Marie Skłodowská Curie,
Research and Innovation Staff
Exchange (RISE)

European Commission
Horizon 2020
European Union funding
for Research & Innovation

Security ECONomics service platform for smart security investments and cyber insurance pricing in the beyonD 2020 netwOrking era

SECONDO

WP3 – Quantitative Risk Analysis and Data Analytics
Deliverable D3.1: Pricing Methods and Risk Modelling

Editor(s):	Sakshyam Panda (SURREY), Emmanouil Panaousis (SURREY/UoG)
Author(s):	Sakshyam Panda (SURREY), Emmanouil Panaousis (SURREY/UoG), Emily Parsons (UoG), Alexandra Dritsa (UPRC), Farnaz Mohammadi (UPRC), Sirivianos Michael (CUT), Nikolaos Salamanos (CUT), Constantinos Papadamou (CUT), George Kalantzantonakis (LST), Nikolaos Episkopos (FOGUS)
Dissemination Level:	**PU-Public**
Type:	**R-Report**
Version:	2

Project Profile	
Contract Number	823997
Acronym	SECONDO
Title	Security ECONomics service platform for smart security investments and cyber insurance pricing in the beyonD 2020 netwOrking era
Start Date	Jan 1st, 2019
Duration	48 Months

Partners

	University of Piraeus Research Centre	Greece
	UNIVERSITY OF SURREY	United Kingdom
	Cyprus University of Technology	Cyprus
	UBITECH LIMITED	Cyprus
	LSTech Espana	Spain
	Cromar Insurance Brokers LTD	Greece
	Fogus Innovations & Services P.C.	Greece
	University of Greenwich	United Kingdom

Document History

Version	Date	Author	Remarks
v1	12/08/2019	SURREY/CUT	Asset pricing methodologies
v1.1	03/12/2019	SURREY/UPRC	Table of Contents
v1.2	10/01/2020	SURREY	Asset characteristics
v1.3	23/01/2020	SURREY/UoG	Asset pricing literature
v1.4	03/02/2020	UPRC/UoG/LST	SEAM
v1.5	20/03/2020	UPRC/UoG/SURREY	Risk Model
v1.6	28/04/2020	SURREY/UoG	Asset identification and classification
v1.7	15/05/2020	SURREY/UoG/FOGUS	Asset Valuation methodology
v1.8	15/05/2020	UPRC	RAOHM-implementation
v2	15/06/2020	ALL	Initial draft
v3	24/06/2020	ALL	Final version

Table 1: Document History

Executive Summary

A particular aim of the SECONDO project is to design, analyze an implement a Quantitative Risk Analysis Metamodel (QRAM). In particular, QRAM will utilise advanced security metrics to quantitatively estimate the exposed cyber risks, taking into account important parameters not currently considered by existing risk analysis tools. Deliverable 3.1 "Pricing Methods and Risk Modelling" is dedicated to the documentation of the work done in the two tasks in WP3 "Quantitative Risk Analysis and Data Analytics", which are detailed below.

- **Task 3.1**: Methods for pricing tangible and intangible digital assets task focuses on the valuation of digital assets that might be categorized as both tangible and intangible. The outcome from this task will feed the Econometrics Module (ECM) (to be carried out in the phase of the project). The valuation of an asset is carried out considering both the tangible and intangible metrics such as cost of an asset, reputation damage and impact to business continuity. We develop a model to compute the lower and upper bound of the value of an asset to be used in Task 3.2 to determine the associated cyber risk from an asset.
- **Task 3.2**: Risk modelling task provide the Quantitative Risk Analysis Metamodel (QRAM) that quantitatively estimates the exposed cyber risks through the Risk Analysis Ontology and Harmonisation Module (RAOHM). As part of the task all concepts of the SECONDO ecosystem (e.g. risk, thread, attack type, behaviour, exploitation impact) have been represented formally expressing their need and applicability in risk assessment. The output of the QRAM and the Social Engineering Assessment Module (SEAM), which is used to obtain the likelihood of Phishing attack on employees of an organisation, are fed as inputs to RAOHM, which then harmonises them using a common vocabulary. This task provides a description of concepts and relationships to be used during the formulation of the formal SECONDO ontology, which is a shared conceptualisation of terms that interrelate with each other. For example, the manifestation of a threat would relate with a set of attack types (e.g. spearphishing).

Contents

List of Figures

List of Tables

1 Introduction

This section provides a brief overview on the most important objectives and tasks of the SECONDO project. It further provides a description on the scope of the tasks carried out in this deliverable together with a summary of the future tasks and objectives to meet the proposed outcome of the project.

1.1 Role of the Deliverable

The role of this deliverable is to provide detailed report on the design, implementation and evaluation of the modules and methods utilised by the SECONDO project for pricing methods and cyber risk modelling.

The present document has two main purposes:

- Describe methods to valuate assets considering both the tangible and intangible matrices.
- Provide the Quantitative Risk Analysis Metamodel (QRAM) that quantitatively estimates the exposed cyber risks.

1.2 Relationship to other Deliverables

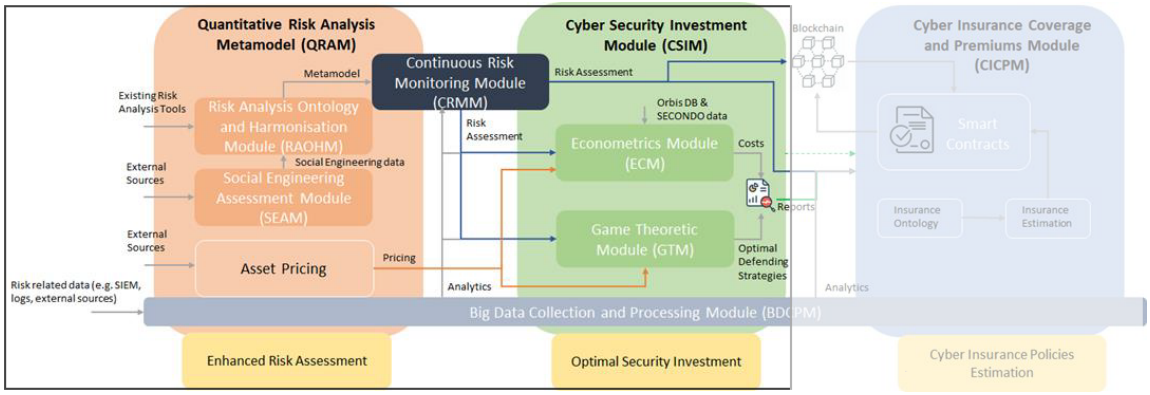


Figure 1: SECONDO Framework

- **D4.1 : Econometrics** – The Econometrics Module (ECM) that provides estimates of all kinds of costs of potential attacks as well as costs will respect the scenarios defined in Section 5 and the architecture described in Section 6 of the Deliverable D2.1.

- **D4.2 : Continuous Risk Monitoring and Blockchain** – The CRMM module will assess on a continuous basis the risk levels, including the performance of the implemented cyber security controls will respect the scenarios defined in Section 5 and the architecture described in Section 6 of the Deliverable D2.1.
- **D4.3 : Cyber Security Investments** – The CSIM module that will be responsible for inferring optimal investment plans will respect the scenarios defined in Section 5 and the architecture described in Section 6 of the Deliverable D2.1.

1.3 Structure of the document

Chapter 2 presents an overview on digital assets, their characteristics, and existing asset valuation methods. Chapter 3 presents the **Use Case 3**: *Cyber Insurance for innovative SME* in detail where we identify and prioritise valuable assets, and determine an valuation model to calculate the upper and lower bounds of the asset value using tangible and intangible metrics of an asset. Chapter 4 demonstrates the importance of assessing user behaviour against social engineering attacks and how the SECONDO aims to utilise this information for achieving better results. Utilising the asset values obtained in Chapter 3, attack likelihood obtained in Chapter 4, and results from OLISTIC, Chapter 5 calculates the associated risks for the SME while providing a detailed analysis of the RAOHM's functions and its architecture. Chapter 6 concludes this deliverable D3.1 and discusses the future work of the SECONDO project and the contribution of this deliverable to other modules.

2 Asset Pricing

In the remainder of the review, we shall cover literature that will aid us in choosing the best method for this task. Firstly, we cover asset management and then asset valuation. After this, we discuss the importance of valuation and pricing before exploring many of the different methods of valuation that could be used. Finally, we summarise what we have learnt.

2.1 Assets

An asset is an economic source of value [1], that is, it is any resource of economic value that an entity owns or controls with an expectation that it will lead to a benefit in future. The International Accounting Standards Board (IASB) defines an asset as "a resource controlled by the entity as a result of past events and from which future economic benefits are expected to flow to the entity" (IASB, 2015). The management of an asset throughout its life-cycle is crucial for guaranteeing a favourable return and ensuring defined services and operations [3]. The management of assets includes planning and support for investment decisions, access, acquisition and maintenance throughout its life-cycle, with an objective to optimise the economic value of the asset while minimising the associated service and operational cost.

ISO/IEC 27001 is an industry-standard publication that observes the different management systems that could be used within businesses. The key factor this highlights is that the assets must be managed in the correct way, considering different aspects. In the case of valuation, we see that from management, responsibility, class and handling of assets are important to make sure that risks are controlled and mitigated. Within this section, the value of these assets can play a key part in what would be lost, and these groups can affect how an asset is valued, showing that management is different depending on the business. Another industry standard publication (ISO/IEC 27005) that targets information security and risk management, within this is the risk identification phase, where the identification of risk and assets is key. It addresses the valuation of different assets providing more information into the sub process of identifying assets. The standard explores the process of asset valuation and what should be considered when asset valuation takes place. Firstly, the classification of assets that regard the importance of fulfilling business objectives must be determined, then the valuation is determined with the use of two different measures. The first is the replacement value of the asset, the cost of replacing information, and the second is the consequence of loss which is the impact from a successful breach of the asset. ISO/IEC 27005 splits the assets into 2 main groups, primary and sup-

porting. The primary assets contain information and business progress, which are critical to achieve core business functionalities. The supporting assets are assets that the primary assets rely on in some way, for example, software and hardware.

2.1.1 Digital Assets

Digital and traditional (those not digital) assets can be broadly categorised into tangible and intangible assets. Tangible assets are physical and measurable and come in two main forms, current and fixed. Current assets are short term where fixed assets are long-term. On the other hand, intangible assets are expressed in discrete numerical form [1] to be used by computing devices and thus can be non-physical and hard to value due to the uncertainty in the value at a given point of time. Information being an intangible asset (not possessing a physical form) has relevant attributes to provide potential services and being able to bring economic benefits to the owning entity [4]. Oppenheim et al. [5] describes information asset as an umbrella category which includes data, information and explicit knowledge managed as a single unit so that it can be understood, shared, protected and exploited. On the other hand, a digital asset is any asset that exists digitally and has defined rights of use. With this, according to [1], the digital assets include, digitised assets, born digital assets and digital operational assets.

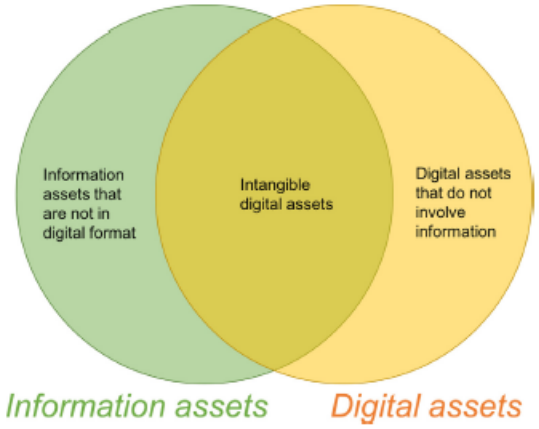


Figure 2: Categorisation of information asset and digital asset adapted from[1]

Ruan [1] provides a distinction between information assets and digital assets as illustrated in Figure 2. Intangible digital asset category act as a juncture between the information assets and digital assets. For example, a company's reputation is

an intangible information asset, but not a digital asset. Whereas, a company's online reviews and ratings are intangible digital assets and not information asset. The other distinction between information assets and digital assets is that the later facilitate the capability to deliver services, improve performance through better decisions, achieve competitive advantage and can be sold as a product [4].

Since information technologies are more critical than ever, it has become a pivotal responsibility of organisations to adequately manage assets for i) protecting them [6], ii) effective information security management [7] and iii) the sustainability of the organisation [8]. As assets can be at risk of being compromised, there is a need to identify important assets of the organisation before validation and proper risk management procedures can be practiced. There are many methodologies that can be used for asset identification and risk management [9], [10] shows that some of these methods can provide a limited perspective on the assets leading to inaccurate risk assessments. Thus, there is a need for proper asset identification and valuation method leading to better risk management.

2.1.2 Digital Asset Management

Over the years, asset management, in general, has evolved from just trying to describe how to manage assets to more as a business strategy. Information security professionals are asked to balance the cost of controls against the value of the assets that the controls protect. Valuing assets for this purpose is important and asset management assists in attaining this valuation. Asset management is a procedure to operate a group of assets throughout their technical lifecycle to attain a suitable return while ensuring defined services and security standards [3]. It assist organisations to identify the value of assets in achieving their organisational objectives. The standardised fundamentals of asset management is detailed in ISO/IEC 55001[1].

Weinstein [11] explores asset management and demonstrates that companies should push their needs and wants when it comes to the management of different assets. The author further states that "In the end, it's all about business". We see the importance and need for not only the management of assets but also the valuation of the assets to aid business improvements and goals. Further, there is a necessity to understand the different types of digital asset management; such as Digital Asset Management (DAM), Media Asset Management (MAM) and Content Management (CM); and management styles related to non-digital assets. These are important because the designed model needs to function efficiently with the chosen methods of

[1] https://www.iso.org/standard/55089.html

management. DAM focuses on the electronic management of any form of digitally stored information. MAM focuses on managing media type assets such as audios, videos and imagery which can be either digital or non-digital. While, CM deals with the use of stored digital and media asset. Both digital and media asset management handle long-term stored contents that is used for archiving, preservation and reuse. Content management, on the other hand, uses contents for a specific period of time. Wager [12] have detailed the benefits of using DAM strategies in effectively managing digital assets. An appropriate asset strategy allows linking of business initiatives so that assets can be determined, viewed and reported for quantification.

2.1.3 Attributes of Digital Asset

Assets identified can have attributes that impact the effectiveness of an organisation. The first step is thus the identification of assets those are valuable to the organisation. Locating the attributes of information assets and identifying the use of it across the organisation can enhance the organisation's effectiveness. However, the attributes of assets must be identified based on predefined context. Digital assets exhibit attributes of traditional assets as discussed in [4]. But, unlike traditional assets, digital assets are not necessarily scare, can be instantly scalable and are non-rival in nature. Some of the unique characteristics of digital assets are:

1. *Digital value increase with increase in usage.*
   Unlike many traditional assets, the value created by digital assets does not decrease with increase in usage rather it increases with increase in usage. In other words, more the people use it, more economic benefits can be derived from it [4]. For example, the economic value of online platforms such as Wikipedia, Facebook and Google increases with increase in people using it.

2. *Duplication does not increase digital value.*
   Duplication of information does not increase the value of a digital asset. As no new information is created through duplication, the assets have same value as a single copy. However, duplication does add additional cost [4]. For example, multiple copies customers' personal data can cause significant additional management cost while the value being the same.

3. *Production and distribution of digital value involve higher fixed cost and lower variable cost.*
   The production and distribution of digital assets mostly involve lower variable cost and higher fixed cost when compared to traditional assets. For example digital services through software requires significant investment in the design,

development, testing and deployment phases. However, once the product has been developed, it can be sold, distributed and maintained at a low marginal cost. Not that there are a range of nonrival goods such as e-books, music and software that can be reproduced at a zero marginal cost [13].

4. *Digital value can be distributed through multi-sided markets.*
   In a traditional single-sided market, sellers interact with only one specific set of customers. In contrast, the digital economy has given rise to "platform economy" facilitating multi-sided markets [13]. In a multi-sided market, the seller deals with more than one set of customers acquiring different services and products. For example, Amazon, Airbnb and Uber provide similar services to various customer segments through their platform.

5. *Digital value is limitless.*
   Unlike with traditional assets, the limitation of the resources is no longer a constraint in digital value creation and distribution. The limitless utility to the owner and the limitless opportunities to distribute and consume digital value defines the limitless nature of digital value. The limitless utility to the owner describes that intangible digital assets cannot be consumed by use and its utility is maintained regardless of the change in ownership directing towards the first attribute in the list. Whereas, the limitless opportunity to distribute and consume digital value describes the multi-sided marketing opportunities.

2.1.4 Categorisation of Assets

From an information security perspective [14], assets can be categorised into

1. *People* - This category includes employees and non-employees of an organisation. Employees of an organisation can be further classified based on their role and access privileges. Non-employees includes contractors, consultants, and third-party collaborators with which the organisation has business relationships.

2. *Procedures* - This category captures all the standard, and IT and business related procedures through which might introduce security risks to the organisation.

3. *Data* - Data assets are the information assets in digital format which is possibly the most important asset of an organisation. They include Intellectual Property (IP), system and application output files, databases, web pages, digital records and documents valuable to the organisation. This category accounts

for management of information involved in transmission, processing and storage.

4. *Software* - Software assets comprises of applications and programmes that are used to operate computers and related devices provisioning the interfaces for the desired services. Software assets can be further classified into applications, operating systems, and security components. Security components can be applications or operating systems and needed to be protected more thoroughly than other system components.

5. *Hardware* - Hardware assets include the physical technology that accommodates and executes the software. Examples of hardware assets includes devices that host software providing services such as laptops, servers, laboratory equipment, smart devices and personal assistive devices. Hardware assets can be further classified into system devices and peripherals, and networking components. Networking components must be prioritised since they are often the focal points of attacks against an organisation.

A finer categorisation of digital assets has been provided by [1] which includes:

- *Software Assets* - Software assets comprises of applications and programmes that are used to operate computers and related devices provisioning the interfaces for the desired services. Examples of software assets include decision support systems, websites, HR management applications, network file sharing applications, asset management platforms, payment applications and financial management applications.

- *Hardware Assets* - Hardware assets include the physical technology that accommodates and executes the software. Examples of hardware assets includes devices that host software providing services such as laptops, servers, laboratory equipment, smart devices and personal assistive devices.

- *Service Assets* - Service assets includes any resources or capabilities that could contribute to the delivery of a service. Service assets can involve hardware, digitally enabled devices and software. These assets can also be outsourced and delivered by a third-part contractor/supplier. Examples of service assets include digital media platforms, Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS).

- *Network Assets* - Network assets include connected information systems that rely on each other to provide services. Network assets can include hardware assets such as routers, switches, printers, servers and wireless access points to

software assets such as operating systems, proprietary software, client facing websites and applications to service facilitated through the network assets.

