[seal]
EPFL . HEIG-VD . UNIL

Projet d'innovation

dans le cadre de l'appel à projets [seal] sur la thématique
« Donner aux petites organisations les moyens de lutter contre la croissance exponentielle de la cybercriminalité »

Combattre le phishing : quelles innovations apporter ?

Décembre 2024 - Novembre 2025

Partenaires associés au projet :

Auteur·e·s du document :
Prof. Sylvain Pasini (HEIG-VD), porteur de projet
Alexis Martins (HEIG-VD)
Axel Vallon (HEIG-VD)
Pablo Saez (HEIG-VD)
Valentin Diaz (Unil)
Carine Dengler (EPFL)
Linus Gasser (EPFL)

Version et date du document :
Version 1.0, 16 décembre 2025

Résumé du projet

La grande majorité des incidents de cybersécurité exploitent une attaque par «phishing». Cette technique se situe aux frontières des aspects sociaux, techniques, et légaux. Ce projet s'inscrit dans une démarche exploratoire visant à répondre à ce **besoin sociétal** : *réduire l'impact du phishing et renforcer la résilience des organisations et de la société.*

Ce travail a adopté une approche interdisciplinaire et collaborative, permettant de croiser les expertises en mobilisant les acteurs institutionnels, les industriels, les universités partenaires, ainsi que les utilisateur·rice·s finaux·ales.

Le travail s'est articulé en quatre phases :

1. Afin d'identifier les points défaillants, il a été nécessaire de récolter des **retours de terrains** et d'en extraire les modes opératoires. Un **état de l'art** a permis de lister les mesures de protections (SPF, DKIM, filtres anti-spam, et autres), les pratiques organisationnelles (politiques, formation, campagnes, etc.) et le cadre légal suisse.
2. Des **ateliers collaboratifs** ont permis de partager les connaissances et croiser les regards sur la situation. Ils ont également permis de générer des **idées innovantes** répondant au besoin sociétal. Les idées ont été approfondies et analysées afin d'en extraire les plus pertinentes à recommander.
3. Parmi les idées innovantes, deux se sont révélées réalisables dans le cadre de ce projet des **prototypes fonctionnels** ont pu être réalisés et testés. Ces derniers visent à **réduire les clics sur des liens frauduleux** à l'aide d'un retour visuel à l'utilisateur suite à une évaluation automatique.
4. La finalité du projet était de **diffuser** largement les résultats, notamment à l'aide du présent rapport, mais également avec d'autres actions de **sensibilisation et dissémination** dans l'objectif de maximiser l'impact auprès du grand public, des professionnel·le·s, et des institutions.

Au final, les ateliers collaboratifs se sont révélés être une approche très pertinente.

Le projet a donc permis de dresser un état des lieux complet de la lutte contre le phishing. Il a permis de lister des recommandations concrètes pour renforcer la résilience des petites structures et de la société. En plus des idées innovantes proposées dans ce document, il ressort un fort besoin pour la Suisse d'une plateforme centralisée et unique rassemblant les éléments d'information, formation, annonce et accompagnement.

Table des matières

Chapitre 1

Introduction

Table des matières du chapitre

La cybersécurité constitue aujourd'hui un enjeu majeur pour les organisations de toutes tailles, en particulier pour les plus petites, qui disposent de moyens limités pour se protéger. Les retours d'expérience récents soulignent que la majorité des incidents de sécurité découlent encore d'attaques de «phishing», de mauvaises pratiques en matière de mots de passe ainsi que de l'absence d'authentification multi-facteurs.

Face à ce constat, le projet présenté ici s'inscrit dans une démarche exploratoire visant à répondre à un besoin sociétal clair : réduire l'impact du phishing et renforcer la résilience des organisations. Il s'agit d'établir un état des lieux complet des mesures existantes – techniques, organisationnelles et juridiques, d'identifier les opportunités d'amélioration et de proposer des pistes de solutions concrètes.

Ce travail adopte une approche résolument collaborative, mobilisant les écoles partenaires, les acteurs institutionnels (notamment la Police cantonale vaudoise (PCV) et la Direction Générale du Numérique et des Systèmes d'Information (DGNSI)), les industriels et les utilisateur·rice·s finaux·ales. L'objectif est de croiser les expertises afin de formuler des recommandations pertinentes et applicables.

La finalité du projet est de diffuser largement les résultats – par exemple sous forme de rapport de synthèse, de communications scientifiques et médiatiques, de guides, de vulgarisation – afin de maximiser l'impact auprès des professionnel·le·s, des institutions et du grand public.

1.1 Le programme d'innovation [seal]

Le **programme d'innovation [seal]** est une initiative stratégique dont l'objectif est de réunir les acteurs afin de favoriser le développement de projets répondant à des problématiques concrètes dans le domaine de la confiance numérique et de la cybersécurité.

