Computer Science and Information Systems 22(4):1533–1553 https://doi.org/10.2298/CSIS250131060G

Defining the Attractiveness Concept for Cyber Incidents Forecasting

Javier García-Ochoa, Alberto Fernández-Isabel, Clara Contreras, Rubén R. Fernández, Isaac Martín de Diego and Marta Beltrán

Rey Juan Carlos University
Department of Computing, ETSII
C/ Tulipán, s/n, 28933, Móstoles, Madrid (Spain)
{javier.garciaochoa, alberto.fernandez.isabel, clara.contreras, ruben.rodriguez, isaac.martin, marta.beltran}@urjc.es

Abstract. Cyber incident forecasting has several applications within the security field, such as attack projection, intention recognition, attack prediction, or situational awareness. One of the main challenges of these issues lies in analysing the proneness of an entity to be attacked by an adversary evaluating the relevance of different target features or behaviours. This paper presents a methodology that defines the *Attractiveness* concept to address this issue. *Attractiveness* is the possession of features or the exhibition of behaviours in entities that raise interest for potential adversaries. Thus, the more significant the *Attractiveness* value is, the greater the proneness of attacking could be considered. The concept is decomposed into three main branches: *basal attractiveness* (relevance of the entity in the world), *online reputation* (the opinion of the individuals and the reach of the entity), and *potential victimisation* (the interest that the entity arouses for potential attackers). Machine Learning (ML) methods in combination with Information Retrieval (IR) and text mining techniques have been proposed to gather relevant information and identify hidden patterns and relations in past security incidents. With this approach, potential targets could reduce their *Attractiveness*, focusing on those aspects that can be remedied. Alternatively, future risky situations could be predicted to better prepare for proactive protection, detection, and response. The proposal has been validated through several experiments.

Keywords: Attractiveness, Cyber incidents, Victimisation, Online reputation, Forecasting.

1. Introduction

In the current digital era, entities (companies, associations, public organizations, and organisations) are constantly in the spotlight for possible malicious intentions. Thus, they must be able to identify risky situations in the scope of cybersecurity to face potential threats coming from the Internet. This strengthens their defences to protect their information and systems from attackers if these risky situations materialise [15].

Risky situations are defined as those where a potential threat could materialise, resulting from a cybersecurity failure. This idea is closely related to incident forecasting, a widely addressed topic [6].

In the case of cyber-incident forecasting, the methods defined in this domain are beneficial for organisations to protect their digital assets and ensure business continuity. By

providing a quantifiable approach to cyber risk assessment, these methods enable these entities to estimate potential threats and vulnerabilities with a certain degree of confidence. Although forecasting does not imply absolute certainty, it offers valuable insights that support the development of a proactive security strategy, allowing for better strategic planning and resource allocation. This approach helps managers and decision-makers at different levels to understand the overall security posture, take preventive actions, identify weak points to reinforce them before they are exploited and prioritise countermeasures [23].

The cyber incident risk varies according to several variables to consider. For instance, specific sectors like energy, financial services, manufacturing, technology, or pharmaceuticals are usually more targeted by current threat actors [3]. Small to medium-sized businesses are also often targeted due to their lack of resources to defend themselves [11]. Moreover, entities holding valuable data (e.g. financial or personal) usually awake more interest, and those with a weak cybersecurity posture could become easier targets [25].

Consequently, the research presented here proposes a novel methodology to analyse the proneness of an entity to be attacked by an adversary using an evolution of the *Attractiveness* concept (firstly introduced in [5]). Thus, the main novelty of the proposal lies in considering static (i.e. firmographic features such as the entity sector, size, or revenue) and dynamic factors (i.e. reputation and dynamic factors such as the value of its information, the number of visible vulnerabilities, or the potential impact of an incident). A higher *Attractiveness* value indicates a higher probability of being targeted by adversaries. It is important to note that this analysis focuses on scenarios where attackers do not have a predefined target and look for easy targets to attack. In this point, it is important to remark that *Attractiveness* is an estimation, not an exact measure.

The proposed methodology uses three different branches to build the *Attractiveness* concept: *basal attractiveness* (relevance of the entity in the world), reputation on the Internet (the opinion of the individuals and the reach of the entity), and *potential victimisation* (the interest that the entity arouses for potential attackers and the possible Common Vulnerabilities and Exposures (CVE) detected). The main contribution of the proposal is the definition and combination of these branches to produce the final *Attractiveness* estimation.

Several data were collected for the experimental setup to validate the proposal. These data consist of information from cyber incidents reported by entities from multiple sectors. The dataset is completed with the static and dynamic features from these entities. DeNexus Inc. in the frame of the DICYME project (Ref: CPP2021-009025), provided support for this issue.

It is important to highlight that the only feasible approach relies on confirmed incidents and organisations that have publicly disclosed them. Notice that attempts cannot be quantified, as they are not publicly reported, making it impossible to collect and aggregate them across different organisations for analysis. Similarly, incidents affecting organisations that have not disclosed them remain unknown and, therefore, cannot be considered. Ultimately, anything that is not known cannot be accounted for in the analysis, which poses a significant challenge for comprehensive risk assessment.

This approach does not address aspects of an organisation's security posture, which is undoubtedly crucial in cyber risk quantification. Instead, it focuses on the organisation's posture as an entity, considering its more static characteristics, such as its scope, presence,

and engagement across networks and social media platforms. By analysing these factors, the proposed approach provides insights into the organisation's external exposure and perceived attractiveness to potential attackers. Consequently, when *Attractiveness* is combined with other methods and quantification approaches that incorporate internal security measures, a more comprehensive and accurate cyber risk estimation can be achieved.

The rest of this paper is organised as follows. Section 2 overviews the related work categorising existing cyber incident forecasting and similar methods. Section 3 sets out the motivation for this work and the research questions addressed. Section 4 presents the estimator for *Attractiveness* and introduces the proposed method. Section 5 details the development process and the dataset features used in the proposal. Section 6 validates the proposal, while Section 7 discusses the results focusing on strengths and limitations. Finally, Section 8 concludes and proposes future research lines.

2. Related Work

Forecasting cyber incidents consists of predicting future risk situations produced by attackers evaluating a specific set of features gathered from entities analysed in the present.

It is a complex task and faces at least two main challenges [21]. The first is the inaccessibility of adequate data and observations about past incidents. When available, the challenge is to extract relevant and reliable signals to treat sporadic and seemingly random acts of adversaries. This involves dealing with imbalanced ground truth labels and unconventional signals [20] gathered from public sources (incident databases, news, social media). The second is the ever-changing threat landscape [3], making it difficult to keep up with the latest threats and adapt forecasting models accordingly.

Intrusion Detection Systems (IDS) play a crucial role in this context by monitoring network traffic for suspicious activity and known threats, providing real-time alerts to potential security incidents [28]. Despite their effectiveness, IDS data can be overwhelming and often contain a high rate of false positives, adding to the complexity of accurately forecasting cyber incidents.

