# Unraveling the Mind behind the APT

## Analyzing the Role of Pretexting in CTI and Attribution

Speaker: Sanne Maasakkers

BlackHat USA 2024 briefings

black hat®
USA 2024

$
?
12
?

# Contents

01 Introduction

02 Research concept

03 Analyzing content

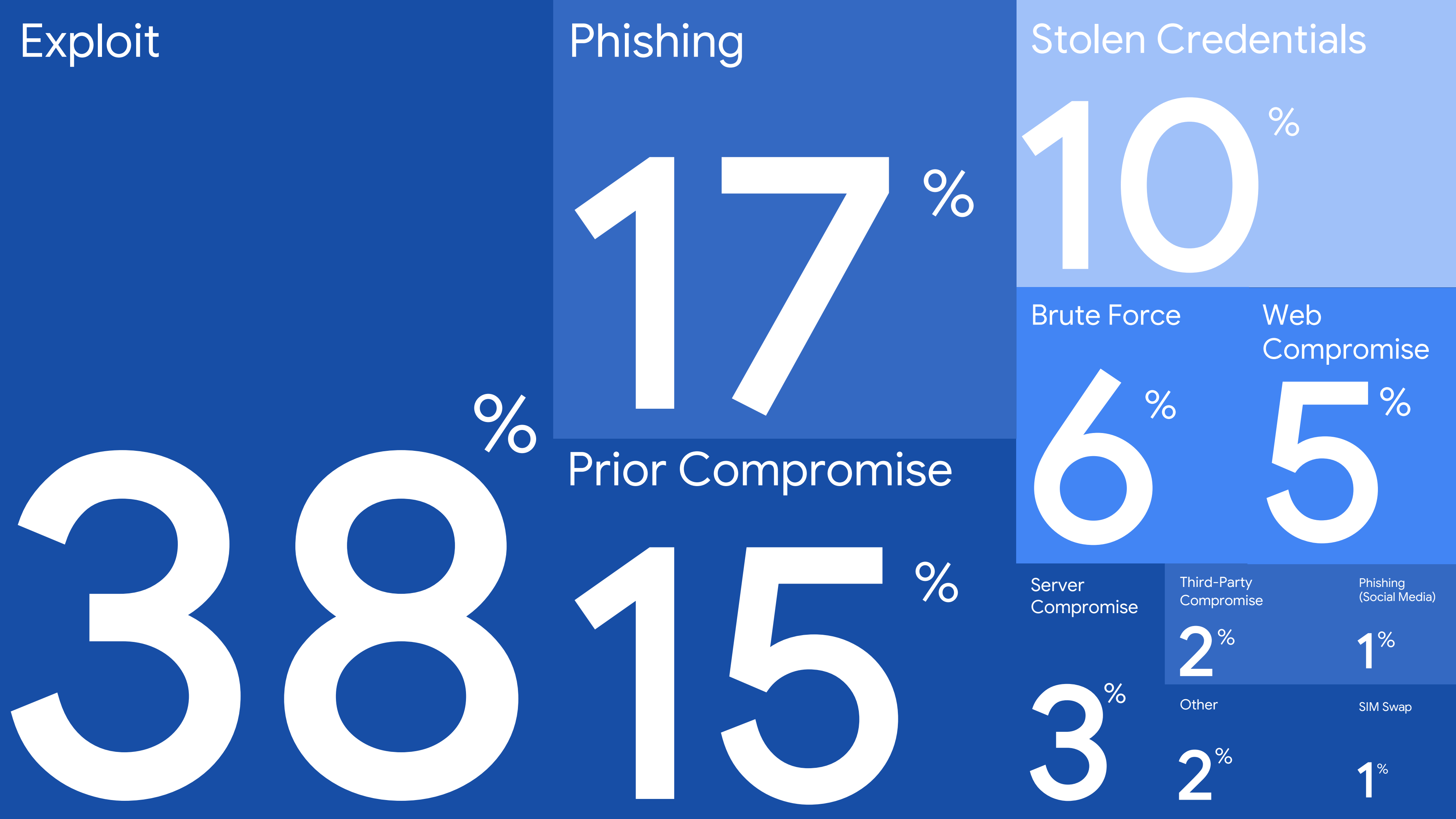
04 Analyzing context

05 Result & demos

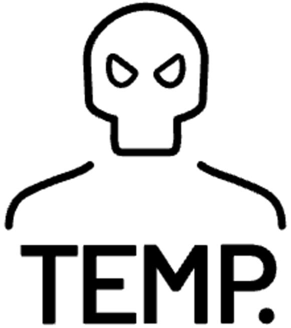
06 Conclusion & outlook

# Sanne

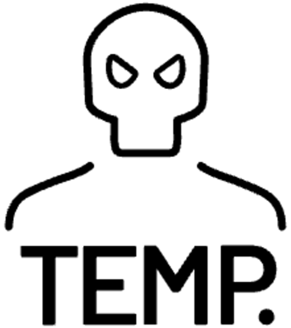
- Joined Mandiant Intelligence / Google Cloud in 2023 as Senior Analyst
- Previously worked in Red Team / Research & Intel Fusion Team (Fox-IT) and Fusion Centre (NCSC-NL) analyzing threats against The Netherlands
- <3 malware and being creative with (actor/threat) data
- Coach of the European CTF team, creator of Hackchallenges
- EU lead at (DEFCON’s) Adversary Village


Exploit
38%
Phishing
17%
Stolen Credentials
10%
Prior Compromise
15%
Brute Force
6%
Web Compromise
5%
Server Compromise
3%
Third-Party Compromise
2%
Phishing (Social Media)
1%
Other
2%
SIM Swap
1%

