Tracking Threat Actors Using Images and Artifacts

When tracking adversaries, we commonly focus on the malware they employ in the final stages of the kill chain and infrastructure, often overlooking samples used in the initial ones.

In this post, we will explore some ideas to track adversary activity leveraging images and artifacts mostly used during delivery. We presented this approach at the FIRST CTI in Berlin and at Botconf in Nice.

Hunting early

In threat hunting and detection engineering activities, analysts typically focus heavily on the latter stages of the kill chain – from execution to actions on objectives (Figure 1). This is mainly because there is more information available about adversaries in these phases, and it's easier to search for clues using endpoint detection and response (EDR), security information and event management (SIEM), and other solutions.

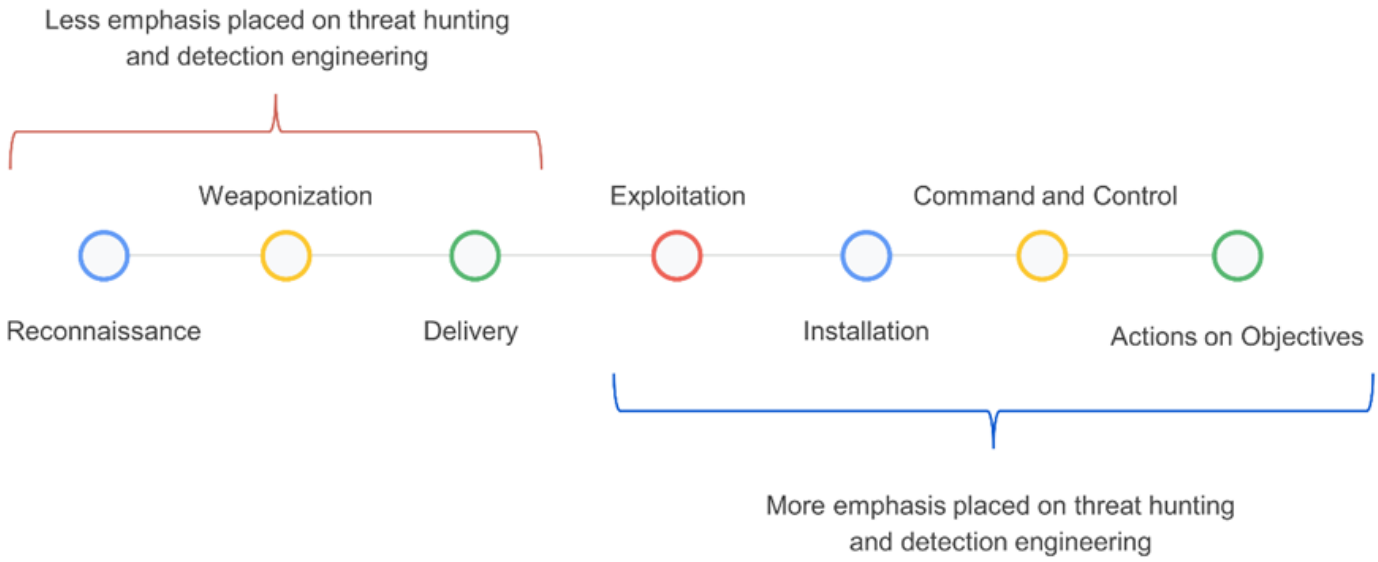


Figure 1: Stages of the kill chain categorized by their emphasis on threat hunting and detection engineering.

We have been exploring ideas to improve our hunting focused on samples built in the weaponization phase and distributed in the delivery phase, focused on the detection of suspicious Microsoft Office documents (Word, Excel, and PowerPoint), PDF files, and emails.

In threat intelligence platforms and cybersecurity in general, green and red colors are commonly used to quickly indicate results and identify whether or not something is malicious. This is because they are perceived as representing good or bad, respectively.

Multiple studies in psychology have demonstrated how colors can influence our decision-making process. VirusTotal, through the third-party engines integrated into it, shows users when something is detected

and therefore deemed "malicious," and when something is not detected and considered "benign." For example, the sample in Figure 2 belongs to a Microsoft Word document distributed by the SideWinder group during the year 2024.

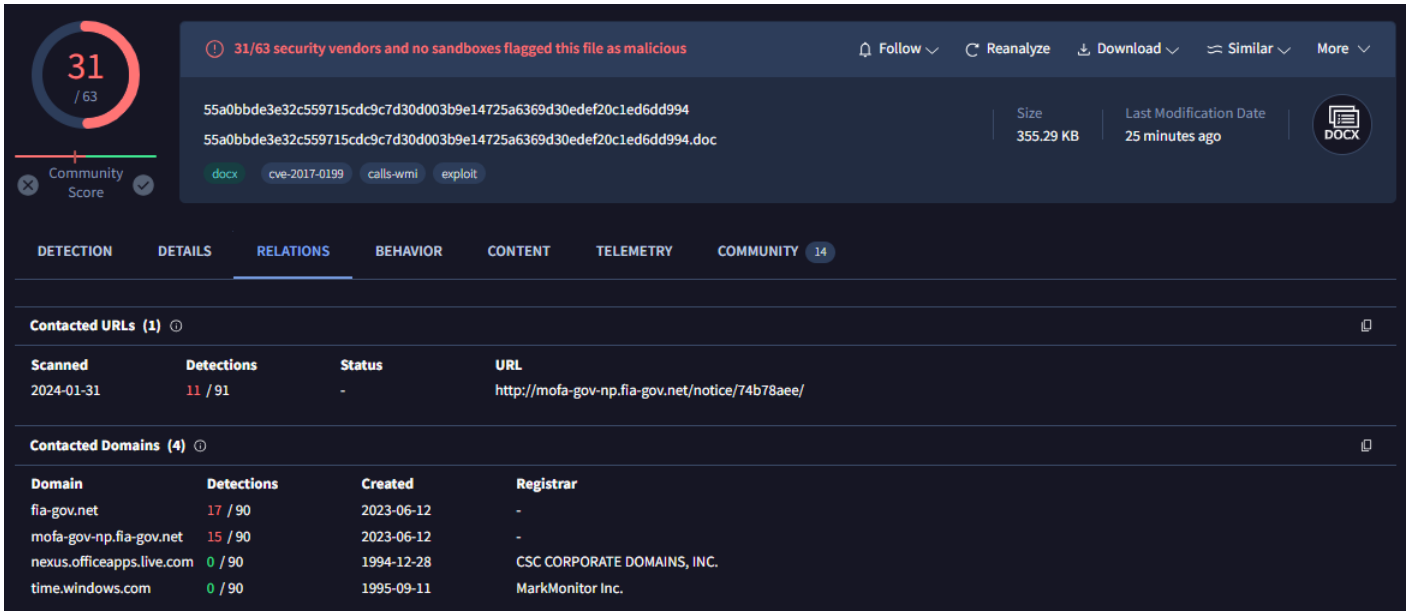


Figure 2: Document used by the SideWinder APT group

The sample in question was identified at the time of writing this post by 31 antivirus engines, leaving no doubt that it is indeed a real malware sample. In the process of pivoting to identify new samples or related infrastructure, starting with Figure 2, the analyst will likely click on the URL detected by 11 out of the 91 engines, and the domains detected by 17 and 15 engines, respectively, to see if there are other samples communicating with them. The remaining two domains (related to windows.com and live.com) in this case are easily identified as legitimate domains that were likely contacted by the sandbox during its execution.

Compressed Parents (1)

Scanned	Detections	Type	Name
2024-01-30	52 / 65	ZIP	fdeef34eae3d21f099a347716aa0869104704ff60150cdbd98aeb5ae11870f4a

Bundled Files (13)

Scanned	Detections	File type	Name
2024-02-17	8 / 60	XML	word/_rels/document.xml.rels
2024-04-23	0 / 61	XML	_rels/.rels
2024-02-12	0 / 60	XML	[Content_Types].xml
2024-02-12	0 / 60	XML	docProps/app.xml
2024-02-12	0 / 60	XML	docProps/core.xml
2024-02-12	0 / 60	XML	word/document.xml
2024-03-21	0 / 60	XML	word/fontTable.xml
2024-02-12	0 / 59	JPEG	word/media/image1.jpeg
2024-02-12	0 / 60	Text	word/media/image2.jpg
2024-02-12	0 / 60	XML	word/settings.xml
2024-04-16	0 / 59	XML	word/styles.xml
2024-04-19	0 / 59	XML	word/theme/theme1.xml
2024-04-23	0 / 61	XML	word/webSettings.xml

Dropped Files (3)

Scanned	Detections	File type	Name
2024-02-12	0 / 60	Text	word/media/image2.jpg
2024-01-30	4 / 60	Windows shortcut	C:\Users\user\AppData\Roaming\Microsoft\Office\Recent\design.LNK
?	?	file	2c7adcff8bde0f7325ec069c0274435c962e0b4f2c3a76e4bdaebaa6925a8388

Figure 3: Relationships within the SideWinder APT group document

In the same sample, if you go down in the VirusTotal report (Figure 3), the analyst will likely click on the ZIP file listed as "compressed parent" to check if there are other samples within this ZIP besides the current one. They may also click on the XML file detected by 8 engines, and the LNK file detected by 4 engines. The remaining files in the bundled files section probably won't be clicked, as the green color indicates they are not malicious, and also because they have less enticing formats — mainly XML and JPEG. But what if we explore them?

XML files generated by Microsoft Office

When you create a new Microsoft Office file, it automatically generates a series of embedded XML files containing information about the document. Additionally, if you use images in the document, they are also embedded within it. Microsoft Office files are compressed files (similar to ZIP files). In VirusTotal, when a Microsoft Word file is uploaded, you can see all these embedded files in the embedded files section.

We have mainly focused on three types of embedded files within Office documents:

- **Images:**Many threat actors use images related to the organizations or entities they intend to impersonate. They do this to make documents appear legitimate and gain the trust of their victims.

- **[Content_Types].xml:**This file specifies the content types and relationships within the Office Open XML (OOXML) document. It essentially defines the types of content and how they are organized within the file structure.
- **Styles.xml:**Stores stylistic definitions for your document. These styles provide consistent formatting instructions for fonts, paragraph spacing, colors, numbering, lists, and much more.

Our hypothesis is: If malicious Microsoft Word documents are copied and pasted during the weaponization building process, with only the content being modified, the hashes of the [Content_Types].xml and styles.xml files will likely remain the same.

Office documents

To check our hypothesis, we selected a set of samples used during delivery and belonging the threat actors listed in Figure 4:

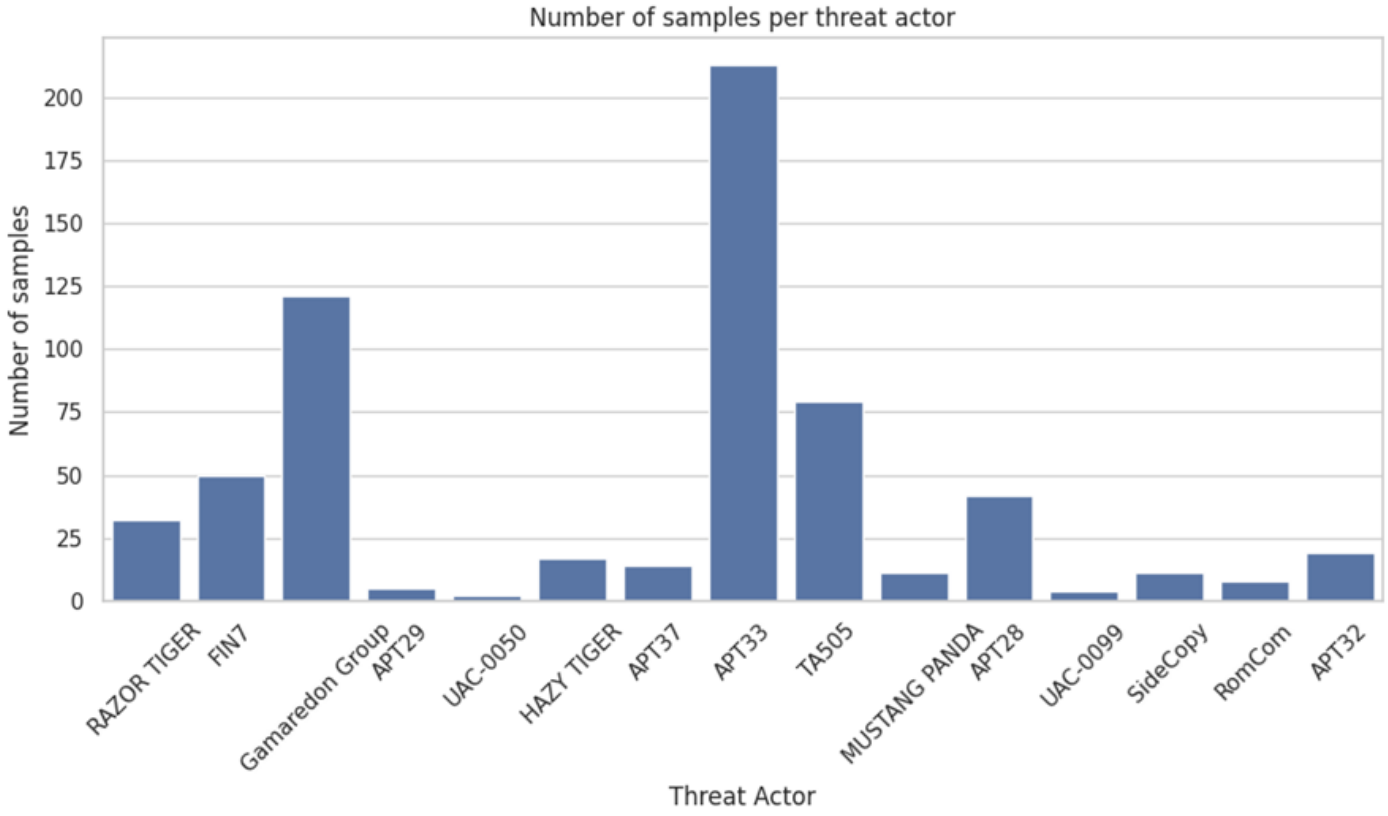


Figure 4: Number of samples per actor within the scope

Let’s analyze some of the results we obtained per actor.

APT28 – Images

We started by focusing on images APT28 has reused for different delivery samples (Figure 5).

