# Advanced Detection of Cybersecurity Threat Mutations through Machine Learning and Behavioural Analysis

Jordi Domenech Fons
i2CAT Foundation
Barcelona, Spain
jordi.domenech@i2cat.net

Nil Ortiz Rabella
i2CAT Foundation
Barcelona, Spain
nil.ortiz@i2cat.net

Albert Calvo Ibañez
i2CAT Foundation
Barcelona, Spain
albert.calvo@i2cat.net

Saber Mhiri
i2CAT Foundation
Barcelona, Spain
saber.mhiri@i2cat.net

***Abstract*—For years, Security Operation Centres (SOC) have relied on detection tools that are becoming less effective in the cybersecurity industry, where sophisticated campaigns made by cybercriminals are not being noticed. Particularly, the detection of cybersecurity threat mutations – where attackers modify their techniques to evade detection – has emerged as a key challenge for organizations seeking to protect their data and systems. Through an extensive analysis of cybersecurity incidents and real network data, we propose a novel methodology and taxonomy in the field to detect threat mutations by combining a supervised machine learning algorithm with behavioural analysis. Our findings reveal the likelihood of a threat being a mutation of a known threat, including a novel representation of user behaviour profiles and an extended analysis of their properties. This study contributes to advancing detection and prevention techniques in the cybersecurity domain, paving the way for more resilient and adaptive defence systems.**

***Index Terms*—Cybersecurity, Threat Mutations, Behavioural Modelling, Cyber Threat Intelligence, Machine Learning.**

**Tipo de contribución:** *Investigación en desarrollo*

## I. Introduction

Cybersecurity threats are evolving rapidly, with attackers constantly developing new techniques to bypass security systems and gain unauthorized access to sensitive data [1]. With the rise of sophisticated cyberattacks, detecting and mitigating cybersecurity threats has become a critical area of research and development for cybersecurity professionals. In particular, the detection of cybersecurity threat mutations has emerged as a key challenge for organizations seeking to protect their data and systems. However, as attackers become more sophisticated, they often mutate their techniques, making them harder to detect and defend against using traditional tools [2].

The National Human Genome Research Institute defines a mutation as “the change in the DNA sequence of an organism” [3]. It is easy to relate a biological mutation with a cybersecurity mutation, where the change is not in the DNA sequence of an organism, but in the behaviour of entities affected by a potential cybersecurity threat. Therefore, detecting cybersecurity threat mutations requires a deep understanding of the tactics and techniques used by threat actors, as well as an ability to analyse large volumes of data to identify suspicious activity. As traditional detection methods are no longer sufficient in detecting these evolving attacks, organisations must leverage advanced technologies such as machine learning and behavioural analytics to detect emerging threats in real-time [4].

Currently, most environments use atomic indicators observed in the wild to detect known threats and heuristics to detect unknown threats, leading to analysts spending a lot of time on triage and contextualization. This research proposes a detailed methodology and a taxonomy to detect cybersecurity threat mutations on known threats within a network, based on specific indicators of behaviour that characterize the behaviour of users against those threats. Thus, multiple techniques are evaluated to estimate the likelihood of a threat being a mutation of a known threat and the information which both threats share. With this, malicious threats within a network are correctly exposed and related to others.

The scope of the study is limited to the data captured from over ten thousand users of a regional university in Spain. The data gathered correspond to anonymized application logs which describe specific actions carried out by the entities (e.g., email sent, web activity, or the number of established SSL connections) during 4 months of activity. Moreover, the data collected to identify possible malicious activity in the network is limited to the indicators of compromise gathered from the ICARO[1] feed. With this in mind, the main goals of the current study are:

- **Collection of data:** IoCs (Indicators of Compromise) must be gathered from external data sources to find matches across activity logs collected from users of a regional university in Spain.
- **Study of the features associated with user’s behaviour:** A deep analysis of ML weights associated with user’s behaviour, that were exposed to different malicious threats, must be conducted.
- **Analysis and classification of threat mutations:** An evaluation of a threat’s likelihood of being a mutation of a known threat using Spearman’s Rank Correlation Coefficient must be assessed with the information acquired in the points exposed above.

The remainder of this paper is structured as follows: Section II defines the main theoretical knowledge required to understand the work explored during the research. The core methodology followed to achieve the main goal of the paper is described in Section III. Moreover, Section IV illustrates

[1] ICARO website: https://www.incibe-cert.es/servicios-operadores/icaro

the results obtained. Finally, the conclusions of this research are discussed in Section V.

## II. Background work

Our study to detect cybersecurity threat mutations relies on Machine Learning and User and Entity Behaviour Analytics (UEBA) mechanisms. To the best of our knowledge, there is a gap in the current literature on how to solve this problem: no previous work in the field has been found that studies the importance of relating cybersecurity threats to improve current security defence systems. Within the landscape, a novel framework based on UEBA is established, allowing the automatic analysis of heterogeneous logs of entities and users to profile behavioural patterns in the network and calculate exposure to specific threats [5]. By employing a data-driven approach modeling user behaviour, as in [6], user risk scores can be computed enabling better anticipation and giving time to take preventive measures. However, current studies lack a methodology and a taxonomy to detect threat mutations on known threats to improve the reliability and effectiveness of those systems. As follows, we present an overview of the cybersecurity framework that covers the whole study, together with a review of UEBA.