- *Data Assets* - Data assets are the information assets in digital format which is possibly the most important asset of an organisation. They include Intellectual Property (IP), system and application output files, databases, web pages, digital records and documents.
- *Metadata Assets* - Metadata is referred as data about data. Metadata assets are descriptive information applied to data assets to support tasks. For example a descriptive file assisting users to locate data assets through searches, workflow documentations, short descriptions and keywords.

Further, based on the criticality or importance of an asset to the organisation they can be classified into: (i) Critical assets - assets which are necessary to accomplish core business functionalities of an organisation such as a critical system or database containing sensitive data with no backups; (ii) Important assets - assets which compromised would not affect the organisation from achieving its core business functionalities, but would affect in a long-run if the assets are not restored such as a server used as a backup for sensitive data; and (iii) Supportive assets - assets which supplement in accomplishing critical functionalities in daily basis and would affect the effectiveness of the business if compromised. Examples of supportive assets include web pages, data for organisational management for example daily catering requirements and stationery.

2.2 Asset Valuation Methods

2.2.1 Tangible Assets Valuation

There are many approaches to valuation, Spencer [15] evaluates these approaches as being the result of one or more of the follow processes: (i) market forces, where there is a history of trading an entity of value; (ii) official value established by someone of authority thus being a recognised value; (iii) expert opinion and appraisal; (iv) bilateral agreement or contract; and (v) cost of creation or re-creation. One asset may also have more than one value, in the case of information, values may be different depending on the purpose or reason of the valuation. With this, the idea of valuing within a "risk" context means some general conditions should be met, Spencer also provides the basis for valuing information within risk management which includes: (i) exclusive possession; (ii) utility; (iii) cost of creation or re-creation; (iv) potential liability; (v) convertibility or fungible nature; and (vi) operational impact.

1. Binary asset valuation:
   This method involves a simple decision of classifying an asset to either the first group or the second group. This type of asset valuation is preferred in situations where specific controls are required for strictly defined data [9].

2. Classification-based valuation:
   In this method, assets are classified into one of the several value classifications. This method is an extension of the binary valuation method and is commonly practised in a security risk assessment. For example, a critical asset can be classified as a high value asset, important assets have medium value and supportive assets have low value [9].

3. Rank-based valuation:
   This method employs a ranking scheme where an asset is ranked against all other assets [9]. For example, if an organisation has identified 10 assets to perform risk assessment on, then each asset will be ranked between 1 to 10 and then the assets will be analysed based on their rank.

4. Consensus valuation:
   This method involves determining the value of an asset through a consensus estimation by a group of experts such as the Delphi method [9].

5. Valuation based on intrinsic value:
   Valuation models based on intrinsic value does not consider the business value of a digital asset rather focuses on the intrinsic value of the digital asset [16]. The intrinsic value of a digital asset is determined through fundamental analysis and does not include its market value. These models quantify digital assets by considering characteristics such as quality, completeness, and accessibility.

6. Valuation based on direct conversion of financial value
   Valuation models based on direct conversion of financial value define digital assets value based on their financial value such as cost of production, licensing, and patents [1]. This means that it is directly related to the proportional value which equates to approximately the same value.

7. Valuation based on business and performance value
   Valuation models based on business and performance value define digital assets value in relation to one or more business processes. These models consider business functions, processes and decisions in which information acts as a critical driver [16]. Performance value can also be defined as a type of business value that measures the impact of digital assets on one or more Key Performance

Indicators (KPIs) over time. Examples of KPIs include profit, cost, revenue vs targets, cost of goods sold, expenses vs budgets.

8. Cost-based valuation
   Valuation models based on cost value define digital assets value based on the cost of producing, buying, reproducing, changing, maintenance along with the cost of acquiring, replacing and regulatory obligations cost. Cost-based models include an adjustment for depreciation of the digital asset over time and generally do not include the potential future benefit that can be derived from the digital asset. These models also do not capture the full impacts of legal aspects of intangible digital asset management. When considering digital assets, cost-based models often do not address the future benefits of an asset but consider the reduction in increase and decrease in value. Not only this, but different companies will need to incorporate different costs into their model, thus cost-based model will vary from company to company [16]. The example is given within [1]) is "Total Cost of Ownership" which was created by Gartner in 1987, containing a "comprehensive assessment of IT" in order to consider a digital asset's costs. Moody and Walsh [4] discussed the cost-based method, explaining that the main advantage was that it is the easiest to collect and was "arguably" the most reliable at the time, with no other models being proven better than it. However once again the main issue is that it is likely to not reflect the current value.

9. Market-based valuation
   Market-based models estimate the value of digital assets by considering the marketplace and comparing them to existing assets to estimate a new value. Where this is easy to apply and can be accurate, the biggest issue is that this model becomes ineffective if there is no appropriate comparable asset. As discussed in [1] the Market for Personal Data, which contains data which is subjective to the owner and therefore becomes hard to compare, with chances of being underestimated. Valuation models based on market estimate the value of digital assets from the marketplace. Existing assets in the marketplace those are comparable to the digital asset under valuation are identified. The revenue derived from the identified assets is used is an estimate of the value of the new digital asset. When comparable intangible assets can be easily identified, these models are relatively easy to apply and can present accurate projections. Like cost-based models, these models also do not capture the full impacts of legal aspects of intangible digital asset management.

10. Utility-based valuation:

Utility-based models are based on the present value of an asset's future economic benefits, this therefore shows the value in use [17]. The main weakness to this model is that it is difficult to determine future cash flows relating to an asset, thus becoming subjective [4] thus most assets would be difficult to assign a monetary equivalent to. This type of valuation is mostly used for long term monetary assets that are bind by contracts, therefore for assets that are non-monetary, it would be nearly impossible to determine what the future cast flow would be.

11. Income-based valuation:
    Income-based models measure the contribution of an asset to the revenue of a company, questioning what the loss would be if this asset was at risk [1]. This considers the future of the assets, however much like market-based models, the issue of having accurate supporting information can nullify this model. Valuation model based on income estimate the value of a digital asset based on its contribution to the revenue of the entity. These models forecast future revenues to place an estimation on the value of the digital asset. The future forecast is primarily subjected to future earnings and cash flow projections.

12. Option based valuation:
    Another given model in [1] is the use of an option model, where owners of assets have many different choices which can be valued. These options can be compared, showing an analytical asset value outlook. These types of models work best when options can easily be identified and valued and are options that are stable, not subject to any dramatic shifts when changes occur. Valuation models based on options aim to define economic value of digital assets for each available option. An option is a choice that the owner can take at a specific time but may not be required to exercise the choice. Available options may include what rights to invoke, how to price the asset and when to apply legal means to enforce rights associated with the digital asset. Advanced forms of option models could capture costs associated with legal aspects of intangible digital assets.

2.2.2 Intangible Asset Valuation

One of the key areas lacking in current papers is the idea of valuing digital assets that are intangible. This idea is considered in Saunders and Brynjolfsson [18], noting that in general, companies lack in the capturing of intangible assets, even outside of the digital landscape. Using an econometric model relating to the market value of

a company to its assets, the authors expand what is considered to be an IT asset, grouping them into different categories such as Hardware, Prepackaged Software, Custom and Own-Account Software, Other Internal IT Assets, IT Consulting, IT-Related Training, IT Capabilities, Research and Development Assets, Brand Assets, Ordinary Assets (non-IT), and Other Assets. Using these, Saunders and Brynjolfsson [18] create an asset measures that are based on the IT -related intangible spending at a firm level, where they estimate the value a comprehensive set of IT assets in a market value equation. This also again confirms the idea of different companies having different values and assets, therefore documenting and managing all types of assets is an important feat, putting a price on assets requires a company to have good management in place.

If a model for asset pricing within big data was developed, it would transform the existing data market, making data science more efficient [19]. The key issue is that personal data is dependent on different variables, thus making it one of the intangible assets that need to be valued. Within the work of Shen et al. [19], a model is created that uses based on tuple granularity and considers the areas that affect data value, these including, cost and uncertainty (entropy). The authors express that their model, if successful, can be the model that is beneficial to any person or company that uses data, forming a data trading environment that scales. Where we have established that digital assets can range from hardware to information, it is also important to understand that before pricing and valuing can take place, companies must have a good management system. Grimaila [20] found that having a "lack of a rigorous, well-documented, information asset-based risk management process" means that there will be more uncertainty when assessing how to value information and the impact that it can have.

2.3 The Value of Data

Information assets are defined as "data that is or should be documented, and which has value or potential value." according to Wilson [21]. This creates a wide scope that must be considered, as information comes within different forms, being seen as raw material from knowledge, which has been formalised in some way. When considering the value of information, the idea of "value-in-use" explains the value of information from both the information user and use standpoint [22] information. This means that the value depends on who owns/uses this data, as well as the current landscape that this data dwells in. With their study, Engelsman shows some different valuation methods that can be found within the literature that we haven't already discussed.

The first method is called "Usage over time valuation" which is based on two main principles. The first is that information value is reflected by its use and the second is that the value can change over time, thus we see two measurable metrics, use and time[22]. Therefore, it infers that information is more valuable if is it used recently and used more. The issue here is that this method does not show a direct financial cost but functions as a scaling system to know what data should be more valuable without taking other details into account. The second method is called "Valuation of knowledge assets", here rather than considering the use of information, it considers the sum of the cost added to the sum of all relevant processes in which it is a resource [22]. Therefore, the production which involves that data is important, and the cost of using it is added to find the value. This means that the asset will have a combined value of its linked processes and its cost. The author has also presented "Valuation in risk perspective" the valuing method that focuses on the risk that each asset can bring, this relates to other work that uses Value-at-Risk (VaR) to price assets. The idea of risk is directed towards the discovery of "effective and defensible techniques" that control risk, rather than provide cost, therefore its purpose is aimed towards risk management, by assessing the controls within a business. On the other side of this, is the use of VaR, which is associated with the loss that can occur over a certain time frame and level of confidence [23]. This theory is also found within risk management, therefore asset valuation and pricing based on this could be possible, where prices are assigned based on the cost of risk.

2.4 Challenges with Digital Asset Valuation

Due to the nature of digital assets in the current climate, there are many challenges that must be considered.

a. Inherent Challenges
   Due to rapidly changing digital technologies, it is much harder to estimate the future value of a digital asset. The unpredictability of future return is another major challenge in aptly valuing a digital asset in the fast-growing digital economy.

b. Market Challenges
   The inherent challenges described above causes minimal transparency between the seller and consumer hindering the growth of digital assets market. Taking data asset as an example, there is little transparency between the provider and the consumer regarding how the data has been collected, process and manipulated pre-sale and how it will be used post-sale. This lack of transparency

introduces information asymmetry in the market (Shen, et al. 2016).

c. Taxation Challenges
The current international taxation framework, which is based on traditional economy, is unable to take into consideration new models for value creation introduced by the digital economy. As a result, public budget and social fairness are affected (European Commission 2018).

d. Regulatory and Standardisation Challenges
Current regulations lack a standard or commonly agreed taxonomy for digital goods and services. Further, existing taxonomy only categorises digital assets based on their technical functions rather than their economic functions causing inconsistency in the valuation.

2.5 Relevant Valuation Frameworks

Ruan [24] introduces a framework to study the requirements for the development of risk analytics solutions for the valuation of digital assets and the risk exposure of these. The paper's method contains the use of economic theory which also shows the categorisation of digital assets. One of these groups is the "Core Value assets" which contain assets that have been digitised and assets that are born digital, showing the different types of digital assets on a business level. The other side to this is the assets that are operational, meaning that they support how the entity is being digitally run. The future of this work is aimed at implementing the model and showing it within a case study. Within Tatar and Karabacak,[7] the work contains the creation of a hierarchy-based digital asset valuation method that had been created to minimise mistakes that occur in asset management. It starts with the identification of hardware, then software then information and assigns a confidentiality, integrity and availability value to each. The authors express drawbacks to the model, with the main part being that it is only for digital assets, however, the bigger issue is that this is valuation system that does not consider the prices of the assets.

As we know, information in a fundamental asset within businesses and must be treated accordingly. The work of Batini [25] provides insight into an information value assessment based model that is based on the concept of information capacity, information utility, and information management costs. The model breaks down key areas within IT management, considering, for example, who uses the asset and the type of information. Where there is uncertainty about the type of method to be used for pricing of digital assets, both the capital asset pricing model (CAPM and its extensions) and arbitrage pricing theory (APT) could be used. Both of these

methods are different and calculate costs based on different factors.

The Capital Asset Pricing Model (CAPM) was first developed by Sharpe [26] and Lintner [27] independently of one another by using portfolio theory to show a market equilibrium. According to the work of Goyal [28], CAPM and is the most "celebrated" approach for asset pricing. The market portfolio is the only factor that is common, where the exposure to this determines the expected returns of the asset [28]. Goyal explores some the additional version that have been developed in recent years:

- Merton's [29], inter-temporal capital asset pricing (ICAPM) version, considers variables that can predict future investment, for example inflation, using these as key components.
- Breeden's [30] consumption capital asset pricing (CCAPM) version, which relates assets returns to their covariances with the utility of consumption.

Arbitrage pricing theory (APT) was first developed by Ross [31] and is a multi-factored model that is based on the idea that the return of an asset can be predicted by using the linear relationship between the expected return and variables that consider risk. Within this are 2 main assumptions, competitive and friction-less markets, this is where the market has no transaction cost and no restriction to trade (friction-less) and another where the market has unlimited quantities of the relevant security, without this price changing (competitive) [32]. As will be shown later on, when considering digital assets, the idea is to have non-arbitrage values, thus having asset prices in an equilibrium [1].

Having detailed the fundamental characteristics and methodologies of asset pricing, the following section demonstrates a use case and the methodologies applied to determine the assets and their values to the business considering both the intangible and tangible impacts of an asset.

3 SECONDO Use Case: Cyber Insurance for Innovative SME

CloudAndTech is an innovative SME that offers Business-to-Business (B2B) solutions for big data analytics and intelligent algorithms as well as professional services related to cloud computing and development. CloudAndTech has no physical infrastructure. All the development and production environments are hosted in the Google Cloud. It also uses tools for development such as Gitlab and Kubernetes for containerised applications. Further, it uses four Kubernetes nodes, two Kubernetes masters and two Kubernetes clusters to assist with the deployment, scaling and management of the application. Apart of the virtual infrastructure for building and deploying its solutions, it uses professional accounts for hosting the company's accounts, such as email and storage, collaboration environment for the team to work remotely. The email service is out-sourced to Google. CloudAndTech employees 42 members of staff at four different roles which are (i) Executives (ii) Upper Management (iii) Management, and (iv) Contributors as detailed in section 4.4.

Having no physical infrastructure (e.g., data centres) offers flexibility and does not require special security measures to be taken, apart of course for securing access to its virtual infrastructure. CloudAndTech implements Virtual Private Network (VPN) and secure access-authentication procedures to ensure that access to its resources is not granted to unauthorised people and only authenticated devices such as personal computers and mobile phones of the employees can only access it through two-factor authentication. The cost for running all its business in the Google cloud and the other services is in the below 8,000 EUR per month. Losing the online infrastructure will cause the pause of its business. The customers will not be affected as their applications are running on their premises, but development, testing and some support tasks will not be able to continue. Code, customer related information and files will be lost (backups are being taken of course from time to time).

3.1 Identification and Prioritisation of Assets

Asset identification is a key step in an information security risk assessment of an organisation. Assets are an important element of a security risk assessment as identifying the number of assets helps to scope the security risk assessment and valuation of the identified assets helps to determine the countermeasures to be employed to mitigate the potential loss. Based on the detailed use case, assets of CloudAndTech are categorised as shown Table 2.

Asset Category	Asset Type	Assets	No. of assets	Critical	Susceptible	
					Human	Technology
Equipment	Hardware	Physical Backup Device	1	*		*
		Computer	19	*		*
		Mobile Phones	15			*
	Software	GitLab 3	1	*		*
		Google Cloud	1	*		*
		Windows OS	19			*
		VPN	4			*
		Collaborative Env	1			*
		Kubernetes Node	4	*		*
		Kubernetes Master	2	*		*
		Kubernetes Cluster	2	*		*
	Network	Gateway	1			*
		HAProxy	1			*
		Host-based Firewall	2			*
		Boundary Firewall	2	*		*
		Connection to Internet		*		*
Data	Sensitive	Code		*		*
		Customer Information		*		*
		Service Details				*
		Employee Details		*		*
Procedure	Necessary	Access-authentication		*	*	*
		Backup Procedure		*	*	*
Personnel	Staff	Contributor			*	
		Management			*	
		Upper management			*	
		Executive			*	

Table 2: Asset identification and categorisation for CloudAndTech

3.1.1 Asset Classification

Using the commonly practised "Classification-based asset valuation" method we have identified assets for CloudAndTech. The "Critical" column indicates assets that are absolutely necessary to support business continuity and carry out business critical operations of CloudAndTech and loss of these would have adverse consequences and disruptions to the continuity of its business. Similarly, assets can be classified into categories as illustrated in Table 3. The "Susceptible" column in table 2 indicates the assets which are vulnerable to cyber attacks primarily exploiting people or technology. We distinguish between human and technology related attacks to highlight that human play a critical role in the success of an attack as more than 99 percent

of threats observed by Proofpoint[2], over 18 months, require human interactions to execute. Taking this as a motivation, we have performed a case study on social engineering as discussed in section 4.4.

Importance Level (Qualitative Value)	**Semi-quantitative Value**	**Rating Criteria**
Critical	100	Indicates that Compromise of the asset would have devastating consequences, leading to loss of life or serious injury to people and disruption to continuation of critical business operations
High	75	Indicates that compromise of the asset would have serious consequences affecting the adequate continuation of business critical operations
Medium	50	Indicates that compromise of the asset would have moderate consequences affecting the continuation of business critical operations
Low	25	Indicates that compromise of the asset will have little or no impact on the human life or on the continuation of business critical operations

Table 3: Asset classification categories based on importance level.

3.1.2 C.I.A Triad

Table 7 and Table 8 lists the identified assets for the detailed use case. Each identified asset has been classified based on its importance using Table 3. Each asset is further assessed based on the C.I.A triad, using rating scale in Table 4, as the value of the information comes from the characteristics it possesses. When the characteristics of the information changes the value of the information and the asset possessing it also changes. Each critical characteristics, defined as the C.I.A triad [14], is described as:

- **Confidentiality (C)**: Confidentiality is preserved when information is protected from disclosure or exposure to unauthorised agents throughout its life-cycle. Confidentiality ensures that only authorised agents who have the rights and privilege to access the information can access it. Confidentiality is breached when unauthorised agents can access the information.
- **Integrity (I)**: Integrity is preserved when the information is whole, complete and uncorrupted. Integrity is breached when the information is exposed to corruption, destruction, damage, and other disruption to its authentic state throughout its life-cycle.

