










An Ontological Model of the Phishing Attack Process

Ítalo Oliveira¹, Gerd Wagner², Glenda Amaral¹, Tiago Prince Sales¹,
Jan-Willem Bullée³, Marianne Junger³, Dipti K. Sarmah¹, Maya
Daneva¹, and Giancarlo Guizzardi¹

¹ Semantics, Cybersecurity, and Services Group
University of Twente, Enschede, The Netherlands
{i.j.dasilvaoliveira, g.c.mouraamaral, t.princesales, d.k.sarmah,
m.daneva, g.guizzardi}@utwente.nl

² Chair of Internet Technology Institute of Informatics
Brandenburg University of Technology, Cottbus, Germany
wagnerg@b-tu.de

³ Industrial Engineering & Business Information Systems Group
University of Twente, Enschede, The Netherlands
{j.h.bullee, m.junger}@utwente.nl

Abstract. Phishing attacks are common social engineering cyber attacks in which threat actors masquerade as reputable entities to mislead recipients into performing specific actions, such as revealing financial information, system login credentials, or installing malware. Grasping the phishing attack process is crucial to prevent and counteract this type of scam. Although useful, current conceptual models describing phishing attacks do not provide an unambiguous characterization to support human understanding, communication, and computational tasks. They are informal drawings, diagrams, data models, or schemata of application-focused RDF/OWL ontologies. Instead, we approach the problem by leveraging the Unified Foundational Ontology (UFO) and OntoUML modeling language to propose a Phishing Attack Process Ontology (PAPO), making ontological commitments explicit. We show that this ontological model supports risk identification, according to ISO 31000, and satisfies important quality requirements, including domain adequacy, transparency, logical and ontological coherence, generality, as well as the FAIR principles. By providing ontological foundations for the investigation and fight against phishing attacks, PAPO paves the way for rigorous representation of corresponding real-world scenarios and enhanced applications, such as systems interoperability, data modeling, knowledge-based systems, discrete event simulations, design of phishing detection systems, and evaluation of security mechanisms' effectiveness.

Keywords: phishing attack · phishing attack process ontology · unified foundational ontology · OntoUML

1 Introduction

Social engineering is a type of cyberattack through which threat agents exploit *human vulnerabilities* to breach security goals, such as confidentiality, integrity, and availability of sensitive data [36]. Phishing is a common social engineering attack in which threat actors masquerade as reputable entities to mislead targets into performing specific actions, such as revealing financial information, system login credentials, or installing malware.

Because phishing attacks are scams that enable other crimes, such as extortion and identity theft, they lead to multiple negative impacts, namely direct damage from phishing scams (e.g. financial loss, loss of intellectual property, and sensitive customer information), impact on employee productivity, costs of business disruption, costs of business email compromise, costs of ransomware, damage to the company reputation and brand value, drop in the stock price, general weakening trust, and even compromise to national security [28,10,3].

To understand the complex dynamics of phishing attacks and design suitable preventive and control measures, such as phishing awareness training and phishing detection systems, researchers have proposed several domain conceptual models, lightweight ontologies, and informal descriptions of phishing. Although these attempts provide vocabularies and applications, they have limitations. For example, because the most detailed descriptions of phishing attacks are in natural language plain texts, drawings, informal diagrams, and schemata, they lack formal semantics, upper ontological distinctions (e.g., objects, intrinsic properties, events), serialization in multiple formats, and the capacity of being computationally queried. The phishing attack process is yet to be explicitly structured for human and machine understanding and communication.

To address this gap, we approach the problem by leveraging the *Unified Foundational Ontology* (UFO) [14] and the UFO-based OntoUML modeling language (ibid.) to propose a Phishing Attack Process Ontology (PAPO) making ontological commitments explicit. We show that this ontological model supports risk identification, according to ISO 31000 [17], and satisfies important quality requirements, including domain adequacy, transparency, logical and ontological coherence, generality, and FAIR principles. By providing ontological foundations for the investigation and fight against phishing attacks, PAPO paves the way for rigorous representation of corresponding real-world scenarios and enhanced applications, such as systems interoperability, data modeling, knowledge-based systems, discrete event simulations, design of phishing detection systems, and evaluation of security mechanisms' effectiveness.

In what follows, Section 2 discusses related works and their limitations, considering descriptions of the phishing attack process. After defining ontology requirements, Section 3 presents our main contribution: a well-founded ontological model of the phishing process, summarizing domain-specific knowledge, and accounting for the role of trust. Section 4 discusses the ontology evaluation according to the requirements described in Section 3. Section 5 concludes with final considerations, limitations, and future work.

2 Related work

Numerous works describe the phishing attack process and propose definitions. We discuss part of this related literature to ground our ontological analysis and identify their shortcomings. The word “phishing” is a variation of the term “fishing” because the act of phishing resembles that of fishing: the attacker lures a victim by employing some bait, then fishes for certain assets from the victim [10]. A 2014 study examines 113 different definitions to propose that phishing is a “scalable act of deception whereby impersonation is used to obtain information from a target” [22]. In 2021, researchers claimed that *tricking the recipient into taking the attacker’s desired action* is the *de facto* definition of phishing attacks [4]. Based on that, they define phishing as a “socio-technical attack, in which the attacker targets specific valuables by exploiting an existing vulnerability to pass a specific threat via a selected medium into the victim’s system, utilizing social engineering tricks or some other techniques to convince the victim into taking a specific action that causes various types of damages” [4]. The attacker can utilize different communication channels (emails, instant messages, voice calls, social networks, etc.) to either deceive the victim directly by a scam or to deliver a payload through an indirect manner, such as by using steganography techniques to hide malicious code within seemingly harmless files, to obtain personal or confidential information (login, passwords, bank account number, etc.) from the victim.