Different previous research has attempted to overcome these challenges with several approaches. Table 1 summarises the most significant prior work in this area. The *Goal* column captures the kind of proposed forecasting and can be used to predict the next adversary's move (attack projection, AP), to infer the adversary's motivation and goals (intention recognition, IR) or to anticipate upcoming cyber attacks (attack prediction and risk quantification, RQ). The *Data signals* column specifies the kind of data signals used to perform the forecasting, therefore, on which aspects the prediction depends. Finally, the *Model* column summarises the kind of Artificial Intelligence (AI) approach selected to achieve the forecasting.

Some studies emphasise the unpredictable nature of cybersecurity threats, suggesting that certain information about an attack can be used to predict subsequent attacks. For instance, network attacks are analysed using dependency graphs and intrusion responses [17]. Other approaches utilise honeypot data combined with probabilistic models like Markov Chains to identify patterns in attack propagation and target areas [7].

Real-time attack intention recognition is another focus area, employing neural network models to analyse known attack patterns and network behaviour [2]. Time series

Table 1. Summary of previous work on forecasting cyber attacks or incidents

Ref.	Goal	Data signals	Model
[17]	RQ	IDS alerts, intrusion responses and dependency graphs	Graph
[7]	AP	Honeypot evidences	Markov chain
[1]	RQ	IDS alerts and logs	Time series
[2]	IR	Known attacks patterns and signatures	Neural network
[19]	AP + RQ	Asset graphs, vulnerabilities and IDS evidences	Bayesian model
[26], [27]	RQ	Incidents landscape, geopolitical context, social mentions and sentiment	Bayesian model
[16]	RQ	Honeypot evidences	Neural network
[30]	RQ	CVE and Twitter	Neural network
[29]	AP	Asset graphs, vulnerabilities, attacker location and capability	Graph
[8]	RQ	IDS alerts and logs	Neural network
[12]	RQ	IDS alerts and logs	Neural network
[22]	RQ	IDS alerts	Neural network

analysis and dynamic risk assessment are also applied to predict incidents in critical cloud infrastructures [1].

In industrial systems, AI models such as Bayesian networks are usually used to evaluate the cyber risk and physical impact of potential attacks [19]. ML models, such as Support Vector Machines, Multi-layer perceptron, and k-Nearest Neighbours, are leveraged to forecast various types of cyber incidents based on past data [26].

Deep Learning frameworks, including Recurrent Neural Networks, capture long-term dependencies and non-linearity in the data to predict attack rates [16]. In addition, big data from social networks and vulnerability databases is used to identify cyber risks, offering strategies to mitigate these risks in critical infrastructures [30].

Nevertheless, a structural limitation is common to most of these approaches: they rely on internal telemetry (e.g., IDS alerts, raw network flows, honeypot traces or sensors) that presuppose a mature monitoring infrastructure and a willingness to share data. Many organisations, especially small and medium-sized enterprises (SMEs) and peripheral entities in supply-chain ecosystems, lack such instrumentation. This reliance introduces several biases: partial coverage, because entities without telemetry are systematically excluded; high noise levels, as IDS outputs can contain false positives, rendering costly preprocessing indispensable before model training; and a short prediction horizon, since the models become effective only seconds or minutes before (or during) the intrusion, offering little value for strategic planning or proactive investment decisions.

Moreover, these methods tend to focus on technical signs of threats rather than the intrinsic nature of potential victims. They assess anomalies in network activity but rarely incorporate information about the entity, its representation, or its visibility and appeal to

an adversary [5]. As a result, they overlook critical contextual or firmographic variables that often shape attackers' target selection.

Furthermore, signature- or pattern-based models are highly vulnerable to concept drift as attackers adapt their tactics, leading to rapid performance degradation. However, while attacker behaviours may evolve quickly, their target selection patterns tend to be more stable over time [9]. These shortcomings motivate the need for a complementary perspective that leverages externally observable, universally available signals, remains interpretable, and is still applicable even in the absence of network instrumentation.

These limitations are the foundation for introducing the concept of *Attractiveness*, a generalist approach to evaluate the proneness of an entity to be attacked. A comprehensive methodology is developed to integrate various features collected from entities and generate the final estimation. ML models are used in this process, mainly considering descriptive variables of entities such as entity country, category, or financial worth into a comprehensive measure designed to assess this proneness.

3. Motivation and Research Questions

This section illustrates the motivation of the proposal. The approaches previously presented have put into the spotlight different methods and techniques that can be applied to predict or forecast cyber incidents, attacks, or events. They could be organised according to some data signals: those related to the targets (IDS alerts, logs, asset graphs, and vulnerabilities), and those related to the attackers and their known behaviour (honeypot pieces of evidence, attack patterns, and signatures).

There is agreement regarding the convenience of applying predictive mechanisms to the targets instead of applying them to the adversaries. This approach makes sense, assuming that targets are not actively being novel or creative, trying to evade security controls as the adversaries are. Therefore, it is more likely that predictions adjust to reality to a greater extent when they are made on the targets than when they are made on the adversaries (there is much less reliable data available on their techniques, motivations, or objectives).

Even so, it is observed that there is great difficulty in deciding which signals should be taken into account about these targets because they are the most significant for making predictions. As shown in the previous section, the largest body of work focuses on predicting whether an attack is imminent, which can be considered an early warning solution. This prediction is based on internal target data collected and stored in IDS-type systems or logging solutions. In recent years, neural networks have proven to be suitable tools for this type of prediction (deep neural networks, recurrent neural networks, and long short-term memory), based on learning associations between different alerts and contextual information.

However, the primary research question is different in this proposal. Which specific high-risk features (static) or behaviours (dynamic) allow predicting the occurrence of an incident? In a less imminent time frame, not because indicators are being observed (IDS alerts, logs) that would allow for early warning. Furthermore, how can this prediction be made once identified or selected? Is it possible to define the *Attractiveness* concept so that it can be used to forecast the future number of cyber incidents?

This research question (RQ) leads us to more specific ones:

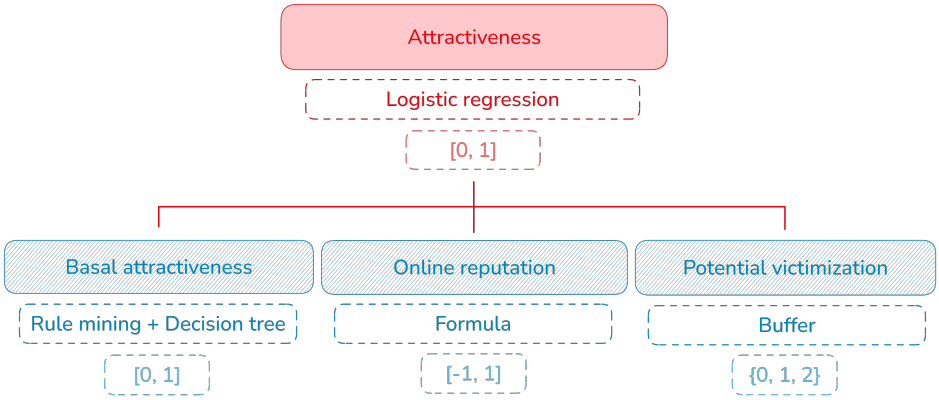

Fig. 1. Proposed high-level methodology

- RQ1: What specific features and behaviours enable the estimation of *Attractiveness*?
- RQ2: Can entities be grouped according to this *Attractiveness* to understand the obtained estimations? Can be the relationship between the entity *Attractiveness* or group *Attractiveness* and cybersecurity incidents identified?
- RQ3: How can this *Attractiveness* be used to forecast cybersecurity incidents? Are they inevitable, or is it possible to influence some of the aspects that influence *Attractiveness* to try to avoid them?
- RQ4: Are public data sources available to build the required data sets to work with this approach?