[2] `https://www.proofpoint.com/us/resources/threat-reports/human-factor`

- **Availability (A)**: Availability ensures that the information is accessed by authorised agents without interference or obstruction and to receive it in the required format. Availability is breached when the authorised agents have hindrance in timely accessing the information.

Semi-quantitative Impact Value	**Confidentiality**	**Integrity**	**Availability**
5 (very Severe)	The unauthorised disclosure of information could be expected to have an exceptionally grave adverse effect on organisation, individuals, or the nation.	The unauthorised modification or destruction of information could be expected to have an exceptionally grave adverse effect on organisation, individuals, or the nation.	The disruption of access to or use of information or computer system could be expected to have an exceptionally grave adverse effect on organisation, individuals, or the nation.
4 (Severe)	The unauthorised disclosure of information could be expected to have a serious adverse effect on organisation, individuals, or the nation.	The unauthorised modification or destruction of information could be expected to have a serious adverse effect on organisation, individuals, or the nation.	The disruption of access to or use of information or computer system could be expected to have a serious adverse effect on organisation, individuals, or the nation.
3 (Moderate)	The unauthorised disclosure of information could be expected to have some adverse effect on organisation, individuals, or the nation.	The unauthorised modification or destruction of information could be expected to have some adverse effect on organisation, individuals, or the nation.	The disruption of access to or use of information or computer system could be expected to have some adverse effect on organisation, individuals, or the nation.
2 (Low)	The unauthorised disclosure of information could be expected to have a limited adverse effect on organisation, or individuals.	The unauthorised modification or destruction of information could be expected to have a limited adverse effect on organisation, or individuals.	The disruption of access to or use of information or computer system could be expected to have a limited adverse effect on organisation, or individuals.
1 (Negligible)	The unauthorised disclosure of information could be expected to have negligible effect on organisation, or individuals.	The unauthorised modification or destruction of information could be expected to have negligible effect on organisation, or individuals	The disruption of access to or use of information or computer system could be expected to have negligible effect on organisation, or individuals.

Table 4: Rating criteria used to mark the C.I.A triad.

3.1.3 Impact to Business Continuity

We model the impact to business continuity as a function of the *confidentiality, Integrity,* and *Availability* values illustrated by the "Impact to Continuity" column in Table 7 and Table 8 where:

$$\text{Business Impact}(a) = \sum \Big\{ f(C,a), f(I,a), f(A,a) \Big\}, \quad \forall a \in \mathcal{A} \tag{1}$$

The impact to business continuity by an asset is determined using the Table 5.

Qualitative Value (Impact on Continuity)	**Semi-quantitative Range** (Business Impact)	**Semi-quantitative Value** (Impact on Continuity)	**Description**
Very High	13-15	10	Severe or devastating consequences to continuation of business critical functions.
High	10-12	8	Compromise of the asset will have high impact on the continuation of business critical functions.
Moderate	7-9	6	Compromise of the asset will have moderate impact on the continuation of business critical functions.
Low	4-6	4	Compromise of the asset will have little impact on the continuation of business critical functions.
Very Low	1-3	2	Compromise of the asset will have negligible or no impact to continuation of business critical function.

Table 5: Rating criteria used to mark the C.I.A triad.

3.1.4 Reputation Impact

The most visible assets of a modern corporation are its structure, physical property, hardware and services, while the most valuable assets are intangible and quantifying them are extremely challenging. Aon's 2019 Global Risk Management Survey report[3] reveals that reputation damage is the biggest cyber threat to business. We model the impact on reputation of an organisation due to a successful breach use a binary-classification approach as detailed in Table 6 and can be expressed as :

$$\text{Reputation Damage} = \begin{cases} 10, & \text{if security configurations of the asset is completely controlled by the organisation.} \\ 5\ , & \text{if security configuration of the asset is partially controlled the organisation.} \end{cases}$$

[3] https://www.aon.com/unitedkingdom/insights/global-risk-management-survey-uk-2019.jsp?utm_source=Aon&utm_medium=website&utm_campaign=Cyber&utm_term=JimTGRMS

Semi-quantitative Value (Reputation Damage)	**Description**
10	Severe consequences of an incident to the reputation of the organisation due to negligence or misconduct or irresponsible security behaviour in the part of the organisation.
5	Moderate to low consequences to the reputation of the organisation due to a mono-culture vulnerability incident or an incident to third part service providers such as Google and Microsoft. Managing these incidents are beyond the abilities of the organisation unless and until a fix has been provided by the service providers.

Table 6: Rating criteria used for classify the damage to the reputation.

3.2 Asset Valuation Model

In order to calculate the value of an asset (a_v), we use the following estimation equation:

$$a_v = \text{Asset Importance} \times \Big\{ \beta_0 + \beta_1(\text{Reputation Damage}) + \beta_2(\text{Impact on Continuity}) + \epsilon \Big\} + \text{Annual Cost} \tag{2}$$

where $\beta_1,\ \beta_2$ are coefficients which represents the ratio between the benefits and the cost of the asset. For our evaluation we use:

$$\beta_1 = \beta_2 = \begin{cases} 0.1, & \text{for lower bound of asset value range} \\ 1\ , & \text{for upper bound of asset value range} \end{cases}$$

and $\epsilon = 0$ is the error rate and the constant $\beta_0 = 1$.

3.3 Threat Scenario

CloudAndTech rents a physical office in Spain where it has been recently reported that a group of cybercriminals has launched a social engineering attack targeting innovative SMEs. CloudAndTech has decided to undertake the Cybersecurity Risk Assessment using the SECONDO platform. Its result will indicate how CloudAndTech must spend their limited cybersecurity budget and whether they must outsource some of the risks to a cyber insurer. It is usually the case that SMEs prefer to treat cybersecurity investments and cyber insurance negligently so that they can prevent charges.

Asset	Asset Type	Importance	CIA Triad			Reputation Damage	Impact on Continuity	Monthly Cost(€)	Asset Value
			C	I	A				
Google Cloud	Software	Critical	5	5	5	5	10	1400	1650-3000
Kubernetes Master 1	Software	Critical	5	5	2	5	10	100	350-1700
Kubernetes Master 2	Software	Critical	5	5	3	5	10	100	350-1700
Kubernetes Node 1	Software	Critical	5	3	2	5	8	70	300-1470
Kubernetes Node 2	Software	Critical	5	3	2	5	8	70	300-1470
Kubernetes Node 3	Software	Critical	5	3	2	5	8	70	300-1470
Kubernetes Node 4	Software	Critical	5	3	2	5	8	70	300-1470
Kubernetes Cluster 1	Software	Critical	5	5	3	5	10	50	300-1650
Kubernetes Cluster 2	Software	Critical	5	5	3	5	10	50	300-1650
GitLab 3	Software	Critical	5	5	3	5	10	260	510-1860
Collaborative Service	Software	Critical	2	1	1	5	4	200	390-1200
VPN Service 1	Software	Critical	3	1	5	10	6	175	370-1450
VPN Service 2	Software	Critical	3	1	5	10	6	175	370-1450
VPN Service 3	Software	Critical	3	1	5	10	6	175	370-1450
VPN Service 4	Software	Critical	3	1	5	10	6	175	370-1450
Source Code	Data	Critical	5	5	5	10	10	-	300-2100
Costumer Information	Data	Critical	5	1	1	10	6	-	260-1700
Service Information	Data	Critical	5	1	1	10	6	-	260-1700
Employees Detail	Data	Critical	5	1	1	10	6	-	260-1700
Host-based Firewall 1	Network	High	3	5	1	10	6	150	410-1850
Host-based Firewall 2	Network	High	3	5	1	10	6	200	460-1900
Boundary Firewall 1	Network	Critical	5	5	1	10	8	250	530-2150
Boundary Firewall 2	network	Critical	5	5	1	10	8	250	530-2150
Connection to Internet	Network	Critical	1	1	5	5	6	-	210-1200
HAProxy	Network	Medium	3	1	2	10	4	150	390-1650
Gateway	Network	Medium	3	1	2	10	4	150	270-900
Backup Procedure	Procedure	Critical	5	5	5	10	10	700	1000-2800
Access-authentication	Procedure	Critical	5	5	1	10	8	600	880-2500
Windows OS	Software	Critical	5	5	5	10	10	700	850-1750
Computer 1	Hardware	Medium	2	2	2	10	4	70	190-820
Computer 2	Hardware	Medium	2	2	2	10	4	70	190-820
Computer 3	Hardware	Medium	2	2	2	10	4	70	190-820
Computer 4	Hardware	Medium	2	2	2	10	4	70	190-820

Table 7: CloudAndTech's Asset List

Asset	Asset Type	Importance	CIA Triad			Reputation Damage	Impact on Continuity	Annual Cost(€)	Asset Value
			C	I	A				
Computer 5	Hardware	Medium	2	2	2	10	4	70	190-820
Computer 6	Hardware	Medium	2	2	2	10	4	70	190-820
Computer 7	Hardware	Medium	2	2	2	10	4	70	190-820
Computer 8	Hardware	Medium	2	2	2	10	4	70	190-820
Computer 9	Hardware	Medium	2	2	2	10	4	70	190-820
Computer 10	Hardware	Medium	2	2	2	10	4	70	190-820
Computer 11	Hardware	Medium	2	2	2	10	4	70	190-820
Computer 12	Hardware	Medium	2	2	2	10	4	70	190-820
Computer 13	Hardware	Medium	2	2	2	10	4	80	190-820
Computer 14	Hardware	High	3	3	2	10	6	80	340-1780
Computer 15	Hardware	High	3	3	2	10	6	80	340-1780
Computer 16	Hardware	High	3	3	2	10	6	80	340-1780
Computer 17	Hardware	High	3	3	2	10	6	80	340-1780
Computer 18	Hardware	High	3	3	2	10	6	80	340-1780
Computer 19	Hardware	High	3	3	2	10	6	80	340-1780
Android 1	Hardware	Medium	2	1	1	10	4	25	145-775
Android 2	Hardware	Medium	2	1	1	10	4	25	145-775
Android 3	Hardware	Medium	2	1	1	10	4	25	145-775
Android 4	Hardware	Low	1	1	1	10	2	10	65-335
Android 5	Hardware	Low	1	1	1	10	2	10	65-335
Android 6	Hardware	Low	1	1	1	10	2	10	65-335
Android 7	Hardware	Low	1	1	1	10	2	10	65-335
Android 8	Hardware	Low	1	1	1	10	2	10	65-335
Android 9	Hardware	Low	1	1	1	10	2	10	65-335
Android 10	Hardware	Low	1	1	1	10	2	10	65-335
iPhone 1	Hardware	Medium	2	1	1	10	4	25	145-775
iPhone 2	Hardware	Medium	2	1	1	10	4	25	145-775
iPhone 3	Hardware	Medium	2	1	1	10	4	25	145-775
iPhone 4	Hardware	Low	1	1	1	10	2	10	65-335
iPhone 5	Hardware	Low	1	1	1	10	2	10	65-335

Table 8: Continuation of CloudAndTech's Asset List

3.3.1 Meta-models using CORAS

The assets identified for the usecase and their respective impact values have been provided to OLISTIC[4] which in return provides the various meta-models used to represents the dependency among assets and their threats. Moreover, OLISTIC provides information not only on the various roles in the organisation (their unique ID) but also associates each asset with its users (see section 4). In other words, OLISTIC can be used to deduce the dependency among the assets (e.g., Computer 1 is connected to VPN Service 1). This information will be used in the later stages of SECONDO such as while developing the Econometrics Module.

OLISTIC uses CORAS language[5] to develop the meta-models. CORAS is a method for conducting security risk analysis. It provides a customised language for threat and risk modelling and comes with detailed guidelines explaining how the language should be used to capture and model relevant information during the various stages of the security analysis. The language consists of five different kinds of diagrams: asset diagrams, threat diagrams, risk diagrams, treatment diagrams, and treatment overview diagrams. An example of the five different diagrams for the selected use case is provided below.

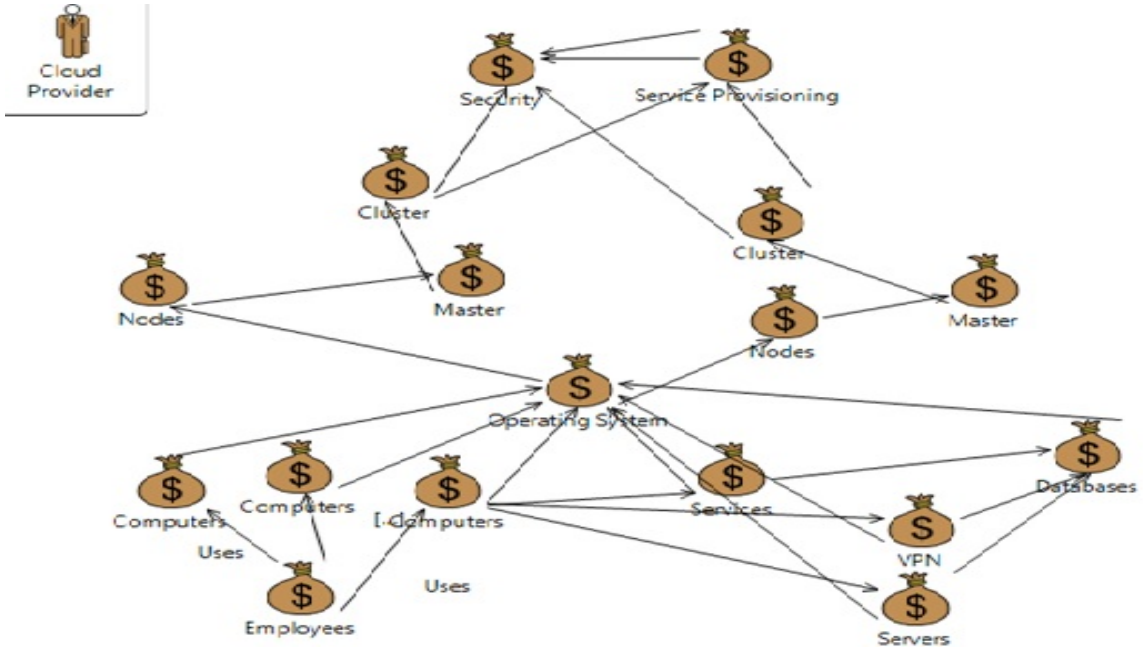
The **Asset Diagram** depicted in the Figure 3a represents in high-level the physical and non-physical components and assets of CloudAndTech's cloud infrastructure. The diagram concludes with the two most precious assets of CloudAndTech's which are Security and Service Provisioning.

The **Threat Diagram** presented in the Figure 3b includes software, applications, systems, networks, distributed systems, etc. In CloudAndTech threat modelling involves mostly vulnerabilities on the cloud infrastructure.

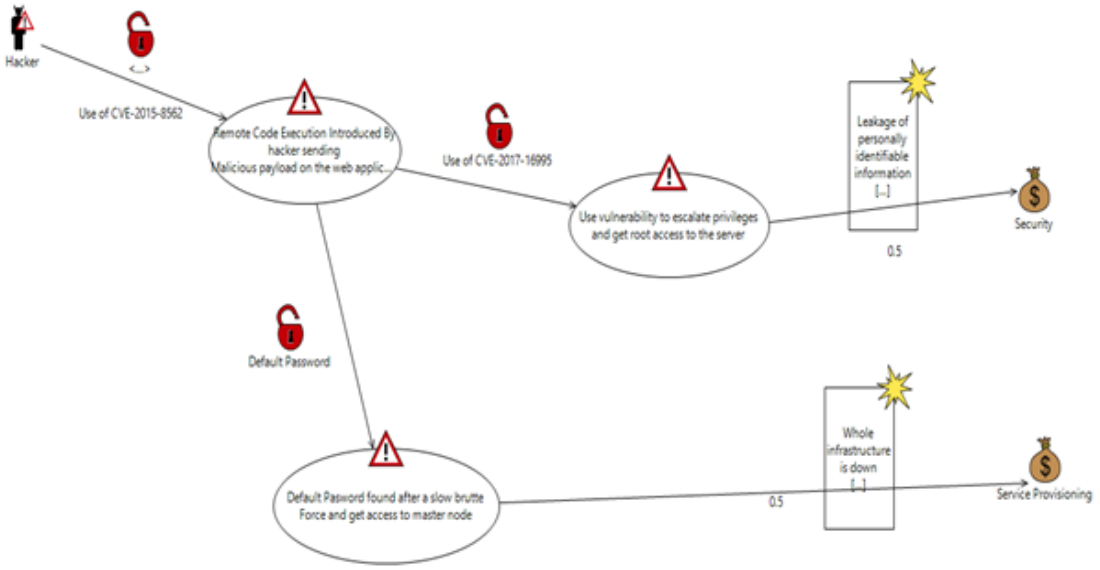
The **Treatment Diagram** in the Figure 4 rely on previous analysis which consists of determining the different ways a threat may initiate an unwanted incident. We have already done that by placing threat scenarios each describing a series of events, between the threats and unwanted incidents and connecting them all with initiative and leads-to relations. The graphical syntax has been designed to maximise the usability of the language. Although helpful in practical modelling situations, the graphical syntax is rather cumbersome to work with when defining the semantics and rules for the CORAS language.