Le programme est né en 2023 d'une ambition commune de trois Hautes Écoles – l'École Polytechnique Fédérale de Lausanne (EPFL), la Haute École d'Ingénierie et de Gestion du canton de Vaud (HEIG-VD) et l'Université de Lausanne (UNIL) – et soutenu par le canton de Vaud, via le Service de la promotion de l'économie et de l'innovation (SPEI).

1.2 La thématique de l'appel à projets ciblant la cybersécurité des PMEs

Les **appels à projets** constituent l'un des principaux instruments du programme [seal]. Ils invitent les acteurs intéressés à soumettre des propositions de projets innovants sur des thématiques prédéfinies. Cette approche vise à encourager l'expérimentation, l'exploration de nouvelles idées et la collaboration entre les différents acteurs de l'écosystème de la cybersécurité.

La **thématique de l'appel à projets** dans lequel s'inscrit ce travail est :

> « Donner aux petites organisations les moyens de lutter contre la croissance exponentielle de la cybercriminalité. »

Cette problématique met en lumière la nécessité de développer des approches accessibles, efficaces et adaptées aux ressources limitées des petites structures. Elle englobe aussi bien la prévention (sensibilisation, formation), la protection (solutions techniques, bonnes pratiques) que la réaction face aux incidents (détection, réponse et résilience). L'objectif est de réduire l'exposition de ces organisations aux menaces numériques et de renforcer leur capacité à faire face aux attaques, afin de protéger non seulement leurs activités, mais également l'ensemble de l'écosystème dans lequel elles évoluent.

1.3 Naissance de l'idée

L'idée de ce projet est née à l'occasion du premier workshop [seal], organisé en février 2024 dans le cadre de l'appel à projets. Lors de cet événement, plusieurs intervenant·e·s – notamment Julien Cartier (PCV) et Guillaume Fumeaux (DGNSI) – ont partagé leurs observations de terrain sur les menaces qui touchent les petites organisations.

Du point de vue d'un expert en cybersécurité, un constat s'impose : de nombreux incidents auraient pu être évités si des règles élémentaires d'hygiène numérique avaient été respectées.

En résumé, les discussions ont mis en évidence trois faiblesses majeures à l'origine de la plupart des incidents :

1. les attaques de «phishing» et la proportion d'utilisateur·rice·s prêt·e·s à cliquer sur des liens ;
2. les problèmes liés à la gestion des mots de passe (mots de passe faibles, réutilisation, absence de bonnes pratiques, absence de gestionnaires) ;
3. l'absence fréquente d'activation de l'authentification multi-facteurs, rendant les systèmes vulnérables aux attaques sur le premier facteur.

Les constats soulignent que **les attaques par phishing** représentent un risque majeur. Son impact se fait sentir de manière transversale, affectant aussi bien les entreprises que les institutions publiques et les particulier·ère·s, et exige une réponse coordonnée à l'échelle de l'écosystème numérique. Malgré les efforts de sensibilisation et les mesures de protection existantes, cette menace demeure largement non maîtrisée et constitue un **enjeu majeur pour la sécurité de la société**.

Cette prise de conscience a été un déclencheur. En tant qu'expert·e·s dans le domaine, nous pourrions initialement nous attendre à explorer une piste très technique ou fortement innovante. Il était cependant nécessaire de réaliser que le véritable enjeu réside dans un problème beaucoup plus fondamental : le «phishing».

Après quelques réflexions et recherches, l'idée de ce projet est donc née et il est nécessaire de répondre à la question :

> Que peut-on mettre en oeuvre pour adresser la problématique sociétale majeure du «phishing» ?

Pour parvenir à répondre à cette questions, plusieurs questions liées méritaient d'être abordées :

- Quelles mesures, techniques, organisationnelles, et légales, existent actuellement ?
- Sont-elles utilisées ? Si oui, sont-elles efficaces ?
- Quels sont les modes opératoires des attaquant·e·s ? Quelles sont les vulnérabilités exploitées ?
- Comment améliorer la situation ?
- Quels seraient les moyens que nous pourrions imaginer pour améliorer la situation dans sa globalité ?
- Parmi ces idées, lesquelles seraient efficaces et réalistes ?

Ces réflexions ont conduit à imaginer un projet exploratoire, visant non pas à développer immédiatement une solution technique, mais d'abord à dresser un état des lieux complet de la situation, à identifier les points faibles et à fédérer différents acteurs autour d'une réflexion collective. Chaque acteur de ce projet a été soigneusement sélectionné pour sa pertinence, son apport, et sa complémentarité dans le regard de la thématique. L'objectif est de poser les bases d'initiatives plus ambitieuses, éventuellement innovantes, dans un second temps, en s'appuyant sur les enseignements de ce travail préliminaire.

1.4 Innover autour de la thématique mais également dans la manière de l'approcher

L'innovation de ce projet réside à deux niveaux :

L'innovation thématique. l'objectif est d'imaginer des solutions nouvelles pour réduire l'impact du phishing, qu'il s'agisse de renforcer la sensibilisation, d'améliorer les outils de détection ou de développer de nouvelles pratiques organisationnelles.