4. Estimation of Attractiveness

This proposal presents a methodology to estimate the *Attractiveness* value of each entity based on high-risk features (static) or behaviours (dynamic). This estimator considers the following critical aspects which can be evaluated in parallel: *basal attractiveness* (firmographic data), *online reputation*, and *potential victimisation*.

For each one of the aspects, the methodology gathers information about publicly confirmed cybersecurity incidents.

As a result of the methodology, a Logistic Regression algorithm joins the three aspects of the *Attractiveness*, producing a normalised value. This value represents the *Attractiveness* estimation for an entity being 0 the least attractive and 1 being the most attractive.

Figure 1 illustrates a general overview of this proposal. The next sections provide details about the estimation of the three considered aspects. Section 4.1 introduces the *basal attractiveness*, while Section 4.2 tackles the *online reputation*. Finally, Section 4.3 addresses the *potential victimisation*.

4.1. Basal attractiveness

The concept of *basal attractiveness* is based on the idea that certain entities are inherently more appealing to adversaries due to their static characteristics, commonly referred to as firmographic data (e.g., location, operational criticality, data sensitivity, and size).

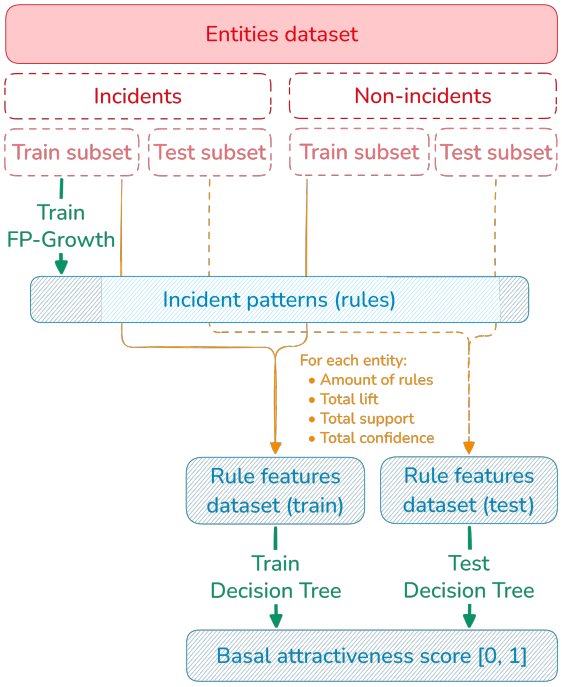

Fig. 2. Workflow of the *Basal attractiveness* model development

To model this, a dataset has been compiled containing both incidents and non-incidents involving various entities. Then, as shown in Figure 2, an association rule mining technique is used, specifically, the FP-Growth algorithm [32]. It is applied to uncover patterns within the incident data. For this, only a training subset of records corresponding to actual incidents is used, allowing the model to identify firmographic traits associated with increased risk.

Once the algorithm generates a set of association rules, these are used to evaluate both the incident and non-incident training subsets. For each observation, rule metrics such as support, confidence, lift, and the number of satisfied rules are computed and aggregated to build a new feature set for each entity.

This enriched dataset is then used to train a Decision Tree classifier, which outputs a binary prediction. The test subsets are subsequently applied to evaluate performance metrics on previously unseen entities.

The final output is a *basal attractiveness* score, ranging from 0 to 1. A score close to 0 indicates a low inherent risk of being targeted, while a score near 1 suggests a high baseline risk based on the entity's static attributes.

4.2. Online reputation

Online reputation tackles the acceptance and recognition that the entity has among the majority of people. The intuition behind the following proposition is explained by the

Table 2. Social media and networks considered for the *online reputation* estimation

Social media or network	
X (Twitter)	Tripadvisor
Facebook	Reddit
Instagram	Website domains
TikTok	Comments from website domains
YouTube	Forum domains

fact that an entity may be more attractive to an adversary based on its *online reputation* defined as the result of what users, customers, or employees write, communicate and share anywhere on the Internet based on their perceptions and experience in any moment of their relationship, direct or indirect, with the entity [26], [27], [30]. Thus, the positive and negative opinions or comments exchanged on social media and social networks can make an entity more attractive to certain attackers. This situation could change over time, finding an entity attractive during a period, and later unnoticed.

Social media is the dynamic content shared through corporate means (corporate website sections, corporate podcasts, corporate blogs). In these media, interaction is possible, but not expected or frequent. On the other hand, social networks allow the entity to share dynamic content using external platforms that are not controlled by the entity (e.g. Twitter, Reddit, or LinkedIn). Users may comment and chat about the shared content. This interaction is expected and more frequent.

Delving into reputation estimation, the Determ tool [13] has been used to gather specific information. Moreover, a formulation to produce the final indication has been built. In this sense, the *online reputation* (OR) indicator is based on the engagement (E), which can be defined as the ratio between the interaction (I) and reach (R) [10]. Reach represents the estimated number of people who read or mention a publication, while interaction (I) is the estimated number of people who answer it. Thus, if only one person has commented on a post with 100 views, the Engagement is $1/100$.

In this proposal, the *online reputation* is time-dependent because it represents a static picture in a specific moment, influenced by previous periods. Thus, OR^t is the temporal indicator computed per social media and network, where t is the current time. This indicator is built by using a weighted average of two other indicators: the entity engagement (EE^t) and the user engagement (UE^t):

$$OR^t = \alpha \cdot EE^t + (1 - \alpha) \cdot UE^t \,. \tag{1}$$

The EE^t indicator considers the engagement generated by the content published by the entity through its social media and networks. It is defined as:

$$EE^t = \frac{1}{E} \cdot \sum_{j=1}^{E} \left[\frac{EE_j^t}{\max\limits_{z=1\ldots t} EE_j^z} \right] , \tag{2}$$

where E is the number of social media and networks where the current entity interacts, and j represents each one of these media and networks. Therefore, EE_j^t means the entity engagement in the current time t in the social media or network j, and $\max\limits_{z=1\ldots t} EE_j^z$ is the maximum entity engagement considering the estimations previously achieved.