Having identified the assets used by the examined organisation, the next section

[4] https://www.olistic.io/
[5] http://coras.sourceforge.net/coras_language.html



(a) Asset diagram



(b) Threat diagram

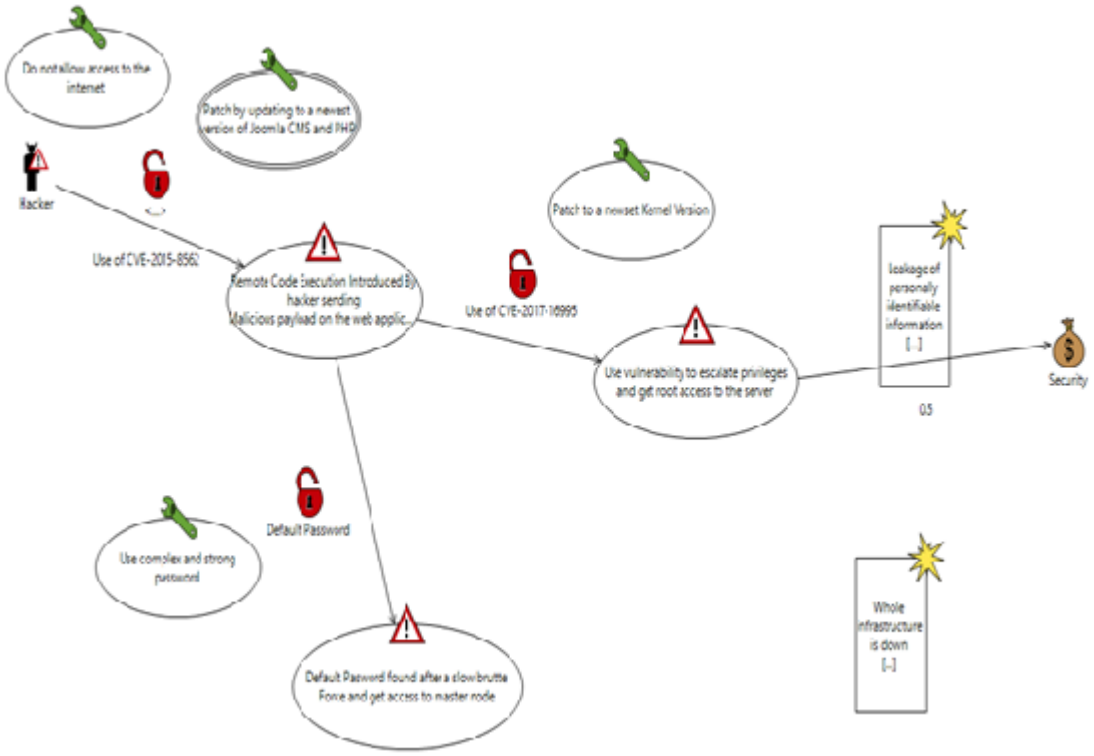


Figure 4: Treatment diagram

aims to demonstrate how SECONDO will calculate likelihoods that determine the possibility of an employee being targeted and exploited by a social engineering attack based on her role using the attack scenario detailed in the previous section.

4 SEAM

Social engineering is the act of manipulating a person to take an action that may or may not be in the "target's" best interest [33]. While the most common cyber-attacks try to exploit information systems, social engineering is an exploitation method targeting the human factor. Although the technical protection measures of preventing cyber-attacks have evolved over time, the current measures are inefficient against this kind of attacks.

The types of social engineering attacks are the following [2]: a) *Physical approach*: The adversary performs some physical actions to father information about a victim. A common method is the dumpster diving [34], where an adversary is willing to dig into victim's trash to look for important data e.g. password written on a paper; b) *Social approach*: The adversary relies on socio-psychological techniques to manipulate her victim. A usually method is the spear-phishing attack [35]; c) *Reverse social engineering*: The adversary instead of contacting directly with her victim, endeavours to convince her that she is a trustworthy entity so that the victim will approach her. This way contains three steps, sabotage, advertising and assisting; d) *Technical approaches*: The adversary accomplishes these kind of attacks via internet. A common type of attack is to harvest passwords since victims may use the same passwords for many accounts and e) *Socio-technical approaches*: Adversaries blend the previous categories. A common attack that belongs to that type is the baiting attack. Moreover, each social engineering attack is launched by an operator [2], this can be human or software. Furthermore, the above attacks can be executed via channels [2]: a) Email; b) Instant messaging; c) Telephone, Voice over IP; d) Social networks; e) Cloud and f) Website. The Table 5 briefly explains the most common and famous social engineering attacks based on their channel, operator and type.

There is plethora of data breaches that occurred the latest period due to social engineering attacks. Toyota Boshoky Corporation, a major supplier of Toyota auto parts was victim of a social engineering attack, in August 2019, losing the amount of $37 million [36]. The adversaries achieved to convince an employee with financial authority to change the account information on an electronic funds transfer. Moreover, Roblox was victim by a social engineering attack [37]. Firstly, the adversary bribed a worker to grant access to him and then he phished an unaware customer. Finally, the adversary was able to see customer's email address, change passwords, remove two-factor authentication from their accounts and even to ban users. Furthermore, the security company Proofpoint issued on 2020 an annual report including analysis of data from a variety of sources, including 5,400 working adults, 600 IT security

		Phishing	Shoulder surfing	Dumpster diving	Reverse social engineering	Waterholing	Advanced persistent threat	Baiting
Channel	E-mail	✓			✓		✓	
	Instant Messenger	✓			✓			
	Telephone, VoIP	✓			✓			
	Social Network	✓			✓			
	Cloud	✓						
	Website	✓				✓	✓	
	Physical	✓	✓	✓	✓			✓
Operator	Human	✓	✓	✓	✓			✓
	Software	✓		✓	✓	✓	✓	
Type	Physical		✓	✓				✓
	Technical					✓	✓	
	Social				✓			
	Socio-technical	✓			✓	✓	✓	✓

Figure 5: Classification of social engineering attacks based on our taxonomy [2]

professionals, 50M simulated phishing attacks and more than 9M reported suspicious emails [38]. From this report the results are very interesting since only 49% of U.S workers answered correctly the following question "What is phishing?". In addition, the impact of successful phishing attacks are presented in this report and the result is the the most possible impact is the loss of data while the less possible impact is the financial loss. Also, based on Verizon report [39] issued on 2020 the 22% of the occurred breached involved Phishing attack. Moreover, in 23,619 incidents social engineering attacked were less than 20%. Finally, the most frequent social engineering attack is the Phishing with almost 85% of frequency.

The **Social Engineering Assessment Module (SEAM)** interacts with users to devise their behaviour using penetration testing approaches and provides specific numeric results on endanger actions (i.e. percentage of users that open suspect files or execute Trojans). In more technical detail, the readiness of the employees of any organisation against social engineering attacks will be evaluated. This module include *planning*, *targeting*, *means* and *evaluation* of replicating such type of attacks. The *planning phase* will be the decisive part of which technological methods and axes will be used, such us email social engineering, client-side attacks like web browsing to a web page or even physical phone calls. The second part is *targeting*, which requires us to find the target audience for our specific purpose. The third part is the *means*, which is a guessing of the correct attack for the correct target using also the most efficient tools. Finally, all the collected data will be *evaluated* and provide the Social Engineering data output that feed directly to the **RAOHM** module.

4.1 GoPhish

In this section we aim to present the Go-Phish [40] which is integrated in our scheme as an external social engineering assessment tool. Choosing the GoPhish utilisation we complete the two phases of social engineering, *planning* and *means* are these are presented above.

4.1.1 Go-Phish Features

GoPhish [40] is an open-source email phishing attack simulation and phishing user awareness promotion platform, that provides organisations the ability to perform counterfeit phishing campaigns, in order to collect cyber-security intelligence. GoPhish can be used as an assessment tool, logging information about the users and fetching users' data relative on how individuals would respond to real-word scenarios, revealing human-oriented security gaps. It automates the process of bulk creating and sending fraudulent emails to unsuspected users, tracking, and yielding comprehensive, thorough, and detailed statistical results about one's behaviour.

The aforementioned actions [41] are the following:

- **Email Reported**: Email was successfully sent to the victim. However, she reported the email as malicious, without opening it.
- **Email Sent**: Email was successfully sent to the victim. However, she ignored the email without opening it or reporting it.
- **Email Opened**: Email was successfully sent to the victim. In addition, she opened the email without taking any other action.
- **Clicked Link**: Email was successfully sent to the victim. Then, she opened the email and finally, she clicked the malicious link attached to email's body.
- **Submitted Data**: The victim successfully receives the email, opened it, clicked the linked and submitted data on the landing page the malicious webpage.

The logging of these actions is performed via the following methods [41]:

- An embedded Tracking Image hidden in the body of the email. A tracking image (also known as pixel tracking) is an HTML tag styled as an invisible 1x1 image with a hyperlink reference pointing to the GoPhish server. The hyperlink has a GET parameter with the id of the victim. Once a user opens an email, the tracing image is requested via an HTTP request from the GoPhish server.

In that way GoPhish can identify who, when and where a simulation email was opened.

- The GET parameter of the link attached to the body of the email.
- The landing page, designed by default in such a way that it logs all submitted credentials.

4.1.2 Go-Phish Functionalities

- **Sending Profile Creation**: This functionality provides the ability to the phisher to choose the emails "from" field, and to connect GoPhish to the SMTP email server. Moreover it provides the ability to modify email headers as well as adding custom ones.
- **Email Templates**: This functionality provides the ability to the phisher to create several email templates. These constitute the actual emails that will be sent through the attack simulation. The user can choose the title of the email, add a tracking image and modify the HTML content of the email. In the case of traditional phishing, phishers can use an email template that is a replica of a social media or a trustworthy email. In the case of *spear phishing* the phisher can create a proper-business looking email targeting specific victims.
- **Landing Pages**: This functionality provides the ability to the phisher to create several landing pages on which the URLs attached to the email lead to. These landing pages are plain HTML pages and data submitted to this form are stored in GoPhish database. In addition, GoPhish provides the ability for the redirection of the user to another webpage after her successful data submission.
- **Users & Groups**: This functionality provides the ability to the phisher to create, modify or delete individuals or group targets. The targeted groups can also be bulk inserted via a *csv* file. Phishers can choose the *name*, the *surname*, the *email* and the *position* of each victim.
- **Campaigns**: This functionality provides the ability to the phisher to create and send the actual phishing campaigns, by choosing an email template, the landing page the template leads to and the targeting group. The data which can be emerged form a campaign are the following.
  - **id**: a unique alphanumeric identification for each victim.
  - **status**: the last action performed by the victim.

- **ip**: the victim's IP.
- **latitude**: latitude is an angle (defined below) which ranges from 0° at the Equator to 90° (North or South) at the poles.
- **longitude**: longitude is the measurement east or west of the prime meridian. Longitude together with latitude specify the precise location of features on the surface of the Earth.
- **send_date**: send_date expressed the date when the phishing email delivered to the victim.
- **reported**: this parameter receives two value *True* or *False* in order to declare if this emails is reported by the victim.
- **modified**: this parameter presents the date when the email changes status (e.g. the date when the victim opened the email).
- **email**: victim's email where the email sent.
- **first name**: victim's first name.
- **last name**: victim's last name.
- **position**: by position we mean the victim's level in the organisational structure, (e.g. Chief Executive Officer).

The above functionalities can either be accessed via the UI or manually through the provided API.

4.2 SEAM Architecture

The SEAM architecture contains the five following entities as these are presented in the the Figure 6. Through this proposed architecture we mainly focus on how to run and evaluate a social engineering campaign via GoPhish. Specifically, we consider a SEAM Control Center, an SMTP Server, a GoPhish Server, a Landing Page as well as a Target Group.

First and foremost, the *end-user* who operates the **GoPhish Server** by sending commands (launch, terminate a campaign, evaluate results) via the **SEAM Control Center** to ti has to identify her **Target Group**. Once, this has been identified and are known the emails, first names, last names and positions of the victims who constitute it then she can upload them to the GoPhish platform. Consequently, the end-user via the **GoPhish Server** has to develop the **Landing Page**.

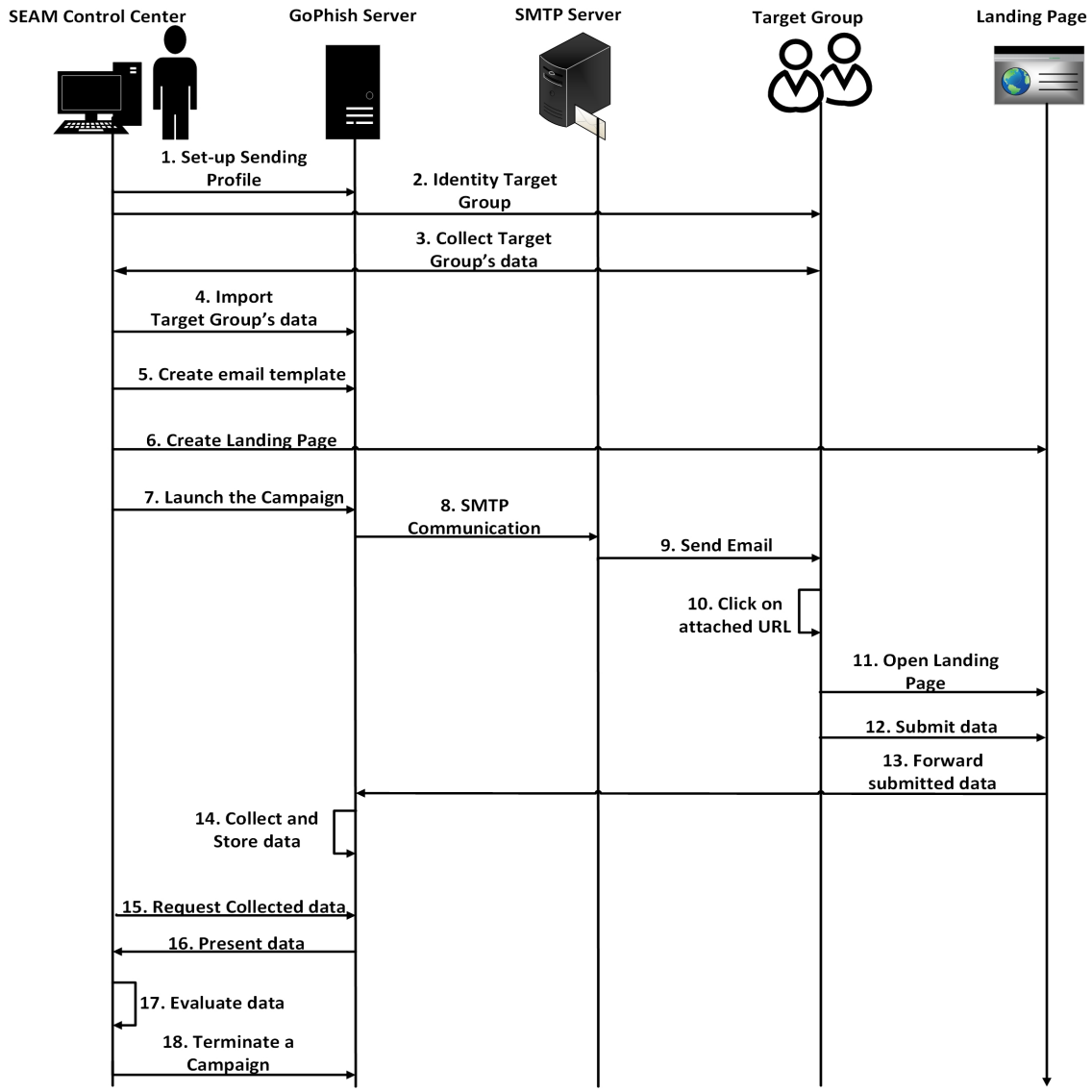


Figure 6: SEAM Architecture

The **GoPhish Server** is an entity that has already installed the *GoPhish* application. Moreover, it stores information about the participated **Target Group** as well as visualises the results that emerge from each campaign. This server once receive from the **SEAM Control Center** the respective command to launch a social engineering campaign directly communicates with the **Emails Server** to send email to the chosen Target Group. Once, the Target Group receives the email then starts

interacting with the **GoPhish Server** based on her action. If she *clicks* on the attacked malicious link then the she is redirected to the **Landing Page**. The **Landing Page** is responsible to monitor the **Target Group**'s submitted data (username and password) which are forwarded to the **GoPhish Server**.

Whenever, the campaign collects values that are enough, then the end-user of the **SEAM Control Center** is responsible to terminate the campaign. Once the termination occurs, then the **GoPhish Server** visualises the collected results. Based on these results the *end-user* can evaluate the examined organisation.

4.3 Install GoPhish and Run A Campaign

In this section we aim to present the steps in order to run a successful campaign via the GoPhish application which is the chosen one for our proposed scheme. We present the process from the scratch including how we installed GoPhish in our system. The steps are enlisted below:

1. Install GoPhish in a Linux distribution. For SECONDO purposes we utilised UBUNTU 20.04 LTS.
   (a) Download the latest GO archive https://golang.org/dl/
   (b) Extract the download GO archive *sudo tar -C /usr/local -xzf go$VERSION.$OS-$ARCH.tar.gz*
   (c) Run the command *sudo nano ~/.bash_profile*
   (d) Write at the end of the file and then save, *export PATH=$PATH:/usr/local/go/bin*
   (e) Run the command *sudo source ~/.bash_profile* to activate the changes.
   (f) Run the command *sudo apt install gcc* to install the GCC compiler.
   (g) Run the command *go get github.com/gophish/gophish* to install GoPhish.
   (h) Navigate to *$GOPATH/src/github.com/gophish/gophish*
   (i) Run the command *go build*
   (j) Run the command *openssl req -newkey rsa:2048 -nodes -keyout gophish.key -x509 -days 365 -out gophish.crt* to start the certificate and key generation process.
   (k) Update config.json
   "db_name" : "mysql",

"db_path" : "root:@(:3306)/gophish?charset=utf8&parseTime=True&loc=UTC",
The format for the db_path entry is
username:password@(host:port)/database?charset=utf8&parseTime=True&loc=UTC

(l) Add the following line at the bottom of the file
[*mysqld*]
sql_mode=ONLY_FULL_GROUP_BY,STRICT_TRANS_TABLES,
ERROR_FOR_DIVISION_BY_ZERO,NO_AUTO_CREATE_USER,
NO_ENGINE_SUBSTITUTION

(m) Log into MySql and run the command *CREATE DATABASE gophish CHARACTER SET utf8mb4 COLLATE utf8mb4_unicode_ci;*

(n) Run GoPhish *gophish@gophish.dev:~/src/github.com/gophish/gophish$ ./gophish*

2. Run a GoPhish campaign

(a) After Gophish starts up, you can open a browser and navigate to https://XXX.XXX.XXX.XXX:3333 (the page which hosts GoPhish) to reach the login page. The Login Page is depicted in Figure 7.