### A. Cyber Threat Intelligence

Cyber Threat Intelligence (CTI) is based on a cybersecurity discipline that attempts to be a proactive measure of computer and network security, allowing the prevention or mitigation of cyber risks to protect the organizations [9], [10]. There are many different definitions to explain the term Cyber Threat Intelligence, and usually, companies tend to use their customized definition to distinguish their product [7]. In this research, we use the definition proposed by Robert M. Lee [8], which relies on the process and product resulting from the interpretation of raw data into information that meets a requirement as it relates to the adversaries that have the intent, opportunity and capability to harm. Particularly, threat intelligence involves the process of data transformation to information that relates to an adversary. However, Cyber Threat Intelligence is not only the output of the proactive measurements (i.e., the prevention or mitigation of cyberattacks), it is also the process (cycle) followed.

The intelligence life cycle, shown in Fig. 1, outlines the core steps to follow to understand and draw conclusions based on the data gathered, which are the same steps used during the detection method of this paper. The iterative process contains five phases an intelligence analyst must follow in pursuance of processing data to transform it into wisdom (i.e., intelligence), which in the last instance is led to action (i.e., decision) [11].

### B. User and Entity Behaviour Analytics (UEBA)

User and Entity Behaviour Analytics (UEBA) refers to the application of advanced analytics techniques to monitor, analyse, and detect anomalous patterns of behaviour exhibited by users and entities within an organization's network infrastructure. UEBA leverages machine learning, statistical modelling, and other analytical methods to establish baseline behaviours and identify deviations from those patterns, enabling the detection of potential security threats, insider attacks, and other malicious activities [4]. This technique has emerged as a powerful approach in the field of cybersecurity, providing organizations with the ability to proactively identify malicious behaviour within their network environment. There is a growing interest in Behavioural-based techniques, understanding how both the threat and the user behave during an incident, offering a resilient and flexible analysis in comparison with using Signature-based techniques, as the pyramid of pain describes in Fig. 2.

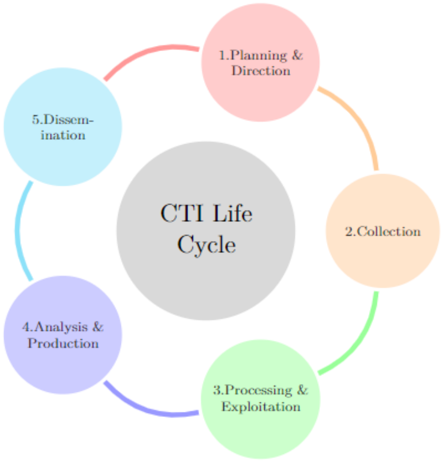


Fig. 1. CTI Life Cycle.

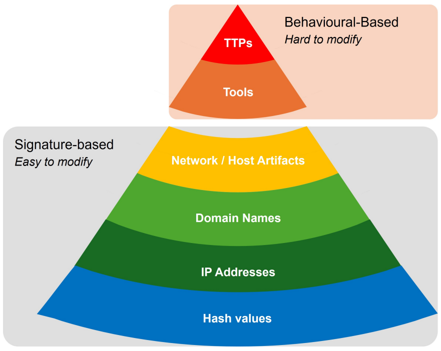


Fig. 2. Pyramid of Pain, targeting the Behavioural-based Methods, which are the hardest to modify by threat actors.

Accordingly, the paper aims to complement the work done in [5], [6] by developing a methodology and a taxonomy to detect cybersecurity mutations of known threats within a network, using classic threat intelligence and advanced threat profiling. Hence, the detection of cybersecurity threat mutations will enable professionals to distinguish between threat families characterized by a set of user behavioural patterns, resulting in a novel classification of potential victims of cybersecurity threats.

## III. Detection method

The methodology used for detecting cybersecurity threat mutations is based on the CTI Life Cycle (Fig. 1). Once the objectives of the investigation are stated, the next parts of the detection method explain in detail the process followed in each step of the cycle.

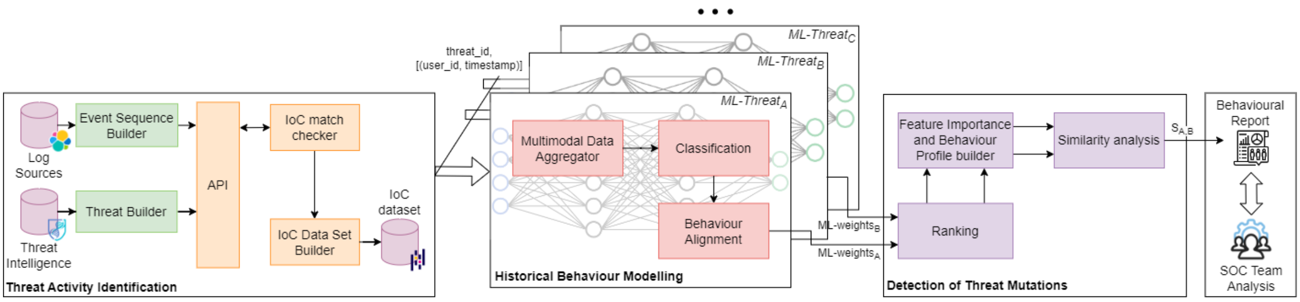


Fig. 3. Proposed architecture for the detection method

### A. Collection of data

The data collection process is a basic step in the research design, where relevant data must be collected to correctly process and analyse it. Therefore, some key design choices should be taken into consideration.