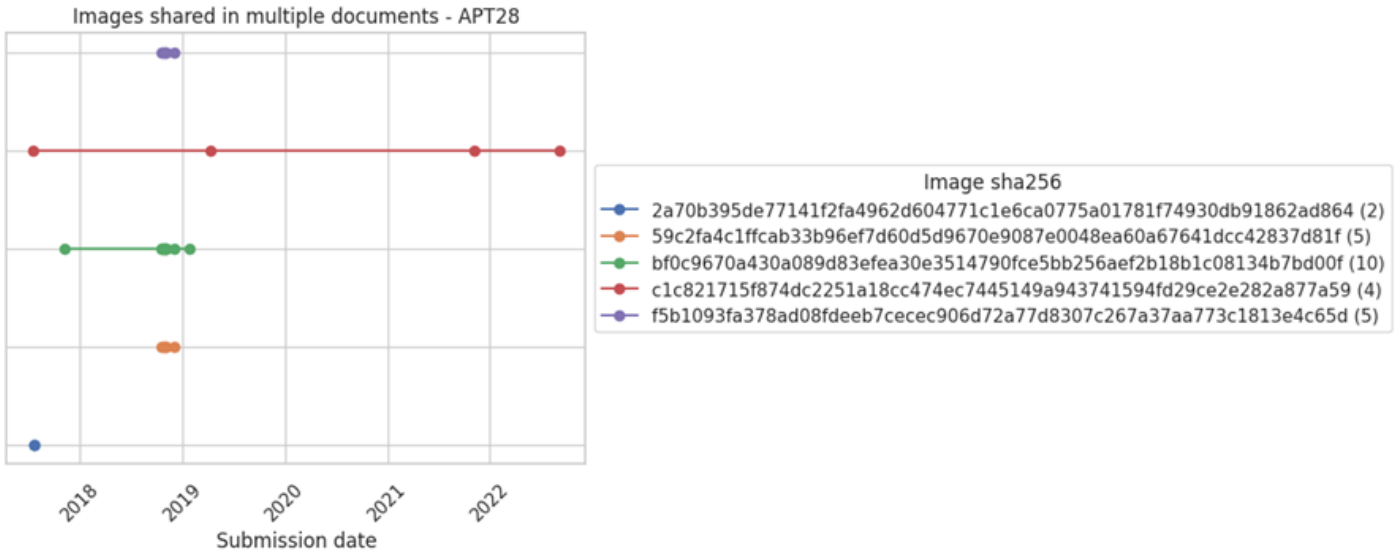


Figure 5: Images shared in multiple documents by APT28

Each line in the Figure 5 graph represents the same image, and each point represents at least two samples that used that particular image.

The second image of the graph shows how it was used by different Office documents at different points in time, from 2018 to 2022 (dates related to their upload to VirusTotal).

Now, the chart in Figure 6 visualizes each of these images.

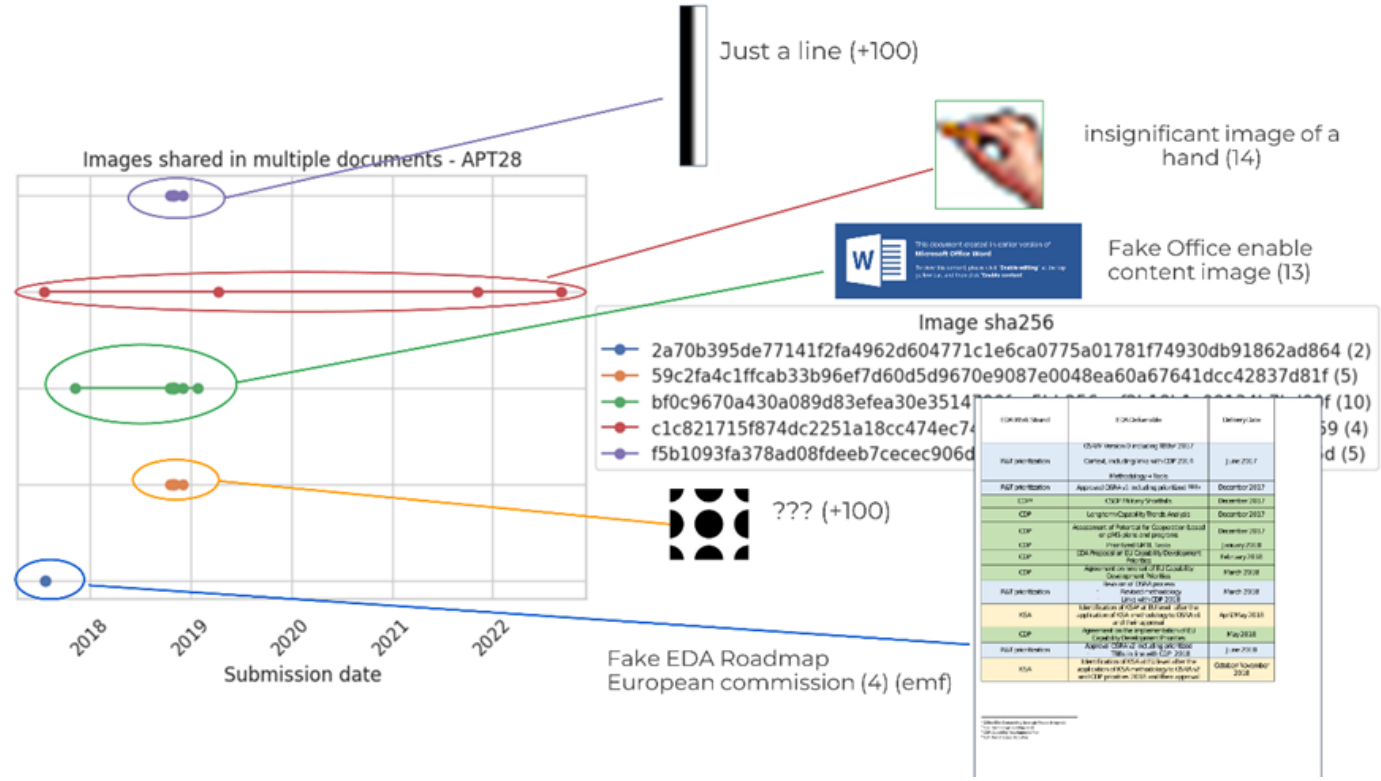
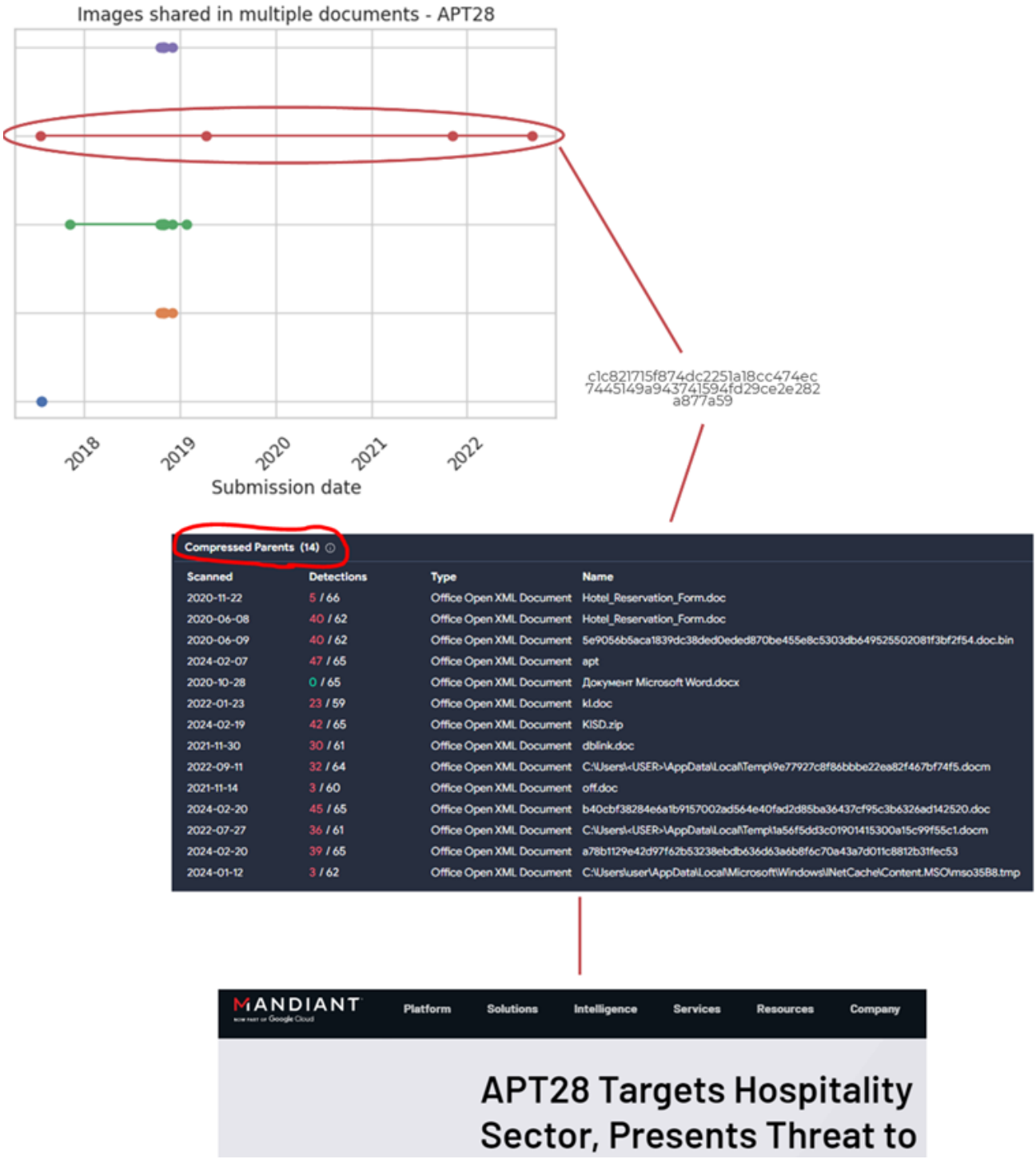


Figure 6: Content of the images shared in multiple documents by APT28

- The first image is just a simple line with no particular meaning. It's embedded in over 100 files known by VirusTotal.

- The second image is a hand and has 14 compressed parents.
- The third image consists of black circles and also has over 100 compressed parents.
- The last image is like a Word page with a table, presenting a fake EDA Roadmap of the European Commission. The image format is EMF (an old format) and it has 4 compressed parents

If we delve into the compressed parents of the second image (the one with the hand), we can see how the image is used in Office documents that are part of a campaign reported by Mandiant attributed to APT28. The image of the hand was used in fake Word documents for hotel reservations, particularly in a small section where the client was supposed to sign.



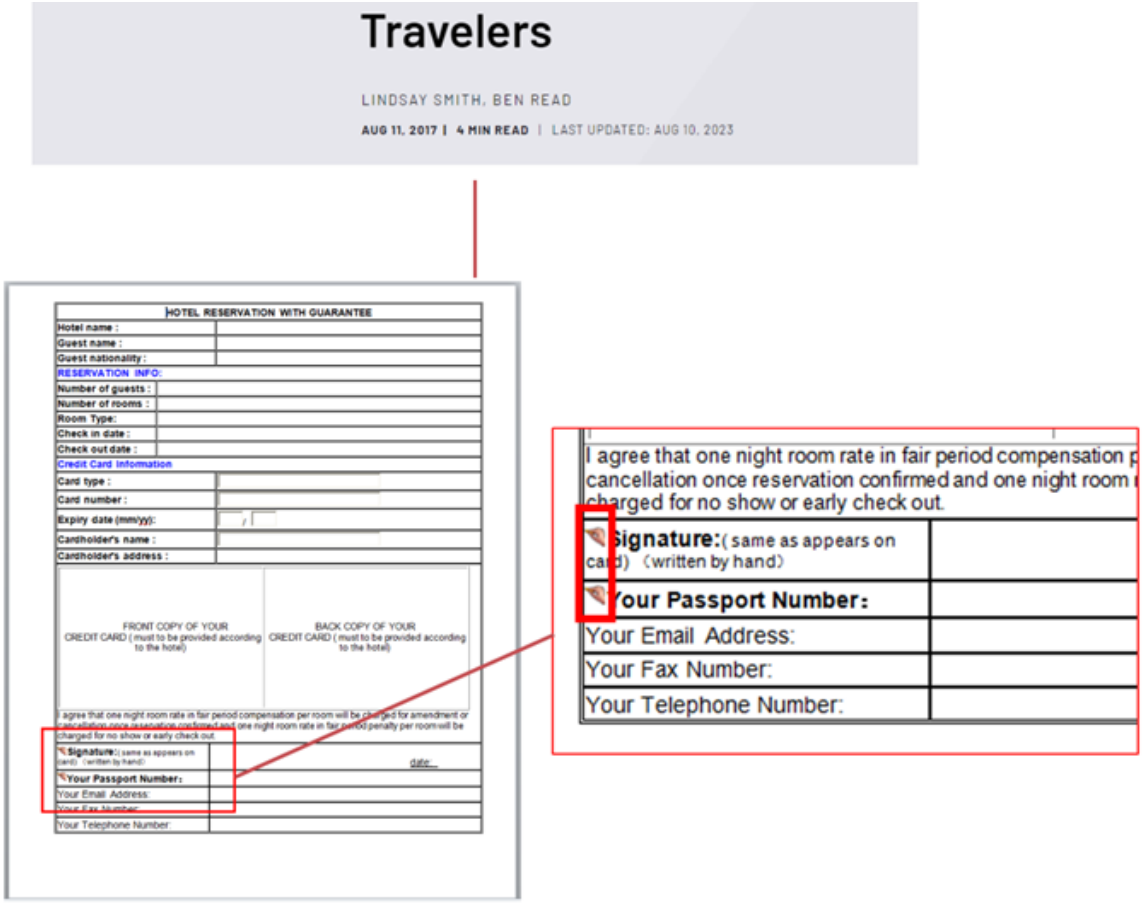


Figure 7: Pivoting through a specific image used by APT28

SideWinder – Images

SideWinder (aka RAZER TIGER) is a group focused on carrying out operations against military targets in Pakistan. This group traditionally reused images, which might help monitoring their activity.