According to Jakobsson [20], the most representative form of phishing displays three key elements: the *lure*, the *hook*, and the *catch*. The lure consists of a phisher spamming many users with an email message that appears to be from a legitimate institution. The message frequently employs a convincing story to encourage the user to follow a hyperlink embedded in the email to a website controlled by the phisher and to provide it with certain requested information. The hook often consists of a website that emulates the appearance of a reputable agent, such as Microsoft’s login website. The goal of the hook is for victims to be directed to it via the lure message and for the victims to disclose confidential information in it. The catch involves the phisher exploiting the collected sensitive data for illegal purposes, such as fraud or identity theft.

In the phishing literature, conceptual models describing the step-by-step attack process are called “anatomy of phishing attacks” [38,4], “phishing attack processes” [3], “phishing attack lifecycle” [10,19,24], root causes analysis [1], and “information flow of phishing attacks” [20]. All these models consist of an informal mixture of natural language texts, drawings, diagrams, and tables. Therefore, they neither possess formal semantics nor a computational representation for software applications.

Current phishing attack ontologies are represented in RDF, OWL, frames, and description logics, focusing on taxonomies and narrow applications, such as detecting phishing emails [21,33,8,27,34], designing warning interfaces [39], and enhanced automatic extraction of hidden semantic information in texts [29]. All these ontologies do not follow upper ontology guidelines, hiding their ontological commitments and making interoperability difficult [13]. Moreover, their respec-

tive artifacts seem publicly unavailable, which hinders any evaluation effort (for example, checking logical consistency and possible unintended instances). These artifacts fail to comply with the FAIR principle [18] because they are not findable or accessible. An initial attempt to propose a well-founded phishing attack ontology [25] was a work-in-progress with no validation. That work does not consider the different compound phases of the phishing attack process, including planning and preparation, nor does it address trust elements. Our work builds upon all these previous initiatives to overcome their limitations.

3 A phishing attack process ontology

As there is no consensus among researchers on the exact divisions of the phishing attack process, which also depends on the type of attack, we propose an original synthesis to be represented by the Phishing Attack Process Ontology (PAPO). In our view, the following elements characterize most, if not all, forms of phishing attacks (not only phishing emails) as compound events: *(1) a phisher impersonates a reputable agent, (2) exploiting the target’s trust in this agent, (3) aiming to trick the target into taking the attacker’s desired action, (4) offering supposedly plausible reasons for this behavior.* The PAPO ontologically unfolds this understanding revealing which entities are involved and how they hang together. The ontology core presents the phishing process as a sequence of events and its participants. These events refer to different phases of the phishing attack from planning and preparation to execution and post-attack phase. We explain the role of trust in phishing attacks by leveraging a UFO-based reference ontology of trust [7]. All files related to PAPO are publicly available under the Apache 2.0 license:⁴

3.1 Ontology requirements

The functional requirement sets out the scope of PAPO. Quality requirements are domain-independent but desirable as they prescribe attributes that enable or strengthen the ontology functionalities.

The purpose of PAPO is to serve as a reference artifact for activities related to risk assessment regarding phishing attacks, particularly *risk identification*. The ISO 31000 [17] describes risk assessment as a process that includes (a) risk identification, (b) risk analysis, and (c) risk evaluation, which informs risk treatment decisions. Risk identification aims to identify and describe risks that might prevent an organization from achieving its objectives. Risk identification should consider factors such as tangible and intangible sources of risk, causes and events, vulnerabilities and capabilities, changes in the external and internal context, indicators of emerging risks, the nature and value of assets and resources, consequences and their impact on objectives, time-related factors, among others. Considering this understanding of risk identification, we define the following

⁴ Permanent link to the repository of the Phishing Attack Process Ontology (PAPO) <http://w3id.org/phishing-process-ontology/git>.

functional requirement for our proposed artifact: *PAPO must model the phishing attack process for risk identification, according to ISO 31000.*

The ontology engineering literature discusses many quality criteria for ontologies [35]. Based on that, we define the following *quality requirements* for PAPO: (1) *Domain adequacy* (appropriateness, accuracy): PAPO shall correctly represent shared real-world conceptual elements in the phishing attack process. This requirement is a necessary condition for the functional requirement satisfaction. Moreover, it is essential to enable interoperability among systems or data in the phishing domain [13]; (2) *Transparency* (intelligibility, clarity): PAPO shall contain explicit definitions for all its concepts and count with corresponding publicly available documentation; (3) *Ontological coherence*: PAPO shall comply with upper ontological distinctions; (4) *Logical consistency*: PAPO shall be free from logical contradictions, that is, it shall be consistent and satisfiable; (5) *Generality*: PAPO shall represent multiple forms of phishing (not only phishing emails, for example), meaning phishing must be characterized in a general way; (6) *FAIR principles* [18]: PAPO shall be findable, accessible, interoperable, and reusable, according to the *International Conference on Formal Ontology in Information Systems's* guidelines for ontology research artifacts [31].

3.2 Phishing-related events and their participants

In UFO, events always exist in the past and, consequently, cannot change [9]. Therefore, phishing-related events correspond to phishing incidents or related previous or posterior occurrences— for example, respectively, preparation and post-attack events. In OntoUML language, an event *b* *depends historically on* an event *a* whenever: (i) *a* (or one of its parts) brings about the situation that triggers *b* (or one of its parts); (ii) *a* (or one of its parts) brings about a situation that is necessary— but not sufficient— to trigger *b* (or one of its parts); (iii) *a* (or one of its parts) brings about a situation that is necessary— and more than sufficient— to trigger *b* (or one of its parts); or (iv) *b* depends historically on an event *z* that depends historically on *a* [6]. We employ this framework to represent the phishing attack process as a succession of orderly events. When used between objects, e.g., object A historically depends on object B, this relation means that the latter must have existed before A, i.e., B's life either *precedes*, *meets* or *overlaps* with A's life - in the sense of Allen's Algebra [6,11].