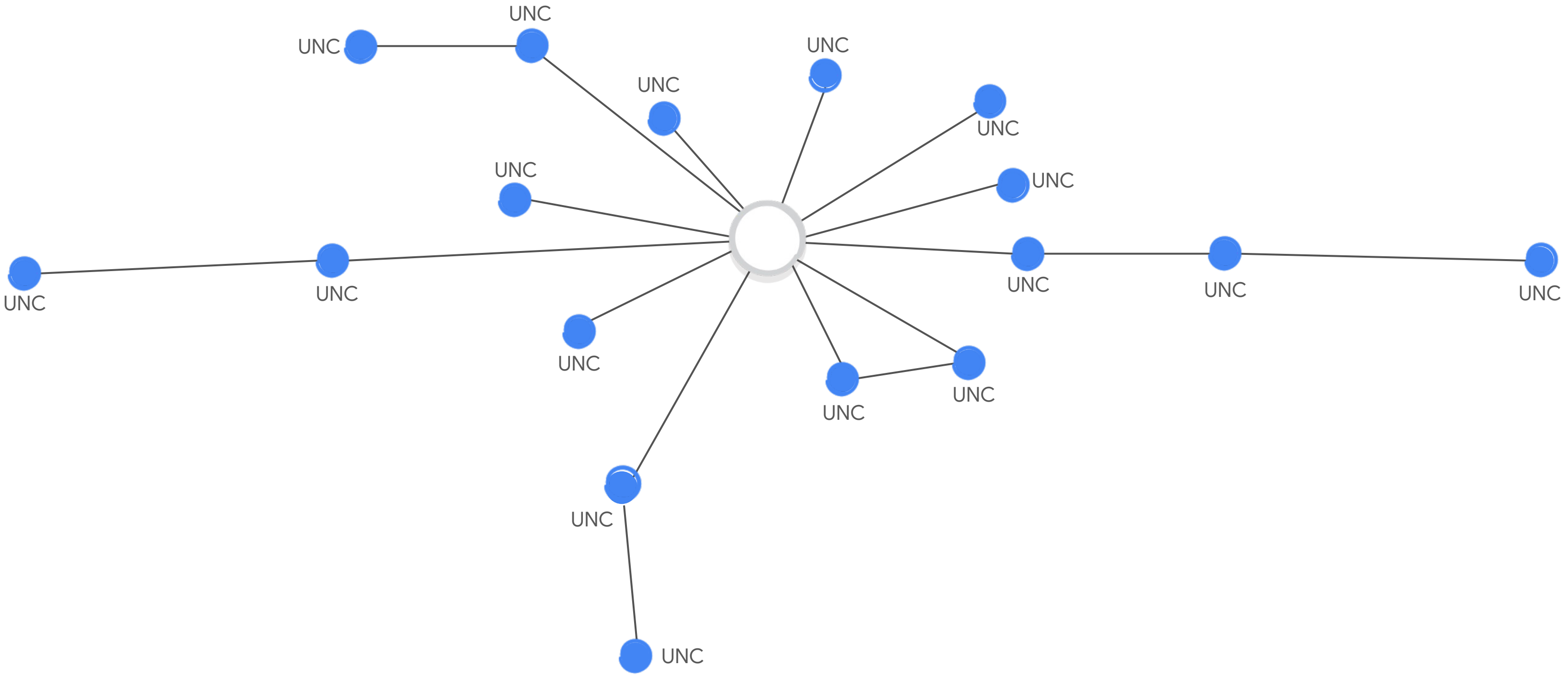
# Threat groups



# Threat groups



# Threat groups



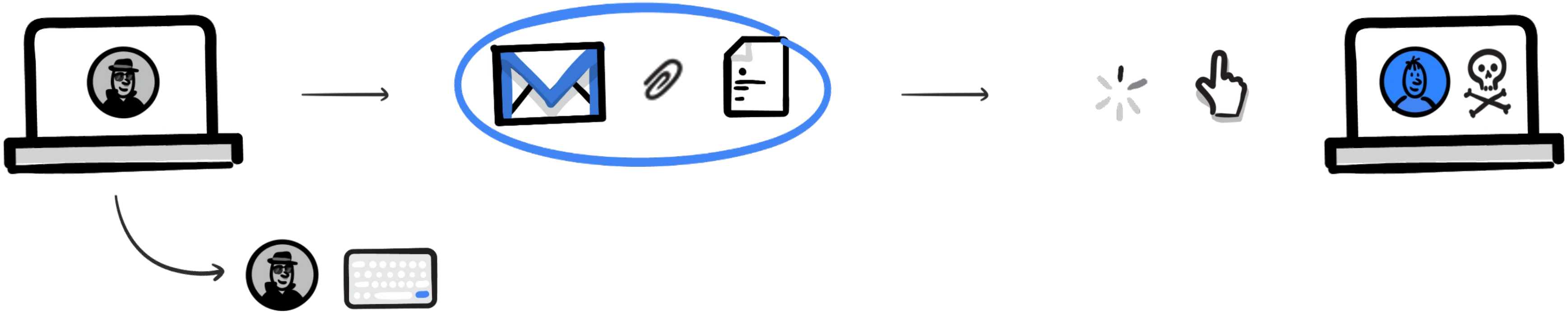
# Clustering

Emails are associated with a threat group mostly through various technical, tactical and strategical indicators, including:

- Technical: reuse of malware or code within malware attachments, reuse of infrastructure, including IP addresses, domains, and hosting providers.
- Tactical: consistent use of specific tactics in the infection chain, patterns in infrastructure.
- Strategical: common geographical and industry targeting.

Behavioral

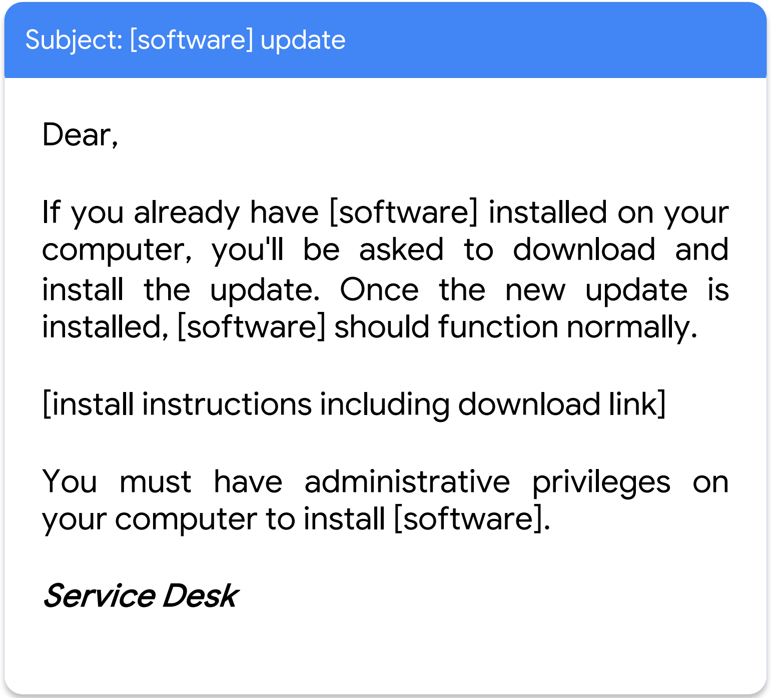
# Spear phishing



# Concept

This research focuses on the behavioral characteristics of APT phishing emails, including the pretext and email scenario, and their importance in linking (new) phishing campaigns to their authors. This includes both the content and context of the email.

# Example



Subject: [software] update

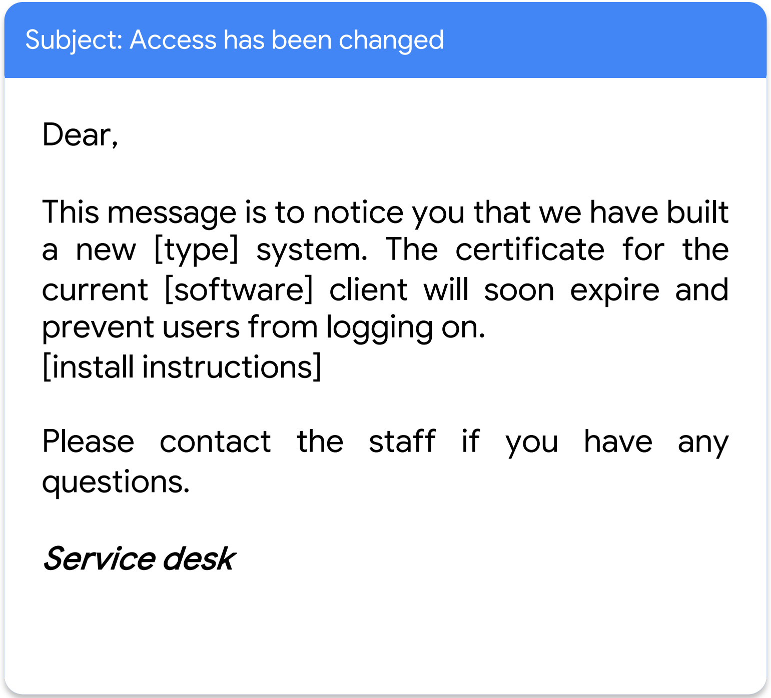
Dear,

If you already have [software] installed on your computer, you'll be asked to download and install the update. Once the new update is installed, [software] should function normally.

[install instructions including download link]

You must have administrative privileges on your computer to install [software].