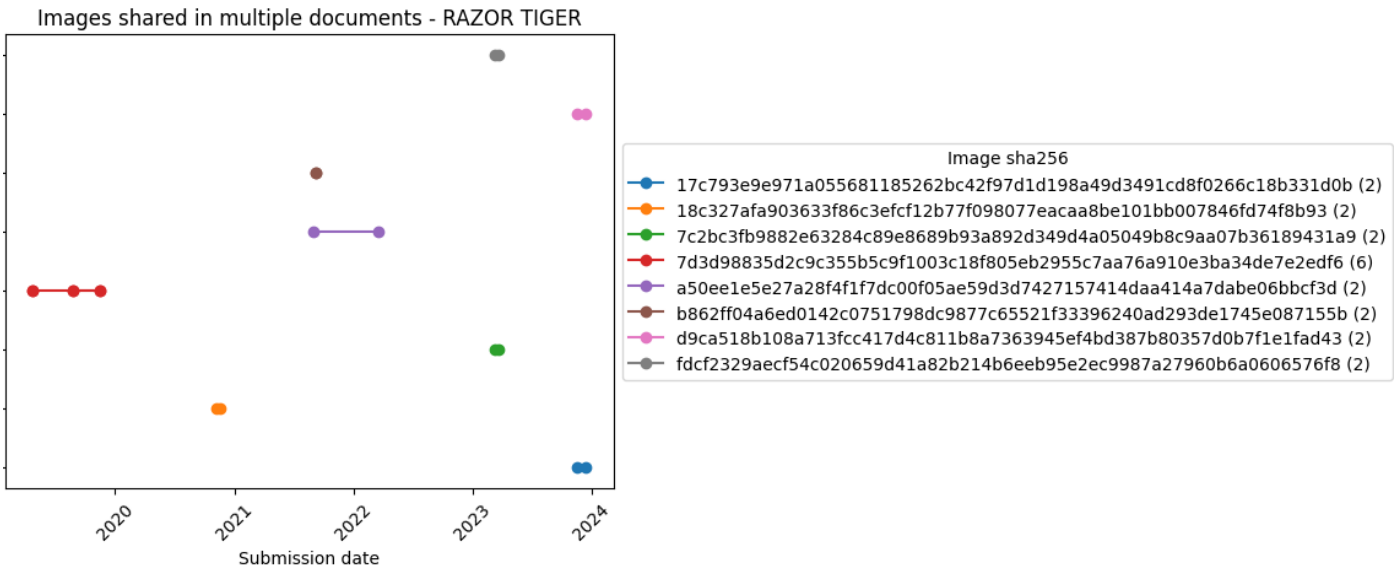


Figure 8: Images shared in multiple documents by RAZOR TIGER

In particular, the image in Figure 9 was used in a sample uploaded in September 2021 and in a second one uploaded March 2022. The image in question is the signature of Baber Bilal Haider.

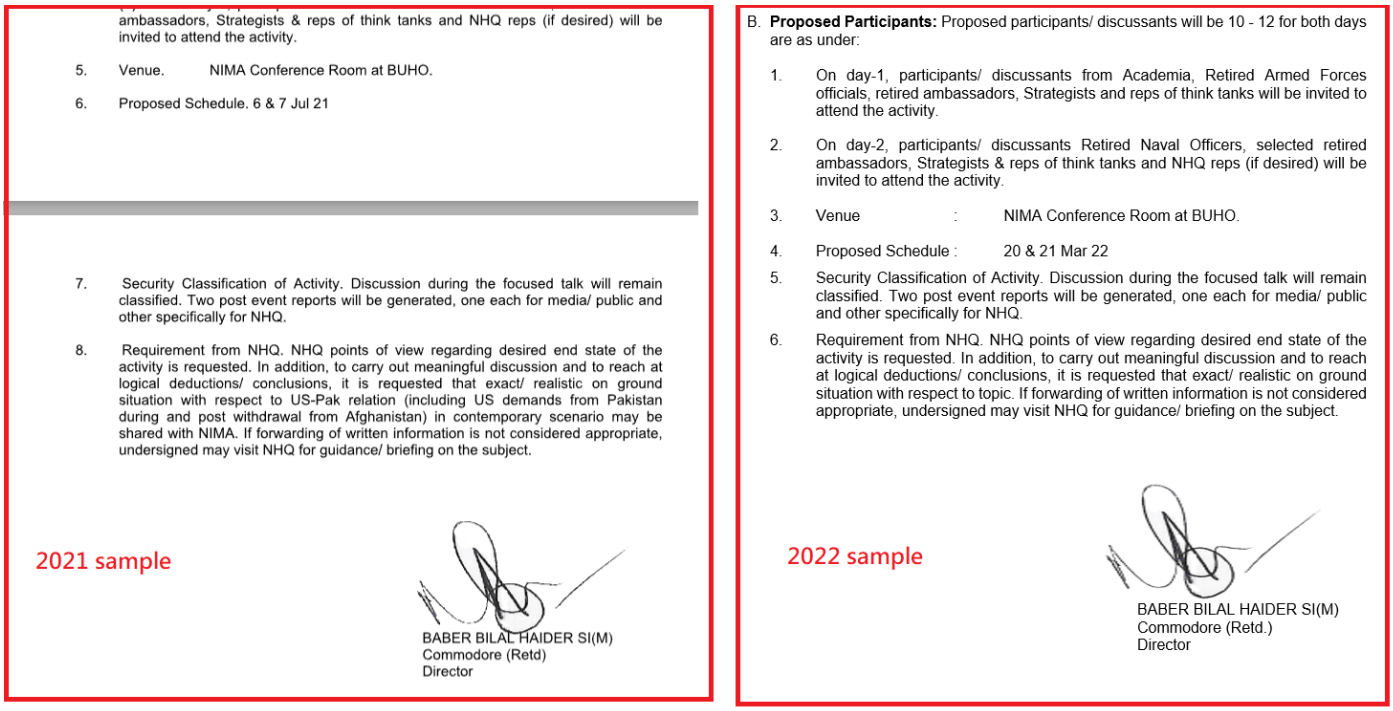


Figure 9: Two different samples of RAZOR TIGER share the same image of a handwritten signature

Gamaredon – [Content_Types].xml and styles.xml

For Gamaredon we found they reused styles.xml and [Content_Types].xml in different documents, which helped reveal new samples.

Figure 10 chart displays all the [Content_Types].xml files from Gamaredon's Office documents.

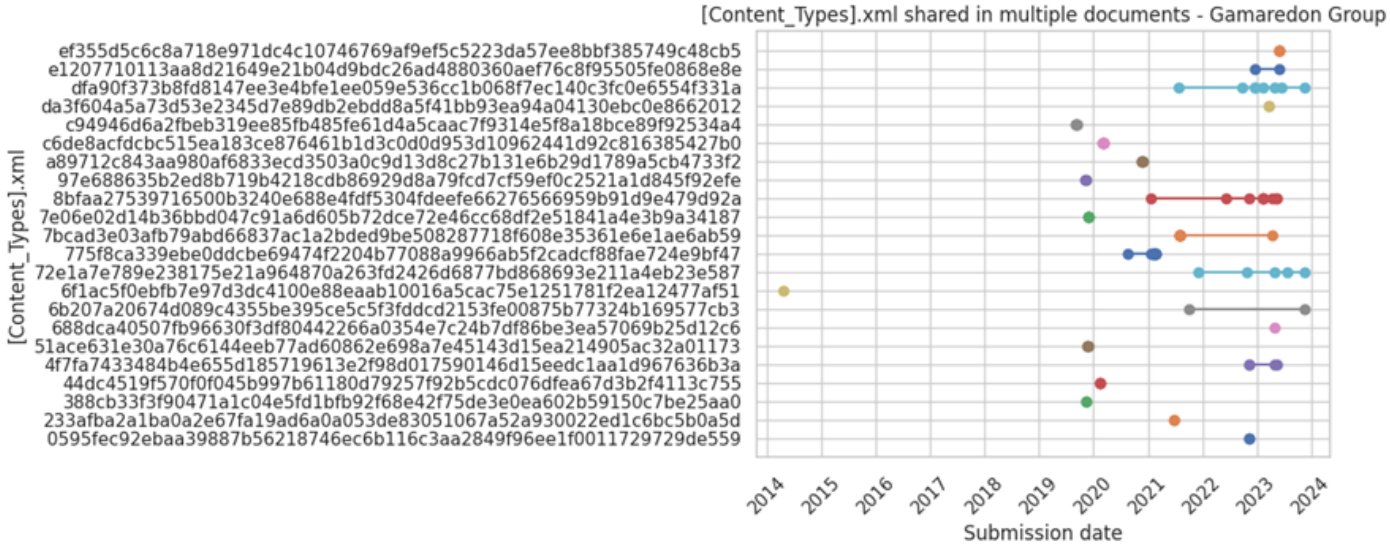


Figure 10: [Content_Types].xml shared in multiple documents by Gamaredon Group

There are a large number of samples that share the same [Content_Types].xml. It's important to highlight that these [Content_Types].xml files are not necessarily exclusively used by Gamaredon, and can be found in other legitimate files created by users worldwide. However, some of these [Content_Types].xml might be interesting to monitor.

Styles.xml files are usually less generic, which should make them a better candidate to monitor:

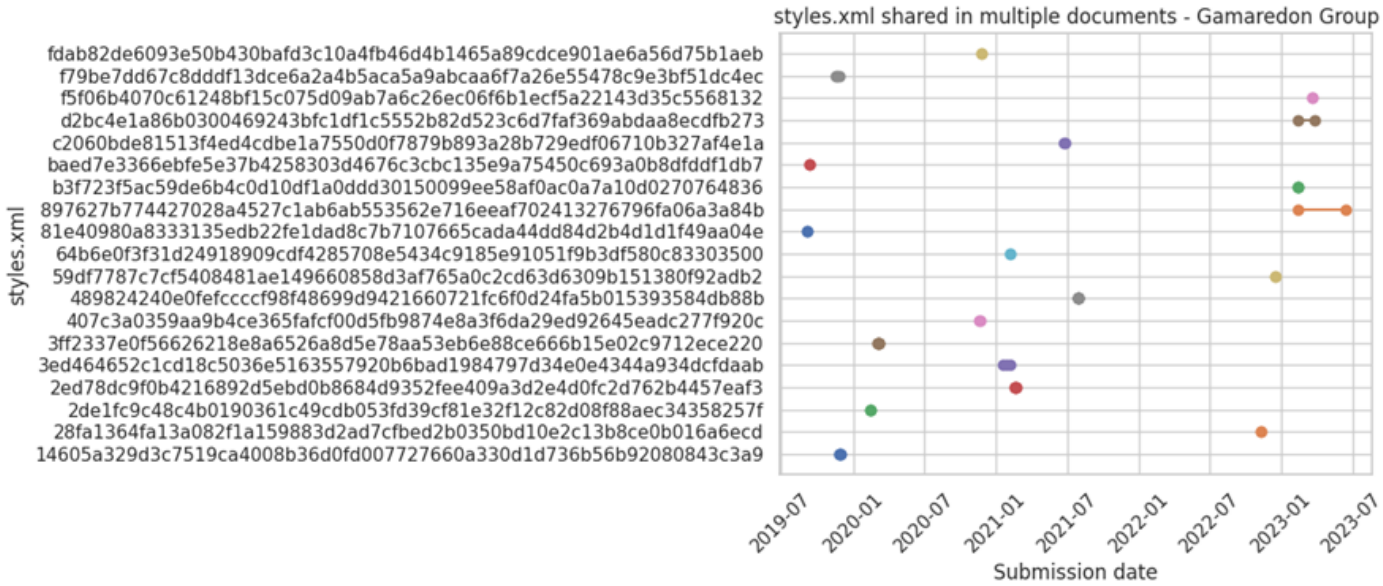


Figure 11: Styles.xml shared in multiple documents by Gamaredon Group

We see styles.xml files are less reused than [Content_Types].xml. This could be because some of the samples used by this actor for distribution are created from scratch or reusing legitimate documents.

We used identified patterns in the styles.xml files to launch a retrohunt on VirusTotal. Figure 12 visually represents the original set of style.xml files (left) and those that were added later after running the retrohunt (right).

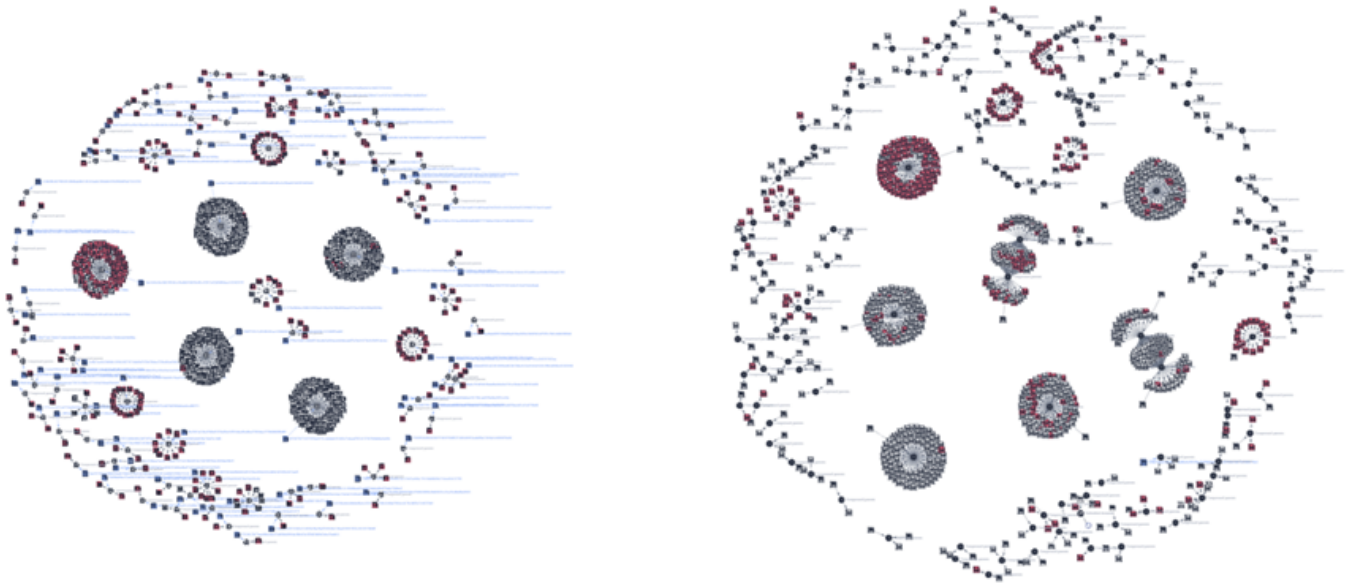


Figure 12: Initial graph of the styles.xml and its parents used by Gamaredon (left). Final graph after identifying new styles.xml and their parents using retrohunt in VirusTotal (right)

One of the new styles.xml files found in our retrohunt has 17 compressed parents, meaning it was included in 17 Office files.

