dell'interazione dell'utente**, il cui rischio può essere ridotto attraverso corsi di formazione sul social engineering.