***Service Desk***



Subject: Access has been changed

Dear,

This message is to notice you that we have built a new [type] system. The certificate for the current [software] client will soon expire and prevent users from logging on.
[install instructions]

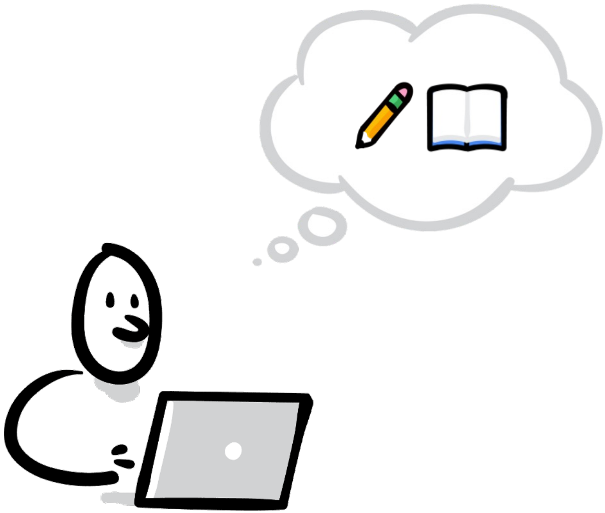
Please contact the staff if you have any questions.

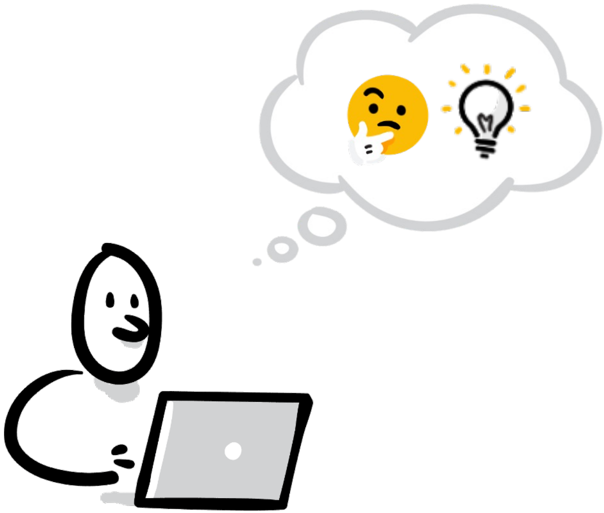
***Service desk***

* Emails are slightly altered for security and privacy purposes

VIBEINT refers to information obtained from a gut feeling or intuition, often based on previous experience. It is mostly unverified and unreliable, but it can sometimes provide insights or lead to further investigation.

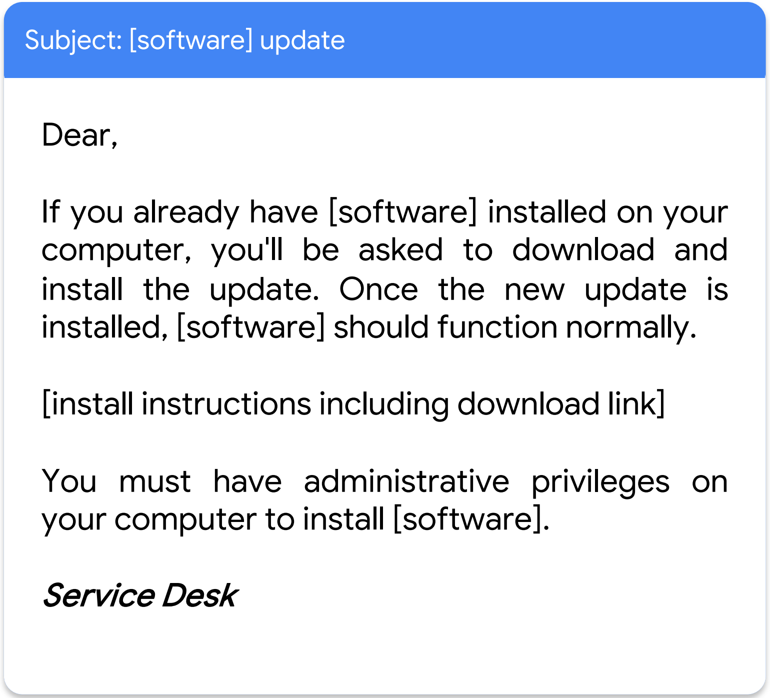
# Scenario





https://blog.sannemaasakkers.com/2021/08/07/adversary-phishing-characteristics/

# Content



Subject: [software] update

Dear,

If you already have [software] installed on your computer, you'll be asked to download and install the update. Once the new update is installed, [software] should function normally.

[install instructions including download link]

You must have administrative privileges on your computer to install [software].

***Service Desk***

Subject

Salutation

Language<br>
Textual features

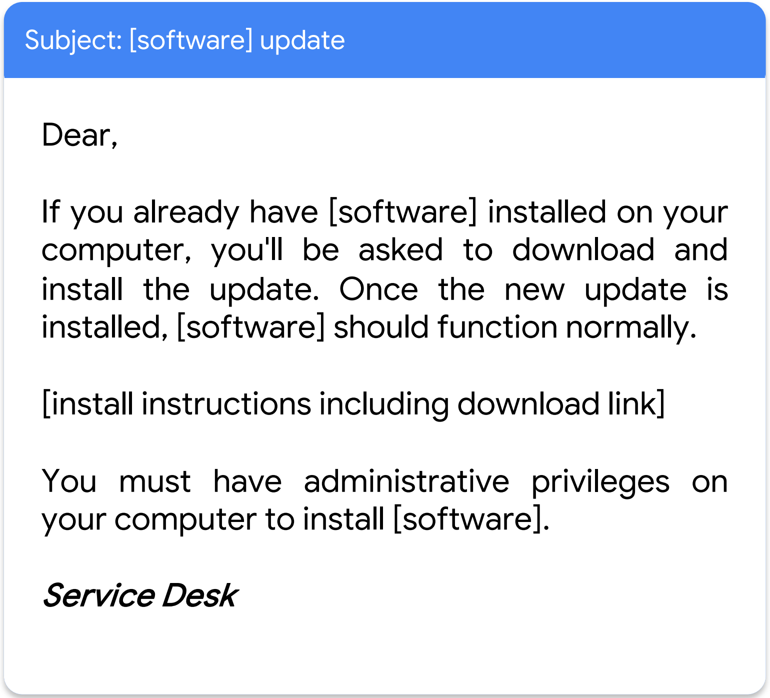
Attachment or URL

Signature

* Email is slightly altered for security and privacy purposes

# Context

Sender type



Subject: [software] update

Dear,

If you already have [software] installed on your computer, you'll be asked to download and install the update. Once the new update is installed, [software] should function normally.

[install instructions including download link]

You must have administrative privileges on your computer to install [software].