PAPO describes a PHISHING ATTACK PROCESS as a complex event necessarily composed of PHISHING ATTACK PLANNING, followed by PHISHING ATTACK PREPARATION, then PHISHING ATTACK EXECUTION with optional subsequent FRAUD and POST-ATTACK PHASE.⁵ The PHISHING ATTACK PREPARATION is historically dependent on PHISHING ATTACK PLANNING, and PHISHING ATTACK EXECUTION is historically dependent on PHISHING ATTACK PREPARATION. Due to the transitivity of the historical dependence relation, a PHISHING ATTACK EXECUTION must be historically dependent on a PHISHING ATTACK

⁵ In this paper, expressions in small caps font, such as PHISHER and PHISHING ATTACK PROCESS, refer to explicit ontology classes.

PLANNING. Although FRAUD and POST-ATTACK PHASE may not occur as part of a PHISHING ATTACK PROCESS, they are historically dependent on PHISHING ATTACK EXECUTION. This represents the expected causal temporal order of the phishing incident events, displayed in Figure 1. Besides this overview, PAPO has three views of the model: (a) Figure 2 detailing planning and preparation, (b) Figure 3 describing the attack execution, and (c) Figure 4 accounting for the role of trust in phishing attacks. Some entities participate throughout the phases of a PHISHING ATTACK PROCESS, for example, a PHISHER. Other participants join in only some events, depending on the type of incident and its particular outcomes, for example, a PHISHING TARGET does not participate in the PHISHING ATTACK PLANNING and a HOOK (for instance, a malicious webpage) may not participate at all due to the attack type.

A PHISHER is a historical role [6] played by an AGENT (an individual person or a criminal organization, hence classified as Mixin), which means an AGENT is a PHISHER *in virtue of having participated in* a PHISHING ATTACK PROCESS or one of its parts. The IMPERSONATED REPUTABLE AGENT does not directly participate in the PHISHING ATTACK PROCESS but via the impersonation performed by a PHISHER. The latter can impersonate a person trusted by the PHISHING TARGET, such as a family member or colleague, or a known organization, such as Microsoft. The PHISHER can also impersonate a made-up AGENT (say, a supposed millionaire) and persuade the PHISHING TARGET to trust it.

A PHISHING TARGET is a historical role and needs to participate in the PHISHING ATTACK PROCESS (or one of its parts). Although possibly a member of a targeted organization, a PHISHING TARGET can only be a PERSON because the attack aims to exploit human vulnerabilities related to trust, beliefs, and intentions. PHISHING ENABLERS are ancillary entities contributing to the PHISHING ATTACK EXECUTION, such as HOOK website or a malicious code embedded into an attachment.

The PHISHING ATTACK PLANNING (depicted in Figure 2), performed by a PHISHER, is an event that creates a PHISHING PLAN. This plan involves multiple decisions made by the PHISHER about whom to attack, how to attack, and for what purpose: (a) the selection of the ATTACK METHOD, which can be a malware-based attack (for example, via keylogger or ransomware embedded into PDF files sent by emails), a webpage-based attack (typically, via a redirected HOOK webpage requesting for sensitive information), a direct request (for instance, in-person or online solicitations), among others; (b) the choice of MESSAGE MEDIUM TYPE, which can be SMS, email, social network, in-person direct communication, etc.; (c) the decision about the FRAUD TYPE to be enabled by the PHISHING ATTACK EXECUTION, such as extortion or identity theft. The PHISHING PLAN also establishes the AGENTS to be impersonated by the PHISHER, the AGENTS to be tricked (who will turn into PHISHING TARGETS once they participate in the attack process), the aimed TARGET ASSETS (for example, login credentials, money, or sensitive bank information), and it considers

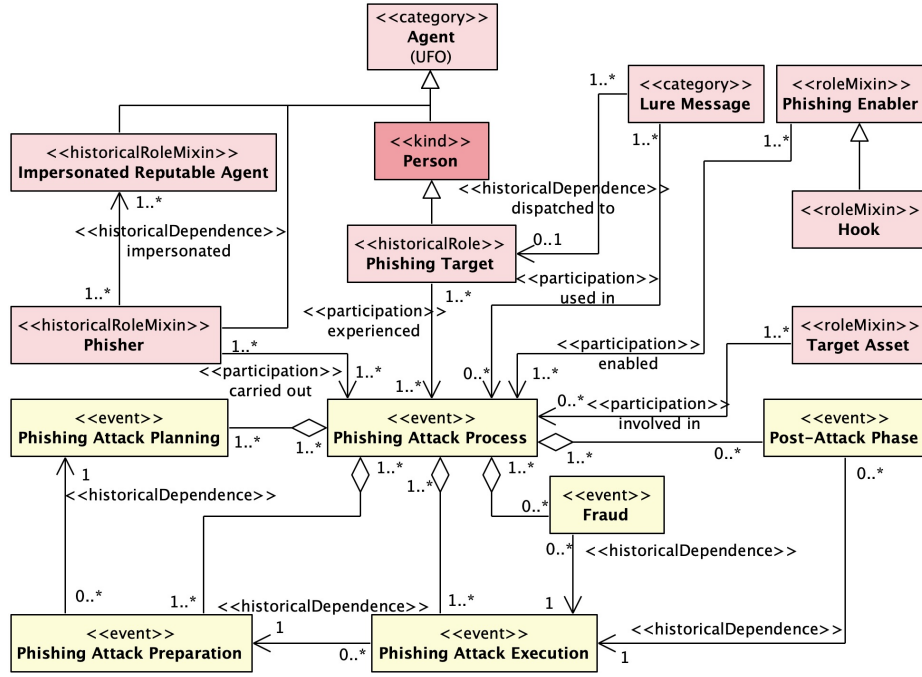


Fig. 1. Phishing attack process overview. The OntoUML stereotypes connect types and relations in these models to ontological categories of monadic and relational universals in UFO, respectively. [14] The colors in these diagrams represent a convention used by the OntoUML community: object types are represented in pinkish, intrinsic aspect types in blue, and event types in yellow, truth-makers of material relations in green, and higher-order types in darker blue.

relevant PHISHING ENABLERS. The PHISHING PLAN specifies what is known as a “phishing campaign” in the cybersecurity community.⁶

The next phase is the PHISHING ATTACK PREPARATION (also depicted in Figure 2). The preparation consists of events wherein the PHISHER acquires the needed PHISHING ATTACK CAPABILITIES for the PHISHING ATTACK EXECUTION. These include gathering hacking knowledge and target information (email addresses, for example), spotting vulnerabilities, setting up HOOK webpages and malware, creating LURE MESSAGES, purchasing phishing kits, etc. The PHISHING ATTACK EXECUTION is the manifestation of those PHISHING ATTACK CAPABILITIES.