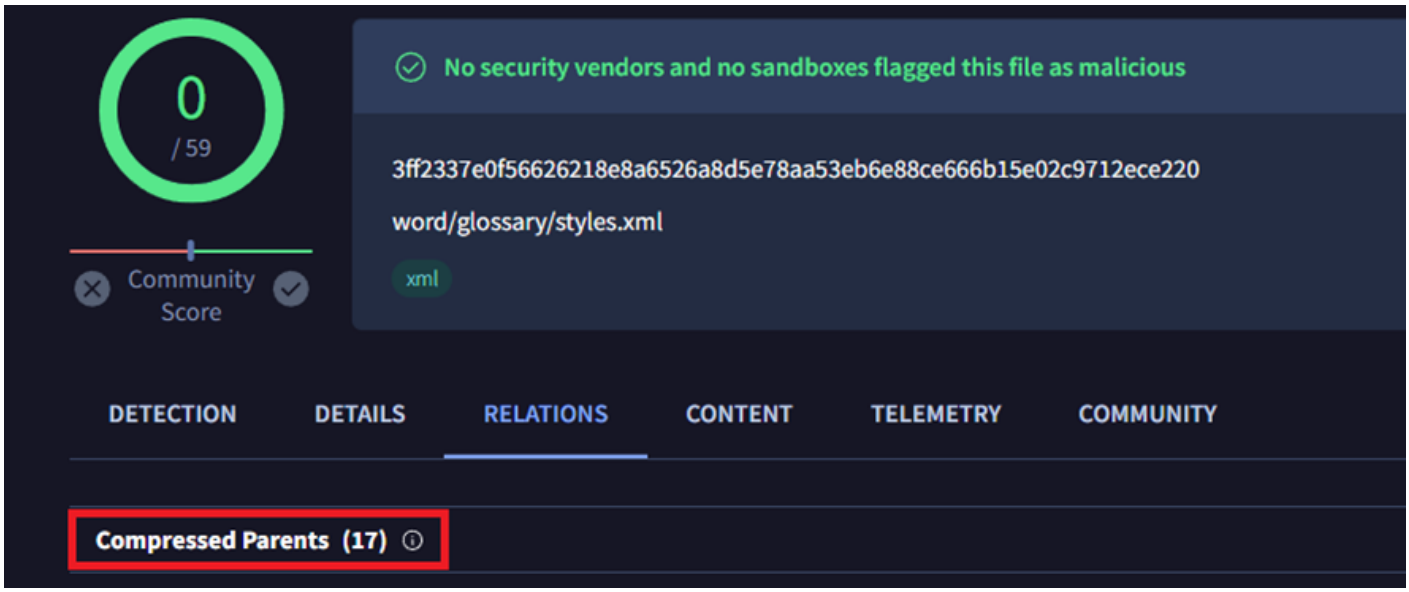
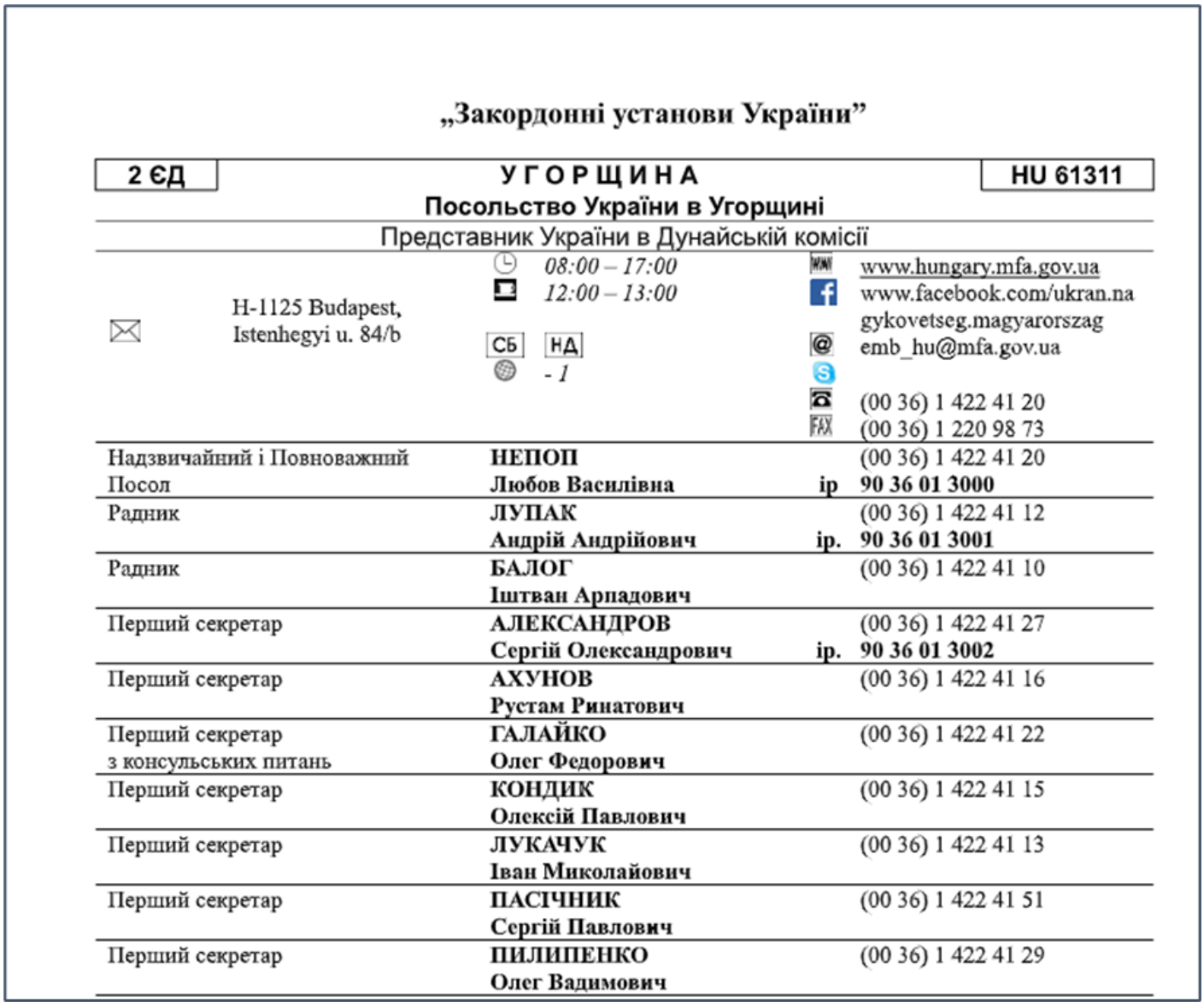


Figure 13: Number of parent documents for a specific styles.xml file used by Gamaredon

All the parents were malicious, some of them identical and the rest very similar between them. The content of many of them referred to "Foreign institutions of Ukraine - Embassy of Ukraine in Hungary," containing a table with phone numbers and information about the embassy, such as social media links and email accounts. Here's an example:



„Закордонні установи України”

2 ЄД	УГОРЩИНА Посольство України в Угорщині Представник України в Дунайській комісії	HU 61311
H-1125 Budapest, Istenhegyi u. 84/b	08:00 – 17:00 12:00 – 13:00 СБ НД - 1	www.hungary.mfa.gov.ua www.facebook.com/ukran.na gykovetseg.magyarorszag emb_hu@mfa.gov.ua (00 36) 1 422 41 20 (00 36) 1 220 98 73
Надзвичайний і Повноважний Посол	**НЕПОП** **Любов Василівна**	(00 36) 1 422 41 20 ip **90 36 01 3000**
Радник	**ЛУПАК** **Андрій Андрійович**	(00 36) 1 422 41 12 ip. **90 36 01 3001**
Радник	**БАЛОГ** **Іштван Арпадович**	(00 36) 1 422 41 10
Перший секретар	**АЛЕКСАНДРОВ** **Сергій Олександрович**	(00 36) 1 422 41 27 ip. **90 36 01 3002**
Перший секретар	**АХУНОВ** **Рустам Ринатович**	(00 36) 1 422 41 16
Перший секретар з консульських питань	**ГАЛАЙКО** **Олег Федорович**	(00 36) 1 422 41 22
Перший секретар	**КОНДИК** **Олексій Павлович**	(00 36) 1 422 41 15
Перший секретар	**ЛУКАЧУК** **Іван Миколайович**	(00 36) 1 422 41 13
Перший секретар	**ПАСІЧНИК** **Сергій Павлович**	(00 36) 1 422 41 51
Перший секретар	**ПИЛИПЕНКО** **Олег Вадимович**	(00 36) 1 422 41 29

Figure 14: Document used by Gamaredon in one of its campaigns that includes multiple images which can be used to monitor new samples

The information for social media includes the logos of these platforms, such as the Facebook logo, Skype logo, an image of a telephone, etc. By pivoting, on the image of the Facebook icon, we find that it has 12 additional compressed parents, meaning it appears in 12 documents, all of them sharing the same styles.xml file.

Visualizing all together, we find a set of about 12-14 images used within the same timeframe by the actor. All of these images can be found in the “Embassy of Ukraine in Hungary” document.

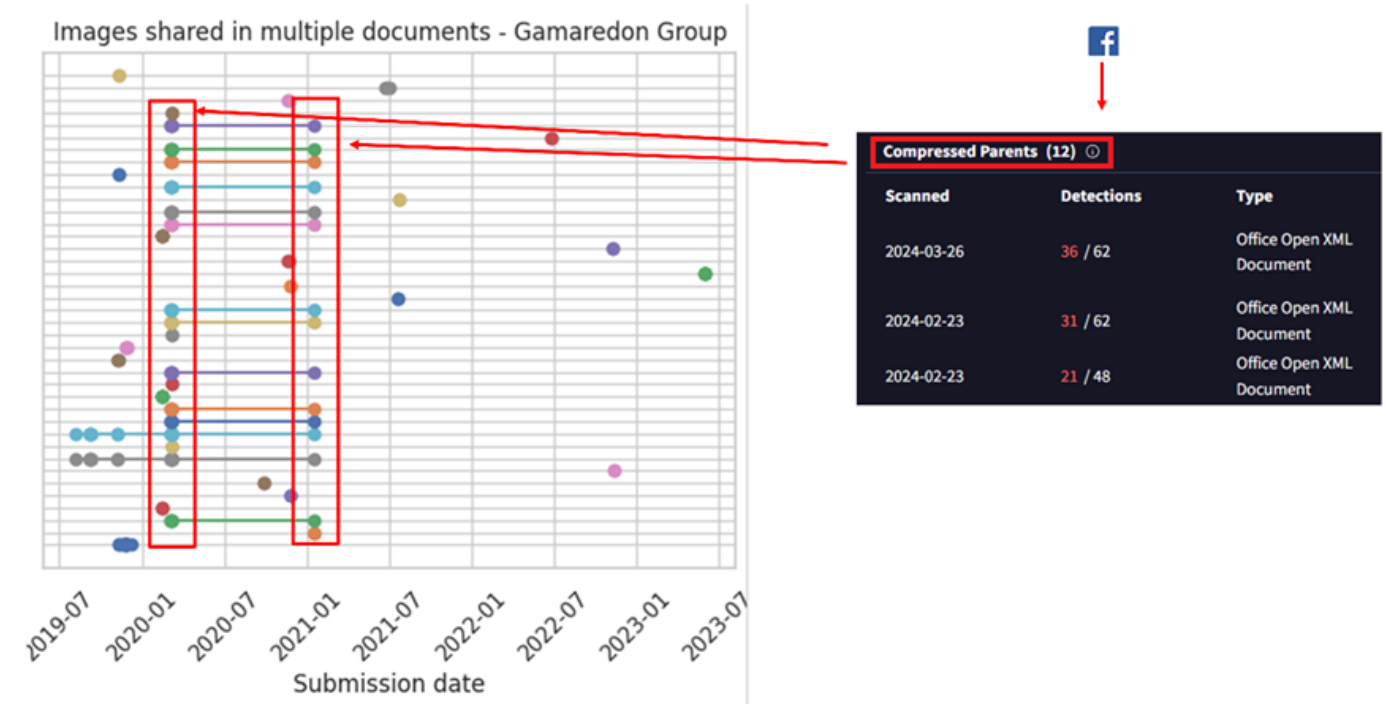


Figure 15: Pivoting through the Facebook image that included the document in Figure 14

There's a pattern evident in the previous image where different images were included in files uploaded simultaneously. This pattern is associated with multiple documents used in the same campaign of the Embassy of Ukraine in Hungary, all of them were using the same social media images explained before.

Styles.xml shared between threat actors

Another aspect we explored was if different threat actors shared similar styles.xml files in their documents. Styles.xml files are somewhat more specific and unique than [Content_Types].xml files because they can contain styles created by threat actors or by legitimate entities that originally created the document and then were modified by the actor. This makes them stand out more and can help in identifying threat actor activity.

This doesn't necessarily imply they share information to conduct separate operations, although in some cases, it could be a scenario worth considering.

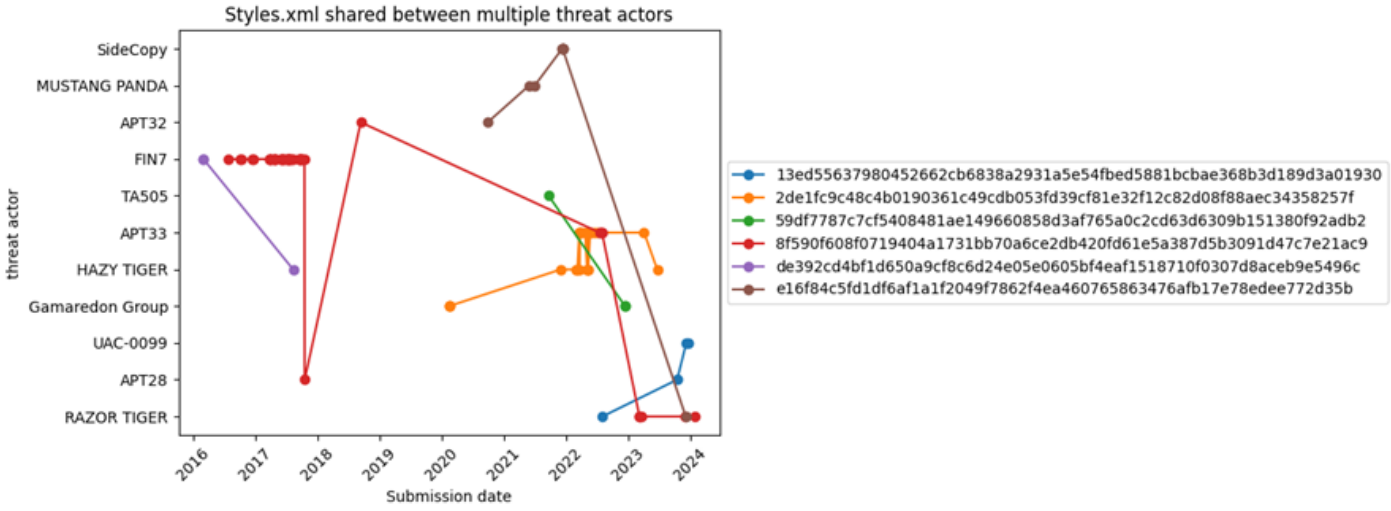


Figure 16: styles.xml shared between different threat actors

Of all styles.xml files related to actors in our initial set, only six of them were found to be shared by at least two actors. Some styles defined by the styles.xml file are very generic and could identify almost any type of file. However, there are others that could be interesting to explore further.