*Service Desk*

Theme

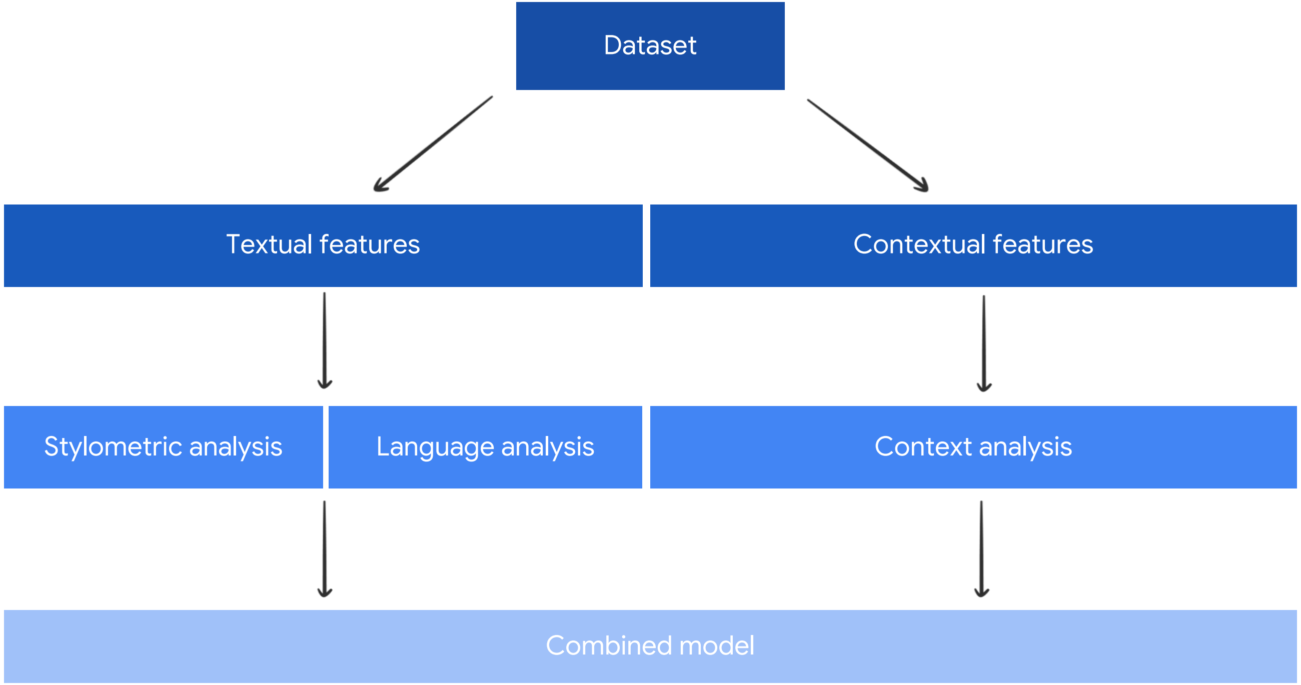
Persuasion

Goal

Design

* Email is slightly altered for security and privacy purposes

# Analysis



Dataset

Textual features

Contextual features

Stylometric analysis

Language analysis

Context analysis

Combined model

Stylometry is the statistical analysis of linguistic style in written or spoken language, aiming to identify patterns and features unique to specific authors. This analysis can be applied to attribute authorship.

# Stylometry

It uses statistics to analyze an author’s lexical and syntactic features.

- Lexical features: word frequencies, word length distribution, Hapax Legomena, vocabulary richness.
- Syntactic features: sentence length, average word length, punctuation usage.

Think of it as identifying someone based on how they talk, not (just) what they say.

It is a common technique and already used to analyze (anonymous) authors, threatening letters or ransom texts.

# Example

### Stylometric 1

Dear Sir,

For your information. See the attach.

### Stylometric 2

[month] Financial Data Table.

Have you got it? Please check it.

Lingua spaCy

| | | |
|---|---|---|
| average_length | short_words | proportion_digits |
| 4.625 | 0.50 | 0.0 |
| proportion_capital | text_richness | hapax_legomena |
| 0.09 | 1 | 8 |

| | | |
|---|---|---|
| average_length | short_words | proportion_digits |
| 4.63 | 0.36 | 0.0 |
| proportion_capital | text_richness | hapax_legomena |
| 0.09 | 0.82 | 9 |

* Emails are slightly altered for security and privacy purposes

# Stylometry

However, stylometry has several limitations, including:

- Semantic understanding: it does not understand the meaning of words or the nuances of language.
- Contextual awareness: it struggles to analyze relationships between non-sequential words, sentences, or paragraphs, missing the broader context of the text.
- Domain-specific knowledge: it lacks understanding of specialized fields or jargon, which can be crucial for accurate analysis in certain types of texts.

# Stylometry

While stylometry has been used for years in authorship attribution, its efficacy on APT emails is limited. A trained model on a relevant dataset results in an overall accuracy of 41%.

01 Text richness

02 Average number of words/sentence

03 Distribution of unicode characters

A language model analyzes text by considering the context of each word, capturing subtle nuances in meaning, and understanding complex word and sentence relationships.

# Language model

Pre-trained models can be used to perform various natural language processing tasks like text classification. They provide a powerful starting point for fine-tuning on specific tasks, saving time and resources compared to training from scratch.

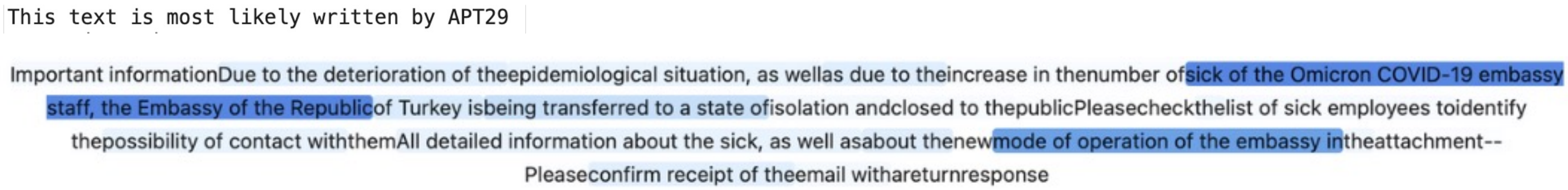
BERT is a pre-trained language model based on the transformer architecture.

Transformers are deep learning models that use multi-head attention to weigh the importance of different words in a sentence, allowing for better understanding of context and meaning.

# Language model

This resulted in an accuracy of 60% on all 33 actors. But *how*?

Machine Learning models can be explained by SHAP (SHapley Additive exPlanations). So; what happens if you try to predict a new text on the fine-tuned language model?



```
This text is most likely written by APT29
```

Important informationDue to the deterioration of theepidemiological situation, as wellas due to theincrease in thenumber ofsick of the Omicron COVID-19 embassy staff, the Embassy of the Republicof Turkey isbeing transferred to a state ofisolation andclosed to thepublicPleasecheckthelist of sick employees toidentify thepossibility of contact withthemAll detailed information about the sick, as well asabout thenewmode of operation of the embassy intheattachment--Pleaseconfirm receipt of theemail withareturnresponse

https://cloud.google.com/blog/topics/threat-intelligence/tracking-apt29-phishing-campaigns

Important information. Due to the deterioration of the epidemiological situation, as well as due to the increase in the number of sick of the Omicron COVID-19 embassy staff, the Embassy of the Republic of Turkey is being transferred to a state of isolation and closed to the public. Please check the list of sick employees to identify the possibility of contact with them. All detailed information about the sick, as well as about the new mode of operation of the embassy in the attachment. -- Please confirm receipt of the email with a return response.

0.047 0.113

Important information. Due to the deterioration of the epidemiological

0.093 0.117 0.03 0.034 1.000

situation, as well as due to the increase in the number of sick of the Omicron

1.000 0.153

COVID-19 embassy staff, the Embassy of the Republic of Turkey is being

0.236 0.047 0.028 0.021

transferred to a state of isolation and closed to the public. Please check the list

0.032 0.014 0.089

of sick employees to identify the possibility of contact with them. All detailed

0.079 0.0127 0.723

information about the sick, as well as about the new mode of operation of the

0.059 0.001 0.033

embassy in the attachment. -- Please confirm receipt of the email with a return

0.005

response.