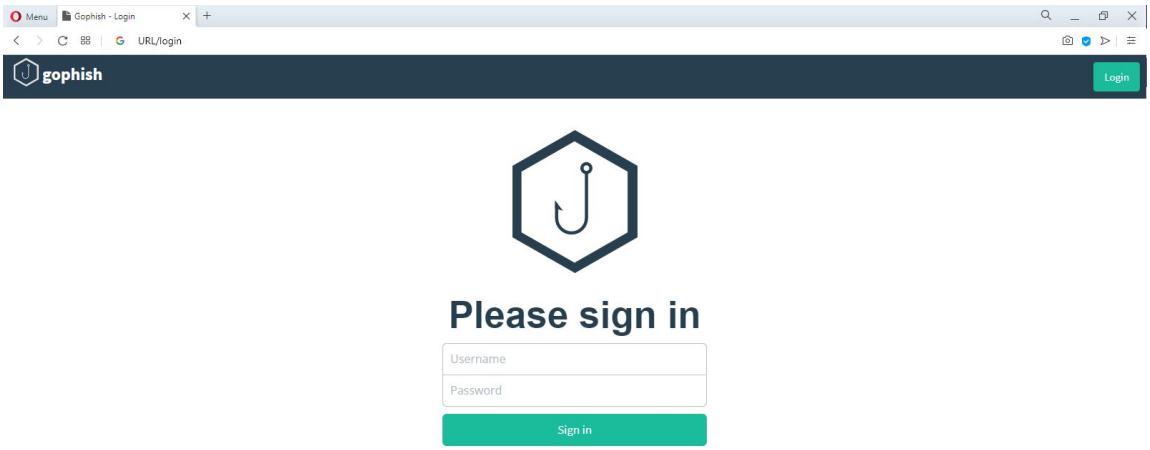


Figure 7: GoPhish - Log-in page

(b) Import target Groups. The first thing thing we have to do before run a campaign is to define our target group (who we aim to target). Assuming we have already obtain an email list containing the emails that belong to our targets, we start importing them into the GoPhish. Firstly, we navigate to *Users & Groups* and then press *+ New Group* and the initial emerged screen is presented in Figure 8.

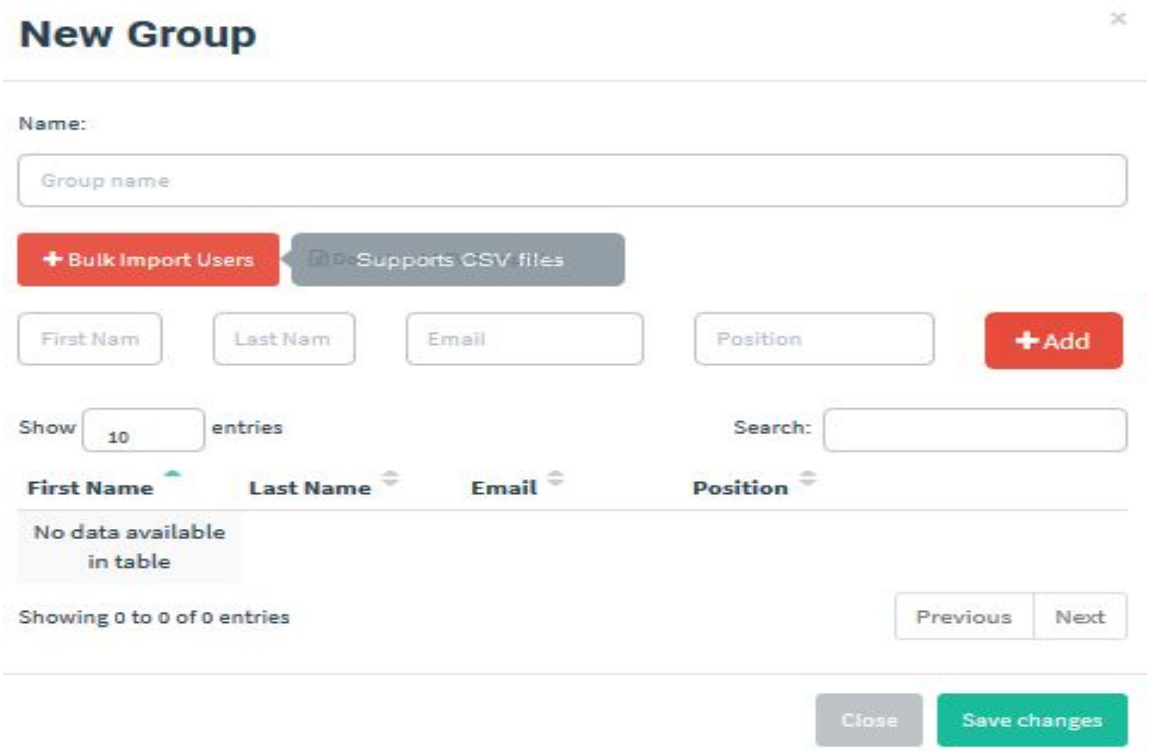


Figure 8: GoPhish - New Group initial screen

i. Firstly, we give a unique name for the User Group which describes the participated target group (e.g. SECONDO Simulation).

ii. Secondly we shall import the information of each victim (Name, Surname, Email and Position) this is a record that represents an employee. For example the record (Harry, Potter, h.potter@org.com, Student) represents an employee whose name is *Harry*, his surname is *Potter*, his email is *h.potter@org.com* and his position is *Student*. The insertion of the targeted emails can be manually or import a

CSV file with the same fields. After inserting the target group then the screen will be transformed as it is presented in the Figure 9 (the presented data in Figure 9 are fictitious and have been generated only for this section in order to avoid the exposure of the real participants' personal data).

iii. Click "Save changes" and confirm the creation of the user group.

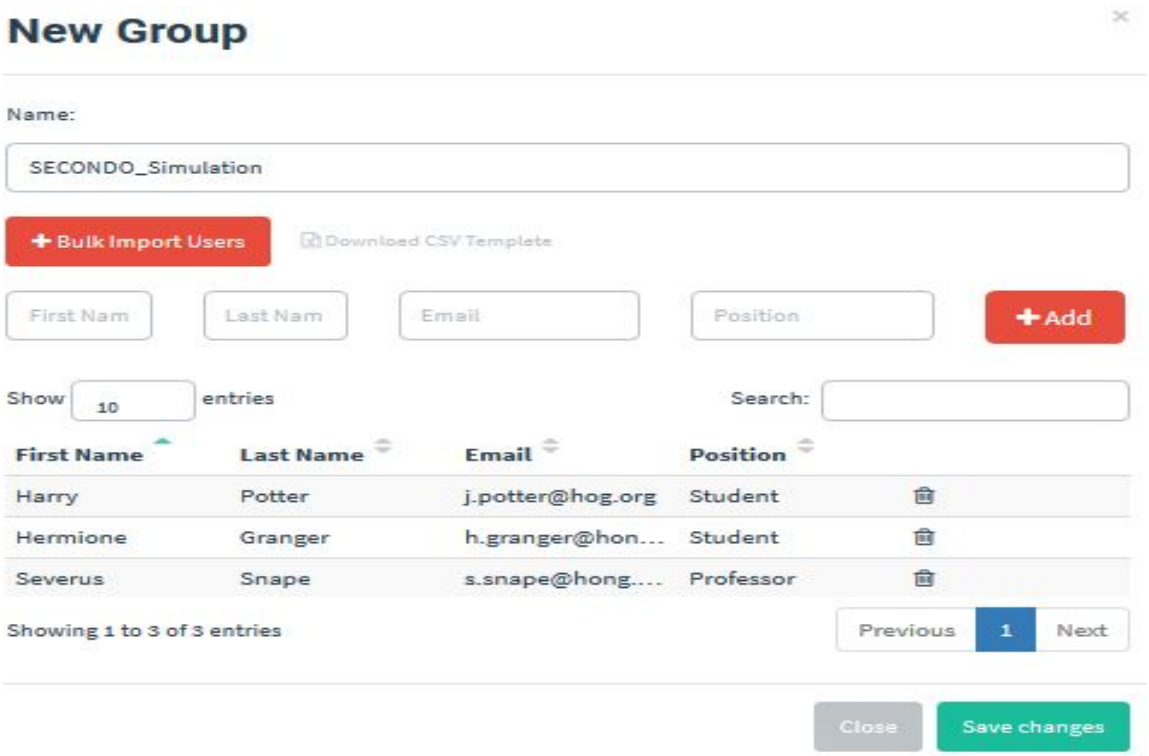


Figure 9: GoPhish - User Group

(c) Create the template that will be presented in the malicious email. We navigate to *Email Templates* and then press *+ New Template* and the initial emerged screen is presented in Figure 10. In this step we can create how the malicious email will appear to our victims. GoPhish provide us with many features which can assist us to create an legible and legitimate email that does not enable a victim to understand that this is a fraud. The better context the email has the more believable will be by the victims. This is the goal we want to achieve through a social engineering attack,

to attract victim's attention in order to monitor her behaviour and educe her personal data. We can write our own code for a the template or we can import an email for better accuracy.

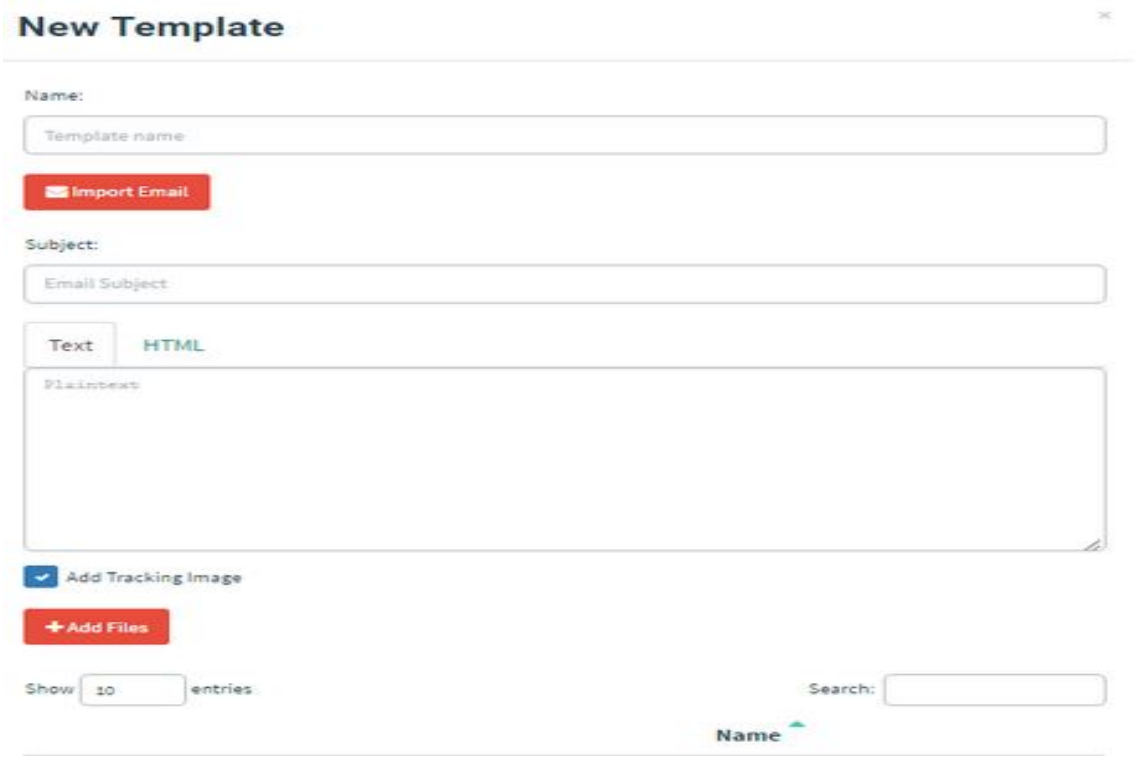


Figure 10: GoPhish - New Template

(d) Having already create the template, now we shall create the landing page. We navigate to "Landing Pages" and then press "+ New Page". The *Landing Page* represents the page where the victim will be redirected in order to submit her data or to download a malicious file. This page can be created by us or we can use an existing one as it is presented in the Figure 11. Finally, we can capture the submitted data and passwords of the victims based on our page.

3. Since, we have already successfully completed the above steps we can launch our campaign. We navigate to "Campaigns" and then press "+ New Campaign". Then we fill the form giving information about the campaign's name, the *email template*, the *Landing Page*, the *URL*, the *scheduled day and time*, the *sending*

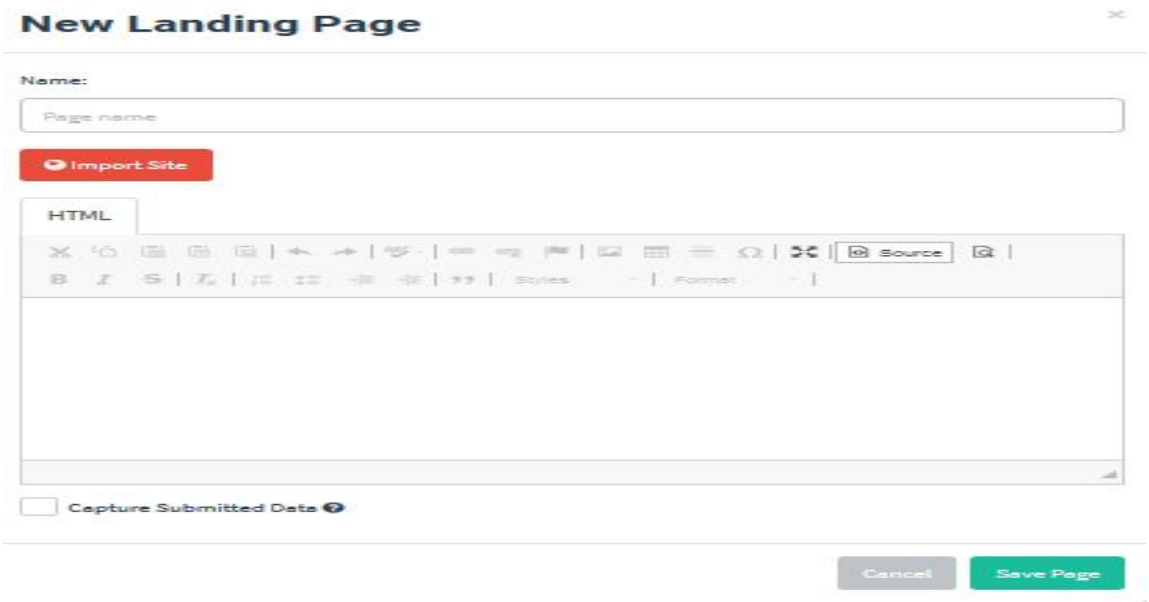


Figure 11: GoPhish - New Landing Page

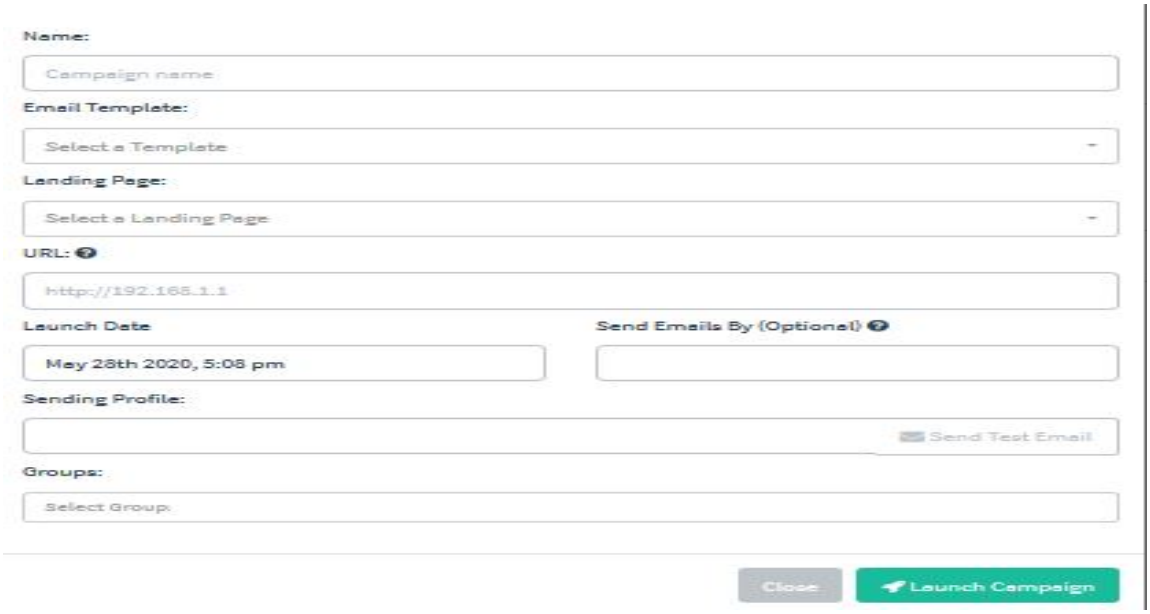


Figure 12: GoPhish - Launch Campaign

profile and finally the *user group* as it is depicted in the Figure 12. After providing the required information (based on the previous steps) then click *Launch Campaign* to start sending the emails to the user group.

4. The final step after launching a campaign is to view and evaluate the results. The results demonstrated briefly in the dashboard. Moreover, the results are available in CSV file which presents who user did which action. The supported action have already been analysed above.

4.4 Likelihood Determination

In order to achieve quantitative values for SECONDO purposes we executed two successful campaigns in one of our partners in order to collect results for this particular Deliverable. Moreover, as it is mentioned in the Table 11 we have to obtain quantitative values in order to express the likelihood of successful phishing attack $R_{\mathcal{O}}$ to an organisation $\mathcal{O}$. In addition, we have to collect quantitative values in order to reflect the likelihood of occurrence of a threat P_i to an employee i.

For achieving a quantitative and precise risk analysis we shall obtain real data results which answer two questions, the first one is "*Who is being attacked in an organisation?*" and the last one is "*Which is the probability for a success full cyber-attack in an organisation?*". The answers of the above questions shall be adaptable to a Small and Medium Enterprise (SME) as well as to a big corporation. This means what we should know who in an organisation is really under attack and the reason they are under attack. This information includes knowledge that uncurls data related to their roles, what data they have access and an estimation of their potential exposure. For the purpose of the project, we separate an organisation in the following four different levels: i) **Executives**: is a group of employees who set the plan for the organisation to succeed and are also responsible for the organisation's failure. The plan for success not only clarifies the following strategy, but also engraves organisation's culture for innovation, employee motivation and management style.

ii) **Upper Management**: is a set of employees within the organisation having one tier of management above (*Executives*), and multiple layers of employees below. They frequently report to the set of *Executives* and are responsible not only to manage the day-to-day activities of the business by setting direction in line with the overall business strategy but also for the emerged financial results within their area. Moreover, they set goals and objectives they are responsible for the spending within their department.

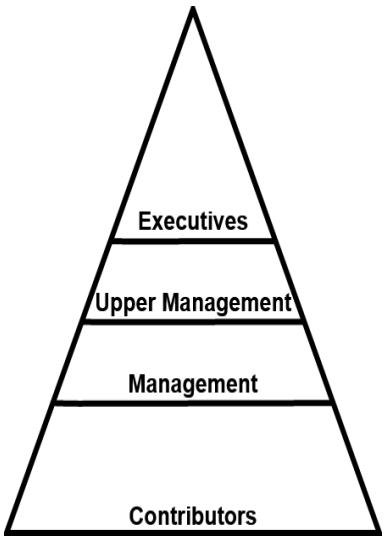


Figure 13: Organisation Hierarchy

iii) **Management**: is a set of employees that run the day-to-day operations, manage the largest number of employees and have the greatest ability to influence success of the company's goals and targets and frequently they report the set of *Upper management*. Moreover, they do not design any strategic direction instead they implement the planned in advance action plan and manages it across their employee base.

iv) **Contributors**: is a group of employees who run the every day tasks of the organisation. They report daily on the *Management* group. They are assigned with a specific type of work. The hierarchy of an organisation based on the aforementioned differentiation is depicted in the Figure 13. Integrating this hierarchy, we pick our target group and fulfil the *targeting* phase of social engineering.