L'innovation méthodologique. l'approche adoptée consiste à rassembler des acteurs issus de domaines complémentaires (technique, organisationnel, juridique) autour d'ateliers collaboratifs, afin de co-construire des recommandations transversales et pragmatiques. Cette démarche favorise la création de solutions adaptées aux réalités du terrain et à la diversité des acteurs concernés.

1.5 «Phishing» : problématique et définition(s)

Le phishing est aujourd'hui l'une des menaces les plus répandues et redoutables en cybersécurité. Exploitant principalement la faille humaine, il sert de porte d'entrée à des attaques plus complexes, comme les ransomwares, la fraude ou l'usurpation d'identité. Les petites et moyennes organisations, souvent moins bien équipées, sont particulièrement vulnérables. Malgré l'existence de nombreuses mesures pour combattre le phishing, qu'elles soient techniques, organisationnelles ou légales, leur déploiement reste inégal et leur efficacité limitée face à des attaques toujours plus sophistiquées.

L'origine du terme *phishing* est issu de la contraction des mots anglais *password*, *harvesting* et *fishing*, signifiant littéralement « aller à la pêche aux mots de passe ». Il désigne donc, à l'origine, la pratique consistant à tromper un·e utilisateur·rice afin de lui soutirer ses informations d'authentification.

De manière plus générale, le terme phishing est souvent utilisé de la manière suivante :

> Une **action malveillante** visant à obtenir des **informations personnelles** (mots de passe, données bancaires, identifiants) en se faisant passer pour (**en trompant**) une entité de **confiance**.

Dans le cadre de ce projet, nous adoptons une définition élargie du phishing, que nous appelons le « phishing au sens large ». Il ne se limite pas au vol de mots de passe, mais englobe l'ensemble des pratiques cybercriminelles qui utilisent la tromperie sociale et des moyens de communication électroniques pour inciter un·e utilisateur·rice à réaliser une action compromettante.

Ainsi, le phishing au sens large inclut :

- **Le vol d'informations personnelles** : identifiants, mots de passe, numéros de carte de crédit.
- **L'incitation à exécuter une action malveillante** : par exemple réaliser une transaction bancaire frauduleuse.
- **L'ouverture de pièces jointes malveillantes** : pouvant conduire à l'installation de malwares (ex. ransomware).
- **Le clic sur des liens frauduleux** : redirigeant vers des sites de collecte de données ou exploitant des failles pour compromettre le système.

Le phishing, dans son acception large, se caractérise par :

- une **tromperie sociale** : l'attaquant se fait passer pour une entité légitime (banque, fournisseur de service, autorité).
- une **communication électronique** : typiquement par e-mail, mais également par les réseaux sociaux, SMS (*smishing*), le téléphone ou la voix (*vishing*), ou via des QR codes (*qishing*).
- un **objectif criminel** : obtenir des informations sensibles, inciter à un paiement ou installer un logiciel malveillant.

Cette approche permet de traiter la problématique du phishing dans toute sa complexité, en

considérant non seulement l'aspect technique, mais également les dimensions organisationnelles et humaines.

1.6 Objectifs du projet

La mission principale de ce projet est la suivante :

Traiter la problématique du phishing, qui est une préoccupation sociétale majeure, afin de diminuer l'impact de ce fléau, en y apportant des regards complémentaires.

Pour remplir cette mission, les objectifs en découlant sont les suivants :

1. Établir un état de l'art des mesures existantes
2. Établir un état de la situation (retours de terrain) et identifier les points faibles
3. Imaginer et élaborer des pistes de solutions innovantes permettant de diminuer l'impact en exploitant la complémentarité des partenaires
4. Développer un ou plusieurs proof of concepts, ainsi que les tester
5. Aider la société à améliorer sa résilience en publiant des guides et bonnes pratiques
6. Informer et disséminer les résultats à destination de différents publics

Le projet souhaite une approche interdisciplinaire, collaborative et orientée vers l'action.

1.7 Méthodologie du projet

La méthodologie adoptée dans ce projet se distingue de celle des démarches d'innovation classiques. L'objectif a été de tirer parti du grand nombre de partenaires et de la diversité de leurs expertises afin de couvrir le plus largement possible les idées et les solutions envisageables.

La méthodologie adoptée repose donc sur une approche progressive et collaborative, structurée en plusieurs phases complémentaires.

1.7.1 Phase 1 - État de l'art et retours de terrain

Afin de permettre de dresser un diagnostic précis des forces et faiblesses actuelles, une analyse approfondie des mesures existantes contre le phishing (techniques, organisationnelles et légales) devait être réalisée

- Recensement et classification des mesures existantes (techniques, organisationnelles, légales et réglementaires).

- Évaluation de leur efficacité et de leur adoption par les petites organisations.
- Étude des tendances et des meilleures pratiques.