# Content analysis highlights 🥁

01 For replies, the language used was not always consistent with the language of the initial email.

02 Similar emails could be written in completely different languages and discuss entirely different topics.

03 It's not just about using theme-specific words; it also focuses on the grammar used, such as "had [adverb] [past participle]" or speaking in the first person.

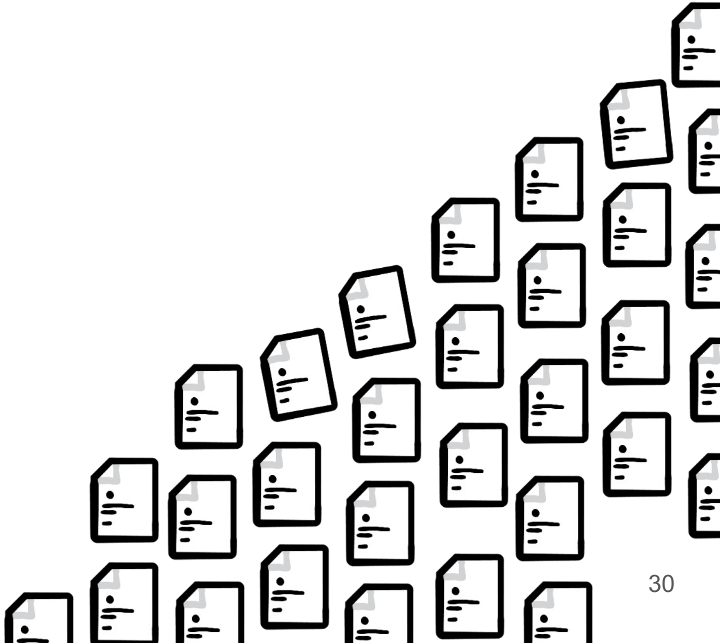
The context of an email includes elements that shape its meaning and purpose, such as theme, goal and the social engineering techniques employed to influence the recipient.

# Extracting these features

Large Language Models (LLMs) can effectively extract key contextual elements from emails, including:

- The inclusion of personal touches or signatures
- The overall theme of the email
- The social engineering techniques used to influence the recipient

The local LLM is given extra training documents to better understand and classify these features from emails and do simple categorization tasks.



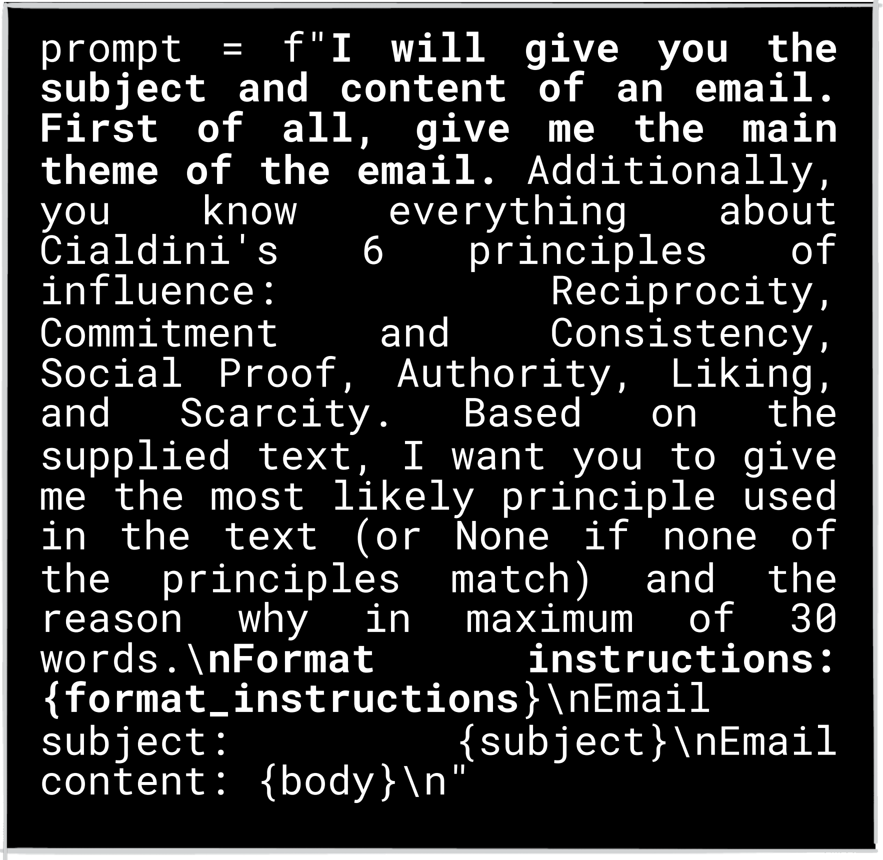
# Theme

Analysis of email themes reveals the most common themes are as follows:

- Invitations or requests (meetings, interviews, events)
- COVID-19 related (absences, changes)
- Account issues (resets, problems, settings)

These are categorized into the following categories:

- A recent event (COVID-19 or global events)
- An important value for the receiver (proposals)
- A timeless and generic theme (please find attached)

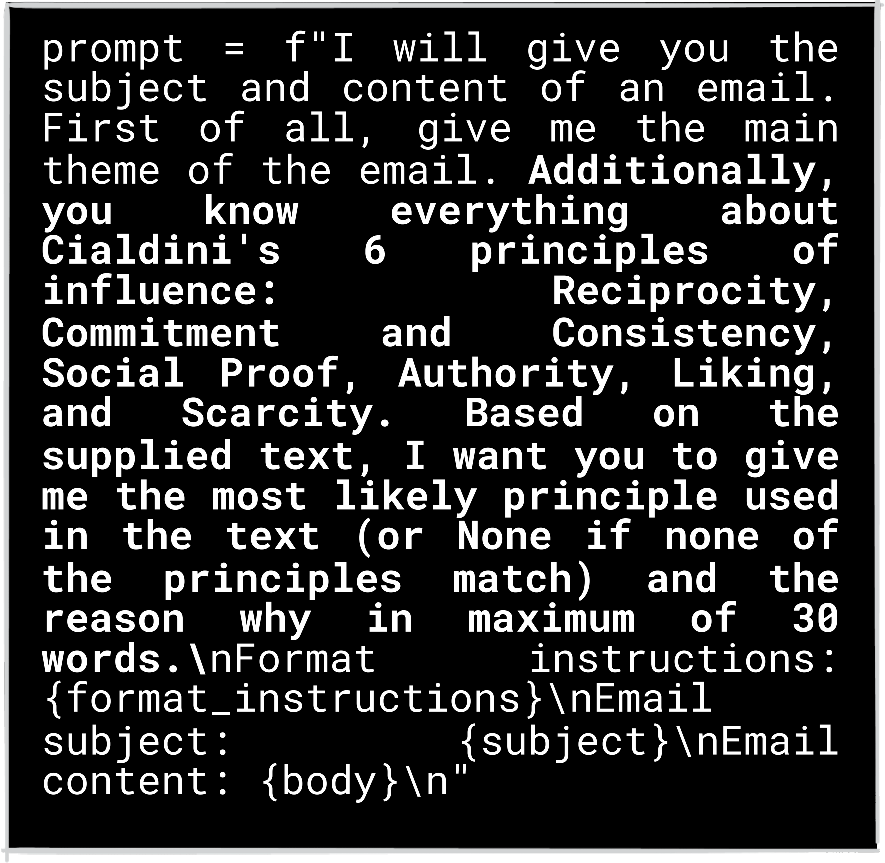


```
prompt = f"I will give you the
subject and content of an email.
First of all, give me the main
theme of the email. Additionally,
you know everything about
Cialdini's 6 principles of
influence: Reciprocity,
Commitment and Consistency,
Social Proof, Authority, Liking,
and Scarcity. Based on the
supplied text, I want you to give
me the most likely principle used
in the text (or None if none of
the principles match) and the
reason why in maximum of 30
words.\nFormat instructions:
{format_instructions}\nEmail
subject: {subject}\nEmail
content: {body}\n"
```

# Social engineering

The principles of influence, defined by Cialdini, are a set of psychological and social phenomena that can be used to influence behavior and decision-making.