An interesting case is the Styles.xml file, which seems to be shared by Razor Tiger, APT28, and UAC-0099. Specifically, the samples from APT28 and UAC-0099 are attract because they were uploaded to VirusTotal within short time frames, suggesting they might belong to the same threat actor.
The file 243bab79863327915c315c188c0589202f64b3500a3fee3e2c9f3d34e8e1f154 attributed to APT28, the file 2c2fa6b9fbb6aa270ba0f49ebb361ebf7d36258e1bdfd825bc2faeb738c487ed attributed to UAC-0099, and the file 61a5b971a6b5f9c2b5e9a860c996569da30369ac67108d4b8a71f58311a6e1f1 attributed to UAC-0099 all share the same styles.xml and [Content_Types].xml.

You can see the list of hashes in the appendix of this blog

[Content_Types].xml shared between threat actors

Like in the previous case, we checked if there were Office documents among different threat actors sharing [Content_Types].xml:

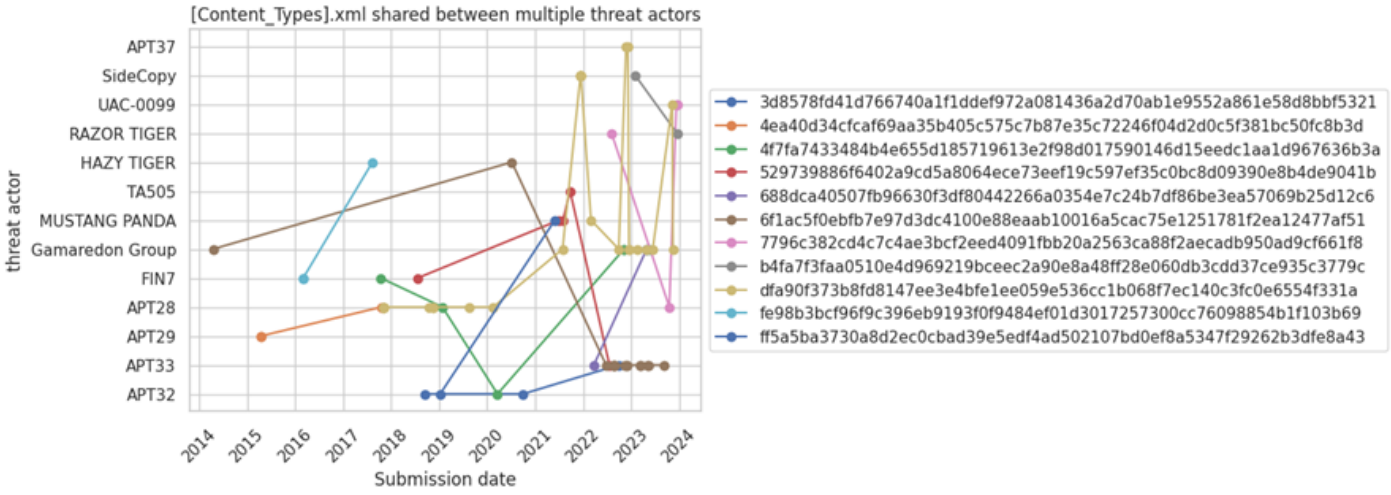


Figure 17: [Content_Types].xml shared between different threat actors

In this case, there are eleven [Content_Types].xml files that are shared by at least two different actors.

An interesting case here is the file dfa90f373b8fd8147ee3e4bfe1ee059e536cc1b068f7ec140c3fc0e6554f331a, which is shared by Gamaredon, APT37, Mustang Panda, APT28, SideCopy, and UAC-0099. Again, there could be different explanations for this.
Another interesting case that is worth analyzing in detail is [Content_Types].xml with hash 4ea40d34cfcaf69aa35b405c575c7b87e35c72246f04d2d0c5f381bc50fc8b3d, which is only shared by APT28 and APT29.

You can see the list of hashes in the appendix of this blog

AI to the rescue

The images reused by attackers seem to be a promising idea we decided to further explore.

We used the VirusTotal API to download and unzip a set of Office documents used for delivery, this way we obtained all the images. Then we used Gemini to automatically describe what these images were about.

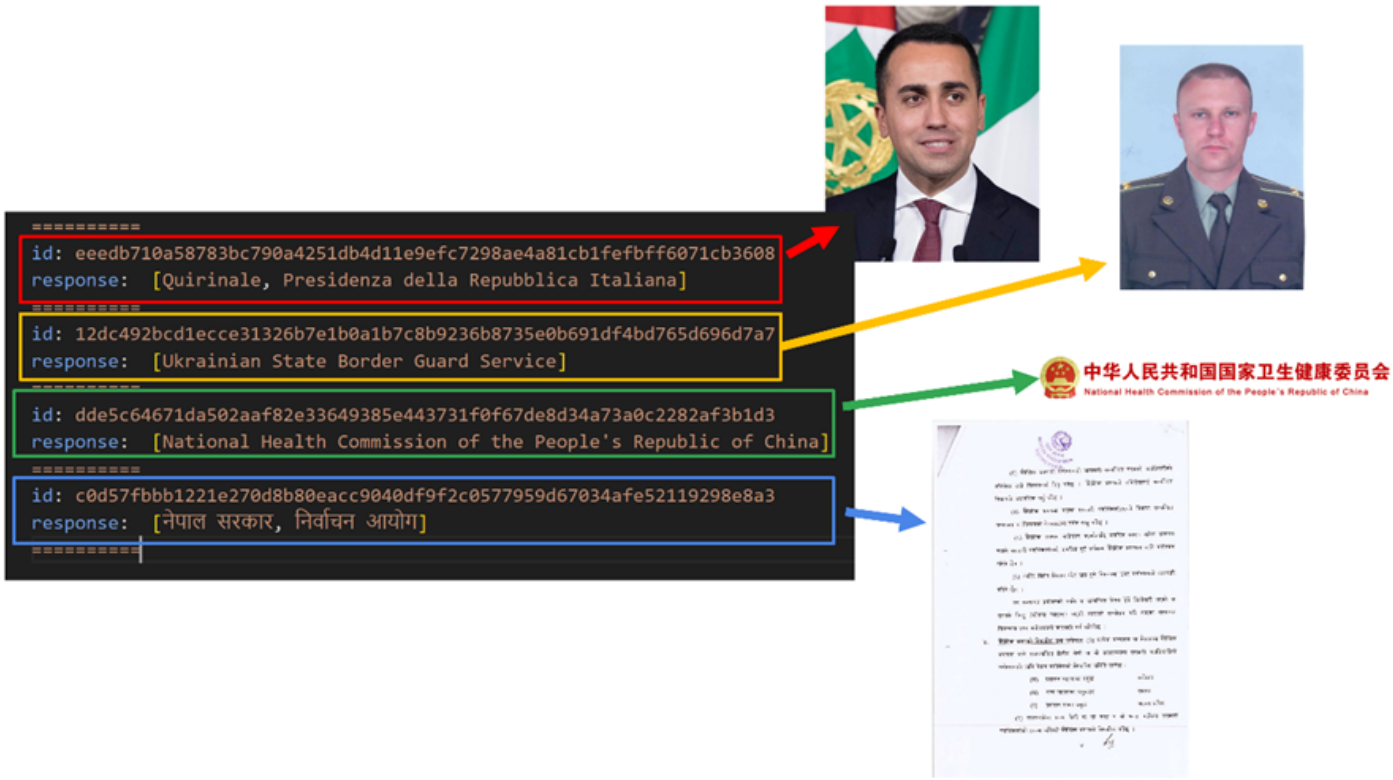


Figure 18: Results obtained with Gemini after processing some of the embedded images in the documents used by the threat actors

Figure 18 shows some examples of images that were incorporated by certain actors. There were also other results that were not helpful, mainly related to images that did not show a logo or anything specific that indicated what they were.

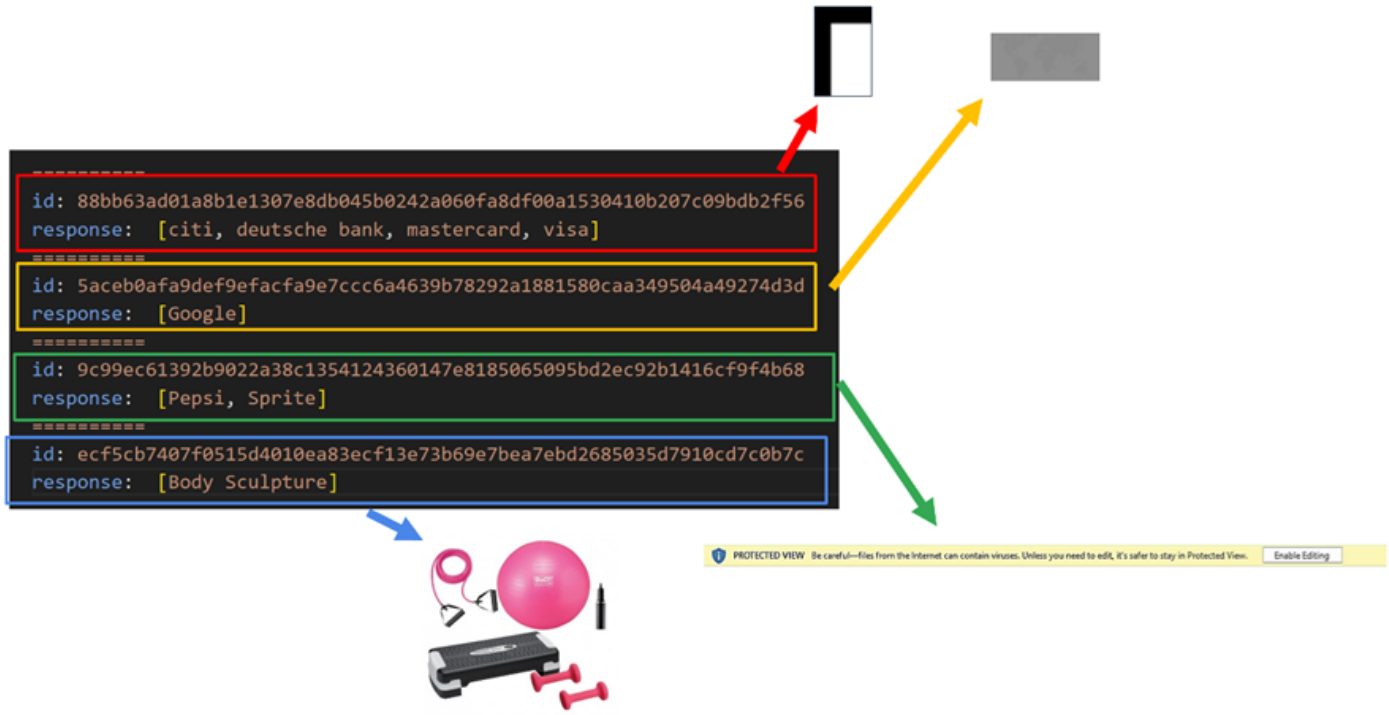


Figure 19: Results obtained with Gemini after processing some of the embedded images in the documents used by the threat actors

Using the VirusTotal API to obtain documents that you might be looking for and combining the results with Gemini to analyze possible images automatically, can potentially help analysts to monitor potential suspicious documents and create your own database of samples using specific images, for example Government images or specific images about companies. This approach is interesting not only for threat hunting but also for brand monitoring.

PDF Documents

Images dropped by Acrobat Reader

Unlike Office documents, PDF files don't contain embedded XML files or images, although some PDF files may be created from Office documents. Some of our sandboxes include Adobe Acrobat Reader to open PDF documents which generates a thumbnail of the first page in BMP format. This image is stored in the directory C:\Users\\AppData\LocalLow\Adobe\Acrobat\DC\ConnectorIcons. Consequently, our sandboxes provide this BMP image as a dropped file from the PDF, allowing us to pivot.

To illustrate this functionality, see Figure 20 attributed to Blind Eagle, a cybercrime actor associated with Latin America.