Afin de comprendre les modes opératoires et identifier les faiblesses exploitées, il était également important de collecter des retours de terrain.

- Entretiens avec des PME, institutions publiques et expert·e·s du domaine.
- Analyse d'incidents réels afin d'identifier les vulnérabilités les plus exploitées et les modes opératoires courants.

1.7.2 Phase 2 - Ateliers collaboratifs

Quatre ateliers d'échange ("workshops") ont réuni les partenaires du projet pour croiser leurs expertises et co-construire des pistes d'innovation :

Workshop #1 - État de la situation actuelle. Les partenaires étaient invités à présenter, discuter et analyser les mesures existantes, confronter les retours de terrain, identifier les failles et défis majeurs.

Workshop #2 - Génération d'idées innovantes. L'idée était de produire un maximum de propositions sans contrainte, en explorant aussi bien les approches techniques qu'organisationnelles et légales.

Workshop #3 - Approfondissement et pré-sélection. L'objectif de cet atelier était d'évaluer la pertinence, la faisabilité (technique, accès aux données nécessaires, budget, maintenance future) et le potentiel d'impact des idées, et retenir uniquement celles à fort intérêt.

A la suite de ce workshop, il a été nécessaire d'investiguer la faisabilité technique de certaines propositions.

Workshop #4 - Retours sur la faisabilité, décision et planification. Examiner la faisabilité des idées retenues et définir les *Proof of Concepts* (PoCs) à développer.

Un dernier atelier ne faisant pas partie de la méthodologie initiale a été rajouté suite à un contact avec l'Office fédéral de la cybersécurité (OFCS). L'objectif de cet atelier était de discuter avec l'OFCS de nos objectifs communs afin de travailler ensemble sur des contenus de sensibilisation et/ou des PoCs.

1.7.3 Phase 3 - Développement et tests des solutions

Les idées retenus des ateliers précédents ont été transformées en PoCs et/ou démonstrateurs.

Ces expérimentations ont permis d'évaluer la faisabilité technique, l'efficacité et l'ergonomie des solutions, ainsi que leur adaptabilité aux différents contextes d'utilisation.

- Conception de prototypes illustrant les solutions les plus prometteuses (ex. : outils de vérification de configuration de messagerie, extensions de navigateur pour détecter les liens suspects).

1.7.4 Phase 4 - Sensibilisation et dissémination

En parallèle, des guides de bonnes pratiques, supports de formation et outils techniques sont préparés pour être diffusés largement auprès des entreprises, collectivités et citoyen·ne·s.

- Publication de guides de bonnes pratiques accessibles à différents publics (employé·e·s de PME, administrateur·e·s, collectivités, grand public).
- Organisation d'ateliers de formation et de conférences pour partager les résultats.
- Diffusion médiatique et vulgarisation scientifique afin de maximiser l'impact sociétal.

1.8 Livrables imaginés

Les livrables attendus visent à produire des résultats concrets et directement exploitables par les acteurs concernés :

- **Rapport de synthèse** : état de l'art, diagnostic des failles, recommandations.
- **Guides pratiques** : fiches techniques, procédures de réponse à incident, bonnes pratiques de sensibilisation, adaptées aux petites organisations.
- **Outils techniques et recommandations** : prototypes ou scripts open source facilitant la mise en place de solutions (vérification de configuration, détection d'e-mails suspects).
- **Publications et communications** : articles scientifiques, communications dans la presse spécialisée et grand public.

1.9 Impacts potentiels du projet

Ce projet vise à générer des impacts multiples, à court, moyen et long terme :

- **Impact sociétal** : contribuer à la confiance numérique de la population et à la protection de l'économie régionale.
- **Renforcement de la résilience des PME et institutions** : réduction de leur exposition au phishing et amélioration de leur capacité de réponse aux incidents.
- **Économie locale** : diminution des pertes liées aux cyberattaques et amélioration de la compétitivité des entreprises locales.
- **Diffusion des connaissances** : mise à disposition de contenus pédagogiques et d'outils permettant aux organisations d'élever leur niveau de maturité en cybersécurité.

- **Dynamisation de l'écosystème suisse de cybersécurité** : favoriser les collaborations entre monde académique, secteur privé et pouvoirs publics.
- **Échanges entre les partenaires, collaborations et connaissance mutuelle des acteurs** : renforcement des relations entre les différents acteurs impliqués, développement d'une culture commune et partage d'expériences utiles pour de futurs projets.
- **Valorisation du programme [seal]** : mise en avant de son rôle dans le soutien à l'innovation et dans le développement de solutions concrètes aux enjeux de cybersécurité.

1.10 Partenaires impliqués et contributions

Pour mener à bien ce projet, il était essentiel de réunir un certains nombres de partenaires. Les différents partenaires seront amenés à collaborer ensemble autour de cette thématique, à confronter leur vision de la situation en apportant des regards complémentaires.