This study was conducted considering two main quantitative sources of information:

- *Log Sources:* The log sources are captured using a visibility agent, allowing the collection of real network logs generated by different entities from the university network; the logs captured are stored in a centralized data lake for the correct analysis. The logs captured correspond to application logs, that is, data collected from the application layer of the OSI Model. Hence, end-user information such as the activity of SMTP protocol, DNS protocol or HTTP protocol, among others, is captured and extracted by the *Event Sequence Builder*.
- *Threat Intelligence Sources:* The Threat Intelligence information, corresponding to real-world IoCs, is extracted using the OpenCTI tool. In particular, the IoCs included in OpenCTI are collected from the ICARO feed; this data feed contains Indicators of Compromise (IoCs) derived from real-world threat incidents (i.e., file hashes, IP addresses, domain names, URLs, email addresses, or hostnames) that come from identifying threats that have attacked a specific entity intending to enhance organizations' detection and prevention systems. INCIBE-CERT[2] is responsible for ensuring high-quality events published by the different entities attached to the ICARO service. Regarding OpenCTI, it uses STIX standard [16] as the reference language to create and share threat intelligence. Therefore, the *Threat Builder* is in charge of extracting and processing all available STIX Bundles in OpenCTI.

The Threat Activity Identification module relates the data extracted from both sources of information; the corresponding data is sent to the *IoC match checker* with the ultimate goal of finding IoC matches associated with real users by comparing each IoC with the information of network logs. Therefore, the final output of the Threat Activity Identification module is a dataset with a list of IoCs extracted from OpenCTI, where each IoC is associated with a vulnerable user and a timestamp, which is the date when the user became affected, as illustrated in Fig. 3.

### B. Processing and Exploitation

The data processing and exploitation step corresponds to the abstraction of the information previously collected and the identification of patterns, which is represented by the Historical Behaviour Modelling module shown in Fig. 3. As a result, the information on the matches found in the collection of data is extracted, which is associated with the vulnerable users and includes: the *threat_id* (STIX ID of the cybersecurity threat), *User ID* (ID of the vulnerable user) and *timestamp* (date where the match has been found). With this information, the module can process the matches found in the data collection stage and exploit them with the *Multimodal Data Aggregator*, the *Classification*, and the *Behaviour Alignment* processes extracted from the work done in [6]. In fact, the XGBoost (Gradient Boosting) supervised classification model is employed for training and testing the framework using XGBoost 1.7.3 library with the following hyperparameters for the *Desktop model*: (objective: binary:logistic, learning rate: 0.05, min child weight: 4, num boost round: 100, max depth: 4), and for the *Smartphone model*: (objective: binary:logistic, learning rate: 0.05, min child weight: 3, num boost round: 100, max depth: 6).

Hence, the format of the data given by the Historical Behaviour Modelling module gives us information about how the vulnerable user was behaving before having a match with an IoC, that is, being the victim of a cybersecurity threat. Specifically, it returns a set of ML weights associated with the behaviour of the vulnerable user (or group of vulnerable users). The Historical Behaviour Modelling module accurately analyses the behaviour of those users within the last 14 days to properly map the habits which caused them to fall into the threat. Accordingly, these features must be correctly interpreted and analysed by the stakeholder to determine risk patterns; if the risk patterns are similar enough between threats, a threat mutation is detected.

The ML weights the Historical Behaviour Modelling module can return are associated with DNS, HTTP, SSL/TLS, and SMTP application-layer protocol features. Therefore, these weights correspond to the feature importance of an ML model, where each ML model is associated with the cybersecurity threat and the vulnerable user or group of vulnerable users that have fallen into it (i.e., we will have as ML models as cybersecurity threats). Each ML weight can take a value from $[0-1]$: meaning high values a high utilization of the feature in front of other features depicted and, meaning low values a low utilization of the feature in front of others.

[2] INCIBE-CERT website: https://www.incibe.es/incibe-cert

### C. Analysis and Production

The information gathered so far must be analysed to generate the proper intelligence. Hence, a comparison between the behaviour of users against particular threats is done.

Using ranking techniques to compare feature importance between models is a practical and informative approach to gain insights into the relative importance of features. It helps in understanding the drivers of model predictions and identifying important features for further analysis or feature selection. In particular, some ranking techniques were studied to analyse and compare similarities and differences between the feature importance of the ML models (i.e., Spearman's Rank Correlation Coefficient [12], Kendall's Rank Correlation [14] and Rank-Biased Overlap [15]). Those methods are the most used nowadays to objectively compare a set of ranked lists and determine the correlation or similarity between them.