At the same time, EE_j^t can be defined as follows:

$$EE_j^t = \frac{1}{n_j^t} \cdot \sum_{i=1}^{n_j^t} \frac{IC_i^t}{RC_i^t}\,, \tag{3}$$

where n_j^t is the total number of mentions in the social media or network, while IC_i^t and RC_i^t are the interactions and reach in the social media j in each of the mentions i at the moment t.

The UE indicator considers the engagement generated by external mentions and comments (not generated or controlled by the entity) through social media and networks. It is defined as follows:

$$UE^t = \frac{1}{U} \cdot \sum_{k=1}^{U} \left[\frac{UE_k^t}{\max\limits_{z=1\ldots t} UE_k^z} \right], \tag{4}$$

where t is the current time, U is the number of mentions or interactions made by users related to the current entity, and k represents each one of these media and networks. Therefore, UE_k^t means the user engagement in the current time t in the social media or network k, and $\max\limits_{z=1\ldots t} UE_k^z$ is the maximum user engagement considering the estimations previously achieved.

At the same time, UE_k^t can be defined as follows:

$$UE_k^t = S_k^t \cdot \frac{1}{n_k^t} \cdot \sum_{p=1}^{n_k^t} \frac{IU_p^t}{RU_p^t}\,, \tag{5}$$

where n_k^t is the total number of mentions in the social media or network, while IU_p^t and RU_p^t are the interactions and reach in the social media k in each of the mentions p at the moment t. S_k^t estimates the sentiment value of each mention related to a social media or network k in a period t.

Then, the estimation of the sentiment value for each social media or network k is defined as follows:

$$S_k^t = \frac{positive_k^t - negative_k^t}{positive_k^t + negative_k^t}\,, \tag{6}$$

where $positive_k^t$ is the number of mentions labelled as positive in the social media or network k in the time t, and $negative_k^t$ is the number of mentions labelled as negative in the social media or network k in the time t. This formula produces a sentimental range between (-1) and $(+1)$.

Notice that the proposal does not consider the sentiment regarding entity engagement because it is associated with content published (and controlled) by the entity. Therefore, it can be assumed that it will always be eminently positive.

As a result, the *online reputation* estimator produces a value between (-1) and $(+1)$, where the lowest value represents a mediocre *online reputation* and the highest value indicates a good *online reputation* in social media and networks.

4.3. Potential victimisation

The intuition behind the *potential victimisation* (also a part of the dynamic behaviour) is explained by the fact that an entity may be more attractive to an adversary if it is often mentioned in underground forums or specific dark websites, or perceived as an approachable victim. Additionally, the entity may become more attractive if there are public data breaches.

Thus, forums, foreground and underground sites are monitored from the point of view of an adversary. There, knowledge and tools, data breaches, intelligence or business, the visibility of the entity infrastructure, and also a possible victim of an attack are analysed to detect if they are shared. These concepts are defined as follows:

- Risky visibility: number of direct mentions in monitored underground forums and dark web sites.
- Perceived ease of success: number of visible open ports and services, number of visible assets connected to the Internet, number of visible remote access protocols, number of visible third-party software dependencies, number of visible CVE, and number of visible wrong settings (i.e. default accounts, default or empty credentials, and valid leaked credentials).

These concepts are collected in two variables: Critic Info (number of mentions in dark web leaks) and Devices (number of devices connected to the Internet). These two variables are transformed into one with a buffer method in the following way:

- If both are 0, the *potential victimisation* value is 0.
- If one is 0 and the other is more than 0, the *potential victimisation* value is 1.
- If both are greater than 0, the *potential victimisation* value is 2.

As a result, the *potential victimisation* estimator produces a result between 0 and 2, where the lowest value indicates a very low *potential victimisation* and the highest value represents a very relevant *potential victimisation*.

5. Dataset development

This section details the development process of the dataset used in the experiments related to the proposal. Multiple primary and secondary data sources were considered, including public databases and data collection platforms, to generate this dataset. The quality and relevance of the data utilised are crucial to ensure the validity and reliability of the results obtained. Therefore, rigorous procedures have been implemented for cleaning, normalisation, and comparability of variables.

Delving into the set of firmographic variables related to *basal attractiveness*, they are defined as follows:

- Country: headquarters location based on country.
- Category: entity sector provided by RocketReach tool (see Table 3 for more details).
- Revenue: annual billing of the entity (USD).
- Earnings: annual profit of the entity (USD).
- Publicly traded: whether it is listed on the stock exchange (true/false).

Table 3. Sectors for entities in the *basal attractiveness* estimation

Sector	
Agriculture & Fishing	Media & Internet
Business Services	Metals & Mining
Chambers of Commerce	Organisations
Cities, Towns & Municipalities	Real Estate
Construction	Research & Technology
Consumer Services	Retail
Cultural	Software
Education	Telecommunications
Energy, Utilities & Waste Treatment	Trade, Supply Chain & Commerce
Finance	Transportation
Government	Healthcare
Hospitality	Insurance
Law Firms & Legal Services	Manufacturing

- Employees: size of the entity regarding the number of employees.
- Profitable: whether it is for-profit (true/false).

This information is gathered from RocketReach [31]. This tool allows consulting the names of entities. In this case, these names are those entities that appear in public databases where confirmed cyber incident victims are reported. These databases are the European Repository of Cyber Incidents (EuRepoC), Hackmageddon, Jam Cyber, TI Safe Incident Hub, KonBriefing, CISSM Cyber Attacks Database, and ICS STRIVE.

The dataset is completed with observations of similar entities that have not reported incidents. RocketReach is also used here to provide a set of these entities for each affected entity. Then, the category variable of the entities is checked and filtered. Finally, from the set of entities that share the category, one of them is randomly considered.

The generated dataset counts 675 observations, collected between September 2023 and May 2024. The proposed methods only consider incidents and non-incidents. Therefore, the data is transformed to a binary classification, obtaining 485 incidents and 190 non-incidents.

Regarding the distribution of incident types, they are organised as follows: 240 observations of ransomware, 120 observations of denial of service, 95 observations of data breach, 27 observations of destruction, and 189 observations of non-incident.

Concerning the properties of the variables, revenue has a wide range, indicating varied sizes of entities, with a mean of approximately $\$4.12$ billion. Notice that this mean is skewed by some entities, as evidenced by a maximum value of $\$271.57$ billion. Earnings also show a large spread and high variability, while most entities (87.7%) are not publicly traded. The mean number of employees is around $43,884$, the gain is highly skewed due to some large organisations. Most entities, however, are smaller as the median number of employees is 341.5, while more entities are profitable (52.6%).