Autour du programme d'innovation [seal], il semblait essentiel que les trois partenaires académiques puissent collaborer sur cette thématique sociétale :

l'EPFL a apporté son expertise reconnue dans les différents domaines de la confiance numérique au travers de son Centre pour la confiance numérique (C4DT) ainsi que de ses différentes chaires de recherche ;

la HEIG-VD a quant à elle mis en œuvre son expertise en sécurité informatique appliquée, participant à la compréhension des attaques, des mécanismes techniques, à la rédaction des guides techniques et au développement des prototypes (PoCs) ;

l'Université de Lausanne (UNIL) a apporté son regard sur la partie légale et réglementaire, notamment grâce à la Faculté de droit, ainsi que son analyse sur la criminalité, les retours de terrain et la vision forensique, notamment avec l'Ecole des Sciences Criminelles (ESC).

Les partenaires académiques, EPFL, HEIG-VD, et Unil, avaient donc pour responsabilité principale de fournir les fondements scientifiques et techniques du projet. Ils ont apporté des compétences pointues dans divers domaines liés à la confiance numérique, tout en mettant à disposition un large réseau de partenaires issus de la recherche et de l'industrie.

Nous avions également besoin de précieux retours de terrains et nous avons inclus les partenaires institutionnels suivants :

la Police Cantonale Vaudoise (PCV), via la Brigade Analyse Traces Technologiques (BATT) et la cellule de Renseignement Criminel Cyber (RCC), permettant d'obtenir des retours concrets, fiables, et détaillés de cas réels, tout en donnant des indications sur les modes opératoires habituels, les types de cibles communes, et les proportions entre individu·e·s et sociétés ;

DGNSI

la Direction Générale du Numérique et des Systèmes d'Information (DGNSI) pour son expertise et sa connaissance du terrain sur l'Etat de Vaud lui-même mais aussi pour avoir suivi des cas concrets d'incidents, notamment auprès de PMEs et communes vaudoises.

Les partenaires institutionnels ont donc permis d'avoir des informations sur la réalité du terrains concernant les attaques de phishing au sein des petites entreprises, des communes, de l'État de Vaud, ainsi que des individu·e·s. Ils ont partagé leur connaissance des modes opératoires des cybercriminels, ainsi que fourni des informations sur la réalité du terrains concernant les attaques sur les particulier·ère·s.

Nous avions également un fort besoin d'acteurs représentant l'industrie et une bonne connaissance des pratiques d'accompagnement et de sensibilisation des entreprises.

Le partenaire industriel Navixia a mis à profit son expérience directe dans la lutte contre le phishing, notamment grâce à son expertise et à la notoriété de leur outil DiagnoPhish. Son rôle a été double : d'une part, apporter une expertise pointue sur les incidents liés à cette menace, et d'autre part, assurer une passerelle vers l'intégration concrète des résultats du projet dans des solutions existantes, favorisant ainsi leur déploiement.

Finalement, il nous fallait un représentant d'une entreprise "cliente" et ciblée par des attaques de ce type, pouvant donner des retours de terrains, mais aussi des solutions existantes ou développées dans le cadre de ce projet.

les Transports Lausannois (TL) ont participé en tant qu'utilisateur pilote. Leur expérience de terrain a permis de confronter les idées et les outils développés à un environnement opérationnel réel.

Ce travail collectif et interdisciplinaire a permis d'enrichir le projet à chaque étape, de la conception à l'évaluation des solutions proposées, tout en assurant leur pertinence pour des cas d'usage concrets.

1.11 Planification du projet

Dates clés administratives :

- 27.03.2024 dépôt booster
- 31.05.2024 dépôt dossier projet
- 26.06.2024 décision seal, projet accepté
- 25.09.2024 contrats signés
- 01.12.2024 démarrage projet
- 31.08.2025 fin officielle
- 31.10.2025 fin réelle
- Par la suite, valorisation

Le tableau ci-dessous donne une vue synthétique de la vision administrative du projet :