The chosen method in the project for detecting threat mutations is the Spearman's Rank Correlation Coefficient, which is a non-parametric technique for evaluating the degree of linear association or correlation between two sets of data [12]. Thus, it is defined as the Pearson correlation coefficient between the rank variables $R(X)$ and $R(Y)$, as illustrated in *Eq. (1)*. The Spearman's Rank Correlation Coefficient is a very easy-to-apply technique to efficiently compare ranked lists and detect basic similarities and differences between them. Moreover, it can be a very useful method for exploratory data analysis, where potential applications are numerous (e.g., analysis of ML data sets) [13].

$$S_{x,y} = \rho_{R(X),R(Y)} = \frac{cov(R(X),R(Y))}{\sigma_{R(X)}\sigma_{R(Y)}} \tag{1}$$

Finally, the design of the data analysis and production stage is illustrated in Fig. 3, which involves the Detection of Threat Mutations module. In particular, the output $S_{A,B}$ of the *Similarity analysis* step conveys the Spearman's Rank Correlation Coefficient between two specific threats, which determines the detection of threat mutations. If the similarity $S_{A,B}$ is high enough, a mutation is detected. Moreover, the *Feature Importance and Behaviour Profile builder* is in charge of generating complementary intelligence to the similarity analysis $S_{A,B}$. This intelligence is also presented in Section IV.

### D. Dissemination

The Dissemination stage is the last step of the CTI Life Cycle, where the information collected, processed and analysed must be shared with the community. Therefore, a Behavioural Report is created, which have the following properties:

- *Indicators of Compromise:* IoCs from which the threat has been created.
- *Indicators of Behaviour:* Ranked feature importance of the model, together with the behaviour profile representation presented in Section IV, to correctly visualize the features.
- *Mutations list:* A list of known threats (children) and their similarity with the parent object. Being the parent, the object is currently selected.

### E. Limitations

As the benefits of doing the research are numerous, it is also important to highlight the limitations we are exposed to. Therefore, the following limitations were encountered during the development of the project:

- Low maturity of the Historical Behaviour Modelling module: The Historical Behaviour Modelling module was not mature enough to fully calculate a user's behaviour to a specific cybersecurity threat. In particular, it needed a minimum number of 20–30 users to train the ML model due to a lack of individual user network data. Consequently, the Historical Behaviour Modelling module cannot accurately predict the behaviour of an individual vulnerable user exposed to a specific IoC; instead, a group of users must be chosen for analysis.
- Usage of ranking techniques to compare users' behaviour: The Historical Behaviour Modelling module was not able to rawly compare the weights between models. This limitation emerges when dealing with more than one ML model: the features of each model are computed locally, and they do not have any validity outside the dataset where they have been calculated. Accordingly, ranking techniques were used, not offering precise quantitative measures.
- Limited IoC feed: The IoCs gathered for finding vulnerable users were limited to the ICARO feed, where a restricted number of IoCs were given.

## IV. Results

### A. Resulting data

The present data corresponds to the information gathered during the IoC collection process, which is the output of the Threat Activity Identification module. However, due to the large data of the investigation, this section only covers the results gathered from IoCs associated with Phishing threats. For more information, the research in [17] covers the whole data associated with the current investigation which corresponds to Malware, Phishing, and TTP-based threats.

The Phishing threats investigated correspond to three different IoCs associated with email phishing campaigns. These phishing campaigns were launched in a controlled environment by the authors in [6] to the users of the regional university in Spain, where their anonymity was fully preserved during the experiment. On the one hand, Campaigns I-II are associated with gaining access, the emails sent are designed to ask the user to enter a third-party service with their credentials. On the other hand, Campaign III is associated with downloading and executing a third-party executable.

Table I presents the results from the phishing campaigns, where four different measures are profiled:

- *Open:* The user opens the phishing email.
- *Click:* The user interacts with a vulnerable artefact.
- *Engagement:* The user exposes credentials or launches the executable file.
- *Hit-rate:* Number of interactions with the email taking into consideration the total population: $Hit - rate = \frac{Engagement}{Population}$.

Therefore, phishing threats are organized in the following six different profiles presented in Table II. As observed, the

TABLE I
RESULTS OF PHISHING CAMPAIGNS I, II AND III.

| Campaign-ID | Population | Open | Click | Engagement | Hit-rate |
|---|---|---|---|---|---|
| Campaign I | 578 | 156 | 77 | 18 | 3.1 % |
| Campaign II | 377 | 87 | 47 | 67 | 17.8 % |
| Campaign III | 410 | 124 | - | 15 | 3.7 % |

phishing behaviour profiles are organized using the behaviour of users that fell into those phishing campaigns while using their desktop or smartphone. Specifically, in Campaign I and III or Campaign II or Campaign I, II and III. The fact that Campaigns I and III are not isolated is because of the low hit rate in those phishing campaigns. Accordingly, when grouping Campaign I and III, a larger number of samples are collected and a more realistic machine learning model is trained.