The PHISHING ATTACK EXECUTION is what we usually recognize as a phishing attack. In this phase, shown in Figure 3, the ultimate attack’s goal is to mislead PHISHING TARGETS into performing specific actions according to the

⁶ “(...) the MITRE ATT&CK team uses the term Campaign to describe any grouping of intrusion activity conducted over a specific period of time with common targets and objectives.” (<https://attack.mitre.org/campaigns/>).

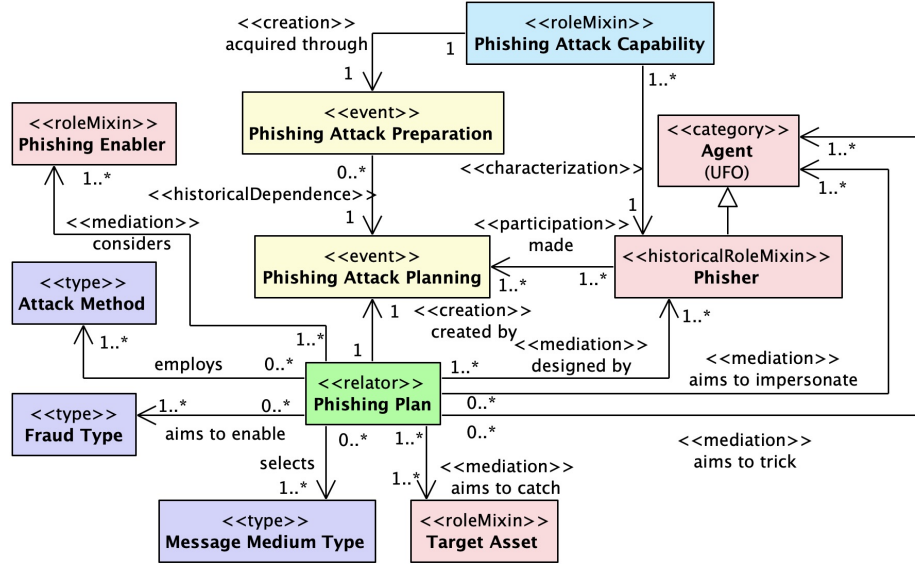


Fig. 2. Phishing planning and preparation

PHISHING PLAN devised by the PHISHER. The PHISHING ATTACK EXECUTION is a complex event composed of several others that occur in succession, varying according to the employed ATTACK METHOD. However, PAPO is designed to capture the essence of this attack execution, that is, what is shared by different forms of phishing. So, a single PHISHING ATTACK EXECUTION is necessarily composed of a single LURE MESSAGE DISPATCH and, contingently, a LURE MESSAGE ARRIVAL event, a LURE MESSAGE PERCEPTION event, and a FULFILLMENT OF PHISHER'S REQUEST event, all of which are temporally ordered by historical dependence relations. For each PHISHING PLAN there may be (and often there is) a large number of PHISHING ATTACK EXECUTION events as manifestations of the PHISHER'S PHISHING ATTACK CAPABILITIES.⁷

In the common case of a phishing email attack, each PHISHING ATTACK EXECUTION corresponds to a single email addressed to a single PHISHING TARGET. It starts when the PHISHER dispatches an email LURE MESSAGE addressed to a PHISHING TARGET's email address. The arrival of this email may be prevented by some security measures like spam filters or due to the wrong email address. A possible automatic system reply would be an event historically dependent on the email dispatch and may be a useful piece of information for PHISHERS. The email arrival in the target's inbox is a LURE MESSAGE ARRIVAL. For some reason, the email may never be seen by the user. The event wherein the PHISHING

⁷ In general, there is a trade-off between scalability and effectiveness of a phishing attack because personalized LURE MESSAGES tend to be more effective, though are harder to scale [23]. However, by leveraging generative artificial intelligence techniques, attackers can produce personalized lure messages at scale [15].

TARGET sees, opens, and reads the email corresponds to the LURE MESSAGE PERCEPTION event. If for whatever reason the PHISHING TARGET clicks the link in the email and downloads malware (PHISHING ENABLER), they are performing a FULFILLMENT OF PHISHER'S REQUEST event under certain conditions. These represent the situations where PHISHING TARGETS are prone to fall for a phish answering the requests. This effect happens due to the presence of certain TARGET FRAGILITIES, which are various kinds of VULNERABILITIES including ignorance, inexperience, prejudice or bias, conformity, hurry, intuitive judgment, laziness, curiosity, tiredness, fear, habits, anger, excitement, tension, happiness, sadness, disgust, guilt, surprise, greed, lust, neuroticism, and many others [37]. The FULFILLMENT OF PHISHER'S REQUEST event may be composed of several sub-events, depending on ATTACK METHOD. For example, in a webpage-based attack, the PHISHING TARGET firstly clicks the link inside the email message, visits a HOOK webpage, and, finally, proceeds with sending the data via it.