By leveraging these principles, phishers can create a sense of urgency, trust, or authority that overrides the recipient's natural caution.



```
prompt = f"I will give you the
subject and content of an email.
First of all, give me the main
theme of the email. Additionally,
you know everything about
Cialdini's 6 principles of
influence: Reciprocity,
Commitment and Consistency,
Social Proof, Authority, Liking,
and Scarcity. Based on the
supplied text, I want you to give
me the most likely principle used
in the text (or None if none of
the principles match) and the
reason why in maximum of 30
words.\nFormat instructions:
{format_instructions}\nEmail
subject: {subject}\nEmail
content: {body}\n"
```

# Principles of influence

## Principle: Authority

Greetings!

On behalf of [important person in policy], I would like to invite you to a briefing with [important person in policy] on [date]. [person] will discuss [topic] and your input will be appreciated.

Kind regards,
[name]

📎 Invite.hta

## Principle: Commitment and Consistency

Dear [name],

As a follow up on our conversation, I'm sending you the job profile of the developer position at [organization] attached. Looking forward to hearing from you soon.

Kind regards,
[name]
[recruiter at organization]

📎 Job profile.doc

* Emails are slightly altered for security and privacy purposes

# Principles of influence

Principle: Liking

Hey [name],

Long time no see and best wishes for the New Year! I hope that you will find good health and luck in the upcoming year. Please find my New Year’s wishes attached on this URL:

[URL]

[name]

Principle: Reciprocity

Hi [name],

Sorry for sending this via [platform], but I’ve had a lot of struggles uploading the files. Hope this is OK! Hope it works for you now, it should only be accessible by you. Let me know if there are problems.

[URL to platform]

[name]

* Emails are slightly altered for security and privacy purposes

# Principles of influence

### Principle: Scarcity

Hi [name],

As mentioned, just wanted to pass this document. The password is 123456. This is a confidential document, so please don’t share it with anyone.

Thank you and we keep in touch.

[name]

📎 Files.rar

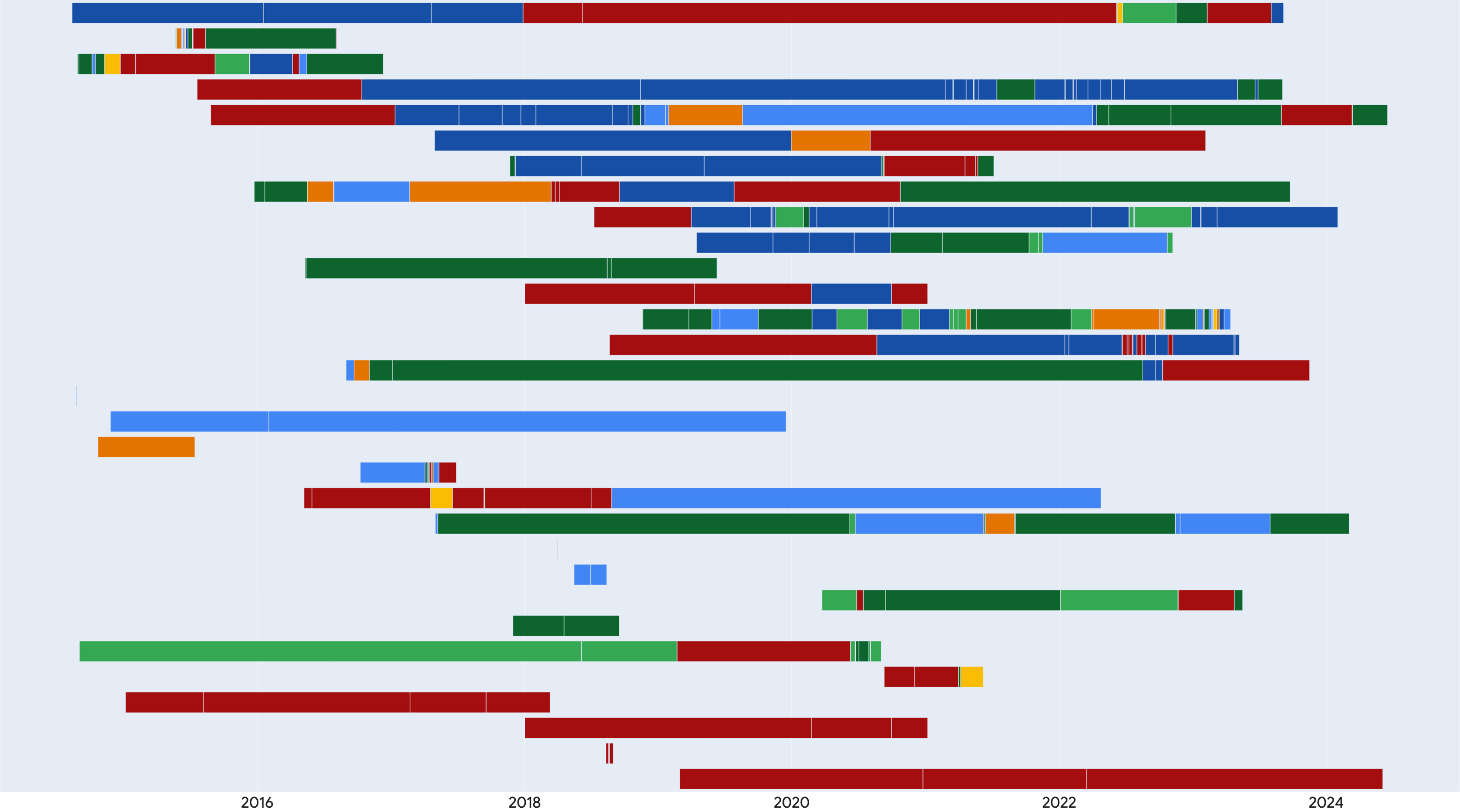
### Principle: Social Proof

Hey [name],

My name is [name] and I’ve recently had a talk with [person X, person Y]. We were wondering if you would be interested in joining [project Z], we definitely think you’re the right person with valuable insights. Please find more details here:
[URL]

Kind regards,
[name]

* Emails are slightly altered for security and privacy purposes


2016
2018
2020
2022
2024

# Context

The prediction model built on contextual features achieved a 67% accuracy across all authors, with the following features being the most prevalent:

01 Principle of influence

02 Sender category

03 Theme

The models are combined with a meta model. A meta-model is a higher-level model that learns how to best integrate the predictions or outputs of the three individual models and is used for this analysis.

# Result

✓ The total accuracy of all three models combined (and tuned) results in an overall accuracy to 88 - 96%, after removing the least performing actors from the set. Insufficient data for certain actors impacts the model's ability to learn their patterns effectively, so the model is not fully able to make predictions for those actors.

The remaining groups represent interesting groups that have been active in the last years.

# Analysis

Let’s dive in some visualizations and examples of how it helps clustering.