TABLE II
PHISHING THREAT BEHAVIOUR PROFILES.

| Phishing behaviour profile | Description |
|---|---|
| desktop_campaign_1_3 | Desktop features corresponding to Campaign I and III |
| desktop_campaign_2 | Desktop features corresponding to Campaign II |
| desktop_campaign_1_2_3 | Desktop features corresponding to Campaign I, II and III |
| smartphone_campaign_1_3 | Smartphone features corresponding to Campaign I and III |
| smartphone_campaign_2 | Smartphone features corresponding to Campaign II |
| smartphone_campaign_1_2_3 | Smartphone features corresponding to Campaign I, II and III |

### B. Experimentation

The proposed analysis encompasses the research done in the area of detection of cybersecurity threat mutations, based on vulnerable user's behaviour. Hence, the following studies are carried out to provide a detailed methodology and taxonomy in the field:

- *Study 1:* Behaviour modelling by threat. Cybersecurity threats are modelled by the behaviour of vulnerable users against that threat, resulting in a novel classification of threats. Each behaviour profile contains the behaviour of users when using the DNS, HTTP, SSL/TLS and SMTP application layer protocols.
- *Study 2:* Taxonomy of threats based on their degree of similarity. Behaviour profiles are compared using Spearman's Rank Correlation Coefficient to detect cybersecurity threat mutations.
- *Study 3:* Feature importance analysis. An expert analysis of the feature importance metrics from each cybersecurity threat is performed. With this, it is studied the importance of each feature when analysing the behaviour of vulnerable users associated with a threat to draw conclusions about suspicious user behaviours.

*1) Behaviour threat modelling:* The behaviour profiles obtained due to the behaviour modelling of users against specific threats of Phishing are presented in Fig. 4 and Fig. 5. In particular, the green area of the figures corresponds to the DNS features, the blue area matches with the HTTP features, the grey area concurs with the SSL/TLS features and the black area corresponds to the SMTP features. Hence, each square of each protocol's area is associated with an ML weight, having a total of 81 features: 32 corresponding to DNS, 20 associated with HTTP, 18 related to SSL/TLS and 11 related to SMTP protocol. The colour of each square indicates the ranking of that feature, suggesting light colours a high-rank feature (feature with a high importance for the creation of the ML model) and dark colours a low-rank feature (feature with a low importance for the creation of the ML model).

The aforementioned characterization gives a novel representation of the behaviour of users against specific cybersecurity threats. With this, it is possible to compare the feature importance rankings of several models side by side. This provides a clear visual representation of the similarities and differences in their ranking. In particular, it is easier for cybersecurity professionals to see which are the most representative features of the different threat models and discover possible relevant behaviour patterns between them. The results explained in this part are related to the similarity analysis presented in Section IV-B2.

On behalf of the phishing behaviour profiles related to smartphone users, several interesting patterns have been discovered. In particular, the DNS behaviour on (0,0), (0,1), (3,2), the HTTP behaviour on (3,4), (4,4), (0,5), (2,5) the SSL/TLS behaviour on (8,5), (6,7) and the SMTP behaviour on (5,8) is very frequent and similar among the profiles.

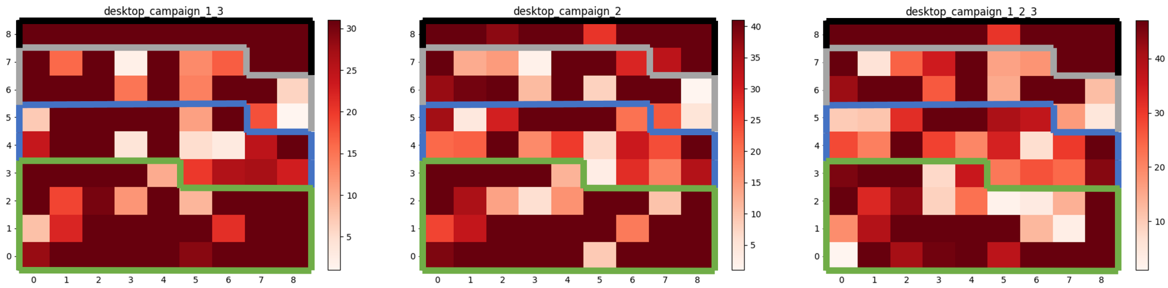
On the other hand, taking into consideration the phishing behaviour profiles related to desktop users, the DNS behaviour on (3,2), the HTTP behaviour on (3,4), the SSL/TLS behaviour on (8,5), (3,6), (5,6), (8,6), (1,7) is also very frequent and similar among the profiles.

*2) Taxonomy of threats based on their degree of similarity:* Taking into consideration the results gathered from this research, the deep understanding of the problem and, the domain knowledge of cybersecurity experts: a phishing mutation is considered when the Spearman's Rank Correlation Coefficient is equal to or exceeds the threshold value of 0.6. However, strong, moderate and weak threat mutations can be considered in dependence on the similarity's value, as presented in *Eq. (2)*.

$$Mutation = \begin{cases} Strong & \text{if } 0.8 < S_{x,y} \leq 1 \\ Moderate & \text{if } 0.7 \leq S_{x,y} < 0.8 \\ Weak & \text{if } 0.6 \leq S_{x,y} < 0.7 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The heatmap corresponding to the correlation, shown in Fig. 6, gives a general idea about the different behaviour between the usage of desktops and smartphones in front of different phishing campaigns. It is seen that the difference between those behaviours is high enough to not include them in comparison, as their correlation coefficient is lower than the threshold value ($S_{x,y} < 0.6$). For that reason, the behaviour between desktop and smartphone usage is not considered as a mutation. We only compare the behaviour profiles using the same type of device (i.e., desktop with desktop and, smartphone with smartphone), where the correlation coefficient is equal to or exceeds the threshold value ($S_{x,y} \geq 0.6$).