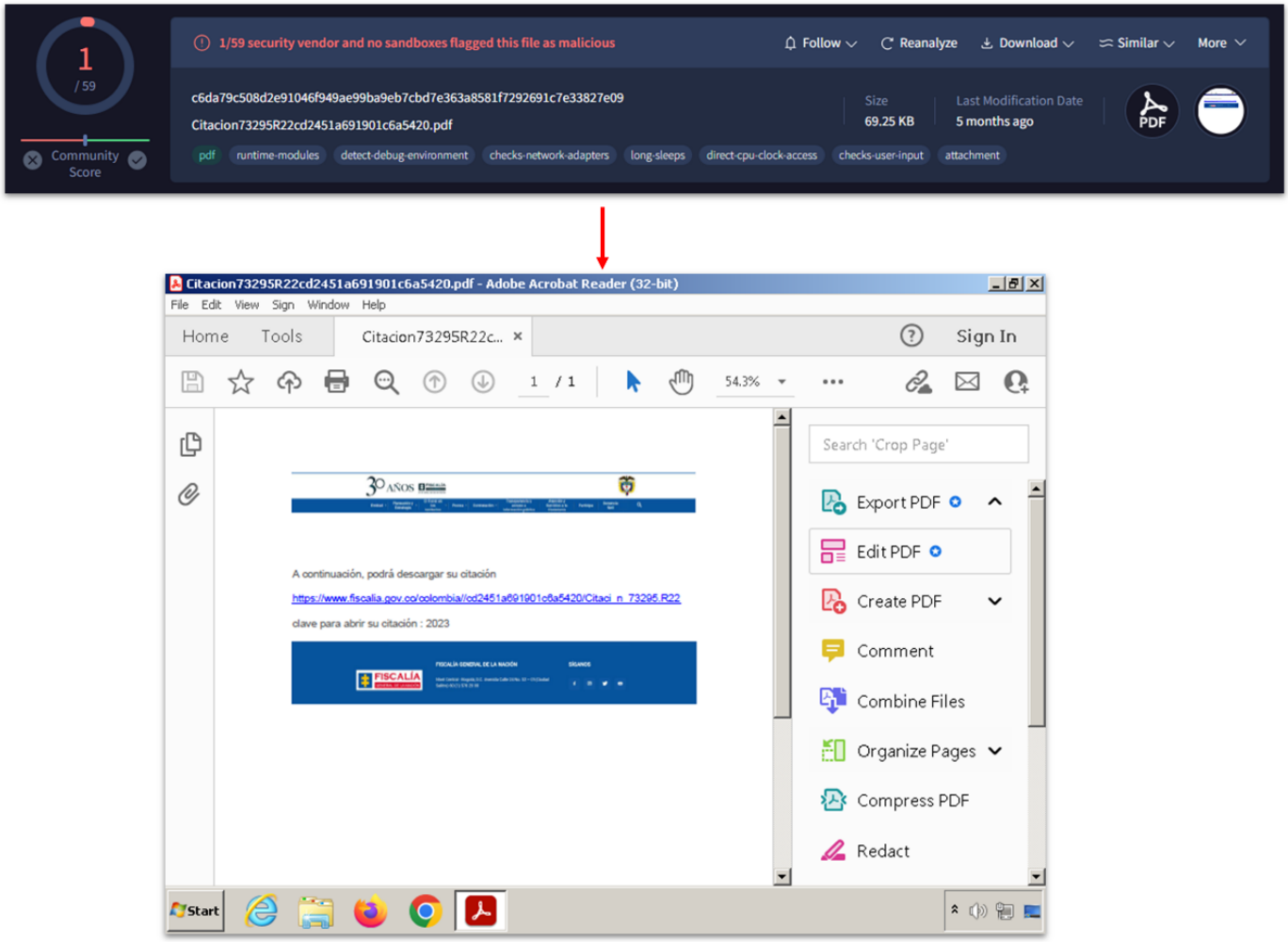


Figure 20: Content of a PDF file related to Blind Eagle threat actor

Figure 20 was provided by our sandbox. In the "relations" tab, we can see the BMP image as a dropped file:

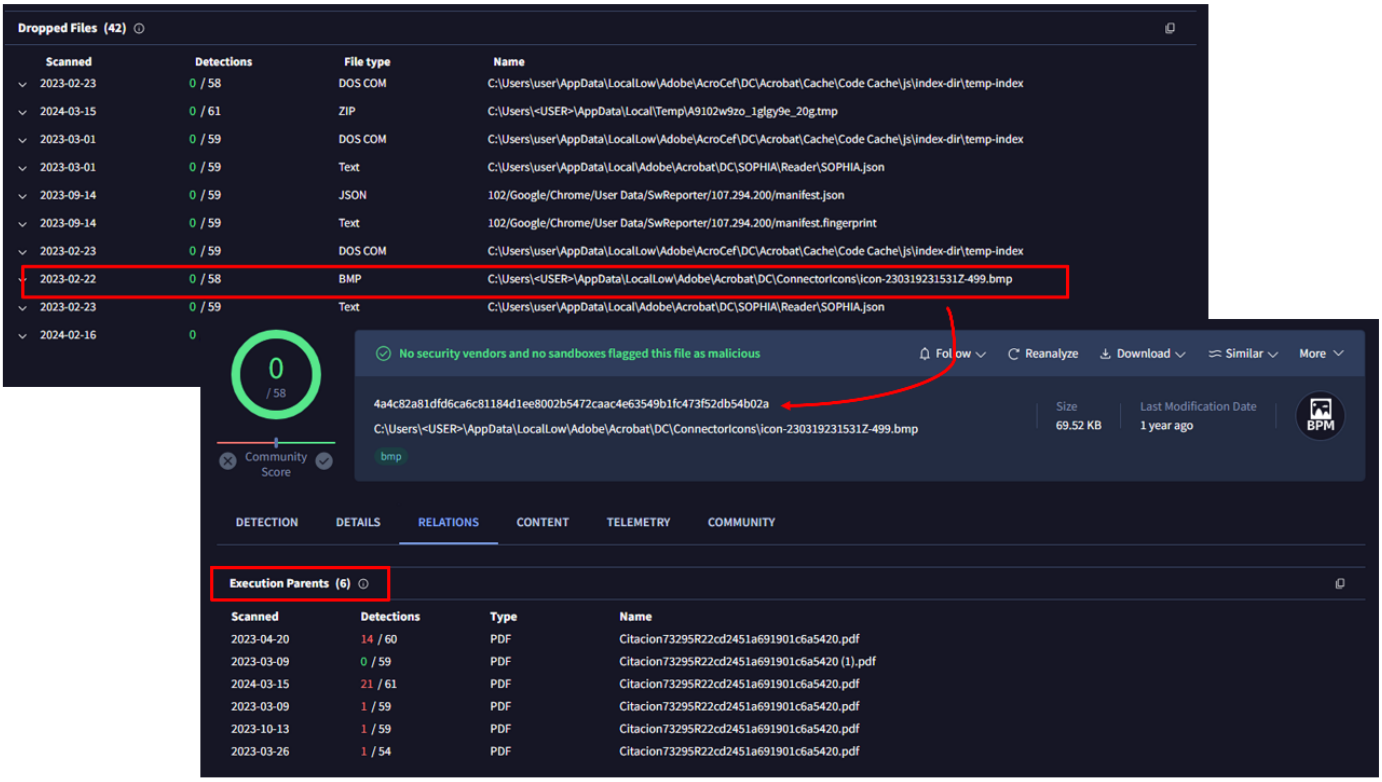


Figure 21: BMP file generated by the sandbox that can be used for pivoting

The BMP file itself also shows relations, in particular up to 6 PDF files in the "execution parents" section. In other words, there are other PDFs that look exactly the same as the initial one.

Typically, many actors engaged in financial crime activities utilize widely spread PDF files to deceive their victims, making this approach highly valuable. Another interesting example we found involves phishing activities targeting a Russian bank called "Tinkoff Bank."

The PDF files urge victims to accept an invitation from this bank to participate in a project.

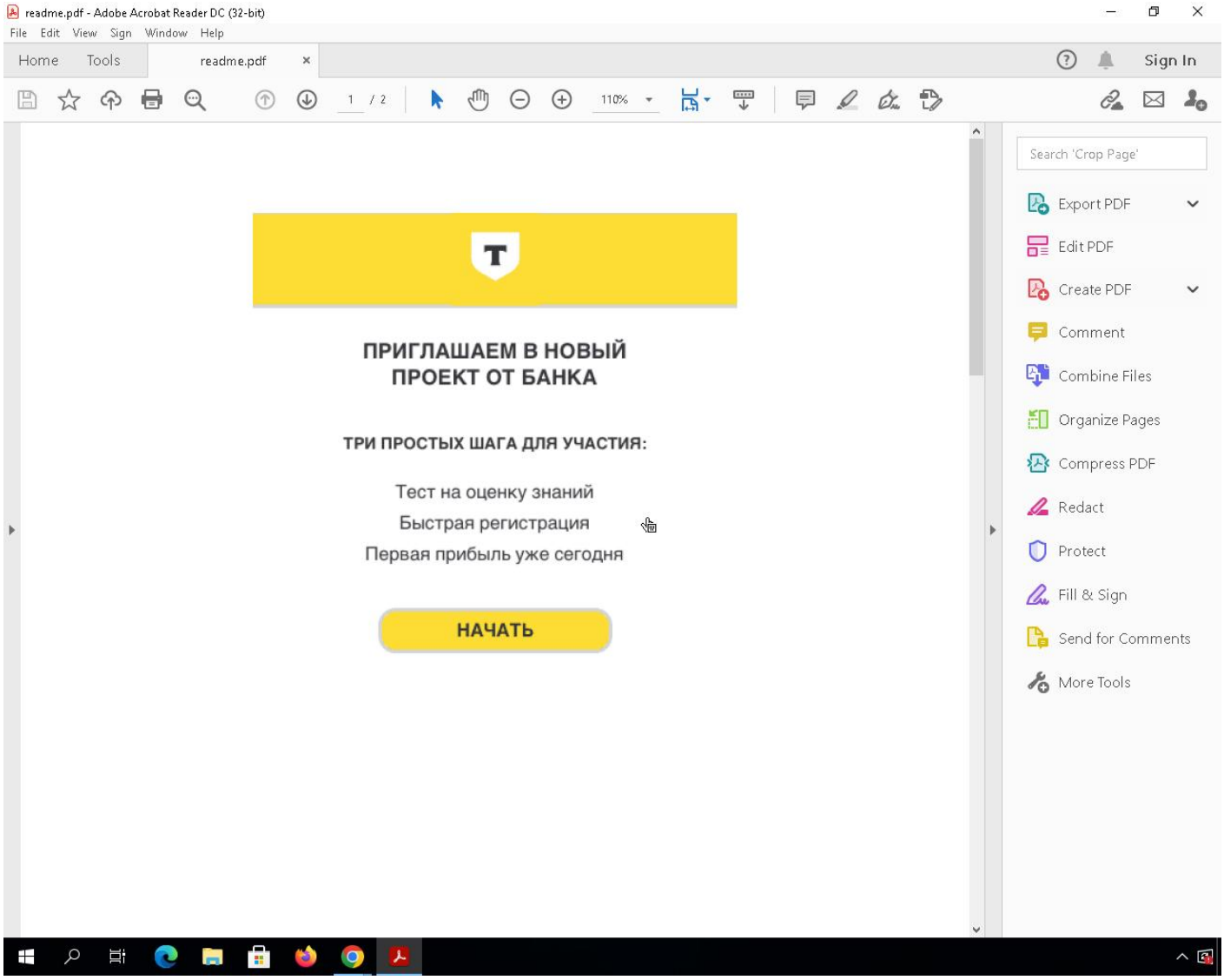


Figure 22: The content of a PDF file used by cybercrime actors

Applying the same approach we identified 20 files with identical content, most of them classified as malicious by AV engines.

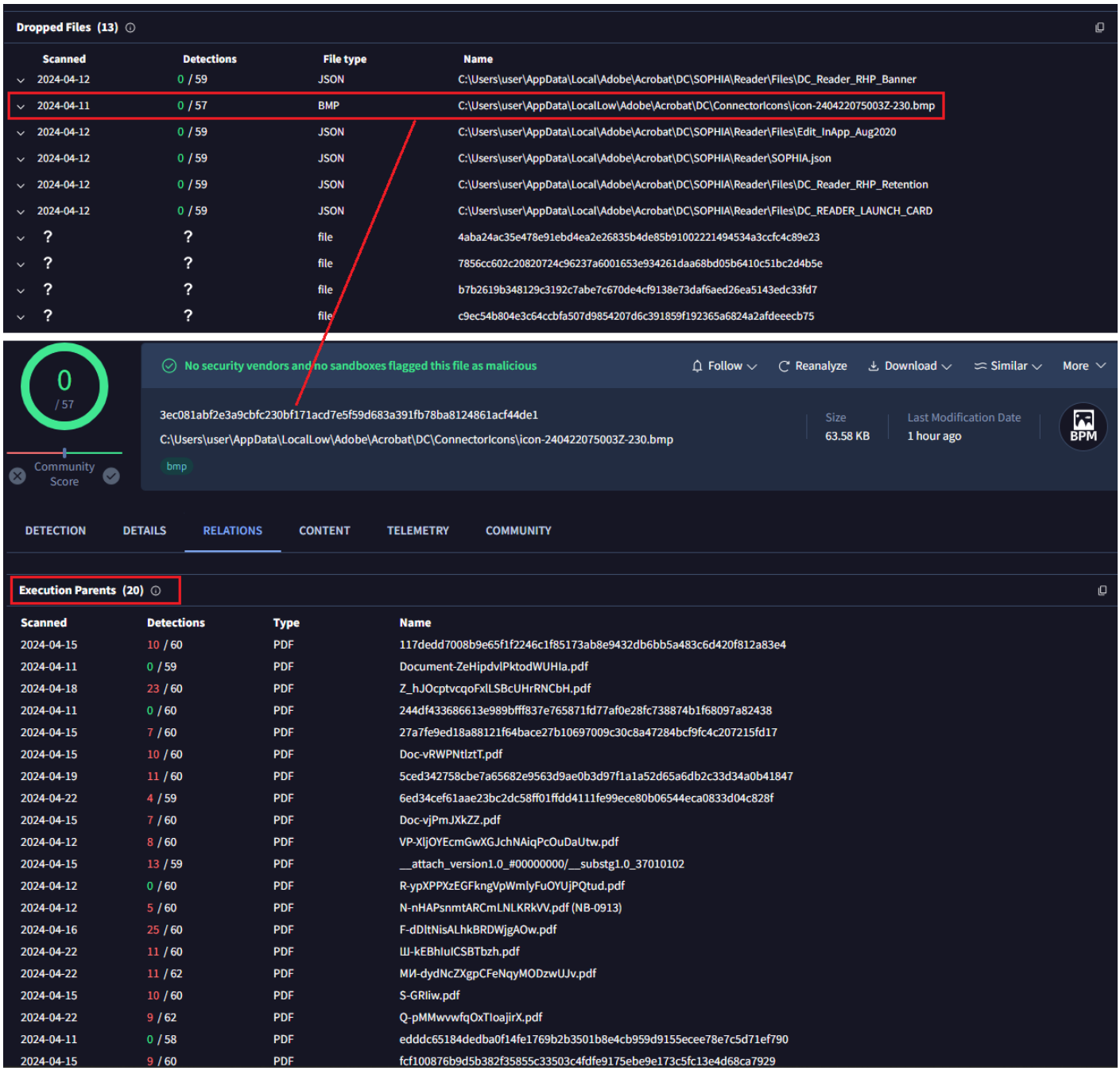


Figure 23: BMP file generated by the sandbox that can be used for pivoting, in this case having other 20 PDF with the same image

There are some limitations to this approach. For instance, the PDF file might be slightly modified (font size, some letter/word, color, …) which would generate a completely different hash value for the thumbnail we use to pivot.

Images dropped by Acrobat Reader

Just like the BMP files generated by Acrobat Reader, there are other interesting files that might be dropped during sandbox detonation. These artifacts can be useful on some occasions.

The first example is a JavaScript file dropped in another PDF attributed to Blind Eagle.