A FULFILLMENT OF PHISHER'S REQUEST event concludes a PHISHING ATTACK EXECUTION. What comes next is the FRAUD phase, which involves committing a *different crime* enabled by successful phishing attacks. The PHISHER can commit criminal identity theft (posing as another person when apprehended for a crime), financial identity theft (using another's identity to obtain credit, goods, and services), identity cloning (using another's information to assume his or her identity in daily life), medical identity theft (using another's identity to obtain medical care or drugs), among other crimes. Before, in parallel, or after the FRAUD phase, the PHISHER can proceed with the *Post-attack* phase by taking measures to protect themselves and assessing the results. This includes deleting HOOK websites, shutting down attack machinery, tracking hunters, money laundering, etc.

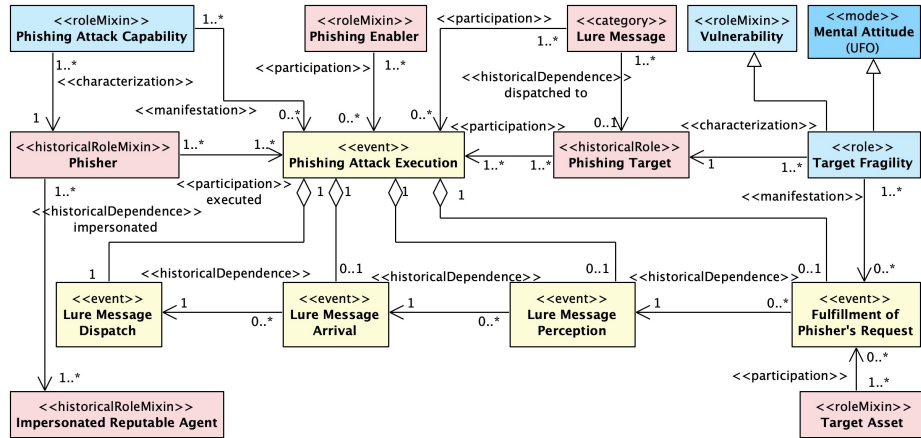


Fig. 3. Phishing attack execution

3.3 Phishing and trust

To succeed, a PHISHING ATTACK EXECUTION requires existing trust relations between the PHISHING TARGET (TRUSTOR) and IMPERSONATED REPUTABLE AGENTS (TRUSTEE). Victims trust people or organizations to satisfy a particular objective. Different entities play the role of TRUSTEE, for example, the IMPERSONATED REPUTABLE AGENTS, the LURE MESSAGE, and the HOOK WEBPAGE. To represent these relations suitably, we employ the *Reference Ontology of Trust* (ROT) [2], a UFO-based ontology. Figure 4 describes the role of trust in phishing attacks by leveraging ROT’s concepts.

ROT defines TRUST as a complex mental state of a TRUSTOR agent, composed of an INTENTION related to a goal, for the achievement of which she counts upon the TRUSTEE, and a set of BELIEFS about the TRUSTEE and its behavior. A TRUSTOR is necessarily an “intentional entity”, that is, a cognitive agent, an agent endowed with goals and beliefs. As for the TRUSTEE, it is an entity capable of impacting one’s intentions by the outcome of its behavior. The TRUST mental state of a TRUSTOR regarding a TRUSTEE and its behavior is composed of (i) an intention of the TRUSTOR; (ii) a set of beliefs about the TRUSTEE’s capabilities, vulnerabilities, and commitments; and (iii) if the TRUSTEE is an agent, beliefs that the TRUSTEE intends to exhibit the expected behavior. Another ontological commitment of ROT is that TRUST implies risks. By trusting, the TRUSTOR becomes vulnerable to the TRUSTEE in terms of potential failure of the expected behavior or outcome.

ROT also provides an ontological account of the factors that can influence TRUST. The ontology categorizes influence relations according to the ontological nature of the factors that explain them, namely: (F1) other TRUST relations; (F2) mental biases; (F3) TRUST CALIBRATION SIGNALS; and (F4) TRUSTWORTHINESS EVIDENCE. F1 represents the situation in which TRUST is influenced by other trust relations. F2 represents situations in which TRUST is influenced by MENTAL MOMENTS. Examples of MENTAL MOMENTS include PERCEPTIONS, BELIEFS, DESIRES and INTENTIONS. F3 categorizes situations in which TRUST can be influenced by trust signals purposefully emitted by the TRUSTEE, to indicate trustworthy behavior. Some examples are uniforms and established brands to create visual identities. In F4, the influence comes from shreds of evidence suggesting that a TRUSTEE could be trusted. They result from TRUSTEES’ trustworthy actions. Examples include third-party certifications and credentials, performance history, track record, recommendations, reputation records, and past successful experiences.

Several factors can influence trust: (i) other trust relations; (ii) TRUSTOR’s mental biases and mental states (perceptions, beliefs, desires, and intentions); (iii) trust signals emitted by the TRUSTEE, such as uniforms, logotypes, and speech style; and (iv) trustworthiness evidence that demonstrates trustworthy behavior on the part of the TRUSTEE. In the case of phishing attacks, trust signals correspond to fake signals emitted by, for example, a HOOK WEBPAGE and LURE MESSAGE, while trustworthiness evidence can be either fake evidence produced by the attacker or true evidence about IMPERSONATED REPUTABLE

AGENTS, such as past successful experiences, performance history, recommendations, and certifications by trusted parties. Because the PHISHING TARGETS trust both a IMPERSONATED REPUTABLE AGENT and a LURE MESSAGE, they follow the instructions in the LURE MESSAGE, according to the PHISHER's expectations.