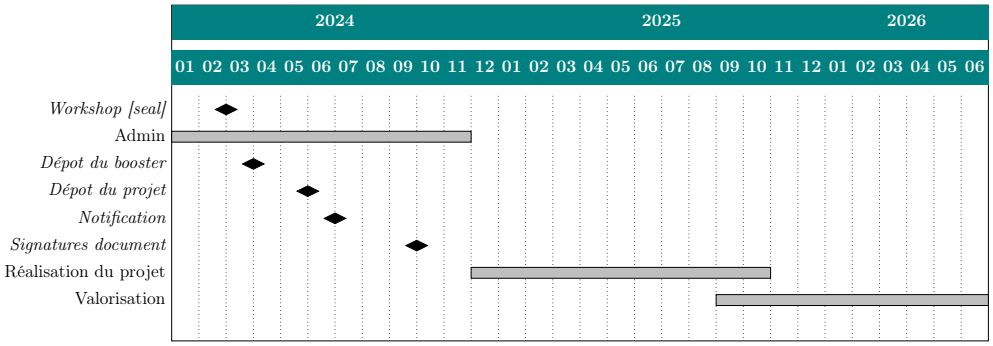

FIGURE 1.1 – Planification administrative : dépôt, réalisation, valorisation

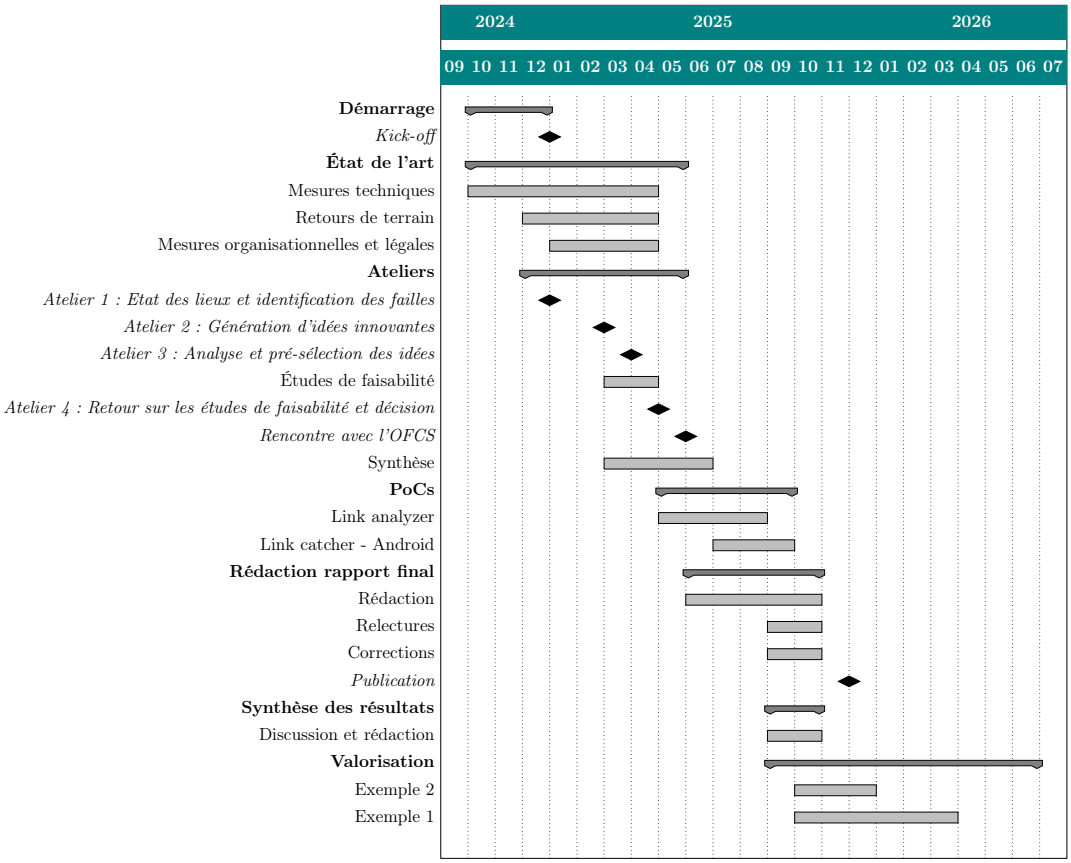

FIGURE 1.2 – Planification du projet

Chapitre 2

Etat de l'art

Table des matières du chapitre

Dans le cadre de ce projet visant à traiter la problématique croissante du phishing, un des premiers objectifs majeurs est d'établir un état de l'art des mesures existantes. Le phishing, une cyberattaque qui consiste à tromper les individu·e·s pour qu'iels divulguent des informations personnelles ou sensibles, représente une menace sérieuse et en constante évolution pour les utilisateur·rice·s et les organisations.

En effet, pour lutter efficacement contre ce fléau, il est d'abord essentiel de comprendre les stratégies et les solutions actuellement mises en place. Cette analyse approfondie permet de lister de manière exhaustive ces stratégies et solutions et de les comprendre. Ainsi, nous pouvons identifier les lacunes à combler, qu'elles soient techniques, organisationnelles ou légales, afin d'améliorer nos défenses face à ce problème sociétal.

En explorant les approches et les technologies déjà mises en œuvre, nous visons à élaborer des recommandations fondées sur des pratiques éprouvées tout en proposant des améliorations ciblées pour une protection plus robuste et adaptable face à cette menace persistante.