Numerical variables have been discretised into four quartiles, resulting in the following ranges:

- Revenue: $[-2.72\text{e+}08, 6.79\text{e+}10)$, $[6.79\text{e+}10, 1.36\text{e+}11)$, $[1.36\text{e+}11, 2.04\text{e+}11)$ and $[2.04\text{e+}11, 2.72\text{e+}11]$

Table 4. Support, Confidence, and Lift for the top 10 Rules from the rule mining algorithm

Rule	Sup.	Conf.	Lift
Earnings=NaN → Employees=0.0	0.845	0.927	0.996
Employees=0.0 → Earnings=NaN	0.845	0.909	0.996
Earnings=NaN → Publicly traded=False	0.871	0.955	1.096
Publicly traded=False → Earnings=NaN	0.871	1.000	1.096
Publicly traded=False → Employees=0.0	0.807	0.926	0.995
Employees=0.0 → Publicly traded=False	0.807	0.867	0.995
Earnings=NaN, Publicly traded=False → Employees=0.0	0.807	0.926	0.995
Earnings=NaN, Employees=0.0 → Publicly traded=False	0.807	0.954	1.095
Publicly traded=False, Employees=0.0 → Earnings=NaN	0.807	1.000	1.096
Earnings=NaN → Publicly traded=False, Employees=0.0	0.807	0.884	1.096

- Earnings: $[-7.84\text{e+}09, -5.40\text{e+}08)$, $[-5.40\text{e+}08, 6.73\text{e+}09)$, $[6.73\text{e+}09, 1.40\text{e+}10)$ and $[1.40\text{e+}10, 2.13\text{e+}10]$
- Employees: $[2.50\text{e+}04, 6.25\text{e+}06)$, $[6.25\text{e+}06, 1.25\text{e+}07)$, $[1.25\text{e+}07, 1.875\text{e+}07)$ and $[1.875\text{e+}07, 2.50\text{e+}07]$

In the case of the *online reputation* case, its estimation involves a series of complex functions designed to extract and analyse data from social media mentions.

The data is pre-processed to segment time intervals into bi-weekly periods from the date the incident happened to the present. After that, the *online reputation* formula is applied to the data collected and included in the dataset.

The results of the formula range from -0.215 to 0.440 with a mean of 0.091, indicating generally positive reputations among entities. The variability (standard deviation of 0.103) suggests differences in how entities are perceived online.

Finally, two variables are included following the methodology for the *potential victimisation*. For devices, most entities do not report device-related incidents, with a 75th percentile value of 0. However, the maximum value of 100 indicates that some entities experience significant device-related incidents. For critical info, the values range from 0 to 314, with most entities (75th percentile) reporting at most 1 critical info incident, suggesting that breaches are not widespread.

6. Experiments

This section illustrates the viability of the proposal through various experiments. The three aspects of the methodology are considered. The relationships between *basal attractiveness*, *online reputation*, and their potential implications on incident occurrences are also explored. Each experiment is structured to validate theoretical assumptions through data analysis.

Delving into the experiments, a complete evaluation of the proposed methodology is presented. The following sections detail the specific experiments conducted to assess the three critical aspects: *basal attractiveness*, *online reputation*, and *potential victimisation*. Section 6.1 tackles the FP-Growth algorithm to estimate *basal attractiveness* based on specific entity features. Section 6.2 explores the relationship between entities regarding *online reputation* and their incidence of cyber incidents. Further analysis of the interaction

between *basal attractiveness* and *online reputation* with incident occurrences is presented in Section 6.3. Section 6.4 details the application of a Decision Tree classifier to predict incidents based on rule mining metrics. Finally, Section 6.5 integrates the insights from the previous experiments into a Logistic Regression model to estimate the combined effect of *basal attractiveness*, *online reputation*, and *potential victimisation* on predicting incident outcomes.

6.1. Rule mining algorithm for estimating basal attractiveness

The primary objective of using the FP-Growth algorithm [18] in this study is to estimate the *basal attractiveness* of entities. This method enables the discovery of underlying patterns within the dataset that contribute meaningfully to the perceived attractiveness, providing a structured approach to understanding the static factors associated with increased risk.

As a preliminary step, continuous variables in the dataset were discretised—i.e., converted into categorical intervals—to simplify the representation of the data and make it more suitable for pattern mining. The discretised data was then processed using the FP-Growth algorithm to identify frequent item sets. A minimum support threshold of 0.01 was used, ensuring that only patterns occurring in at least 1% of the transactions were considered. Other algorithm parameters were maintained at their default settings.

For mining frequent patterns, the dataset was split using an $80/20$ ratio, yielding 385 incident cases for training and a separate test set comprising 97 incident cases and 189 non-incident cases. It is important to highlight that the FP-Growth algorithm was trained exclusively on observations labelled as incidents.

Once extracted, the association rules were ranked by their *support* values, which indicate how frequently the associated item sets appear within the incident data. Rules with higher support are considered more representative and were analysed in detail to assess their contribution to the *basal attractiveness* estimation.

Table 4 presents the top 10 rules with the highest support. For instance, one of the most frequent rules reveals that entities experiencing cyber incidents tend to have few employees and lack available data on their annual earnings. This suggests that smaller organisations with limited financial transparency may be more vulnerable to attacks. Such findings can inform cybersecurity policies by encouraging targeted support for small enterprises that might otherwise lack the resources or visibility to manage cyber risks.

6.2. Relationship between online reputation and cyber incidents

In this experiment, the *online reputation* is estimated to assess the relationship between an entity's reputation in social media and social networks and the occurrence of cyber incidents. The α parameter of Equation 1 is fixed to 0.5 to provide neutral relevance to each part of the equation. The dataset was divided into two groups: those observations associated with incidents and those with non-incidents.

The non-parametric Wilcoxon rank-sum test is selected to assess whether there are statistically significant differences in *online reputation* scores between groups defined by their incident status [14]. This test compares the median of the *online reputation* scores of both groups. This approach is more robust to outliers and non-normal data distributions than mean comparisons used in t-tests.

A statistic W equals 49698 with a p-value of 0.06408 is provided. This result suggests no statistically significant difference in the distributions of *online reputation* scores between the non-incident and incident groups at the conventional 0.05 significance level. However, the p-value is close to the threshold, indicating a potential trend that could be significant with a larger dataset or different grouping methods.

The findings indicate that, while there is a visible difference in the median *online reputation* scores between the groups, it is not statistically significant under the current experimental setup. The marginally high p-value suggests a potential pattern in which *online reputation* could influence incident outcomes, albeit not strongly enough to be deemed significant.

6.3. Relationship between basal attractiveness and online reputation with incident occurrences

This analysis is focused on investigating the association between *basal attractiveness* and incident occurrences, as well as *online reputation* and incident occurrences, utilising the Chi-squared test of independence.

It is a non-parametric test to detect if two categorical variables are independent of each other across different groups [4]. It compares the observed frequencies in the data against the expected frequencies, which are calculated under the assumption that the variables are independent.

The data includes recorded incidents and non-incidents categorised by levels of *basal attractiveness* and *online reputation*. The data is grouped as follows:

- *Basal attractiveness*
  - Level 0: 119 non-incidents, 4 incidents.
  - Level 1: 9 non-incidents, 9 incidents.
  - Level 2: 1 non-incident, 57 incidents.
- *Online reputation*
  - Range $[-1, 0]$: 36 non-incidents, 29 incidents.
  - Range $(0, 1]$: 93 non-incidents, 41 incidents.