Consider the example in which the PHISHER sends an email (LURE MESSAGE) to Mary (PHISHING TARGET) impersonating a trusted airline company X (IMPERSONATED REPUTABLE AGENT). The LURE MESSAGE offers discounted flight tickets to attractive destinations. Airline X is a traditional trusted company and thus Mary trusts X to purchase flight tickets (TRUSTOR's INTENTIONS). She believes that (1) Airline X is capable of taking her to the desired destination if she buys the tickets; (2) the airline will provide her with tickets for the trip after she makes the payment; and (3) Airline X has the INTENTION to do (1) and (2). Mary's trust in Airline X influences her to trust the email. Other factors that can positively influence Mary's trust in the email are her excitement and desire to travel, the email's design according to airline X visual identity (TRUST CALIBRATION SIGNAL), and TRUSTWORTHINESS EVIDENCE about Airline X, such as Mary's past successful experiences flying with them. Based on her trust in the email, Mary clicks on the link to access Airline X's (fake) website (HOOK webpage) to buy the tickets. Mary's trust in the airline along with the TRUST CALIBRATION SIGNALS emitted by the fake website positively influenced her to trust the HOOK WEBPAGE too, leading her to provide her credit card data to pay for the tickets. At this point, the PHISHER has reached his goal, accomplishing a successful phishing attack.

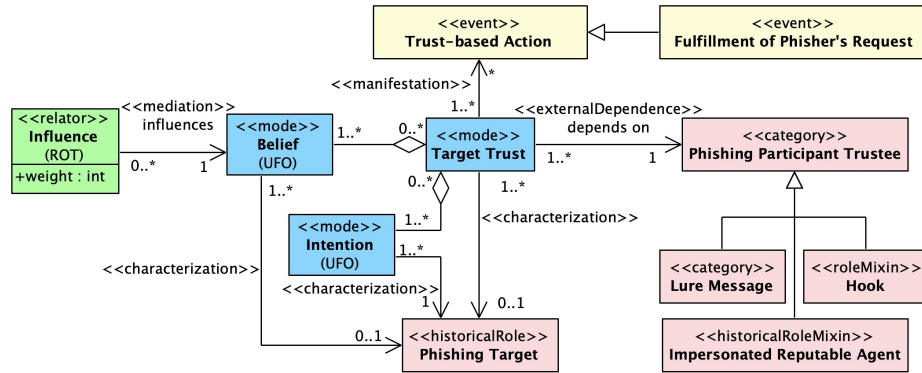


Fig. 4. Trust elements involved in Target answering phisher's requests

4 Ontology evaluation

To show that PAPO can support risk identification activities defined in ISO 31000, explained in Section 3.1, we illustrate a phishing incident through a UML Object diagram that instantiates PAPO, as shown in Figure 5. Incidents are realized instances of risks. Because of this, risk analysis and incident analysis are similar activities. The former investigates possible occurrences, whereas the latter looks into past events, but both share the goal of informing risk treatment decisions.

Our illustration describes a realistic incident wherein the APT16⁸ executes a successful spear phishing attack, that is, a personalized phishing attack directed to a PHISHING TARGET, namely, the SoftBank CTO.⁹ The Spear Phishing Campaign is the PHISHING PLAN, which employs the Hookpage-based Spear Phishing as its ATTACK METHOD, aims to enable Espionage as its Fraud Type, selects Email as its Message Medium Type, to catch SoftBank Secret Information on Cutting-Edge Technology (TARGET ASSET). This campaign also specifies the impersonation of the Japanese Financial Services Agency as its IMPERSONATED REPUTABLE AGENT to trick the SoftBank CTO by exploiting his trust in this authority. The PHISHING ATTACK PREPARATION involves the Set up Hookpage 1 event. The Spear Phishing Execution is composed of Spear Phishing Email Dispatch on day X at time Y, followed by SoftBank CTO submits data on Hookpage 1 (FULFILLMENT OF PHISHER’S REQUEST). This event is the manifestation of the SoftBank CTO’s Fatigue at a given moment and his Trust in Email 1, which is composed of his Belief in the authenticity of the Email 1 and Intention to answer that Japanese authority requests.

Furthermore, we can leverage the *Common Ontology of Value and Risk* (COVER) [30] to expand this analysis. For example, by creating event types for certain events of PHISHING ATTACK PROCESS, we can assign CAUSAL LIKELIHOOD values to these event types and update the assignments according to real or simulated phishing incidents. The interoperability between COVER and PAPO is facilitated by the fact that both are UFO-based reference ontologies. For instance, PHISHING ATTACK EXECUTION can be seen as a subtype of COVER’s RISK EVENT, where LURE MESSAGE DISPATCH is a THREAT EVENT and FULFILLMENT OF PHISHER’S REQUEST is a LOSS EVENT, just like FRAUD. Naturally, these LOSS EVENTS positively impact the PHISHER’S INTENTIONS (objectives), at the same time that they hurt PHISHING TARGET’S INTENTIONS. These considerations show that PAPO can cover all major elements of risk identification related to phishing attacks, according to ISO 31000, as described in Section 3.1, including tangible and intangible sources of risk, causes and events, vulner-

⁸ “APT16 is a China-based threat group that has launched spearphishing campaigns targeting Japanese and Taiwanese organizations.” (<https://attack.mitre.org/groups/G0023/>).

⁹ We use expressions in LaTeX teletype font, such as APT16 and SoftBank CTO, to denote individuals of the UML Object diagram.

abilities and capabilities, assets, consequences and their impact on objectives, time-related factors, etc.

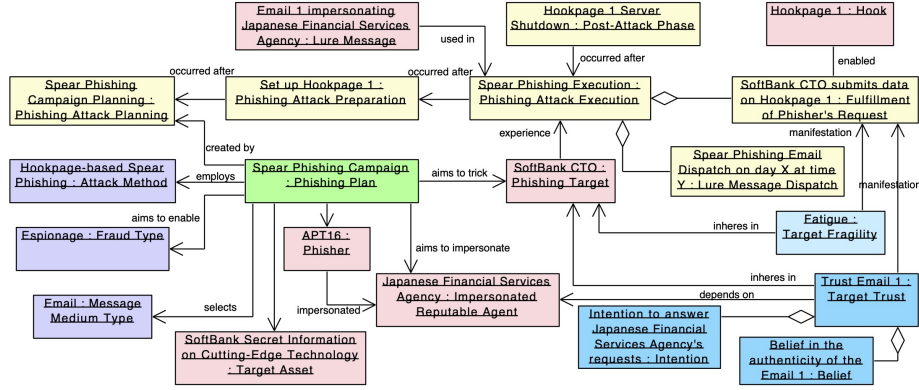


Fig. 5. An illustration of a phishing incident

Let us now consider each quality requirement (QR). The above illustration is evidence of PAPO's domain adequacy (QR1). The consideration of phishing dataset schemata is also a factor that favors domain appropriateness.¹⁰ In fact, common phishing dataset features are easily mapped into PAPO concepts, such as the impersonated brand (IMPERSONATED REPUTABLE AGENT), URL and its status (active/inactive) as attributes of a HOOK webpage, and email content as an attribute of LURE MESSAGE.