Accordingly, a significant similarity between phishing threats corresponding to smartphone usage is seen (*smartphone_campaign_1_2_3*, *smartphone_campaign_1_3*, *smart-*

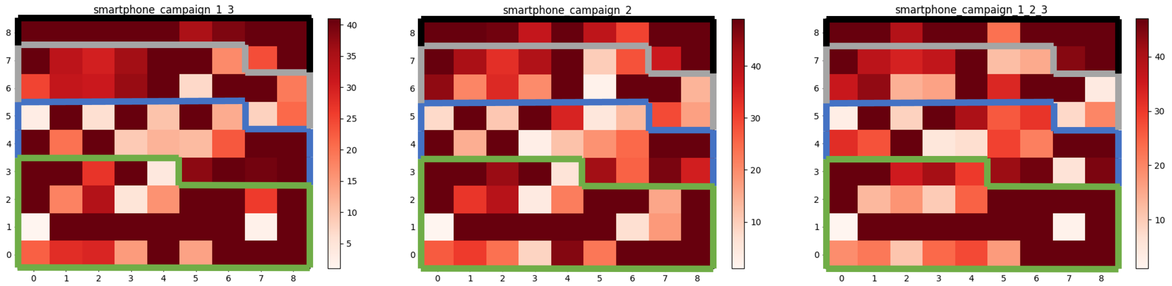


(a) Desktop Campaign I and III.

(b) Desktop Campaign II.

(c) Desktop Campaign I, II and III.

Fig. 4. Behaviour of desktop users corresponding to phishing campaigns.



(a) Smartphone Campaign I and III.

(b) Smartphone Campaign II.

(c) Smartphone Campaign I, II and III.

Fig. 5. Behaviour of smartphone users corresponding to phishing campaigns.

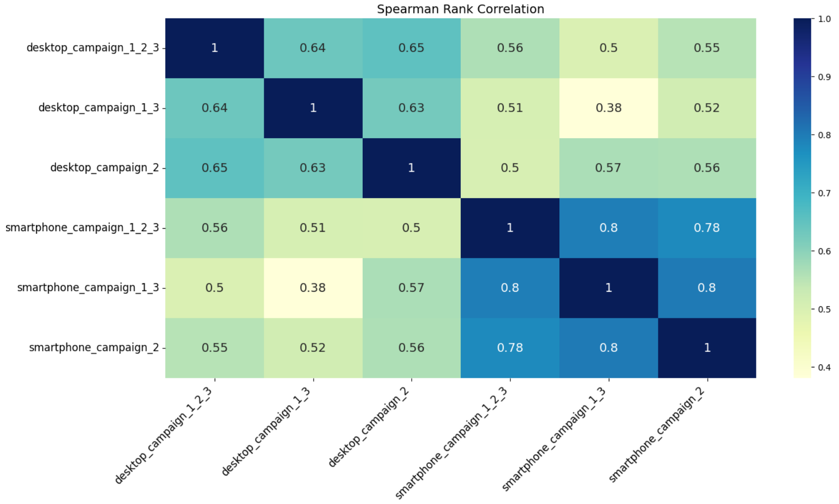


Fig. 6. Heatmap of the correlation among phishing campaigns.

*phone_campaign_2*). This means that the behaviour of users using smartphones against phishing campaigns is almost identical, leading to strong ($S_{x,y} \geq 0.8$) and moderate ($0.7 \leq S_{x,y} < 0.8$) mutations between those behaviour profiles. This performance can be associated with users' restrictive actions while using their smartphones, leading to more similar behaviour among users exposed to phishing threats. On the other hand, the similarity between behaviour profiles using desktops against phishing threats (*desktop_campaign_1_2_3*, *desktop_campaign_1_3*, *desktop_campaign_2*) is lower. In particular, a weak mutation is considered between them, as the highest similarity is 0.65 and the lowest is 0.63. In that case, we can assume that desktop users have a wider number of possible behaviours while using their desktop computer in comparison with smartphone users, leading to a lower similarity between desktop user profiles.

*3) Feature importance analysis:* The feature importance of each type of cybersecurity threat is useful for determining which are the most important characteristics that define each threat. Hence, a future deep analysis can be done to outline and improve the detection of threat mutations.

Fig. 7 and Fig. 8 represent the top-10 feature contributors related to the phishing threats for users associated with desktop or smartphone usage in front of potential risks. The horizontal axes illustrate the sum of ranked values from each behaviour profile, meaning low values for top-ranked features (very important feature: a feature that appears very frequently) and high values for low-ranked features (less important feature: a feature that does not appear frequently). On the other hand, the vertical axes depict the top-10 feature importance metrics. The key findings discovered during the feature importance study of the different threats are discussed below as an expert analysis, which results in a description of a set of user risk behaviours. These uncommon and unsafe practices should be taken into consideration to mitigate or, at least, reduce the impact of future cybersecurity threats.