The Chi-squared test for *basal attractiveness* produced a statistic of $X^2 = 157.98$ with 2 degrees of freedom. The extremely low p-value ($< 2.2e - 16$) indicates a highly significant statistical association between *basal attractiveness* levels and incident occurrences. This suggests that *basal attractiveness* is not independent of incident status, with higher *Attractiveness* levels correlating with a higher frequency of incidents.

The Chi-squared test for *online reputation* produced a statistic of $X^2 = 3.1823$ with 1 degree of freedom. The p-value of 0.07444 suggests that while there is a notable trend, the association between *online reputation* ranges and incident occurrences does not reach conventional levels of statistical significance ($p < 0.05$). However, the p-value close to the threshold indicates a potential mild association that might become significant with a larger sample size or different categorisation of reputation scores.

Although the test for *online reputation* did not reach statistical significance independently, it is worth considering the potential interplay between *online reputation* and *basal attractiveness*. It is possible that *online reputation* could interact with *basal attractiveness* to influence incident occurrences in ways that are not captured when these variables are considered separately, this would be explored in further experiments.

6.4. Prediction of incidents based on rule mining metrics

This analysis consists of constructing a predictive model utilising a Decision Tree classifier. This model aims to forecast incidents by leveraging metrics obtained through rule-mining techniques.

A new dataset with 286 observations was prepared by labelling entities as 0 for non-incidents and 1 for incidents using the rule mining algorithm results for the test observations of the original dataset. The features extracted for the rule mining algorithm included counts of rules and their aggregated measures of support, confidence, and lift.

Once the new dataset is built, two steps are addressed: data preparation and splitting, and model training. The first splits the new dataset into training and testing groups with a 70-30 ratio, ensuring a balance between the learning and validation capabilities of the model. The second trains a Decision Tree classifier with specified hyper-parameters (10 as minimum samples split, 5 as minimum samples leaf, 5 as maximum depth, and 42 as random state). These values have been selected to prevent over-fitting issues while maintaining the model's generalisation ability.

The Decision Tree model achieved robust performance in the test with the following metrics: Precision: 0.87, Accuracy: 0.89, Recall: 0.74, and Kappa Statistic: 0.64.

Therefore, the model effectively classifies entities based on the rule mining metrics, with high Accuracy and a good balance between Precision and Recall. The high Precision rate indicates that the model is reliable in predicting incidents when it classifies an entity as such, while the Recall rate shows that it captures a significant proportion of actual incidents.

6.5. Prediction of incidents based on basal attractiveness, online reputation and potential victimisation

This last experiment aims to demonstrate that integrating the three critical aspects of the methodology improves the obtained results. A new dataset is created containing the three measures previously estimated as follows, with the attributes of the observations from the FP-Growth test:

- *Basal attractiveness*: output of the Decision Tree probability applied to the Decision Tree test dataset and trained with the output of the FP-Growth algorithm.
- *Online reputation*: output of Equation 1 with the α parameter fixed to 0.5.
- *Potential victimisation*: buffer applied to both columns of this category, getting a value in $\{0, 1, 2\}$.

A Logistic Regression model was built using the R package caret, which eases robust model building and evaluation. This package was selected due to its comprehensive array of functions that not only streamline model training but also provide extensive tools for tuning and evaluating model performance [24].

The experiment consists of three steps: data preparation, model training, and evaluation. Firstly, the dataset was read and processed to remove unnecessary entity identifiers and convert key variables into categorical forms suitable for analysis. Thus, the original dataset was discretised, while the information related to *potential victimisation* was transformed into a binary factor indicating the presence or absence of the two evaluated conditions. Then, the data were randomly split into a training set (80%) and a test

Table 5. Comparison of the metrics obtained by the Logistic Regression model for training and testing data

Metric	Training Data	Testing Data
Accuracy	0.931	0.923
Kappa Statistic	0.847	0.835
Recall	0.875	0.929
Precision	0.925	0.867

set (20%), using stratified sampling to maintain the proportion of incidents across these sets. In the next step, the Logistic Regression model was trained on the training set using cross-validation (5-fold) to optimise model parameters and prevent over-fitting. The model included the three outcomes of the aspects as predictors. The Logistic Regression model maps a linear combination of the predictors to a value between 0 and 1, representing the probability of an incident. A threshold is applied to the predicted probability of making a binary decision. If the predicted probability is greater than or equal to 0.5, the outcome is 1 (incident). If it is less than 0.5, it is classified as 0 (non-incident). In the last step, model performance was assessed on training and testing sets using confusion matrices and associated statistics to measure Accuracy, Recall, Kappa, and Precision.

Results are summarised in Table 5, showing key performance metrics for training and testing data. The model maintained high-performance metrics across training and testing datasets, with consistent metrics reported. The model exhibits strong accuracy, maintaining over 0.92 in both datasets. The Kappa statistic, which measures agreement beyond chance, indicates excellent model reliability with values of 0.847 and 0.835 for training and testing, respectively, suggesting that the model is consistent in its predictions across different data sets.

The increase in Recall from 0.875 in training to 0.928 in testing highlights the model's enhanced ability to identify positive cases in unseen data and generalise well without being overly fitted to the training data. This balance is crucial for practical applications where false positives and negatives carry significant implications.

The proposed model demonstrated high Accuracy and Recall. This high performance suggests that the discretisation of *basal attractiveness* and *online reputation*, including the binary variable of *potential victimisation*, provides a strong foundation for identifying patterns associated with incident occurrences.

7. Lessons Learned

This section synthesises the insights and key observations from the experiments designed to forecast possible cyber incidents using the *Attractiveness* concept. These findings contribute to a deeper understanding of the complex dynamics in cybersecurity threat assessment and the relevant features that make entities prone to suffering attacks.

Firstly, it is important to note that the developed methodology can be integrated into cybersecurity platforms to produce more robust predictive tools.

Regarding the selected ML, the FP-Growth algorithm to estimate the *basal attractiveness* revealed significant patterns correlated with cyber incident susceptibility in entities.

The discretisation data process proved invaluable in identifying these patterns, enhancing the understanding of key vulnerability factors.

The Decision Tree model achieved high Accuracy and Precision in classifying potential incident occurrences. This success illustrates the efficacy of ML approaches in extracting actionable intelligence from complex datasets.

The impact of *online reputation* on incident occurrences was evaluated using the Wilcoxon Rank-Sum test. Subtle, yet insightful, differences were found between affected and unaffected groups. This underscores the potential of nuanced statistical methods in identifying marginal trends. A statistically significant association between *basal attractiveness* and cyber incidents was confirmed when the Chi-square test of independence with *basal attractiveness* and *online reputation* was used. Thus, robust evidence is provided on how the physical infrastructure of attractiveness relates to the proneness of suffering possible attacks.

Integrating all the relevant variables into a Logistic Regression model enhanced predictive performance, affirming the value of synthesising multiple data sources and analytical perspectives.