For quantitative probabilities that answer our first question we follow the white paper from Proofpoint [42] that summarises data about cyber attacks and especially from phishing attacks. The results are depicted in the Figure 14. From their report we emerged that *Contributors* and *Management* accounts were highly targeted during malware and phishing attacks. However, these results express the targeted accounts based on the emails the have received not based on frequency they receive phishing emails. We aim to utilise this information for the Risk Determination Process.

Since we successfully achieved to obtain quantitative values for answering the first question "*Who is being attacked in an organisation?*" we should proceed finding

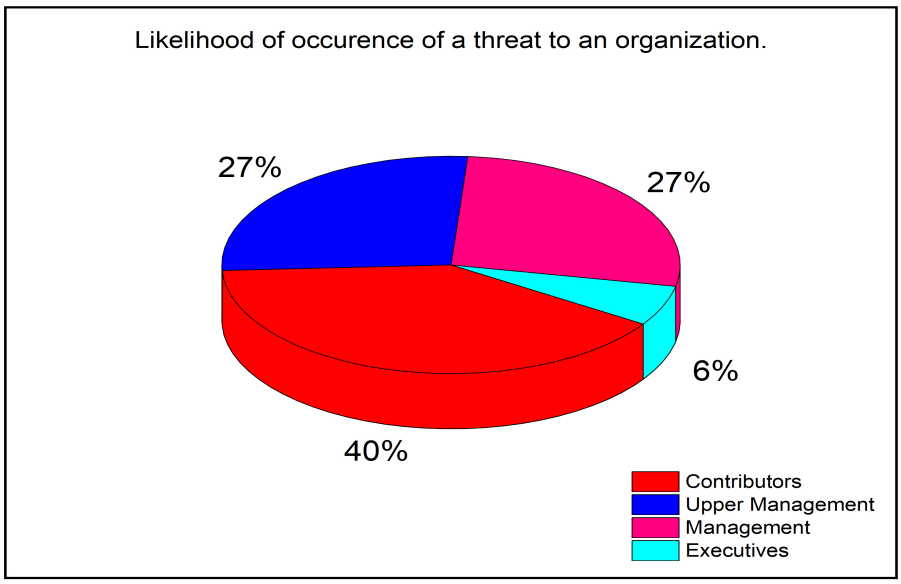


Figure 14: Likelihood of occurrence of a threat.

quantitative values for the second question "*Which is the probability for a success full cyber-attack in an organisation?*". For answering this question we executed two campaigns in one our partners. The GoPhish campaigns provide us with five different activity status as these are presented in Section 4.1.1. As successful attack we declare the following actions: i) **Link Clicked**: The link which is attached on the email was clicked. This is a dangerous action than can jeopardise organisation's confidentiality, integrity and availability. This is a successful phishing-attack and ii) **Submitted Data**: The victim submitted data on the fraud website. This actions sets in danger the organisation's confidentiality, integrity and availability. This is a successful phishing attack. We run the campaign sending a fishing email to 42 employees.

While we define as unsuccessful attack the following actions: i) **Email Reported**: The email received and was reported by the user. This action has no harm to the organisation and the cyber-attack is unsuccessful; ii) **Email Sent**: The email received by the victim however this was not opened. This action has no harm to the organisation and the cyber-attack is unsuccessful; iii) **Email Opened**: The email received and opened by the victim without taking any other actions on it. This

action has no harm to the organisation and the cyber-attack is unsuccessful.

We express $\mathcal{O}$ as the set of organisation which contains the subsets of Contributors $\mathcal{C}$, of Upper Management $\mathcal{U}$, of Management $\mathcal{M}$ and of Executives $\mathcal{X}$, where $\mathcal{O} = \{\mathcal{C}, \mathcal{U}, \mathcal{M}, \mathcal{X}\}$. We recognise that each level of the organisation is an independent set, so that and these subset have not common values, so that $\mathcal{C} \cap \mathcal{U} = \emptyset$, $\mathcal{C} \cap \mathcal{M} = \emptyset$, $\mathcal{C} \cap \mathcal{X} = \emptyset$, $\mathcal{U} \cap \mathcal{M} = \emptyset$, $\mathcal{U} \cap \mathcal{X} = \emptyset$ *and* $\mathcal{M} \cap \mathcal{X} = \emptyset$. Moreover, we denote R_l the likelihood to click the attached link as well as $R_\jmath$ the likelihood to submit data on the malicious web-page. In addition, we denote as R_r the likelihood to report the phishing email, as R_z the likelihood to not open the phishing email an as R_m the likelihood to just open the email. Furthermore, we assume that each probability is independent since an employee cannot be at the same time in two different situation, she can be in online one status. Moreover, we declare that the likelihood of a successful cyber-attack $R_\mathcal{S}$ to a subset of an organisation $\mathcal{S}$ is equal to the union of the likelihood to click the attached link together with the likelihood to submit data to the malicious web-page, this is depicted in the equation 3.

$$R_\mathcal{S} = P_c \cup P_\jmath = P_c + P_\jmath - P_c \cap P_\jmath = P_c + P_\jmath \tag{3}$$

Furthermore, we compute the likelihood of achieving an unsuccessful cyber-attack $R^\complement_\mathcal{S}$ which is equal to the union of the likelihood to report the received email together with the likelihood to not open the phishing email and the likelihood to just open the received malicious email, this is depicted in the equation 4. The Figure 15 depicts through a Venn diagram how the aforementioned probabilities are located in the space of each subset.

$$R^\complement_\mathcal{S} = 1 - R_\mathcal{S} = 1 - (P_c \cup P_\jmath) = R_r \cup R_z \cup R_m = R_r + R_z + R_m - R_r \cap R_z \cap R_m = R_r + R_z + R_m \tag{4}$$

We run a campaign to one of our partners and 42 employees participated. Each level (Contributors, Management, Upper Management and Executives) was represented with many employees, 18 employees were registered as *Contributors*, while the roles of *Management* and *Upper Management* were assigned to 10 and 8 respectively and 6 employees were charged as *Executives*. The results from the campaign are presented in the Table 9 which reflects the action in conjunction with the number of employees of each level who did it.

Since we assume that each subset is independent we have to compote superlatively

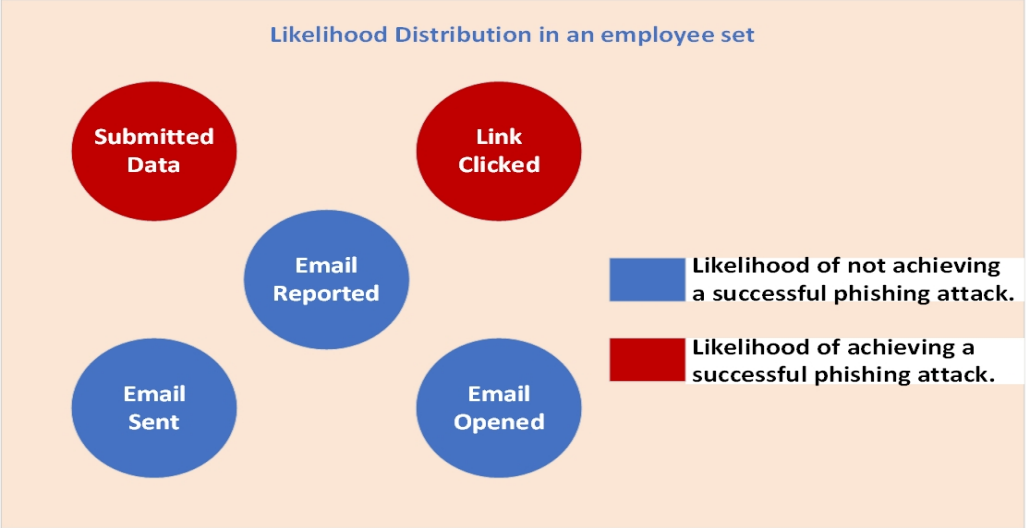


Figure 15: Likelihood distribution.

Action	Contributors	Management	Upper Management	Executive
Email Reported	0	0	0	0
Email Sent	2	0	2	2
Email Opened	6	2	2	2
Link Clicked	2	3	0	0
Submitted Data	8	5	4	2
Total Employees	18	10	8	6

Table 9: GoPhish Campaign

each likelihood of GoPhish actions which will lead us to the calculation of the likelihood of having a successful phishing attack to the organisation. To calculate each likelihood R_m of occurrence of an action m to the subset S then we utilise the total number of cases which is divided by the total number of employees who constitute this subset, this is expressed by the equation (5).

$$R_m = \frac{Total\ number\ of\ cases\ m\ in\ this\ subset}{Total\ number\ of\ employees\ who\ constitute\ this\ subset} \tag{5}$$

Applying the equation 5 to each subset of the organisation, the likelihoods of occurrence of each GoPhish action fluctuate. Moreover, performing the equations (3) and (4) we receive the likelihood for having a successful and an unsuccessful phishing attack to each subset respectively. The results are analysed below and summarised in

the Table 10. For the subset $\mathcal{C}$ which represents the organisation level of *Contributors* we educe the following likelihoods from our campaign.

- **Likelihood to report the phishing email**: $R_r = \frac{0}{18} = 0$
- **Likelihood to not open the phishing email**: $R_z = \frac{2}{18} = 0.111$
- **Likelihood to just open the email**: $R_m = \frac{6}{18} = 0.333$
- **Likelihood to click the attached link**: $R_l = \frac{2}{18} = 0.111$
- **Likelihood to submit data on the malicious web-page**: $R_j = \frac{8}{18} = 0.445$
- **Likelihood of a successful phishing-attack**:
  $R_{\mathcal{C}} = R_l + R_j = 0.111 + 0.445 = 0.556$
- **Likelihood of an unsuccessful phishing-attack**:
  $R_{\mathcal{C}}^{\complement} = R_r + R_z + R_m = 0 + 0.111 + 0.333 = 0.444$

For the subset $\mathcal{M}$ which represents the organisation level of *Management* we educe the following likelihoods from our campaign.

- **Likelihood to report the phishing email**: $R_r = \frac{0}{10} = 0$
- **Likelihood to not open the phishing email**: $R_z = \frac{0}{10} = 0$
- **Likelihood to just open the email**: $R_m = \frac{2}{10} = 0.20$
- **Likelihood to click the attached link**: $R_l = \frac{3}{10} = 0.30$
- **Likelihood to submit data on the malicious web-page**: $R_j = \frac{5}{10} = 0.50$
- **Likelihood of a successful phishing-attack**:
  $R_{\mathcal{C}} = R_l + R_j = 0.3 + 0.5 = 0.8$
- **Likelihood of an unsuccessful phishing-attack**:
  $R_{\mathcal{C}}^{\complement} = R_r + R_z + R_m = 0 + 0 + 0.2 = 0.2$

For the subset $\mathcal{U}$ which represents the organisation level of *Upper Management* we educe the following likelihoods from our campaign.

- **Likelihood to report the phishing email**: $R_r = \frac{0}{8} = 0$
- **Likelihood to not open the phishing email**: $R_z = \frac{2}{8} = 0.25$
- **Likelihood to just open the email**: $R_m = \frac{2}{8} = 0.25$
- **Likelihood to click the attached link**: $R_l = \frac{0}{8} = 0$

Likelihood	Contributors	Management	Upper Management	Executives
Report Email	0	0	0	0
Email Opened	0.333	0.2	0.25	0.334
Email Sent	0.111	0	0.25	0.333
Link Clicked	0.111	0.3	0	0
Submitted Data	0.445	0.5	0.5	0.333
Successful Phishing Attack	0.556	0.8	0.5	0.334
Unsuccessful Phishing Attack	0.444	0.2	0.5	0.666

Table 10: Overall likelihood results

- **Likelihood to submit data on the malicious web-page**: $R_\jmath = \frac{4}{8} = 0.5$
- **Likelihood of a successful phishing-attack**:
  $R_{\mathcal{C}} = R_l + R_\jmath = 0.25 + 0 = 0.25$
- **Likelihood of an unsuccessful phishing-attack**:
  $R_{\mathcal{C}}^{\complement} = R_r + R_z + R_m = 0 + 0.375 + 0.25 = 0.63$

For the subset $\mathcal{X}$ which represents the organisation level of *Executives* we educe the following likelihoods from our campaign.

- **Likelihood to report the phishing email**: $R_r = \frac{0}{6} = 0$
- **Likelihood to not open the phishing email**: $R_z = \frac{2}{6} = 0.333$
- **Likelihood to just open the email**: $R_m = \frac{2}{6} = 0.333$
- **Likelihood to click the attached link**: $R_l = \frac{0}{6} = 0$
- **Likelihood to submit data on the malicious web-page**: $R_\jmath = \frac{2}{6} = 0.334$
- **Likelihood of a successful phishing-attack**:
  $R_{\mathcal{C}} = R_l + R_\jmath = 0.334 + 0 = 0.334$
- **Likelihood of an unsuccessful phishing-attack**:
  $R_{\mathcal{C}}^{\complement} = R_r + R_z + R_m = 0 + 0.33 + 0.33 = 0.666$

Having calculated the likelihoods of an employee being targeted and exploited by a social engineering attack, in the next section, we aim to calculate the risk per asset based on number of employees using it. We achieve this utilising the risk model (proposed in section 5) and the asset valuation method (refer to section 3). Finally, we visualise the obtained results for comprehensive understanding of the exposed risks.

5 RAOHM

The Risk Analysis Ontology and Harmonisation Module is part of the QRAM and receives the outcomes of the existing risk analysis tool as well of the SEAM, and harmonises them using a common vocabulary. RAOHM combines the outcome of the OLISTIC and the SEAM in order to provide an harmonised result that will feed the Econometric Module (to be developed in the later stages of SECONDO).

Risk analysis is a determination of risk to the system, an analysis that requires the consideration of closely interwoven factors, such as the security controls in place for the system under review, the likelihood that those controls will be either insufficient or ineffective protection of the system, and the impact of that failure [43]. It is not feasible to accurately calculate the loss posed by a successful cyber-attack of a specific vulnerability without taking under consideration the effectiveness of the security control that have been implemented to mitigate or eliminate the potential loss for such an exploitation; nor the threat's motivation, opportunity, and capabilities, which contribute to the likelihood of a successful attack; nor the impact to the system and organisation should successful exploitation of a vulnerability occur. Risk analysis is conducted through four steps steps which are executed nearly simulated since due to these are linked to each other: i) control analysis; ii)likelihood determination; iii) impact analysis and iv) risk determination.

Control analysis is a process which defines the controls being used to protect the system under examination. The results are used to reinforce the determination of the likelihood that a specific threat might successfully exploit a particular vulnerability. This step is derived by OLISTIC and SEAM.

Likelihood determination is the process that considers a threat's source motivation and capability to exploit a vulnerability. For example, if a threat is highly motivated and sufficiently capable, and controls implemented to protect the vulnerability are ineffective, then it is highly likely that the attack would be successful. This step is executed by SEAM.

Impact analysis is a process that is used to determine the possible impact. The factors that contribute to impact determination as the impact to the systems, data, and the organisation's mission. This is provided by the examined organisation through OLISTIC.

Risk determination is a process that is executed once the rating for the likelihood and impact have been carefully determined. This step is executed through RAOHM.

Symbol	Description
$\mathcal{E}$	Set of employees of an organisation O.
P_i	The likelihood of occurrence on a threat to an employee i.
$\mathcal{A}$	The set of assets.
L_i	Expected loss from employee i.
I_a	Asset value of the asset a.

Table 11: List of Symbols

5.1 System Model for user behaviour

Our model assists in computing the expected loss L for an organisation due to a successful cyber attack. We assume $\mathcal{U}$ be the set of potential user groups u consisting of employees of an organisation O including in an organisational network, $\mathcal{U} \in \mathbb{Z}_n^+$. We denote $\mathcal{E}$ as the set of employees i where $\mathcal{E} = \{i : i \in \mathcal{U}\}$. In addition, users belonging to a specific user group have the same privileges to the organisation structure.

Each employee u of a user group u has the same degree of susceptibility towards an attack as the user group she belongs. We express P_i as the likelihood of occurrence of a threat to a specific employee i, where $P_i = [0, 1]; P_i \in \mathbb{R}$ [43]. Also, we denote the likelihood of a successful exploitation of a cyber-attack as to an employee i as $R_i = [0, 1]; R_i \in \mathbb{R}$ [44].

Each user is associated with one or more than one assets that belong in the organisation O. We denote $\mathcal{A}$ the set of the assets which belong to an organisation O, the assets as a, where $a \in \mathcal{A}, \mathcal{A} \in \mathbb{Z}_n^+$. The tuple $(i, a_1, \ldots, a_n)$ is called an *employee record* and describes the assets that are used by an employee, each asset can be used by more than one employees. Each asset a is characterised by an Impact value I_a which is a range of values in monetary units; in our scheme we utilise the maximum value of the range. We calculate the risk using the well-known risk assessment formula containing the likelihood of a threat event's occurrence, the successful exploitation of the target likelihood and potential adverse impact should the event occur [45], as it is presented in the equation (6).

Figure 16: General Overview of RAOHM architecture

$$risk = \text{(likelihood of being attacked)} \times \text{(probability of successful attack)} \times \text{(probable loss)} \quad (6)$$

In our scheme the total risk is calculated based on human behaviour for the assets and the relationship between them and organisation's employees. By this, we mean that an employee may use more than one assets as well as an assets may be used by more than one employees. Risk in our scheme is defined as expected loss that derives from the equation 6, where the *probable loss* is expressed in monetary units ranges. Hence, the total expected loss is given by the sum of the maximal expected losses [46].

We express the L_i as the expected loss associated with a specific user i, the L_a as the expected loss associated with a specific asset a. Moreover, we define the $L_{a,i}$ as the expected loss associated with a specific asset a and a specific employee i. Furthermore, we calculate the overall risk associated to a specific asset a utilised by an employee $i \in \mathcal{E}$ based on the equation 7.

$$L_{a,i} = P_i \times R_i \times I_a \quad (7)$$

In addition, we compute the expected loss of the asset a which is used by the employee i as it is presented in the equation 8.

$$L = \sum_{a \in \mathcal{A}} L_a \quad (8)$$

5.2 Workflow Architecture

In order to operate correctly, RAOHM architecture needs multiple components to interact to each other. In Figure 16, a general overview of the RAOHM architecture in the context of data analysis is presented. In Figure 17, we provide a detailed description with the required functionalities of each component.