PAPO accomplishes QR2 (transparency) because every concept is defined and documented in multiple formats (VPP, JSON, TTL, and HTML). Ontological coherence (QR3) is obtained by complying with UFO and OntoUML, verified by a service of the OntoUML plugin [12]. Logical consistency (QR4) has been checked for the TTL version of PAPO, which is automatically generated via the plugin from an OWL implementation of UFO [5]. This verification has been performed in Protégé [32] using the HermiT reasoner. Generality (QR5) has already been shown as our very definition of phishing, captured by PAPO, includes multiple forms of phishing attacks, unlike other models in the literature. Finally, PAPO is publicly available following FAIR principles (QR6): stably findable and accessible in multiple formats, including metadata, reusable, and interoperable with other ontologies, powered by *explicit* ontological commitments—a key requirement for interoperability [13].

¹⁰ Examples of phishing datasets include: (a) https://www.phishtank.com/developer_info.php, the Anti-Phishing Working Group (APWG)'s datasets, datasets on Kaggle (for instance, <https://www.kaggle.com/datasets/arnavs19/phishing-websites-dataset>), a dataset for web page phishing detection [16], and many others.

5 Conclusion

Phishing attacks involve intricate relations among physical, social, technical, and psychological entities. They have manifold negative impacts on people and organizations and are hard to manage and fight against. Understanding and modeling phishing attacks is crucial in this context. However, current models are mostly drawings, informal diagrams, data models, or schemas of application-focused RDF/OWL ontologies. This imposes limitations on what they can do as information tools for human understanding and communication, and supporting machine tasks. We addressed this gap by proposing a well-founded Phishing Attack Process Ontology (PAPO), grounded in the Unified Foundational Ontology (UFO). The resulting artifact is a model described in the OntoUML language, accounting for the phishing attack process, divided into planning, preparation, execution, fraud, and post-attack. This model ontologically unpacks our characterization of the phishing attack process as a complex event wherein: (1) a phisher impersonates a reputable agent, (2) exploits the target’s trust in this agent, (3) aims to trick the target into taking the attacker’s desired action, (4) offering supposedly plausible reasons for this behavior.

This is the first model of the phishing attack process that satisfies several important quality requirements (detailed in Section 3.1). PAPO stands out for complying with FAIR principles for scientific artifacts and a level of generality that allows the representation of different forms of phishing (email, SMS, social network, in-person communication, etc.). Moreover, it accounts for the role of trust and human vulnerabilities in the attack process.

PAPO, however, is limited to the overall attack description. It does *not* include security elements, which can be placed to counteract each phase of the phishing attack process. Moreover, a detailed taxonomy of phishing attacks is not included in PAPO by default, but can be added by subtyping PAPO’s classes.

Once we have established ontological foundations of phishing attacks, we intend to employ them to support risk treatment activities (the design and evaluation of security mechanisms) and phishing research, including experiments, empirical phishing research, simulations, and RDF knowledge graph construction (for example, via phishing dataset integrations). The natural next steps are modeling security aspects via alignment with the Reference Ontology of Security Engineering (ROSE) [26].

References

1. Abroshan, H., Devos, J., Poels, G., Laermans, E.: Phishing attacks root causes. In: Risks and Security of Internet and Systems: 12th International Conference, CRiSIS 2017, Dinard, France, September 19-21, 2017, Revised Selected Papers 12. pp. 187–202. Springer (2018)
2. Akbar, N.: Analysing Persuasion Principles in Phishing Emails. Master’s thesis, University of Twente (2014), <https://purl.utwente.nl/essays/66177>
3. Aleroud, A., Zhou, L.: Phishing environments, techniques, and countermeasures: A survey. *Computers & Security* **68**, 160–196 (2017)