Finally, despite its contributions, this proposal faces some limitations. In the scope of data and the generalisability of some findings, the data collected is limited and may not capture all the dimensions that influence cyber risk and increase or decrease the *Attractiveness*. Another limitation arises primarily because not all entities report cyber incidents when they occur. There can be various reasons for this lack of reporting, such as concerns about reputation damage or financial implications. Consequently, the data available for analysis might skew towards more transparent organisations or those mandated by regulation to disclose cybersecurity issues. Future studies could address this limitation by incorporating methods to estimate unreported incidents or using anonymised data contributions to encourage fuller disclosure from a wider range of entities.

8. Conclusions

This paper has introduced a novel methodology based on the *Attractiveness* concept. The proposed approach facilitates cyber incident forecasting by identifying entities prone to cyber attacks through three perspectives: the entity's relevance, its *online reputation*, and the interest it generates among potential attackers.

Regarding the first research question (RQ1), the study has identified specific static and dynamic features that enable the estimation of *Attractiveness*. Static features, such as firmographic data (e.g., sector, size, and revenue), establish a foundational risk profile by reflecting the inherent characteristics of an entity. Dynamic features, including *online reputation*, the number of visible vulnerabilities, and media exposure, provide a temporal dimension to the evaluation, capturing fluctuations in risk over time. The results indicate that combining these static and dynamic factors produces a more robust estimation of *Attractiveness*, offering insights into both baseline risk and evolving exposure.

For the second research question (RQ2), the findings confirm that entities can be grouped effectively based on their *Attractiveness* scores, revealing clear patterns between these groupings and the likelihood of experiencing cyber incidents. This analysis highlights a strong relationship between *Attractiveness* and cyber incidents, providing organi-

sations with actionable benchmarks for comparing their risk levels to similar entities and prioritising targeted interventions.

The third research question (RQ3) explored the use of *Attractiveness* for forecasting cyber incidents and its potential to influence risk mitigation. The results demonstrate that *Attractiveness* is a powerful predictor when used in conjunction with other data sources, such as threat intelligence and security posture. While certain factors contributing to *Attractiveness*, such as industry sector or size, are fixed and difficult to modify, others, such as reducing exposure by managing *online reputation* or addressing visible devices, can be proactively influenced to lower risk levels. This finding underscores the value of a proactive approach tailored to an entity's specific *Attractiveness* profile.

Finally, addressing the fourth research question (RQ4), the study confirmed the viability of using public data sources to build the required datasets for this methodology. Publicly available information, such as disclosed cybersecurity incidents, financial reports, social media activity, and vulnerability databases, provided the foundation for the analysis. However, the research acknowledges significant limitations due to the reliance on publicly reported data, as undisclosed incidents and unreported attack attempts remain inaccessible. Despite these constraints, the findings demonstrate that publicly available data, when combined with robust analytical techniques, offers a strong basis for cyber risk quantification and forecasting.

Future research could expand on this work by incorporating more observations into the dataset to verify the findings presented here. Furthermore, integrating real-time data streams could significantly enhance the model's performance. Additional efforts could focus on integrating intelligence from cyber threat actors, such as information from underground forums or dark web activities, to enhance understanding of adversarial behaviour and refine the *Attractiveness* concept further. The integration of AI-driven predictive analytics into real-time cyber defence systems could also be explored. These advancements would provide even greater value for organisations seeking to anticipate and mitigate cyber risks.

Acknowledgments. This work has been funded by the Spanish MICINN under the CPP program in the DICYME project (Ref: CPP2021-009025), partially funded by the XMIDAS project (PID2021-122640OB-I00), and supported by DeNexus Inc.

References

1. Abdlhamed, M., Kifayat, K., Shi, Q., Hurst, W.: A system for intrusion prediction in cloud computing. In: Proceedings of the International Conference on Internet of Things and Cloud Computing. pp. 1–9 (2016)
2. Ahmed, A.A., Mohammed, M.F.: Sairf: A similarity approach for attack intention recognition using fuzzy min-max neural network. Journal of Computational Science 25, 467–473 (2018)
3. Ardagna, C., Corbiaux, S., Van Impe, K., Sfakianakis, A.: ENISA ThreatLandscape 2022. European Agency for Cybersecurity (2022)
4. Argyrous, G., Argyrous, G.: The chi-square test for independence. Statistics for Social Research pp. 257–284 (1997)
5. Awan, M.S.K., Dahabiyeh, L.: Corporate attractiveness index: A measure for assessing the potential of a cyber attack. In: 2018 9th International Conference on Information and Communication Systems (ICICS). pp. 1–6. IEEE (2018)

6. Bakdash, J.Z., Hutchinson, S., Zaroukian, E.G., Marusich, L.R., Thirumuruganathan, S., Sample, C., Hoffman, B., Das, G.: Malware in the future? forecasting of analyst detection of cyber events. Journal of Cybersecurity 4(1), tyy007 (2018)
7. Bar, A., Shapira, B., Rokach, L., Unger, M.: Identifying attack propagation patterns in honeypots using markov chains modeling and complex networks analysis. In: 2016 IEEE international conference on software science, technology and engineering (SWSTE). pp. 28–36. IEEE (2016)
8. Ben Fredj, O., Mihoub, A., Krichen, M., Cheikhrouhou, O., Derhab, A.: Cybersecurity attack prediction: a deep learning approach. In: 13th international conference on security of information and networks. pp. 1–6 (2020)
9. Cho, J., Eling, M., Jung, K.: Spatial cyber loss clusters at county level and socioeconomic determinants of cyber risks. North American Actuarial Journal 29(2), 345–389 (2025)
10. Cioppi, M., Curina, I., Forlani, F., Pencarelli, T.: Online presence, visibility and reputation: a systematic literature review in management studies. Journal of Research in Interactive Marketing 13(4), 547–577 (2019)
11. CrowdStrike Holdings Inc: 2023 Global Threat Report. https://www.crowdstrike.com/global-threat-report/ (2023), online accessed: 2024-09-02
12. Dalal, S., Manoharan, P., Lilhore, U.K., Seth, B., Mohammed alsekait, D., Simaiya, S., Hamdi, M., Raahemifar, K.: Extremely boosted neural network for more accurate multi-stage cyber attack prediction in cloud computing environment. Journal of Cloud Computing 12(1), 14 (2023)
13. Determ d.o.o.: Determ - ai media monitoring and analytics software. https://www.determ.com/ (2024), online accesssed: 2024-09-02
14. Divine, G., Norton, H.J., Hunt, R., Dienemann, J.: A review of analysis and sample size calculation considerations for wilcoxon tests. Anesthesia & Analgesia 117(3), 699–710 (2013)
15. Djajasinga, N.D., Fatmawati, E., Syamsuddin, S., Sukomardojo, T., Sulistyo, A.B.: Risk management in the digital era addressing cybersecurity challenges in business. Branding: Jurnal Manajemen dan Bisnis 2(2) (2023)
16. Fang, X., Xu, M., Xu, S., Zhao, P.: A deep learning framework for predicting cyber attacks rates. EURASIP Journal on Information security 2019, 1–11 (2019)
17. GhasemiGol, M., Ghaemi-Bafghi, A., Takabi, H.: A comprehensive approach for network attack forecasting. Computers & Security 58, 83–105 (2016)
18. Han, J., Pei, J., Yin, Y., Mao, R.: Mining frequent patterns without candidate generation: A frequent-pattern tree approach. Data mining and knowledge discovery 8, 53–87 (2004)
19. Huang, K., Zhou, C., Tian, Y.C., Yang, S., Qin, Y.: Assessing the physical impact of cyberattacks on industrial cyber-physical systems. IEEE Transactions on Industrial Electronics 65(10), 8153–8162 (2018)
20. Husák, M., Kašpar, J.: Towards predicting cyber attacks using information exchange and data mining. In: 2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC). pp. 536–541. IEEE (2018)
21. Husák, M., Komárková, J., Bou-Harb, E., Čeleda, P.: Survey of attack projection, prediction, and forecasting in cyber security. IEEE Communications Surveys & Tutorials 21(1), 640–660 (2018)
22. Jain, J.K., Waoo, A.A.: An artificial neural network technique for prediction of cyber-attack using intrusion detection system. Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN) ISSN pp. 2799–1172 (2023)
23. van der Kleij, R., Schraagen, J.M., Cadet, B., Young, H.: Developing decision support for cybersecurity threat and incident managers. Computers & Security 113, 102535 (2022)
24. Kuhn, M.: Building predictive models in r using the caret package. Journal of statistical software 28, 1–26 (2008)
25. Mandiant: M-trends 2023. https://www.mandiant.com/m-trends/ (2023), online accessed: 2024-09-02