Dans le cadre de la lutte contre une menace informatique, les mesures de prévention sont souvent réparties dans les trois catégories suivantes :

- Les mesures techniques englobent l'utilisation de technologies et d'outils pour prévenir, détecter et répondre aux cyberattaques. Cela inclut des dispositifs et logiciels conçus pour protéger les systèmes informatiques et renforcer la sécurité en réduisant les vulnérabilités et en bloquant les attaques potentielles.
- Les mesures organisationnelles se concentrent sur la gestion interne et les pratiques de sécurité au sein d'une organisation. Chaque membre de l'organisation doit comprendre et respecter les pratiques de sécurité pour minimiser les risques internes. Cela inclut la formation et la sensibilisation des employé·e·s, l'établissement de politiques et de procédures, ou encore la gestion des accès et des autorisations.
- Les mesures légales concernent l'adhésion aux lois et réglementations relatives à la cybersécurité et à la protection des données. Cela implique la conformité aux normes légales, la mise en place de politiques de confidentialité, et la collaboration avec les autorités en cas de cyber incident. Le cadre juridique doit viser à dissuader les cybercriminel·le·s par des sanctions appropriées et à fournir des lignes directrices claires pour la protection des informations sensibles et des infrastructures critiques.

2.1 Concepts fondamentaux

Cette section vise à expliquer les quelques concepts clés qui sont nécessaires à la compréhension des mesures techniques présentées dans la présente section. Certaines compétences de base sont nécessaires afin de comprendre le contenu de ce chapitre.

2.1.1 DNS (Domain Name System)

Le DNS est un système essentiel qui joue le rôle d'annuaire pour Internet. Son objectif principal est de traduire des noms de domaine lisibles par les humains, comme `www.google.com`, en adresses IP, telles que `216.58.215.228`, nécessaires pour localiser et communiquer avec les serveurs sur le réseau. Sans DNS, les utilisateur·rice·s seraient contraints de mémoriser les adresses IP exactes des sites web qu'ils souhaitent visiter, rendant l'utilisation d'Internet beaucoup plus complexe. Le port 53 est le port standard utilisé par DNS pour échanger des requêtes et réponses.

2.1.1.1 Zones DNS

Le DNS est un système hiérarchique et distribué, où l'espace des noms de domaine est divisé en zones distinctes. Une zone DNS représente une portion de l'arborescence pour laquelle la responsabilité administrative a été déléguée à un serveur ou un groupe de serveurs. Par exemple, la zone `.edu` délègue la gestion de la zone `yale.edu` à un serveur spécifique, qui à son tour délègue une sous-zone comme `cs.yale.edu` à un autre serveur.

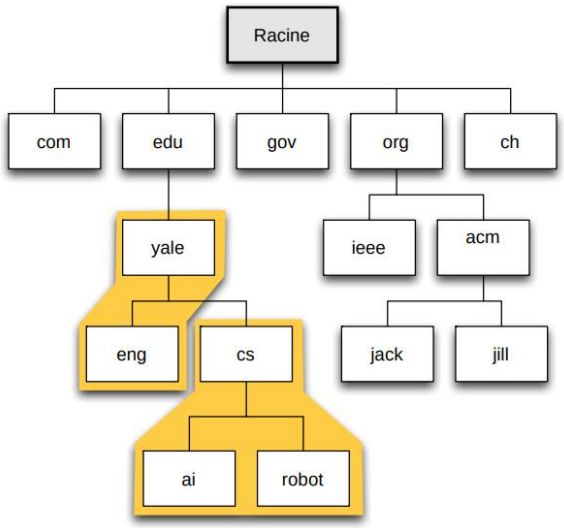

FIGURE 2.1 – Organisation des zones DNS

2.1.1.2 Fonctionnement d'un DNS

Prenons l'exemple d'un·e utilisateur·rice qui saisit `www.heig-vd.ch` dans son navigateur. Voici les étapes principales du processus de résolution DNS :

- **Requête initiale :** Le navigateur envoie une requête DNS au résolveur DNS (souvent celui de son fournisseur d'accès Internet ou interne à un réseau d'entreprise).
- **Interrogation des serveurs racine :** Si le résolveur DNS ne connaît pas l'adresse IP, il envoie la requête à l'un des serveurs racine. Ces serveurs ne retournent pas directement l'adresse IP du domaine demandé ; ils indiquent simplement l'adresse du serveur suivant à interroger, responsable d'une partie plus précise du nom de domaine (serveur DNS autoritaire de la zone `.ch` (TLD, Top-Level Domain)).
- **Interrogation du serveur .ch :** Le serveur des domaines `.ch` est interrogé. Il ne connaît pas l'adresse IP exacte de `www.heig-vd.ch`, mais il fournit l'adresse du résolveur DNS responsable de la zone `heig-vd.ch`.
- **Interrogation du serveur heig-vd.ch :** Le serveur autoritaire de la zone `heig-vd.ch` reçoit la requête et retourne l'adresse IP correspondante à `www.heig-vd.ch`.
- **Réponse au client :** L'adresse IP est transmise au navigateur, qui peut alors établir une connexion directe avec le serveur web de la HEIG-VD et charger la page demandée.
- **Mise en cache :** Pour accélérer les futures requêtes, l'adresse IP de `www.heig-vd.ch` est conservée temporairement dans un cache local (sur le résolveur DNS local ou l'appareil client), pendant une durée spécifiée par le `TTL (Time To Live)` de l'enregistrement DNS.