4. Alkhalil, Z., Hewage, C., Nawaf, L., Khan, I.: Phishing attacks: A recent comprehensive study and a new anatomy. *Frontiers in Computer Science* **3**, 563060 (2021)
5. Almeida, J.P.A., Guizzardi, G., Sales, T.P., Falbo, R.A.: gUFO: A lightweight implementation of the unified foundational ontology (UFO) (2019), <http://purl.org/nemo/doc/gufo>
6. Almeida, J.P.A., Falbo, R.A., Guizzardi, G.: Events as entities in ontology-driven conceptual modeling. In: *Conceptual Modeling: 38th International Conference, ER 2019, Salvador, Brazil, November 4–7, 2019, Proceedings 38*. pp. 469–483. Springer (2019)
7. Amaral, G., Sales, T.P., Guizzardi, G., Porello, D.: Towards a reference ontology of trust. In: *On the Move to Meaningful Internet Systems: OTM 2019 Conferences: Confederated International Conferences: CoopIS, ODBASE, C&TC 2019, Rhodes, Greece, October 21–25, 2019, Proceedings*. pp. 3–21. Springer (2019)
8. Bazarganigilani, M.: Phishing e-mail detection using ontology concept and naive bayes algorithm. *International Journal of Research and Reviews in Computer Science* **2**(2), 249 (2011)
9. Benevides, A.B., et al.: Representing a reference foundational ontology of events in sroiq. *Applied Ontology* **14**(3), 293–334 (2019)
10. Chiew, K.L., Yong, K.S.C., Tan, C.L.: A survey of phishing attacks: Their types, vectors and technical approaches. *Expert Systems with Applications* **106**, 1–20 (2018)
11. Fonseca, C.M., Porello, D., Guizzardi, G., Almeida, J.P.A., Guarino, N.: Relations in ontology-driven conceptual modeling. In: *Conceptual Modeling: 38th International Conference, ER 2019, Salvador, Brazil, November 4–7, 2019, Proceedings 38*. pp. 28–42. Springer (2019)
12. Fonseca, C.M., Sales, T.P., Viola, V., Da Fonseca, L.B., Guizzardi, G., Almeida, J.P.A.: Ontology-driven conceptual modeling as a service. In: *CEUR workshop proceedings*. vol. 2969. Rheinisch Westfälische Technische Hochschule (2021)
13. Guizzardi, G.: Ontology, ontologies and the "I" of FAIR. *Data Intelligence* **2**(1-2), 181–191 (2020)
14. Guizzardi, G., Botti Benevides, A., Fonseca, C.M., Porello, D., Almeida, J.P.A., Sales, T.P.: UFO: Unified foundational ontology. *Applied ontology* **17**(1), 1–44 (2022)
15. Gupta, M., Akiri, C., Aryal, K., Parker, E., Praharaj, L.: From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy. *IEEE Access* (2023)
16. Hannousse, A., Yahiouche, S.: Web page phishing detection (2021). <https://doi.org/10.17632/c2gw7fy2j4.3>
17. ISO: ISO 31000:2018 - Risk management – Guidelines (2018)
18. Jacobsen, A., et al.: FAIR principles: interpretations and implementation considerations. *Data Intelligence* **2**(1-2), 10–29 (2020)
19. Jain, A.K., Gupta, B.: A survey of phishing attack techniques, defence mechanisms and open research challenges. *Enterprise Information Systems* **16**(4), 527–565 (2022)
20. Jakobsson, M., Myers, S.: Phishing and countermeasures: understanding the increasing problem of electronic identity theft. John Wiley & Sons (2006)
21. Kerremans, K., Tang, Y., Temmerman, R., Zhao, G.: Towards ontology-based e-mail fraud detection. In: *2005 portuguese conference on artificial intelligence*. pp. 106–111. IEEE (2005)
22. Lastdrager, E.E.: Achieving a consensual definition of phishing based on a systematic review of the literature. *Crime Science* **3**(1), 1–10 (2014)

23. Lastdrager, E.E.H.: From fishing to phishing. PhD thesis, University of Twente, Enschede, Netherlands (February 2018), available at <https://doi.org/10.3990/1.9789036544795>
24. Mohammad, R.M., Thabtah, F., McCluskey, L.: Tutorial and critical analysis of phishing websites methods. *Computer Science Review* **17**, 1–24 (2015)
25. Oliveira, Í., Calhau, R.F., Guizzardi, G.: Toward a phishing attack ontology. In: *ER-Companion 2023: ER Forum, 7th SCME, Project Exhibitions, Posters and Demos, and Doctoral Consortium*. pp. 10–21. No. 3618 in *CEUR Workshop Proceedings*, Aachen (2023), https://ceur-ws.org/Vol-3618/forum_paper_25.pdf
26. Oliveira, Í., Sales, T.P., Baratella, R., Fumagalli, M., Guizzardi, G.: An ontology of security from a risk treatment perspective. In: *International conference on conceptual modeling*. pp. 365–379. Springer (2022)
27. Park, G., Rayz, J.: Ontological detection of phishing emails. In: *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. pp. 2858–2863. IEEE (2018)
28. Ponemon Institute LLC: The 2021 Cost of Phishing Study. Tech. rep., Ponemon Institute LLC (2021), <https://www.proofpoint.com/au/resources/analyst-reports/ponemon-cost-of-phishing-study>
29. Raskin, V., Taylor, J.M., Hempelmann, C.F.: Ontological semantic technology for detecting insider threat and social engineering. In: *Proceedings of the 2010 New Security Paradigms Workshop*. pp. 115–128 (2010)
30. Sales, T.P., Baião, F., Guizzardi, G., Almeida, J.P.A., Guarino, N., Mylopoulos, J.: The common ontology of value and risk. In: *Conceptual Modeling. ER 2018*. vol. 11157, pp. 121–135. Springer (2018)
31. Bonino da Silva Santos, L.O., dos Santos Vieira, B., Bernabé, C.H.: FAIR FOR FOIS. <https://w3id.org/FAIR-academic/fair4fois> (2024), [Online; accessed on May of 2024]
32. Stanford Center for Biomedical Informatics Research: Protégé (2024), <https://protege.stanford.edu/>
33. Tchakounté, F., Molengar, D., Ngossaha, J.M.: A description logic ontology for email phishing. *International Journal of Information Security Science* **9**(1), 44–63 (2020)
34. Tseng, S.S., Ku, C.H., Lee, T.J., Geng, G.G., Wang, Y.J.: Building a frame-based anti-phishing model based on phishing ontology. In: *International Conference on Advances in Information Technology* (2013)
35. Vrandečić, D.: Ontology evaluation. In: *Handbook on ontologies*, pp. 293–313. Springer (2009)
36. Wang, Z., Sun, L., Zhu, H.: Defining social engineering in cybersecurity. *IEEE Access* **8**, 85094–85115 (2020)
37. Wang, Z., Zhu, H., Sun, L.: Social engineering in cybersecurity: Effect mechanisms, human vulnerabilities and attack methods. *IEEE Access* **9**, 11895–11910 (2021)
38. Wetzel, R.: Tackling phishing. *Business Communications Review* **35**(2), 46–49 (2005)
39. Zahedi, F.M., Chen, Y., Zhao, H.: Ontology-based intelligent interface personalization for protection against phishing attacks. *Information Systems Research* (2023)