26. Okutan, A., Werner, G., Yang, S.J., McConky, K.: Forecasting cyberattacks with incomplete, imbalanced, and insignificant data. Cybersecurity 1, 1–16 (2018)
27. Okutan, A., Yang, S.J., McConky, K., Werner, G.: Capture: cyberattack forecasting using non-stationary features with time lags. In: 2019 IEEE Conference on Communications and Network Security (CNS). pp. 205–213. IEEE (2019)
28. Panigrahi, R., Borah, S., Bhoi, A.K., Mallick, P.K.: Intrusion detection systems (ids)—an overview with a generalized framework. Cognitive Informatics and Soft Computing: Proceeding of CISC 2019 pp. 107–117 (2020)
29. Polatidis, N., Pimenidis, E., Pavlidis, M., Papastergiou, S., Mouratidis, H.: From product recommendation to cyber-attack prediction: Generating attack graphs and predicting future attacks. Evolving Systems 11, 479–490 (2020)
30. Subroto, A., Apriyana, A.: Cyber risk prediction through social media big data analytics and statistical machine learning. Journal of Big Data 6(1), 50 (2019)
31. Szymoniak, S., Foks, K.: Open source intelligence opportunities and challenges–a review. Advances in Science and Technology. Research Journal 18(3) (2024)
32. Wang, K., Tang, L., Han, J., Liu, J.: Top down fp-growth for association rule mining. In: Advances in Knowledge Discovery and Data Mining: 6th Pacific-Asia Conference, PAKDD 2002 Taipei, Taiwan, May 6–8, 2002 Proceedings 6. pp. 334–340. Springer (2002)

Javier Sánchez García-Ochoa was born in Toledo, Spain, in 2000. He holds a bachelor's degree in Cybersecurity Engineering from Rey Juan Carlos University (URJC) and a master's degree in Cybersecurity and Privacy from the Open University of Catalonia (UOC). He worked in the private sector for over a year before joining Rey Juan Carlos University in 2023 as a research staff member. His main research activity is carried out within the public-private collaboration project DICYME ("Dynamic Industrial Cyber Risk Modelling based on Evidence"), focused on dynamic cyber risk quantification. He has also contributed to several research articles and conference proceedings. Additionally, he is involved in various other research tasks and projects related to cybersecurity and data science.

Alberto Fernández-Isabel was born in Toledo, Spain in 1984. He received a PhD in Computer Science from Complutense University of Madrid (UCM) in 2015. He obtained a scholarship at the Spanish National Research Council (CSIC) as a technical assistant. He has been working for several years on European and national projects as a predoctoral and postdoctoral researcher. Since 2019 he is Assistant Professor at the Higher Technical School of Computer Engineering (ETSII) at Rey Juan Carlos University (URJC). He has authored more than 30 scientific articles and books. He completes his background with a Master´s degree in Artificial Intelligence and a Master´s degree in Information Systems. His research interests include intelligent agents, machine learning, data visualization, and natural language processing in various application domains, including distributed programming, sentiment analysis, agent-based collaboration and negotiation, smart cities, and simulations.

Clara Contreras is as an Associate Professor at the Department of Computing, Universidad Rey Juan Carlos, Madrid (Spain) and Cybersecurity Engineer at Siemens. Her research interests are Cybersecurity and Artificial Intelligence.

Rubén Rodríguez Fernández was born in Bembibre, León, Spain in 1973. He received a PhD degree in Artificial Intelligence from Rey Juan Carlos University (URJC). He also received a master's degree in data science from URJC and a master's degree in Artificial Intelligence from the Technical University of Madrid (UPM). He is part of the Data Science Laboratory high performance research group and has been an Assistant Professor at URJC since 2024. His research interests include active learning, explainable machine learning, generative artificial intelligence, and applied machine learning.

Isaac Martín de Diego was born in Campaspero, Valladolid, Spain in 1973. He received a PhD degree in Mathematical Engineering from Carlos III de Madrid University in 2005 (Extraordinary Doctorate Award). Since 2023 he is a full professor at the Higher Technical School of Computer Engineering at Rey Juan Carlos University (Associate Professor from 2018). He is the co-founder of the Data Science Laboratory and Head of the Sports Analytics Master at Rey Juan Carlos University. He has been head of the ERICSSON Chair on Data Science applied to 5G. He is the author of more than 100 articles. His research interests include methods, processes, and tools for Data Science in various application domains: explainability, sampling, complexity, performance evaluation, visualization, recommendation systems and security with a special interest in Machine Learning algorithms and a combination of information methods.

Marta Beltrán received the master's degree in electrical engineering from Universidad Complutense of Madrid (Spain) in 2001, the master's degree in industrial physics from UNED (Spain) in 2003 and the PhD degree from the Department of Computing, Universidad Rey Juan Carlos, Madrid (Spain) in 2005. She is currently an Associate Professor at this department (on leave). She has published extensively in high-quality national and international journals and conference proceedings in the areas of parallel and distributed systems, cybersecurity and privacy. Her current research interests are Cloud computing, Edge/Fog Computing and Internet of Things, specifically, risk management, identity management and privacy-preserving mechanisms for these paradigms.

Received: January 31, 2024; Accepted: August 10, 2025.