The RAOHM reference platform components are the following:

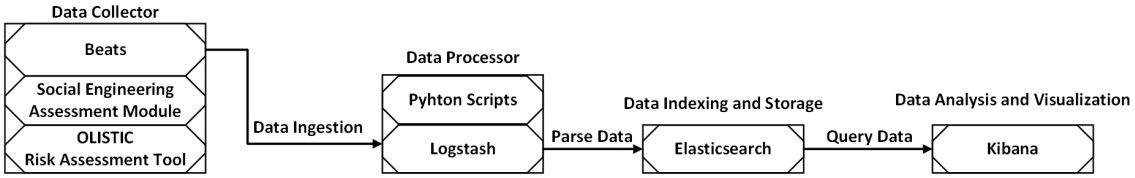


Figure 17: RAOHM components and their functions

- Data Collector
- Data Processor
- Data Indexing and Storage
- Data Analysis and Visualisation

RAOHM is designed to be able to handle big data as well as to support real time analysis in case of new findings. This scheme is applicable to many organisation with different structure.

5.3 Functionalities

Data Collector: The data collector resides in a server, is used by the operators of the risk analysis and are responsible for collecting as well as measuring variables of interest. A Data Collector is able to present results as raw data which are the primary data collected from a source or filtered which are organised, user-friendly and more eligible than the raw data.

Data Processor: The Data Processor resides in a server, different than the Data Collector, is used to transform data from different source. Data Processor has to handle each record in the coming data. Once, the record has been transformed then the Data Processor processes the next record.

Data Indexing and Storage: The Data Index is a data structure that improves the lookup of data a database table. Data Storage is the persistence and management of massive amounts of data in a scalable way that satisfies the needs of applications that require fast access to the data [47].

Data Analysis and Visualisation: The Data Visualisation is the process of graphical representation of the transformed data. Data Visualisation enhances the Data Analysis process with the goal of discovering useful information, information conclusions and supporting decision-making.

The RAOHM architecture is based on the following main pillars:

- SECONDO aims at designing an harmonisation module using a common vocabulary in order to harmonise different data which is output of more than one different sources.
- RAOHM is expected to be adjustable to different scale of organisations by being able to handle small, medium and massive amounts of data.
- RAOHM is expected to provide quantitative and harmonised values to the end-user for a better data analysis.
- SECONDO will design but also implement RAOHM in a way to generate results that will feed the Econometrics Module.
- RAOHM has to receive the output of existing risk analysis tools, as input.
- RAOHM has to receive the output of the Social Engineering Assessment Module (SEAM), as input.
- RAOHM must recognise potential loss regarding both intangible and tangible assets.
- RAOHM should perform asset identification.

5.3.1 Data Collector

The *Data Collector* is an entity that represents the process which is responsible to collect information from different sources and then to forward them to the *Data Processor*. In our scheme, the collector consists of two different technologies that collect data, **GoPhish** as well as **OLISTIC**, and **Beats** to forward them to the *Data Processor*. Each technology is installed in different servers that reside in premises of two of our organisations. Each server is responsible for specific tasks, after the completion of these tasks then they forward their results to Data Processor in order to harmonise the data. Moreover, throughout the interior functions which are used we achieve to monitor the system in case where new values have emerged, e.g. risk analysis of a new asset.

Gophish is responsible to run campaigns against the employees who conclude an organisation. The campaigns, as will be discussed later, provide specific metrics demonstrating the current behaviour of the users who participated. Once, the campaigns are finished then Beats takes on by reading from the results from GoPhish and then by shipping to a particular instance of Data Processor.

Furthermore, there is OLISTIC which is responsible to vulnerability assessment to assets of the examined organisation. OLISTIC receives as input the assets and automatically finds the threats which could occur on each registered asset. Finally, the results from it are processed by Beats which forwards them to a particular instance of Data Processor.

5.3.2 Data Processor

The Data Processor entity is responsible for receiving data from different sources in different form, harmonising them and then send them to the storage place. In our proposed scheme, the processor consists of two different technologies that assist the data process, **Logstash** and *Python Scripts*. Both technologies are installed on the same server and reside to the same premises where the **GoPhish** implementation resides. After the receipt of the data, the *Data Processor* via interior function calculates the overall risk of the examined organisation.

Firstly, Logstash is the Data Processor's instance and is utilised as a data pipeline to fetch data from different sources. Also, it is responsible to send later the transformed data to the **Elasticsearch** which is part of the Data Indexing and Storage entity. The Python Scripts constitute the core of RAOHM. These scripts are responsible for jointly receive files from different sources (GoPhish and OLISTIC), and harmonise them based on a common vocabulary.

5.3.3 Data Indexing and Storage

The Data Indexing and Storage entity is responsible for storing data in the form of records. This entity contains only one technology, the Elasticsearch. Elasticsearch is a scalable engine that is responsible to efficiently search and analyse huge amounts of data. Also, it sends data to Kibana which is a particular instance of the Data Analysis and Visualisation entity.

5.3.4 Data Analysis and Visualisation

The Data Analysis and Visualisation entity is responsible for visualizing the transformed data. So that, only the Kibana is utilised. Kibana is a data visualisation tool for logs and analytics. It offers many visualisation offers such as histograms, line graphs, pie charts, heat maps, and built-in geospatial support. This assists the *end-user* of RAOHM to evaluate the results and feed other parts of SECONDO platform.

5.4 Execution

Here we aim to describe how RAOHM execution will occur including all participants and its operations. The flow diagram of RAOHM is presented in Figure 18. Firstly,

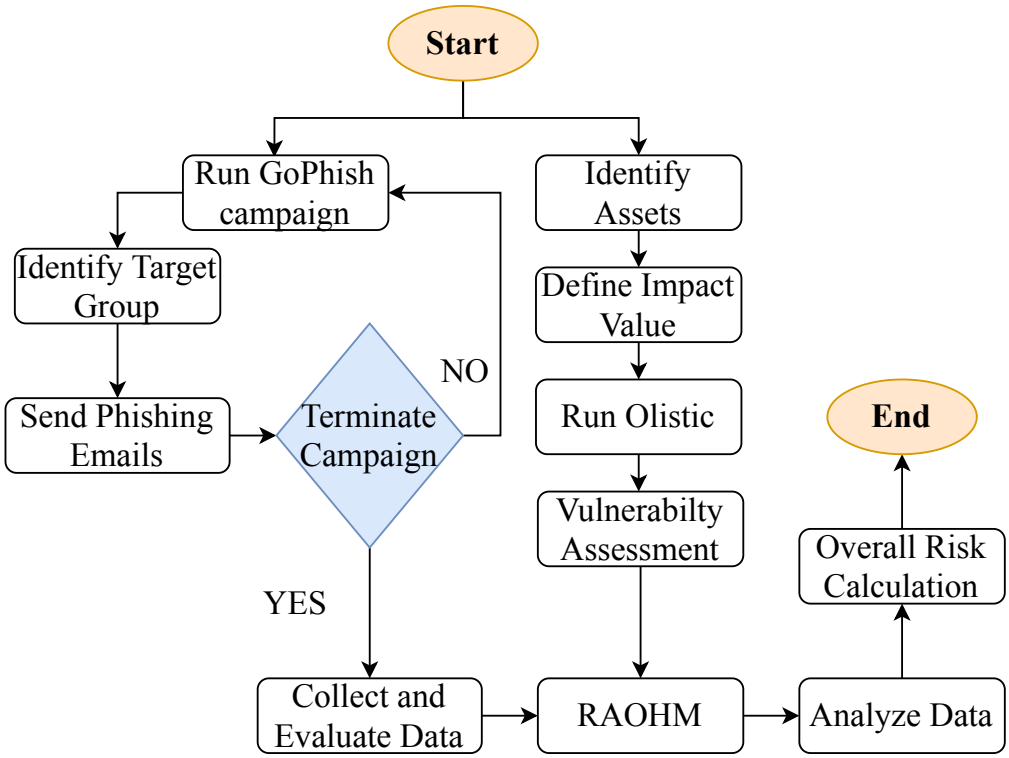


Figure 18: RAOHM Flowchart

the operator of RAOHM shall Run a GoPish Campaign. In order to do it, she must identity its target group and then the phishing emails will be automatically sent. One the emails delivered on the victims and they start interacting with them, then the handle can terminate the GoPhish campaign. If the campaign has been successfully terminated then she collects and evaluates the data. Parallel, the risk assessment team identifies the assets of the organisation which is under examination and the impact value per asset is defines. Once, these task have been terminated then OLISTIC is launched to provide the vulnerability assessment. Since both OLISTIC and Gophsish campaigns have been terminated then their outputs constitute the input of RAOHM. Then the handler of RAOHM can analyse the exported data and calculate the overall risk of the examined organisation.

5.5 Relevant File Structure

Here we aim to describe and present the files that are utilised by by RAOHM to achieve its final goals. Moreover, we describe how these files communicate and assist the achievement of the final result. The utilised files are enlisted below:

1. **constants.py**: contains the required variables which are used to the rest files.
2. **collect.py**: communicates with the *GoPhish API*, pulls and stores all the generated files from the *GoPhish campaigns* to a local folder.
3. **join.py**: reads all the stored campaign files and creates one files containing their join.
4. **seam_process.py**: locates the SEAM.csv file and process by adding two columns: a) the first one is a column that represents the likelihood of an employee to be attacked based on her role (Executive, Upper management, management and Contributor); b) the last one is a column that expresses the likelihood of being exploited.
5. **risk_determination**: reads both the updated SEA.csv file and the olisitic.csv file. Firstly, if separates the column of impact value in two different column in order to distinguish the minimum and the maximum of it. Later, it separates the column that describes the employees who use each asset. Once, this is done, it transfers the values of *likelihood of being attacked* and *likelihood of being exploited* form SEAm.csv file to the *olistic.csv* file. Once, this transfer is successfully done then the overall risk is calculated.
6. **harmonize.py**: call the collect.py, joinh.py, seam_process.py along with the risk_determination.py to start preparing the risk calculation.
7. **SEAM.csv**: contains the results from one or more GoPhish campaigns. The SEAM.csv before the execution of seam_process.py is depicted in the Figure 19a and its state after the execution of seam_process.py is depicted in the Figure 19b.
8. **olistic.csv** contains the results from the OLISTIC tool.

5.6 Usecase Risk Calculation

In order to analyse the data related to risk we should visualise them. To achieve it we use the final *csv* file created by RAOHM containing the information about the risk related to each asset. The final *csv* file by RAOHM contains data which

id	status	First Name	Last Name	email	position
emp1	Email Opened	Firstname1	Lastname1	emp1@email.com	CONTRIBUTOR
emp2	Email Sent	Firstname2	Lastname2	emp2@email.com	CONTRIBUTOR
emp3	Email Opened	Firstname3	Lastname3	emp3@email.com	UPPER_MANAGEMENT
emp4	Clicked Link	Firstname4	Lastname4	emp4@email.com	MANAGEMENT
emp5	Submitted Data	Firstname5	Lastname5	emp5@email.com	UPPER_MANAGEMENT
emp6	Submitted Data	Firstname6	Lastname6	emp6@email.com	CONTRIBUTOR
emp7	Submitted Data	Firstname7	Lastname7	emp7@email.com	CONTRIBUTOR
emp8	Clicked Link	Firstname8	Lastname8	emp8@email.com	MANAGEMENT
emp9	Submitted Data	Firstname9	Lastname9	emp9@email.com	CONTRIBUTOR
emp10	Submitted Data	Firstname10	Lastname10	emp10@email.com	CONTRIBUTOR
emp11	Submitted Data	Firstname11	Lastname11	emp11@email.com	EXECUTIVES
emp12	Email Opened	Firstname12	Lastname12	emp12@email.com	EXECUTIVES
emp13	Submitted Data	Firstname13	Lastname13	emp13@email.com	UPPER_MANAGEMENT
emp14	Clicked Link	Firstname14	Lastname14	emp14@email.com	MANAGEMENT
emp15	Submitted Data	Firstname15	Lastname15	emp15@email.com	MANAGEMENT
emp16	Email Sent	Firstname16	Lastname16	emp16@email.com	EXECUTIVES
emp17	Clicked Link	Firstname17	Lastname17	emp17@email.com	CONTRIBUTOR
emp18	Email Sent	Firstname18	Lastname18	emp18@email.com	UPPER_MANAGEMENT
emp19	Email Opened	Firstname19	Lastname19	emp19@email.com	CONTRIBUTOR
emp20	Email Opened	Firstname20	Lastname20	emp20@email.com	MANAGEMENT
emp21	Email Opened	Firstname21	Lastname21	emp21@email.com	CONTRIBUTOR
emp22	Email Opened	Firstname22	Lastname22	emp22@email.com	CONTRIBUTOR
emp23	Email Sent	Firstname23	Lastname23	emp23@email.com	CONTRIBUTOR
emp24	Email Opened	Firstname24	Lastname24	emp24@email.com	UPPER_MANAGEMENT
emp25	Submitted Data	Firstname25	Lastname25	emp25@email.com	MANAGEMENT
emp26	Submitted Data	Firstname26	Lastname26	emp26@email.com	UPPER_MANAGEMENT
emp27	Submitted Data	Firstname27	Lastname27	emp27@email.com	CONTRIBUTOR
emp28	Submitted Data	Firstname28	Lastname28	emp28@email.com	CONTRIBUTOR

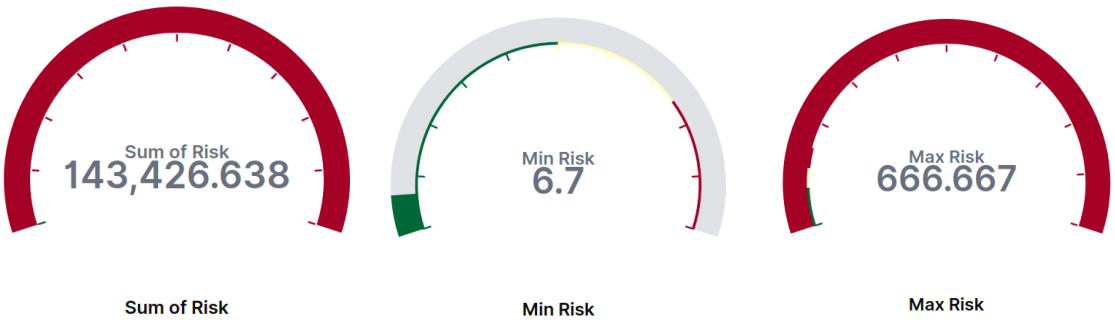
(a) SEAM.csv before seam_process.py

id	status	First Name	Last Name	email	position	pr_attack	pr_exploit_attack
emp1	Email Opened	Firstname1	Lastname1	emp1@email.com	CONTRIBUTOR	0.4	0.5555555555556
emp2	Email Sent	Firstname2	Lastname2	emp2@email.com	CONTRIBUTOR	0.4	0.5555555555556
emp3	Email Opened	Firstname3	Lastname3	emp3@email.com	UPPER_MANAGEMENT	0.27	0.5
emp4	Clicked Link	Firstname4	Lastname4	emp4@email.com	MANAGEMENT	0.27	0.8
emp5	Submitted Data	Firstname5	Lastname5	emp5@email.com	UPPER_MANAGEMENT	0.27	0.5
emp6	Submitted Data	Firstname6	Lastname6	emp6@email.com	CONTRIBUTOR	0.4	0.5555555555556
emp7	Submitted Data	Firstname7	Lastname7	emp7@email.com	CONTRIBUTOR	0.4	0.5555555555556
emp8	Clicked Link	Firstname8	Lastname8	emp8@email.com	MANAGEMENT	0.27	0.8
emp9	Submitted Data	Firstname9	Lastname9	emp9@email.com	CONTRIBUTOR	0.4	0.5555555555556
emp10	Submitted Data	Firstname10	Lastname10	emp10@email.com	CONTRIBUTOR	0.4	0.5555555555556
emp11	Submitted Data	Firstname11	Lastname11	emp11@email.com	EXECUTIVES	0.06	0.3333333333333
emp12	Email Opened	Firstname12	Lastname12	emp12@email.com	EXECUTIVES	0.06	0.3333333333333
emp13	Submitted Data	Firstname13	Lastname13	emp13@email.com	UPPER_MANAGEMENT	0.27	0.5
emp14	Clicked Link	Firstname14	Lastname14	emp14@email.com	MANAGEMENT	0.27	0.8
emp15	Submitted Data	Firstname15	Lastname15	emp15@email.com	MANAGEMENT	0.27	0.8
emp16	Email Sent	Firstname16	Lastname16	emp16@email.com	EXECUTIVES	0.06	0.3333333333333
emp17	Clicked Link	Firstname17	Lastname17	emp17@email.com	CONTRIBUTOR	0.4	0.5555555555556
emp18	Email Sent	Firstname18	Lastname18	emp18@email.com	UPPER_MANAGEMENT	0.27	0.5
emp19	Email Opened	Firstname19	Lastname19	emp19@email.com	CONTRIBUTOR	0.4	0.5555555555556
emp20	Email Opened	Firstname20	Lastname20	emp20@email.com	MANAGEMENT	0.27	0.8
emp21	Email Opened	Firstname21	Lastname21	emp21@email.com	CONTRIBUTOR	0.4	0.5555555555556
emp22	Email Opened	Firstname22	Lastname22	emp22@email.com	CONTRIBUTOR	0.4	0.5555555555556
emp23	Email Sent	Firstname23	Lastname23	emp23@email.com	CONTRIBUTOR	0.4	0.5555555555556
emp24	Email Opened	Firstname24	Lastname24	emp24@email.com	UPPER_MANAGEMENT	0.27	0.5
emp25	Submitted Data	Firstname25	Lastname25	emp25@email.com	MANAGEMENT	0.27	0.8
emp26	Submitted Data	Firstname26	Lastname26	emp26@email.com	UPPER_MANAGEMENT	0.27	0.5
emp27	Submitted Data	Firstname27	Lastname27	emp27@email.com	CONTRIBUTOR	0.4	0.5555555555556
emp28	Submitted Data	Firstname28	Lastname28	emp28@email.com	CONTRIBUTOR	0.4	0.5555555555556

(b) SEAM.csv after seam_process.py

their combination can provide many interesting information. However, here we will present results through them we can deduce interesting data about examined organisation's culture, approach, cyber awareness and employees' behaviour since they are

responsible for the successful exploitation of many attacks form the common history.