- *HTTP request mean body length (http_request_body_len_ratio):* A high request body length implies that the users are submitting a substantial amount of data to the targeted servers. This behaviour might indicate potential data exfiltration, for instance, that users are used to easily interact with forms or input fields on websites; whereas in the case of phishing websites, they can potentially provide sensitive information that can be exploited by attackers.
- *HTTP response mean body length (http_response_body_len_ratio):* A high length of HTTP responses means a high probability of users downloading large amounts of files from web servers. This behaviour could indicate the delivery of malware or malicious content. Attackers might include large files or payloads within the response body to distribute malware or exploit vulnerabilities in the user's system. Users with this behaviour are at risk of these attacks.
- *HTTP POST method connections (http_method_port_ratio):* The high number of HTTP connections using the POST method suggests that users may be attempting to submit data to web servers frequently; this behaviour is typically used for sending data to the server, such as form submissions, file uploads, or other user input. It is a very high-risk behaviour when interacting with phishing websites, where users can submit sensitive information like login credentials or personal details.
- *HTTP GET method connections (http_method_get_ratio):* A high ratio of GET requests implies that the users are predominantly accessing websites and/or retrieving data from web servers. GET requests are commonly used for fetching resources from servers, such as web pages, images, or API data. In the context of a phishing campaign, this behaviour could easily expose users to phishing websites due to its frequent interaction with web servers.
- *HTTP 400 status code response (http_status_400_ratio):* The ”Bad Request” error, is returned when the server cannot process the client's request due to malformed syntax, invalid parameters, or other issues with the request itself. Therefore, this risky behaviour can be associated with accessing unfamiliar websites that no longer exist, which could be potential malicious websites that are only available on the web for a limited amount of time.
- *Non-working hours DNS (non_working_hours_dns):* A high ratio of DNS logs during non-working hours is strongly related to risky behaviour, meaning that users are used to using their smartphone inside the network during non-working hours (i.e., from 7 pm to 8 am).
- *Obsolete DNS Q (query) type labels (dns_qtype_obsolete_ratio):* The high ratio of obsolete DNS Q (query) type labels among the vulnerable users involved in the phishing campaign is a noteworthy characteristic. DNS Q types are used in DNS queries to specify the type of information being requested. Obsolete DNS Q types refer to query types that are no longer widely used or have been deprecated due to security concerns or protocol advancements.
- *Recursion Desired (RD) flag set to 1 (dns_recursion_desired_ratio):* When the RD flag is set to 1 in a DNS query, it indicates that the client making the query desires the DNS resolver to perform recursive resolution on its behalf, that is, perform all the necessary steps to resolve the domain name by recursively querying authoritative DNS servers starting from the root DNS servers until it obtains the final IP address associated with the domain name. Thus, this risky behaviour might be associated with high access to unknown websites, that is, websites not cached in the user's DNS resolver.
- *SSL connections with TLS v1.1 (ssl_version_ratio_v10):* The high number of SSL connections with deprecated TLS versions indicates that these vulnerable users are utilizing outdated or insecure TLS protocol versions. Deprecated TLS versions, such as TLS 1.0 or TLS 1.1, are known to have security vulnerabilities and may not provide adequate protection against cybersecurity attacks.

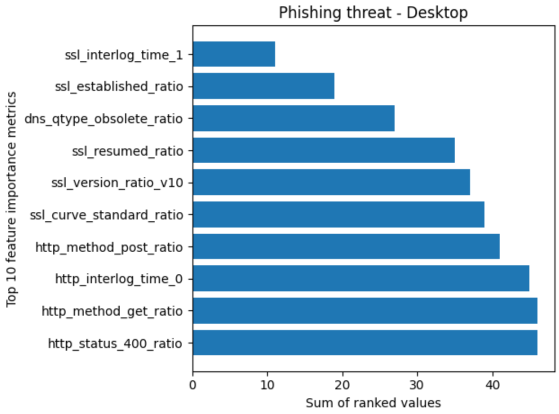


Fig. 7. Top-10 feature importance metrics corresponding to desktop phishing threats.

Our results show a first approach to the detection of cybersecurity threat mutations by providing expert analysis of threats' characterisation to effectively improve security defence systems for enterprise networks.

## V. Conclusions and Future Work

In this research, a novel methodology and taxonomy for identifying cybersecurity threat mutations within a network by combining machine learning algorithms with behavioural analysis were proposed. Our approach enabled us to categorize threats and determine the likelihood of a threat being

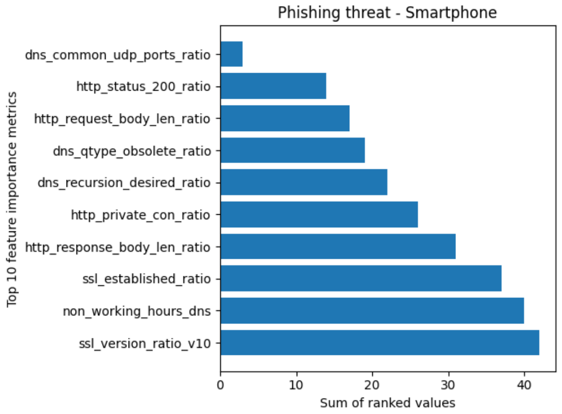


Fig. 8. Top-10 feature importance metrics corresponding to smartphone phishing threats.

a mutation of a known threat based on expert analysis. Additionally, we identified shared features between threats, including a novel representation of user behaviour profiles and a comprehensive analysis of their properties.