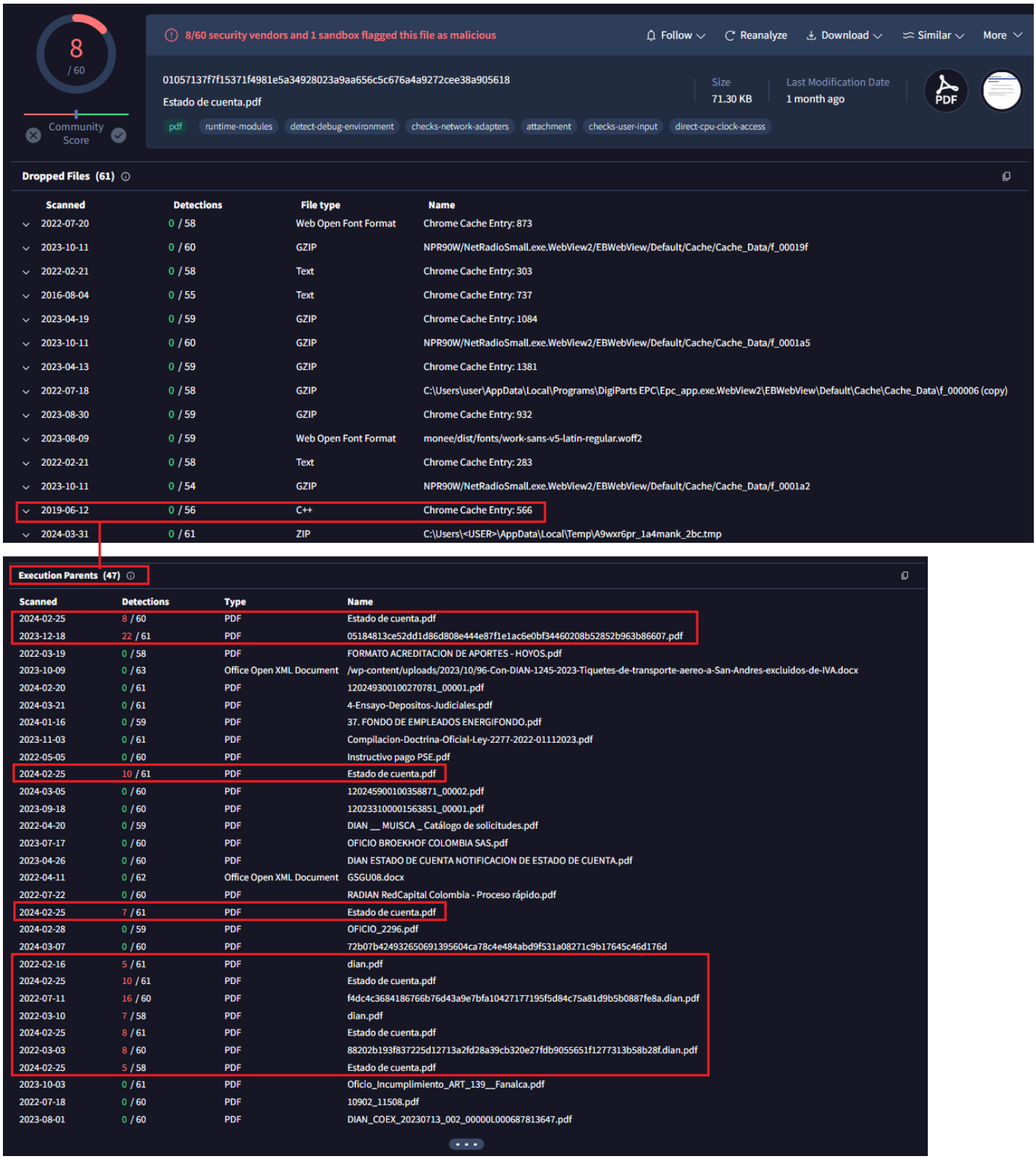


Figure 24: BMP file generated by the sandbox that can be used for pivoting, another example of Blind Eagle threat actor

The dropped JavaScript file's name during the PDF execution was "Chrome Cache Entry: 566" indicating that this file was likely generated by opening an URL through Chrome, possibly triggered by a sandbox click on a link within the PDF. Examining the file's contents, we observe some strings and variables in Spanish.

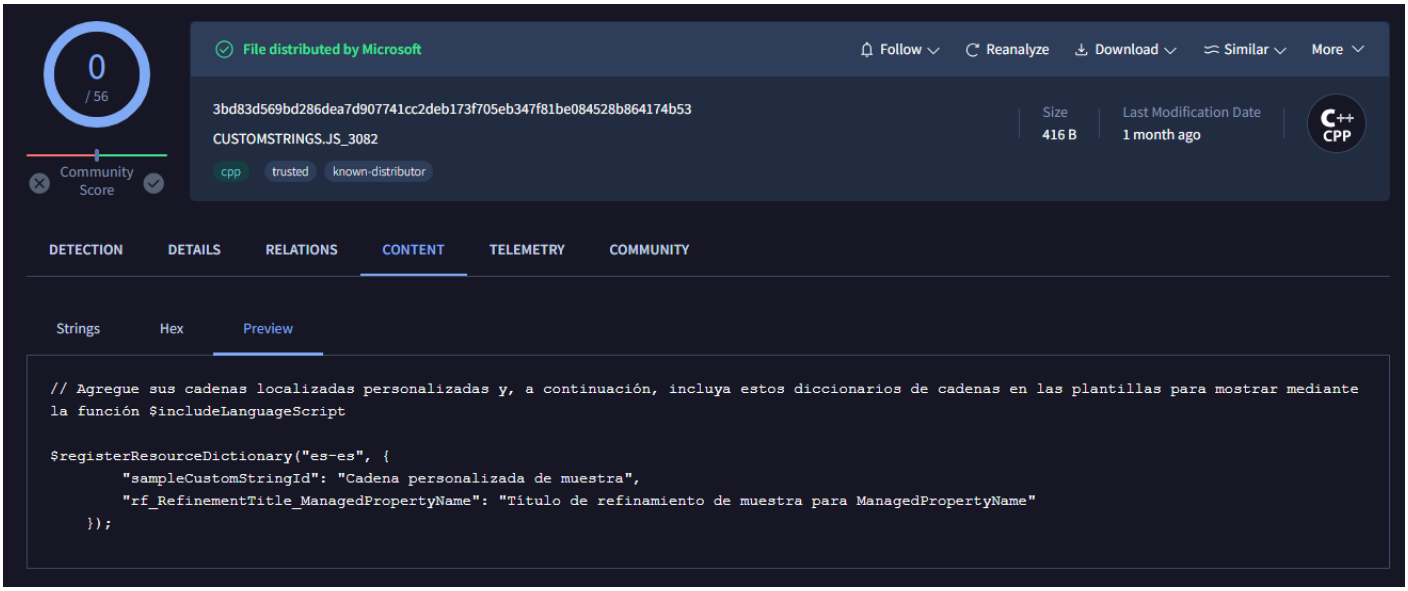


Figure 25: Artifact generated by the sandbox via Google Chrome when connecting to a domain

The strings “registerResourceDictionary”, “sampleCustomStringId”, “rf_RefinementTitle_ManagedPropertyName” are related to Microsoft SharePoint as we were able to confirm. These files were probably generated after visiting sites that have Microsoft Sharepoint functionalities. We found that all the PDFs containing this artifact dropped by Google Chrome came from a website belonging to the Government of Colombia.

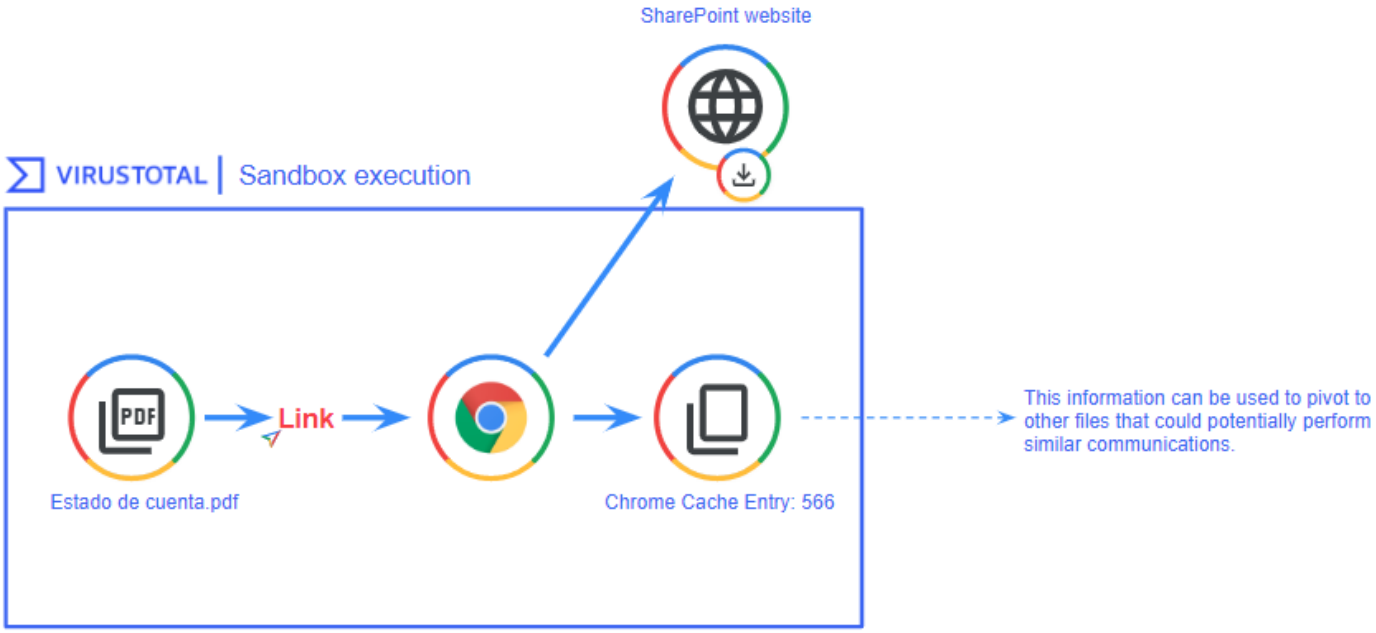


Figure 26: Flow of artifact generation related to Google Chrome that can be used for pivoting in VirusTotal

Email files

Many threat actors incorporate images in their emails, such as company logos, to deceive victims. We used this to identify several mailing campaigns where the same footer was used.

Campaign impersonating universities

On November 13, 2023, we details about a new campaign impersonating universities, primarily located in Latin America. By leveraging the presence of social network logos in the footer, we were able to find more universities in different continents targeted by the same attacker.

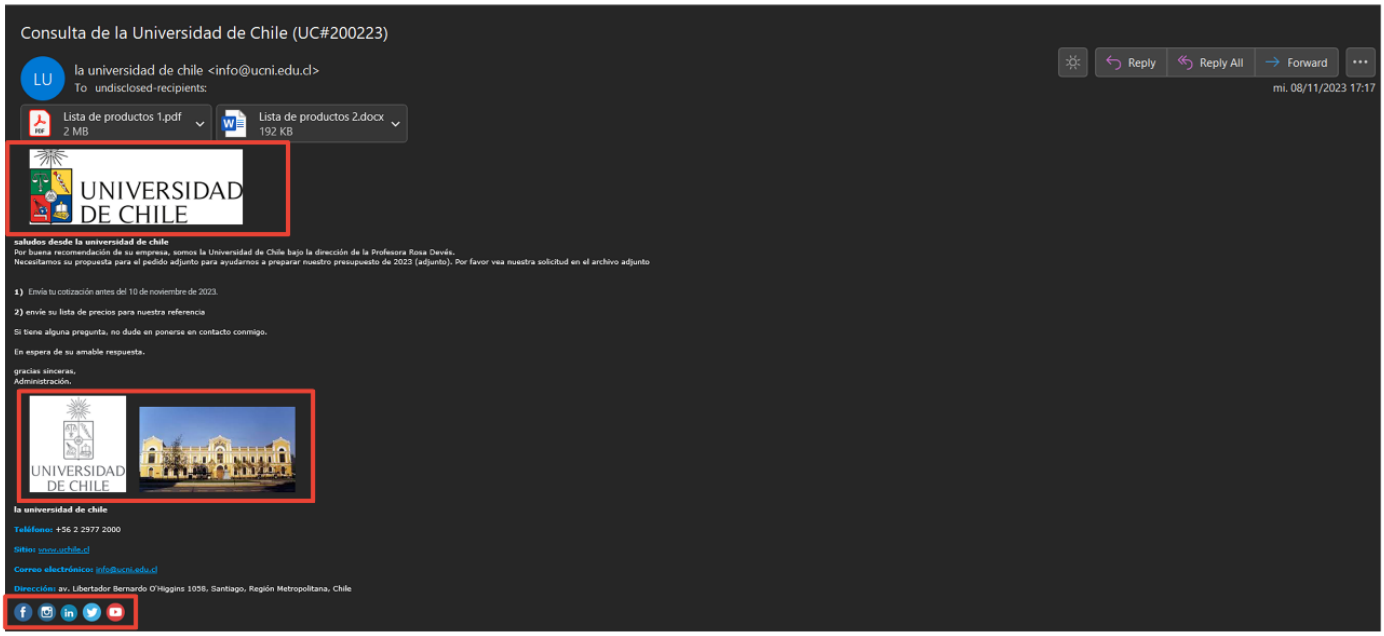


Figure 27: Email impersonating a university that contains multiple images

Pivoting through the images related to the University of Chile doesn't yield good results, as it's too specific. However, if we pivot through the images of the social media footer, represented as email attachments, we can observe multiple files using the same logo.