(a) Overall Risk (b) Minimum Risk (c) Maximum Risk

Figure 20: Risk Visualisations for SME from Kibana

These results are the following:

(i) **Overall risk**: represents the overall risk that characterises the examined organisation and comes from the aggregation of each calculated risk (see Figure 20a);

(ii) **Minimum Risk**: describes the minimum risk that the organisation has to cope with (see Figure 20b);

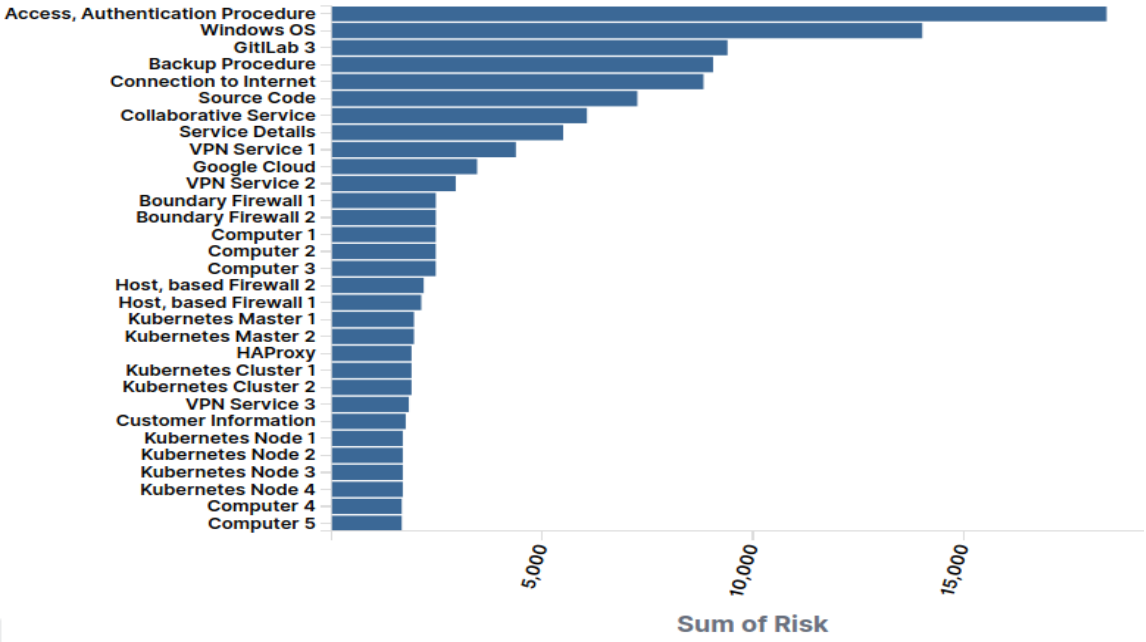
(iii) **Maximum Risk**: describes the estimated maximum risk that the organisation has to cope with (see Figure 20c);

(iv) **Overall Risk per Asset Name**: describes the overall risk per asset taking under consideration that each asset is used by more than one employees realising the asset which is in most risk (see Figures 21a and 21b);

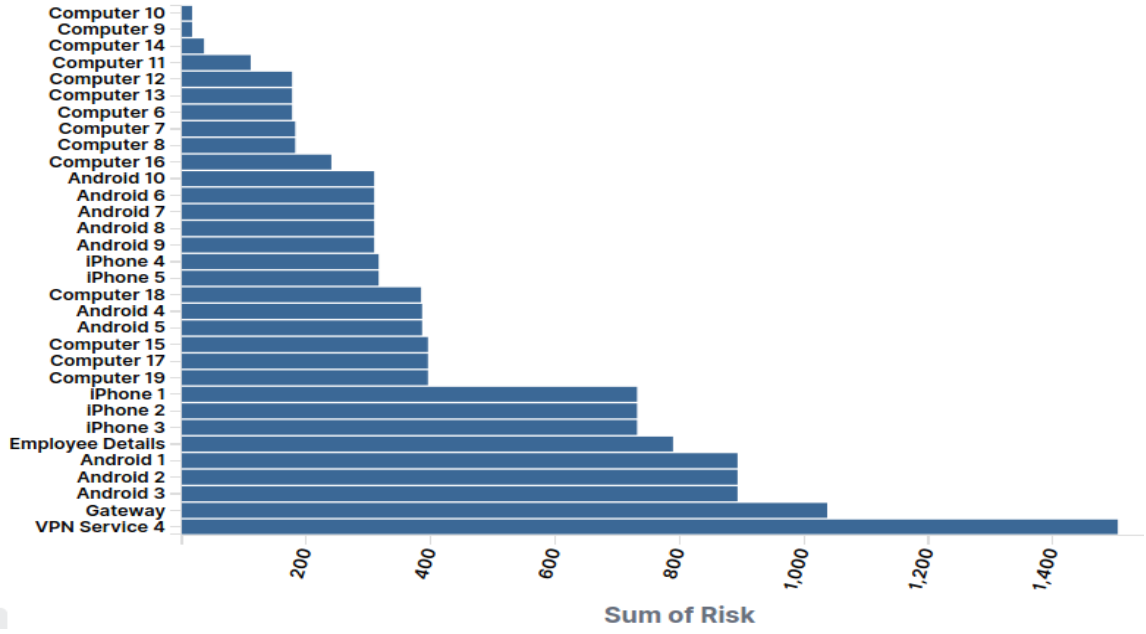
(v) **Overall Risk per Asset Category**: represents the overall risk per asset category taking under consideration that each asset type more than one unique assets (see Figure 22a);

(vi) **Overall Risk per Role**: represents the overall risk per role taking under consideration that each role is responsible for many various operations uses more than one different assets than belong in different asset types (see Figure 22b);

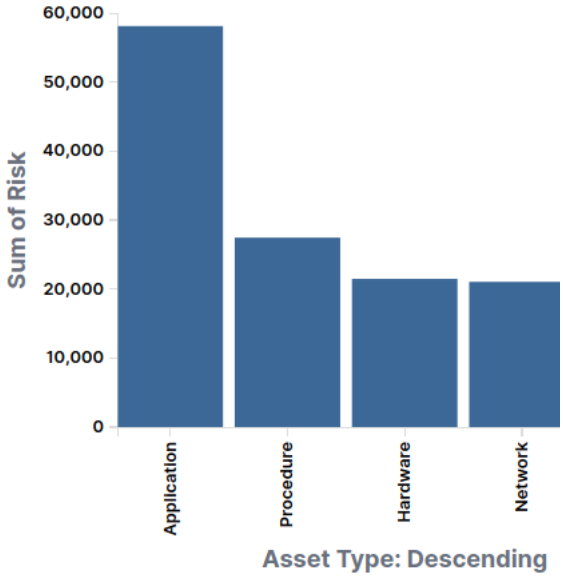
(vii) **Overall Risk per Employee**: depicts the overall risk per employee taking under consideration that each employee utilises assets that belong in different asset types (see Figure 23a and 23b.



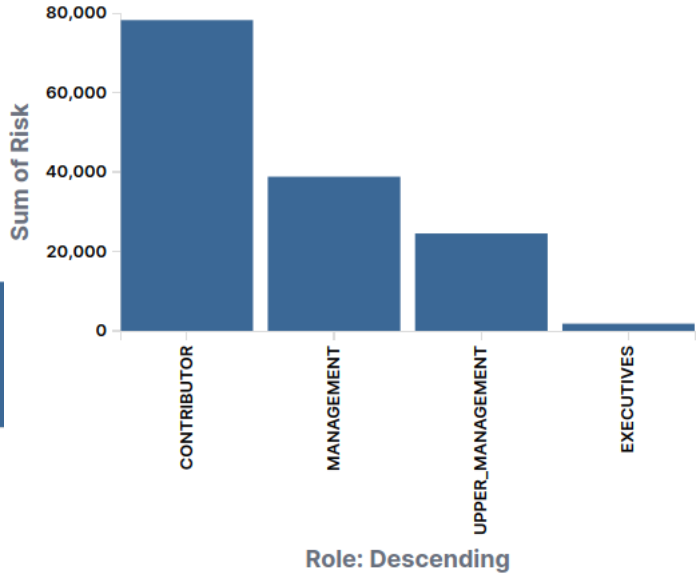
(a) Overall Risk per Asset Name part1 - Kibana



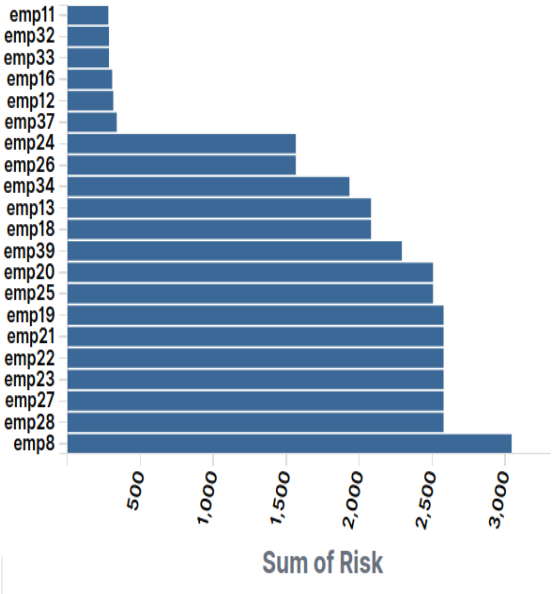
(b) Overall Risk per Asset Name part2 - Kibana



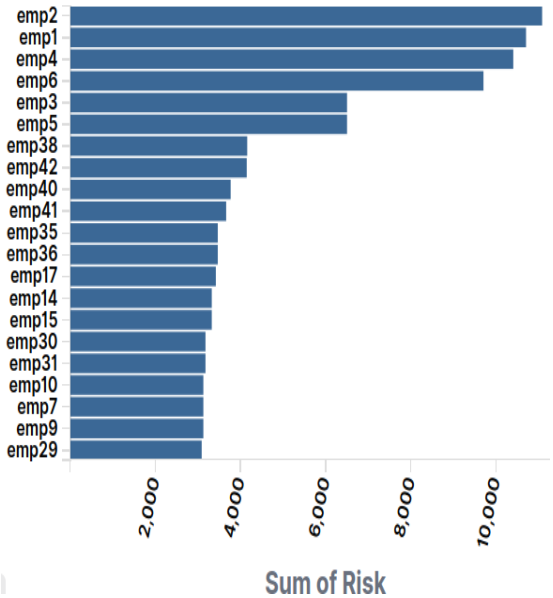
(a) Overall Risk per Asset Type



(b) Overall Risk per Role



(a) Overall Risk per Employee part 1



(b) Overall Risk per Employee part 2

6 Conclusions & Future Work

Deliverable D3.1 "Pricing Methods and Risk Modelling" presents the Risk Analysis Ontology and Harmonisation Module that is a vital module of the SECONDO project as we have already mentioned on the D2.1 "Technical requirements and reference architecture". In this deliverable we presented the stat of the art methods for pricing tangible and intangible digital assets as well as we proposed our innovative formula that is adaptable to many different organizational environments. Moreover, we analysed how we achieve to determine the overall risk of an organisation by providing detailed formulas and explaining in depth each parameter. Utilising, state of the art techniques we harmonised data from different sources (SEAM and OLISTIC) aiming to providing a detailed output information that is valuable to other modules like CRMM, ECM and GTM which will be executed based on this. The final output has been visualised in depth emerging interesting information about the examined organisation. Furthermore, well-know tools like GoPhish and OLISTIC have been directly in order to generated required data. The utilisation of these tools assists RAOHM to provide more accurate and advanced results.

For future work, we aim to successfully develop the ECM providing estimates of all kinds of costs of potential attacks and taking into account costs, (i.e. purchase, installation, execution, etc.), of each possible security control using a set of existing econometric models integrating the results of the Asset Pricing of this deliverable. In addition, the implementation of GTM that models all possible attacking scenarios and defensive strategies, (i.e. available security controls), by employing attack graphs utilising the results from the Asset Pricing of this deliverable. Finally, we will finalise the CRMM that assesses on a continuous basis the performance of the implemented risk-reducing cyber security controls allowing the adaptation of the cyber insurance contract to the changing IT environment and the evolving cyber threat landscape using the output of RAOHM.

References

[1] Keyun Ruan. *Digital Asset Valuation and Cyber Risk Measurement: Principles of Cybernomics*. Academic Press, 2019.

[2] Katharina Krombholz, Heidelinde Hobel, Markus Huber, and Edgar Weippl. Advanced social engineering attacks. *Journal of Information Security and Applications*, 22:113 – 122, 2015. Special Issue on Security of Information and Networks.

[3] Joachim Schneider, Armin J Gaul, Claus Neumann, Jürgen Hogräfer, Wolfram Wellßow, Michael Schwan, and Armin Schnettler. Asset management techniques. *International Journal of Electrical Power & Energy Systems*, 28(9):643–654, 2006.

[4] Daniel L Moody and Peter Walsh. Measuring the value of information-an asset valuation approach. In *ECIS*, pages 496–512, 1999.

[5] Charles Oppenheim, Joan Stenson, and Richard MS Wilson. The attributes of information as an asset. *New library world*, 2001.

[6] Bilge Karabacak and Ibrahim Sogukpinar. Isram: information security risk analysis method. *Computers & Security*, 24(2):147–159, 2005.

[7] Ünal Tatar and Bilge Karabacak. An hierarchical asset valuation method for information security risk analysis. In *International Conference on Information Society (i-Society 2012)*, pages 286–291. IEEE, 2012.

[8] Bomil Suh and Ingoo Han. The is risk analysis based on a business model. *Information & Management*, 41(2):149–158, 2003.

[9] Douglas J Landoll and Douglas Landoll. *The security risk assessment handbook: A complete guide for performing security risk assessments*. CRC Press, 2005.

[10] Piya Shedden, Atif Ahmad, Wally Smith, Heidi Tscherning, and Rens Scheepers. Asset identification in information security risk assessment: A business practice approach. *Communications of the Association for Information Systems*, 39(1):15, 2016.

[11] Perry Weinstein. So, what is asset management anyway? *Journal of Digital Asset Management*, 1(1):67–70, 2005.

[12] Skiff Wager. Digital asset management, media asset management, and content management: From confusion to clarity. *Journal of Digital Asset Management*, 1(1):40–45, 2005.

[13] OECD. *Tax Challenges Arising from Digitalisation – Interim Report 2018*. 2018.

[14] Michael E Whitman and Herbert J Mattord. *Principles of information security*. Cengage Learning, 2011.

[15] Poore R Spencer. Valuing information assets for security risk management. *Information Systems Security, Auerbach Publications*, 9(4), 2000.

[16] Douglas B Laney. *Infonomics: How to monetize, manage, and measure information as an asset for competitive advantage*. Routledge, 2017.

[17] Divya Rao and Wee Keong Ng. Information pricing: a utility based pricing mechanism. In *2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, pages 754–760. IEEE, 2016.

[18] Adam Saunders and Erik Brynjolfsson. Valuing information technology related intangible assets. *Mis Quarterly*, 40(1), 2016.

[19] Yuncheng Shen, Bing Guo, Yan Shen, Xuliang Duan, Xiangqian Dong, and Hong Zhang. A pricing model for big personal data. *Tsinghua Science and Technology*, 21(5):482–490, 2016.

[20] Michael R Grimaila and Larry W Fortson. Towards an information asset-based defensive cyber damage assessment process. In *2007 IEEE Symposium on Computational Intelligence in Security and Defense Applications*, pages 206–212. IEEE, 2007.

[21] Richard MS Wilson and Joan A Stenson. Valuation of information assets on the balance sheet: The recognition and approaches to the valuation of intangible assets. *Business Information Review*, 25(3):167–182, 2008.

[22] Wilco Engelsman. Information assets and their value. In *Proceedings of the 6th Twente student conference on IT. Enschede, Netherlands: University of Twente*, 2007.

[23] Suleyman Basak and Alexander Shapiro. Value-at-risk-based risk management: optimal policies and asset prices. *The review of financial studies*, 14(2):371–405, 2001.

[24] Keyun Ruan. Introducing cybernomics: A unifying economic framework for measuring cyber risk. *Computers & Security*, 65:77–89, 2017.

[25] Carlo Batini, Marco Castelli, Gianluigi Viscusi, Cinzia Cappiello, and Chiara Francalanci. Digital information asset evaluation: A case study in manufacturing. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 49(3):19–33, 2018.

[26] William F Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.

[27] John Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. In *Stochastic optimization models in finance*, pages 131–155. Elsevier, 1975.

[28] Amit Goyal. Empirical cross-sectional asset pricing: a survey. *Financial Markets and Portfolio Management*, 26(1):3–38, 2012.

[29] Robert C Merton. An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, pages 867–887, 1973.

[30] Douglas T Breeden. An intertemporal asset pricing model with stochastic consumption and investment opportunities. In *Theory of valuation*, pages 53–96. World Scientific, 2005.

[31] Stephen A Ross. The arbitrage theory of capital asset pricing. In *Handbook of the fundamentals of financial decision making: Part I*, pages 11–30. World Scientific, 2013.

[32] Umut Cetin, Robert A Jarrow, and Philip Protter. Liquidity risk and arbitrage pricing theory. In *Handbook of Quantitative Finance and Risk Management*, pages 1007–1024. Springer, 2010.

[33] Cristopher Hadnagy. *Social engineering: The art of human hacking*. 2011.

[34] Sarah Granger. Social engineering fundamentals, part i: Hacker tactics. 2003.

[35] Jason Hong. The state of phishing attacks. *Commun. ACM*, 55(1):74–81, January 2012.

[36] Forbes. Toyota parts supplier hit by $37 million email scam. Online: Last accessed: 29/5/2020.

[37] VICE. Hacker bribed 'roblox' insider to access user data. Online: Last accessed: 29/5/2020.

[38] Proofpoint. 2020 state of the phish report. Online: Last accessed: 29/5/2020.

[39] VERIZON. 2020 data breach investigations report. Online: Last accessed: 29/5/2020.

[40] J. Wright. Gophish - open source phishing framework.

[41] Gophish documentation. Online: Last accessed: 20/5/2020.

[42] Inc. Proofpoint. Protecting people: A quarterly analysis of highly targeted cyber attacks, autumn 2018.

[43] Nist. information security handbook: A guide for managers, nist special publication 800-100.

[44] Michael E. Whitman and Herbert J. Mattord. Principles of information security,fourth edition.

[45] Nist. guide for conducting risk assessment, nist special publication 800-30 revision 1.

[46] Carol Taylor, Axel Krings, and Jim Alves-Foss. Risk analysis and probabilistic survivability assessment (rapsa): An assessment approach for power substation hardening. In *Proc. ACM Workshop on Scientific Aspects of Cyber Terrorism,(SACT), Washington DC*, volume 64, 2002.

[47] Jose Cavanillas, Edward Curry, and Wolfgang Wahlster. *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe*. 01 2015.