The results significantly contributed to the cybersecurity domain, providing valuable findings to the field and offering practical implications for cybersecurity professionals, ultimately contributing to the development of more robust and adaptive defence systems. Moreover, our approach to detecting threat mutations offers a practical means for enhancing incident response strategies, thereby strengthening overall security postures for organizations facing evolving cybersecurity threats.

The investigation of the ML features' importance in characterizing users' behaviour can be a promising avenue for future research. In particular, identifying shared features between models that do not add extra knowledge to the detection of cybersecurity threat mutations and substituting them for more valuable features can lead to more specific and strong behaviour profiles. Furthermore, investigating the possibility of creating specific real-time user behaviour profiles to detect mutations among already modelled behaviour profiles could early prevent users from falling into those cybersecurity threats. Finally, this research can be extended to different emerging domains such as the Internet of Things (IoT), where the detection of cybersecurity threat mutations will be important in the foreseeable future.

## Acknowledgements

The research received funding from the *Departament de Recerca i Universitats de la Generalitat de Catalunya* under grant 2021 SGR 01524.

## References

[1] European Union Agency for Cybersecurity, I. Lella, C. Ciobanu and E. Tsekmezoglou: ”ENISA threat landscape 2023 : July 2022 to June 2023”, 2023.

[2] E. Al-Shaer. ”A Cyber Mutation: Metrics, Techniques and Future Directions”. 1-1, 2016.

[3] D. A. Gilchrist: ”Mutation”, in *National Human Genome Research Institute*, 2023. Available: https://www.genome.gov/genetics-glossary/Mutation.

[4] M. A. Salitin and A. H. Zolait, ”The role of User Entity Behavior Analytics to detect network attacks in real time”, in *2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Sakhier, Bahrain*, pp. 1-5, 2018.

[5] A. Calvo, N. Ortiz, J. Guijarro and S. Siddiqui: ”OpenUEBA – A systematic approach to learn behavioural patterns”, in *Investigación en Ciberseguridad: Actas de las VII Jornadas Nacionales (JNIC 2022): 27-29 de junio de 2022, Palacio Euskalduna, Bilbao*, pp. 216–219, 2022.

[6] A. Calvo, S. Escuder, J. Escrig, M. Arias, N. Ortiz and J. Guijarro, ”A Data-driven Approach for Risk Exposure Analysis in Enterprise Security”, in *2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA), Thessaloniki, Greece*, pp. 1-9, 2023.

[7] M. Sahrom, S. R. Selamat, A. Ariffin and Y. Robiah: ”Cyber Threat Intelligence – Issue and Challenges”, in *Indonesian Journal of Electrical Engineering and Computer Science.*, vol. 10, pp. 371-379, 2018.

[8] R. M. Lee: “Intelligence Defined and its Impact on Cyber Threat Intelligence”, 2016. Available: http://www.robertmlee.org/tag/intelligence/.

[9] S. Saeed, S. A. Suayyid, M. S. Al-Ghamdi, H. Al-Muhaisen, and A. M. Almuhaideb: ”A Systematic Literature Review on Cyber Threat Intelligence for Organizational Cybersecurity Resilience”, in *Sensors* 23, no. 16: 7273, 2023.

[10] J. Friedman and M. Bouchard: ”Definitive Guide to Cyber Threat Intelligence”, 2015.

[11] V. Palacín: ”Practical Threat Intelligence and Data-Driven Threat Hunting”, ch. Chapter One - Cyber Threat Intelligence, 2021.

[12] Y, Dodge. ”Spearman Rank Correlation Coefficient”, in *The Concise Encyclopedia of Statistics. Springer, New York, NY*, 2008.

[13] T. D. Gauthier: “Detecting trends using spearman’s rank correlation coefficient”, vol. 2, no. 4, pp. 359–362, 2001.

[14] Kendall, M. G. “A New Measure of Rank Correlation”, in *Biometrika*, vol. 30, no. 1/2, pp. 81–93, 1938.

[15] W. Webber, A. Moffat, and J. Zobel, “A similarity measure for indefinite rankings”, one *ACM Trans. Inf. Syst.*, vol. 28, p. 20, 2010.

[16] STIX Version 2.1. Edited by Bret Jordan, Rich Piazza, and Trey Darley. 10 June 2021. OASIS Standard. Available: https://docs.oasis-open.org/cti/stix/v2.1/os/stix-v2.1-os.html. Latest stage: https://docs.oasis-open.org/cti/stix/v2.1/stix-v2.1.html, 2021.

[17] J. Domenech: ”Detection of Cybersecurity Threat Mutations”, Master Thesis, UPC, Escola Tècnica Superior d’Enginyeria de Telecomunicació de Barcelona, Departament d’Enginyeria Telemàtica, 2023. [Manuscript submitted for publication].