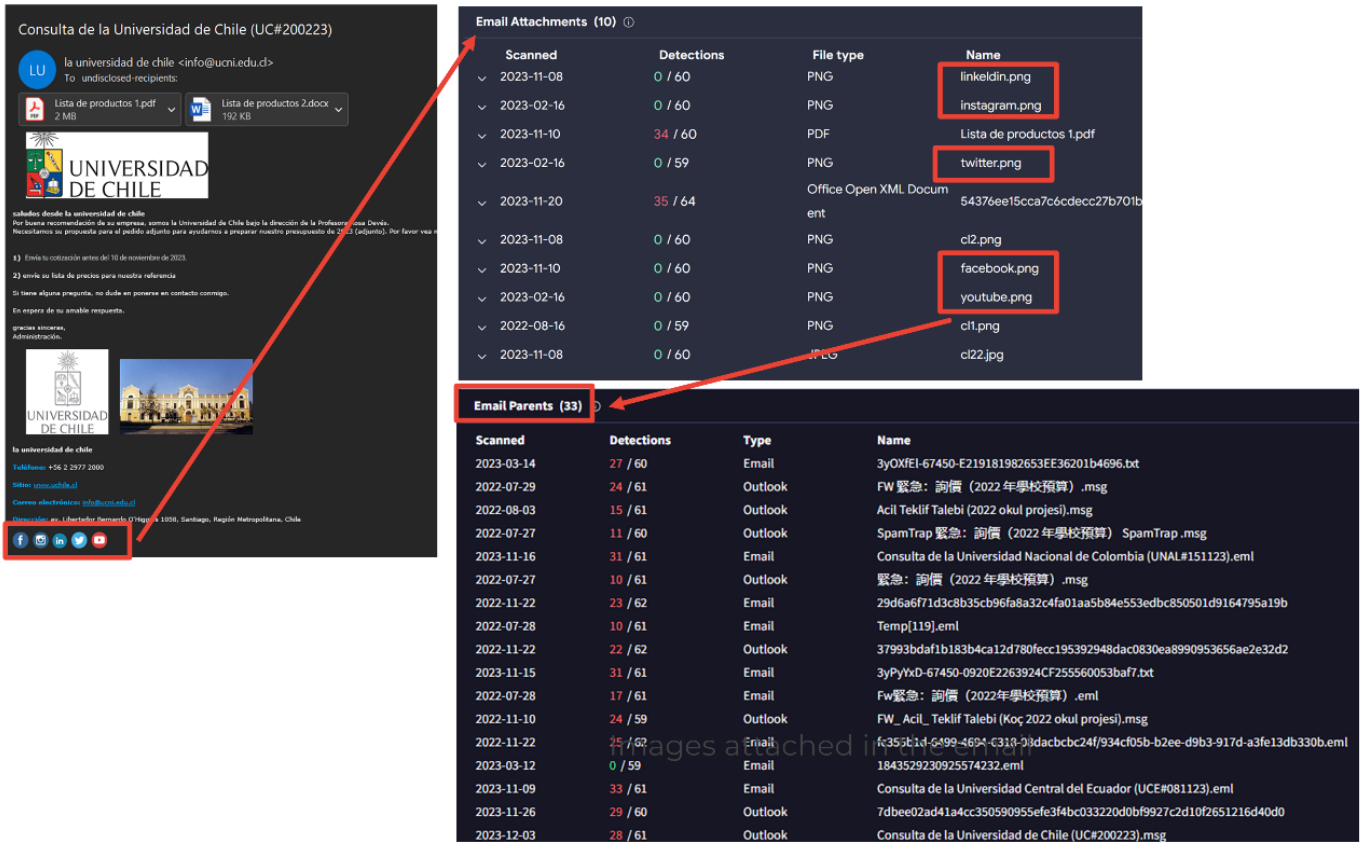


Figure 28: Using the images from the email footer to pivot and identify new emails

Just by analyzing one of the social media logos, we saw 33 email parents, all of them related to the same campaign.

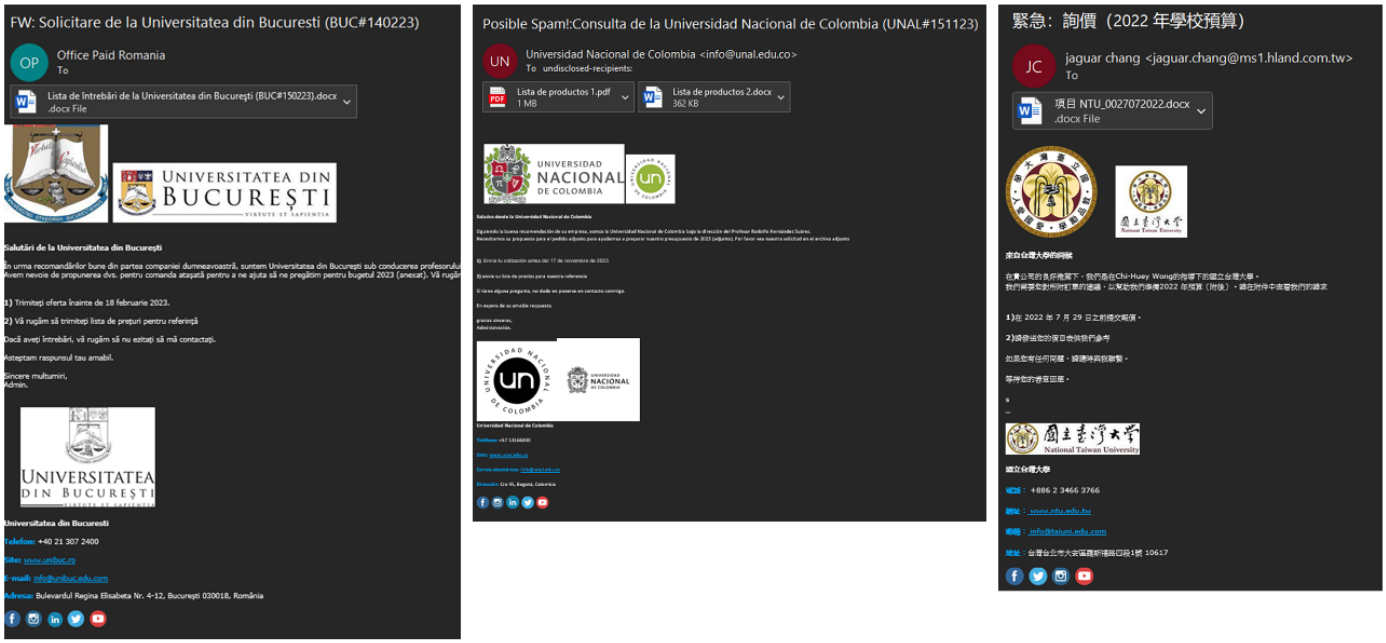


Figure 29: Other emails identified through image pivoting techniques

Campaigns impersonating companies

Another usual case is adding a company logo in the email signatures to enhance credibility. Delivery companies, banks, and suppliers are some of the most observed images during our research.

For example, this email utilizes the corporate image of China Anhui Technology Import and Export Co Ltd in the footer.


[TOP URGRENT] NEW ORDER INQUIRY-TDN001-180424

Kim Chung <sales@ahtech.com.cn>
To undisclosed-recipients:

Reply | Reply All | Forward

Wed 4/17/2024 5:36 PM

TDN001-180424.PDF.arj
12 KB

Dear Supplier,

Good day!

Please provide us with your best offer as per attached inquiry (TDN001-180424).

Hope for your soonest reply..

Thanks & Best regards

Kim Chung / Team Leader

安徽省技术进出口股份有限公司
ANHUI TECHNOLOGY IMP. & EXP. CO., LTD.

Anhui Technology Import and Export Co., Ltd.
Shushan, Hefei, Anhui Yangtze River Road 459
Direct Dial：8655165101188
FAX：+867578360911
Switch Board：+8675786201290
Email: sales@ahtech.com.cn
Web: www.ahtech.com.cn

Figure 30: Email impersonating a Chinese organization using the company logo in the footer

Pivoting through the image we found 20 emails using the same logo.

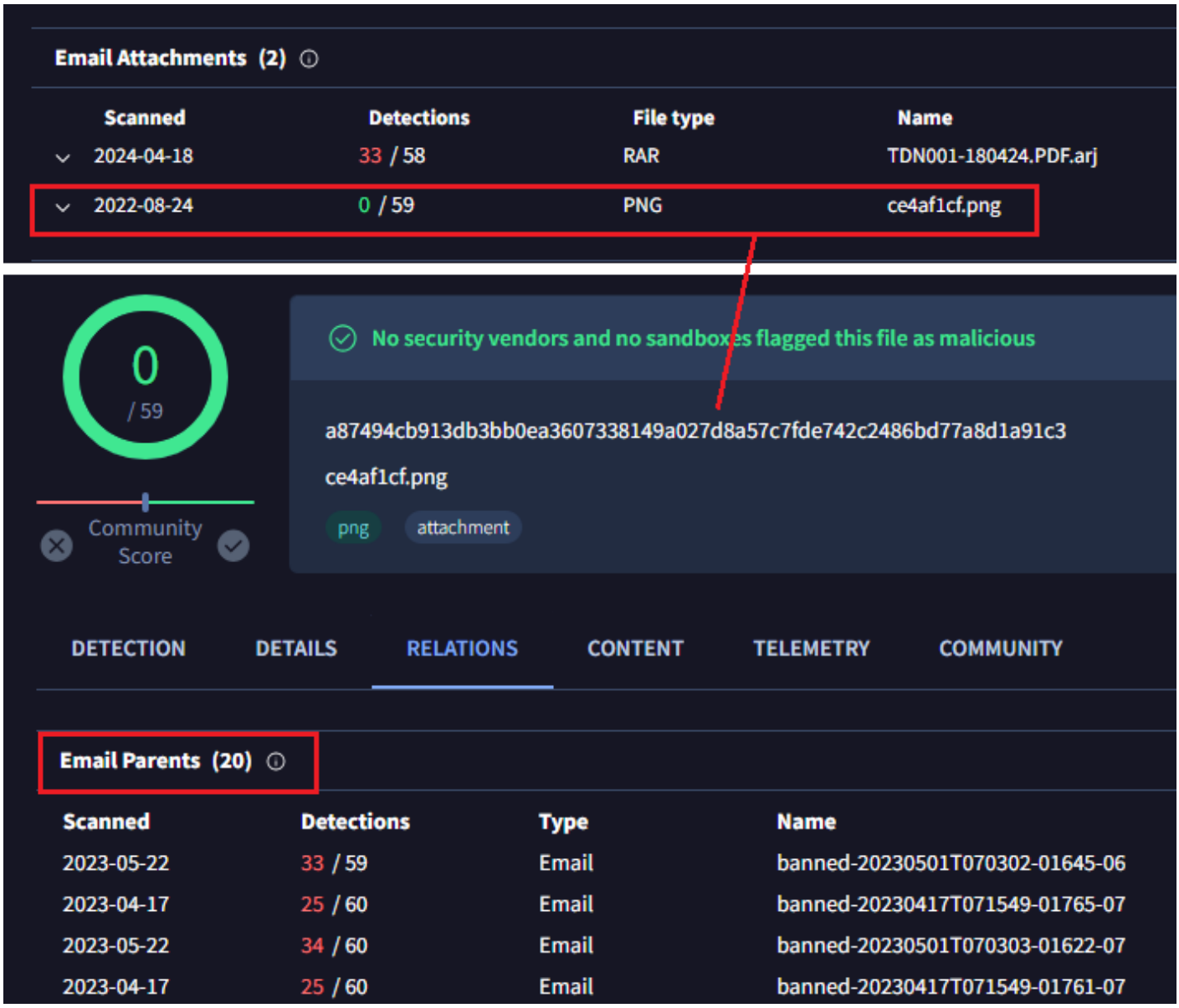


Figure 31: Other emails identified through image pivoting techniques

Wrapping up

We can potentially trace malicious actors by examining artifacts linked to the initial spreading documents, and in the case of images, AI can help us automate potential victim identification and other hunting aspects.

In order to make this even easier, we are planning to incorporate a new bundled_files field into the IOCs JSON structure, which basically will help to create livehunt rules. In the meantime you can use vt_behaviour_files_dropped.sha256 for those scenarios where the files are dropped.

In certain situations, the styles.xml and [Content_Types].xml files within office documents can provide valuable clues for identifying and tracking the same threat actor. The method presented here offers an alternative to traditional hunting or pivoting techniques, serving as a valuable addition to a team's hunting activities.

We hope you found this research interesting and useful, and as always we are happy to hear your feedback.

Stay tuned for the next blog post on PDFs and emails approach!

APPENDIX

[Content_types].xml shared between threat actors

[Content_Type].xml sha256	Shared by
3d8578fd41d766740a1f1ddef972a081436a2d70ab1e9552a861e58d8bbf5321	APT33, APT32
4ea40d34cfcaf69aa35b405c575c7b87e35c72246f04d2d0c5f381bc50fc8b3d	APT29, APT28
4f7fa7433484b4e655d185719613e2f98d017590146d15eedc1aa1d967636b3a	FIN7, Gamaredon, APT28, APT32
529739886f6402a9cd5a8064ece73eef19c597ef35c0bc8d09390e8b4de9041b	FIN7, APT33, TA505, Mustang Panda
688dca40507fb96630f3df80442266a0354e7c24b7df86be3ea57069b25d12c6	Gamaredon, APT33
6f1ac5f0ebfb7e97d3dc4100e88eaab10016a5cac75e1251781f2ea12477af51	Gamaredon, Hazy Tiger, APT33,
7796c382cd4c7c4ae3bcf2eed4091fbb20a2563ca88f2aecadb950ad9cf661f8	Razor Tiger, APT28, UAC-0099
b4fa7f3faa0510e4d969219bceec2a90e8a48ff28e060db3cdd37ce935c3779c	Razor Tiger, SideCopy
dfa90f373b8fd8147ee3e4bfe1ee059e536cc1b068f7ec140c3fc0e6554f331a	Gamaredon, APT37, Mustang Panda, APT28, UAC-0099, SideCopy
fe98b3bcf96f9c396eb9193f0f9484ef01d3017257300cc76098854b1f103b69	FIN7, Hazy Tiger
ff5a5ba3730a8d2ec0cbad39e5edf4ad502107bd0ef8a5347f29262b3dfe8a43	Mustang Panda, APT32

styles.xml shared between threat actors

Styles.xml sha256	Shared by
13ed55637980452662cb6838a2931a5e54fbed5881bcbae368b3d189d3a01930	APT28,

	UAC-0099, Razor Tiger
2de1fc9c48c4b0190361c49cdb053fd39cf81e32f12c82d08f88aec34358257f	Hazy Tiger, Gamaredon, APT33
59df7787c7cf5408481ae149660858d3af765a0c2cd63d6309b151380f92adb2	TA505, Gamaredon
8f590f608f0719404a1731bb70a6ce2db420fd61e5a387d5b3091d47c7e21ac9	APT28, FIN7, Razor Tiger, APT32, APT33
de392cd4bf1d650a9cf8c6d24e05e0605bf4eaf1518710f0307d8aceb9e5496c	Hazy Tiger, FIN7
e16f84c5fd1df6af1a1f2049f7862f4ea460765863476afb17e78edee772d35b	APT32, SideCopy, Mustang Panda, Razor Tiger