1 **Get insights in clusters**
Find similarities and differences

2 **Finding outliers**
Reconsider the links

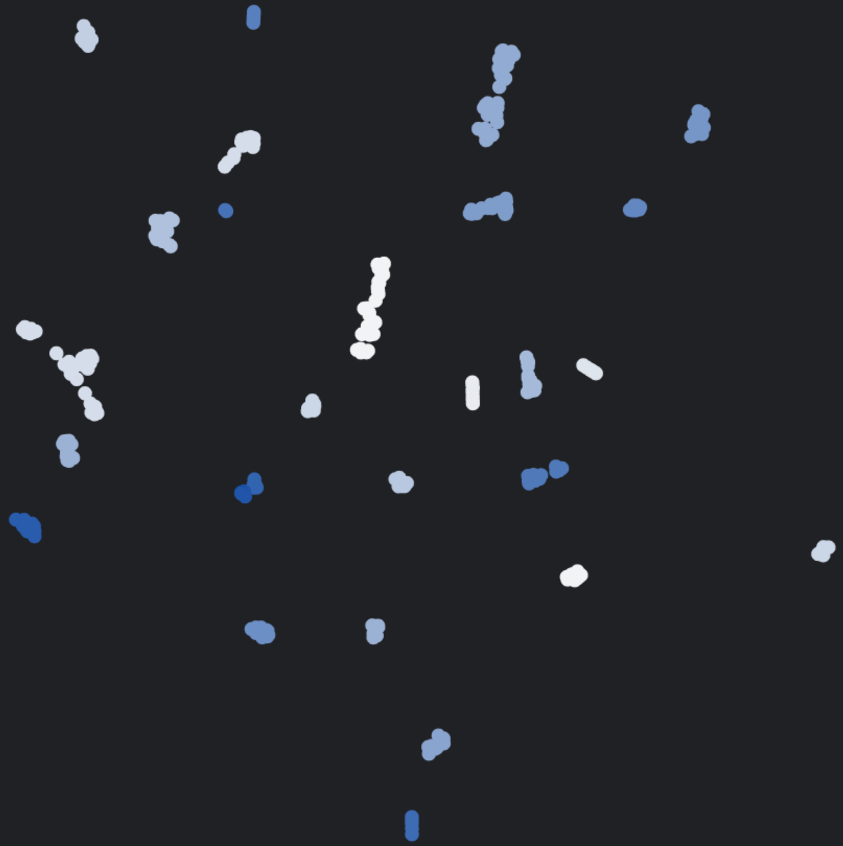
3 **Find author**
Cluster with more confidence

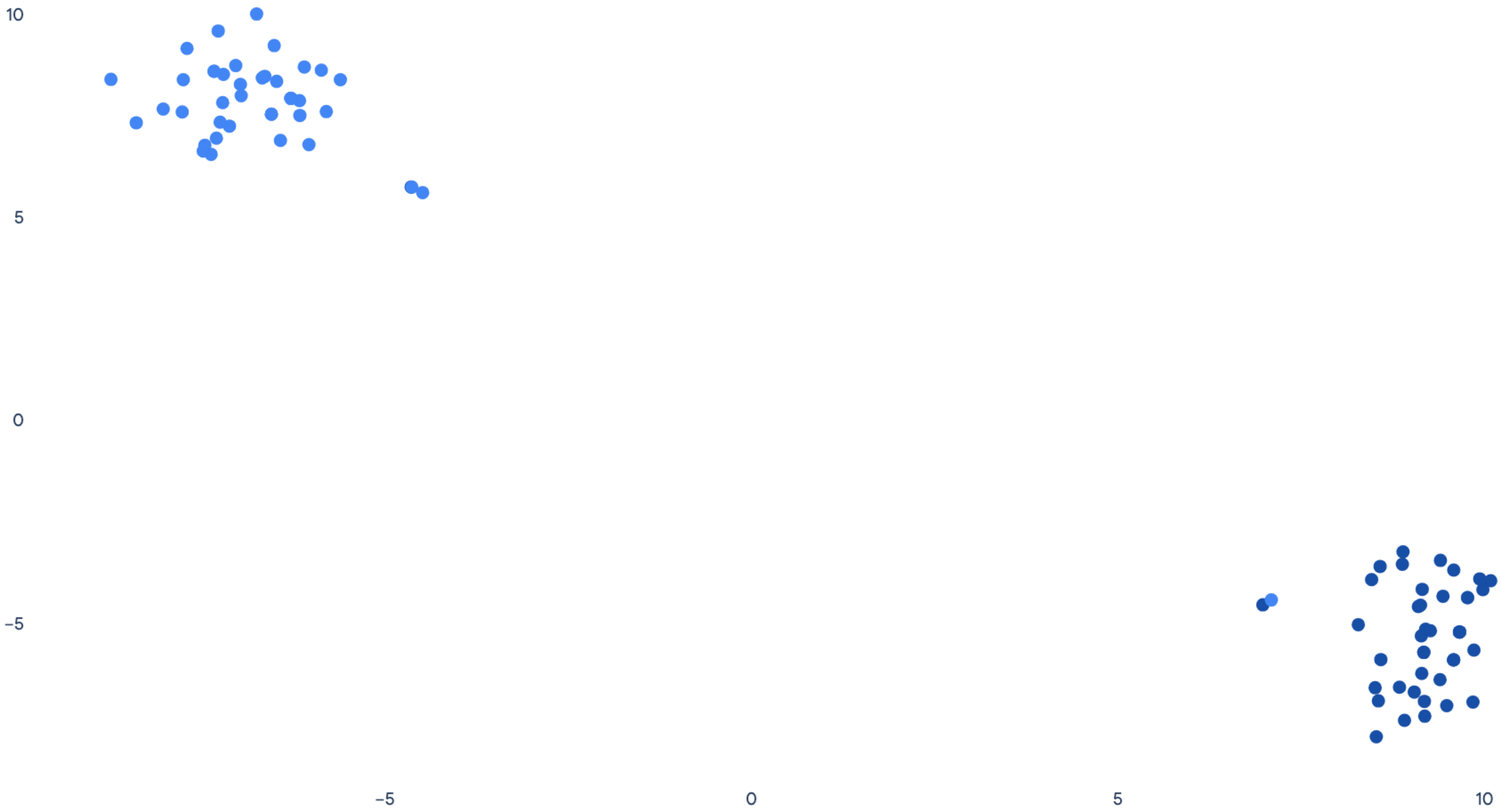
Cluster galaxy 1

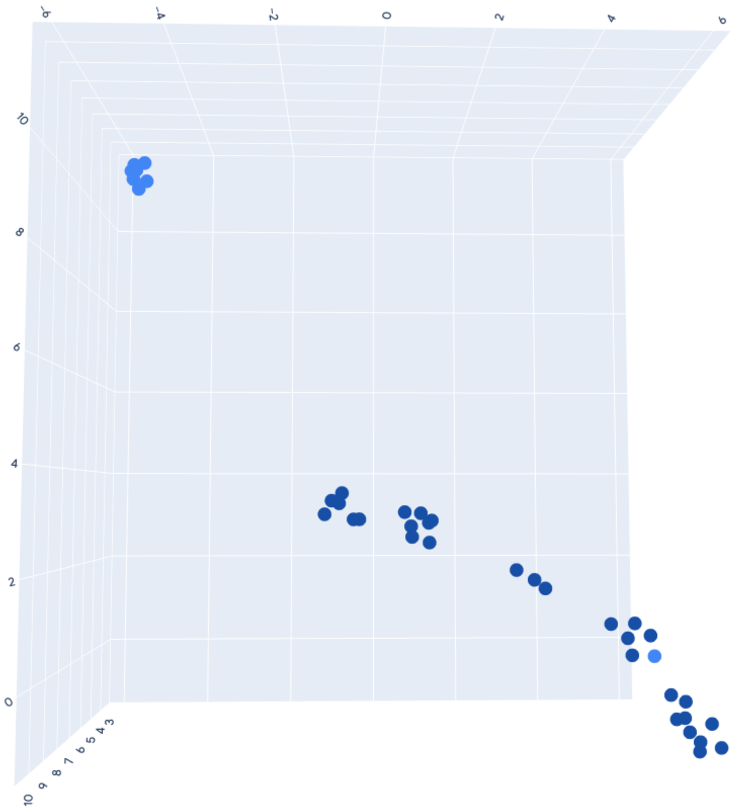
Cluster galaxy 2

Outliers

Subclustering




10
5
0
−5
−5
0
5
10



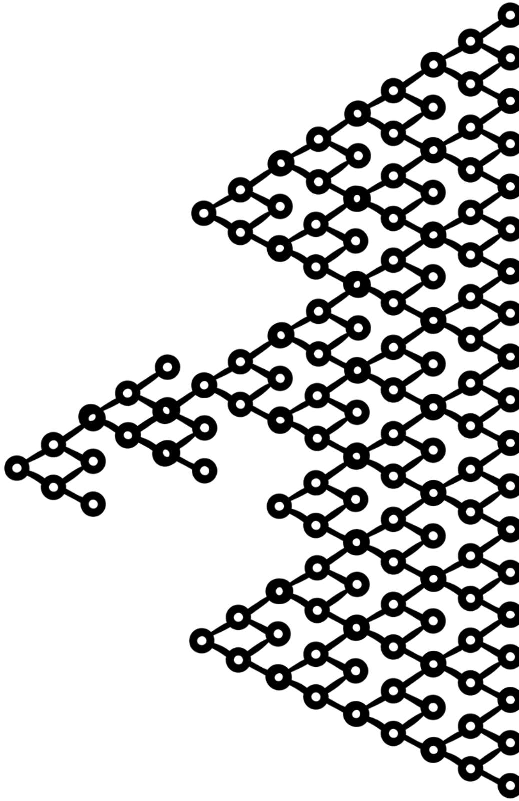
−6
−4
−2
0
2
4
6
10
8
6
4
2
0
10
9
8
7
6
5
4
3

# Subclusters

Multiple actors (labels) have subclusters within their respective clusters.

A review of these subclusters revealed the following:

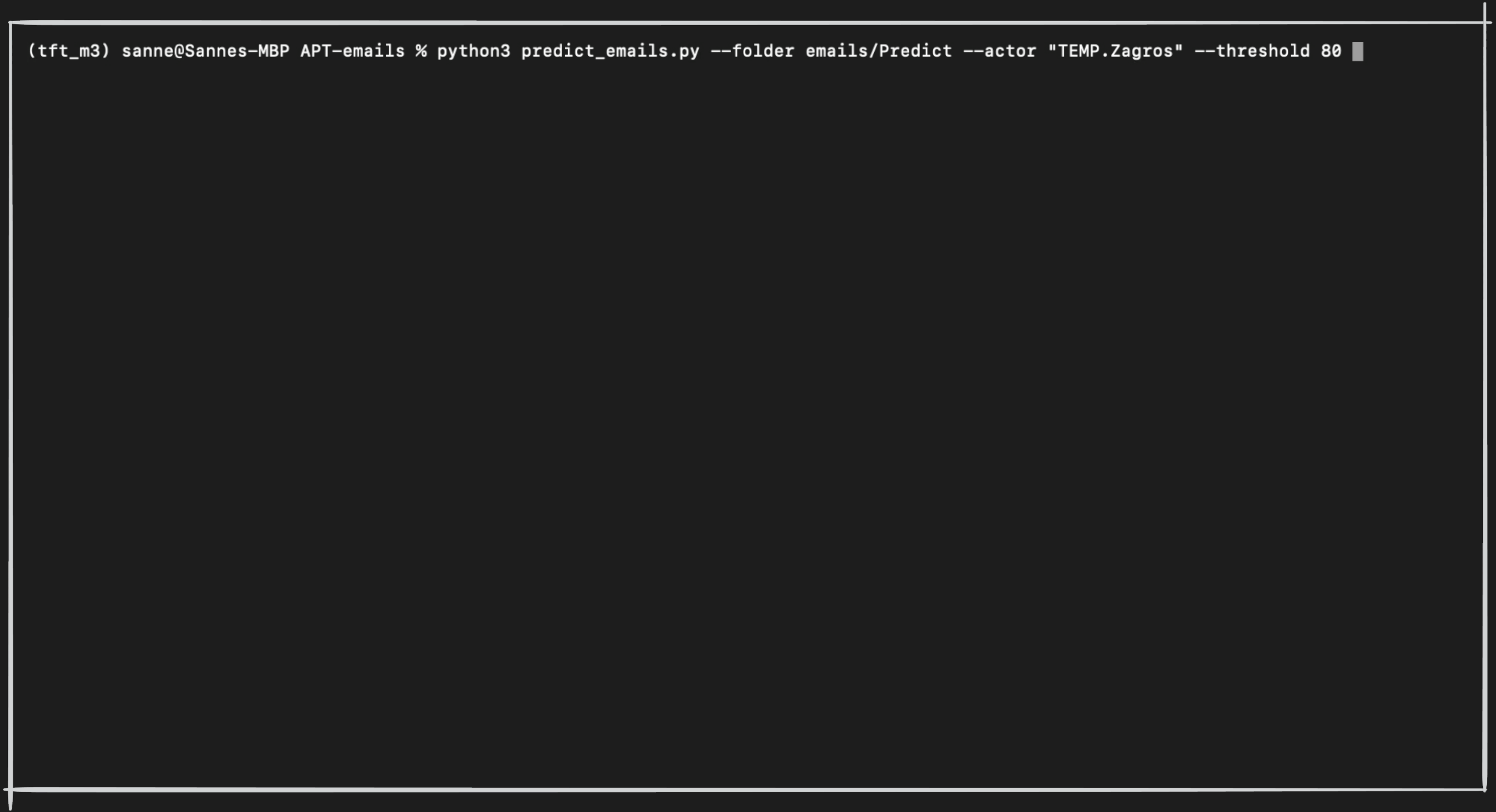
- Change in targeting: actors have adapted their writing styles based on the targets (geographical, industry or person).
- Change over time: subclusters showed emails sent around the same time. Although the content is totally different, these emails might be part of the same campaign.
- Distinct cluster: UNCs are considered part of the actor, but this isn't necessarily the case. They could represent a separate cluster, an affiliate, or simply another individual.



# Find author

```
comment:”#muddywater” has:email_parents
```

18 results



```
(tft_m3) sanne@Sannes-MBP APT-emails % python3 predict_emails.py --folder emails/Predict --actor "TEMP.Zagros" --threshold 80
```

# Conclusion

The proposed model for clustering campaigns based on behavioral features has proven effective in analyzing the majority of emails from both APTs and TEMPs.

This underscores the potential of behavioral analysis to contribute to the accurate clustering of groups or linking new attacks to groups, next to clustering techniques already in place. It can aid threat intelligence analysts to understand trends and new phishing TTPs leveraged by specific threat actors and support in threat hunting.

# Outlook & implications

Further research could involve incorporating technical, tactical and strategical attributes into the model to have a full overview of a campaign. As discussed, those models have limitations, but so does this model:

- LLM usage: The use of LLMs for generating the email text can blur the lines between actors' writing styles.
- Copycats: Actors tend to use the same themes in emails based, like a recent event or even mimic other actors in their emails.

The streets of persuasion
are plated with gold

The Killers - The Rising Tide

# Thank you

| | |
|---|---|
| X | @sannemaasakkers |
| LinkedIn | /sannemaasakkers |