2.1.1.3 Les enregistrements DNS

Les enregistrements DNS sont des entrées spécifiques dans les fichiers de zone qui permettent de gérer les fonctionnalités et la résolution des noms de domaine. Voici les principaux types d'enregistrements :

- **A (Address) :** Associe un nom de domaine à une adresse IPv4 (ex. : `heig-vd.ch` → `193.134.223.20`).
- **AAAA (Address IPv6) :** Équivalent de l'enregistrement A, mais pour les adresses IPv6.
- **MX (Mail Exchange) :** Définit les serveurs responsables de la réception des e-mails pour un domaine (ex. : mail.heig-vd.ch).
- **CNAME (Canonical Name) :** Utilisé pour aliaser un domaine vers un autre (ex. : `home.heig-vd.ch` → `heig-vd.ch`).
- **TXT (Text) :** Contient des informations textuelles, souvent utilisées pour publier des politiques de sécurité (comme SPF, DKIM ou DMARC) ou stocker d'autres données utiles à la configuration.
- **PTR (Pointer) :** Permet une résolution inversée, associant une adresse `IP` à un nom de domaine.

2.1.2 SMTP (Simple Mail Transfer Protocol)

Le SMTP est un protocole standard utilisé pour l'envoi d'e-mails. Son rôle principal est de permettre la communication entre le client mail (ex. : Outlook, Thunderbird) et le serveur de messagerie, ainsi qu'entre différents serveurs pour acheminer les e-mails jusqu'à leur destinataire final. Le SMTP fonctionne principalement sur le port 25 (ou 587 pour les connexions sécurisées avec STARTTLS).

2.1.2.1 Chaîne de réception/envoi de mail

On retrouve de multiples entités et machines dans la chaîne de transmission d'un mail d'un·e utilisateur·rice à un autre :

- L'expéditeur
- Le serveur e-mail d'envoi
- Le réseau Internet (composé de serveurs relais)
- Le serveur e-mail de réception
- Le destinataire
- Serveur web (si un lien est présent dans le mail)

Cette chaîne fait appel à trois protocoles de communication pour distribuer un message à son destinataire. SMTP pour l'envoi, ainsi qu'IMAP (Internet Message Access Protocol) ou POP (Post Office Protocol) pour la récupération des e-mails.

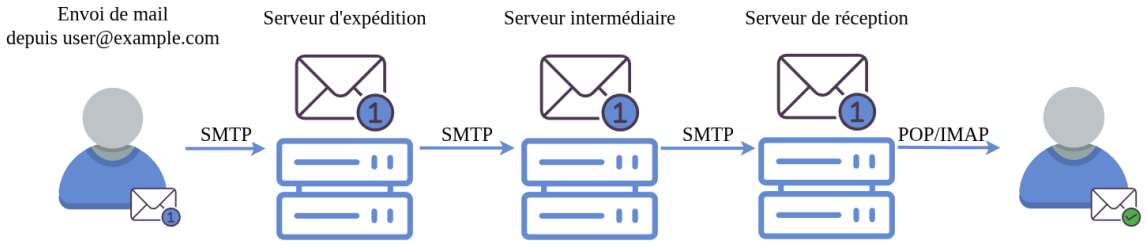

FIGURE 2.2 – Chaîne de réception et d'envoi d'un e-mail

2.1.2.2 Composition d'un e-mail

Un e-mail est constitué de deux parties principales : les en-têtes et le corps du message. Les en-têtes fournissent des informations techniques essentielles pour l'acheminement, l'identification et l'organisation de l'e-mail. Le corps contient le contenu principal, destiné au destinataire.

2.1.2.3 En-têtes principales

Le protocole SMTP permet l'utilisation de nombreux en-têtes pour structurer et transmettre un e-mail, mais nous nous concentrerons ici sur les en-têtes fondamentaux et particulièrement pertinents pour le contexte de ce rapport.

Dans le cadre de la lutte contre le phishing, l'analyse de ces en-têtes est essentielle : ils permettent de retracer le parcours d'un message, de vérifier l'authenticité de son expéditeur et de détecter des manipulations ou falsifications.

- **Return-Path :** Il s'agit de l'adresse utilisée pour renvoyer les notifications de non-livraison (bounces).
- **Delivered-To :** Ce champ indique la boîte de réception finale où l'e-mail a été livré.
- **Received :** Chaque serveur SMTP traversé par le message ajoute un enregistrement Received. Ces lignes permettent de tracer le chemin suivi par l'e-mail, avec des informations comme l'adresse IP de l'expéditeur, le nom du serveur intermédiaire, et la date et l'heure de passage. L'en-tête est donc composée de plusieurs champs `Received` qu'il faut lire de bas en haut pour retracer le chemin parcouru.
- **Subject :** Définit l'objet du message.
- **From :** Spécifie l'adresse de l'expéditeur visible pour le destinataire.
- **To :** Spécifie l'adresse du destinataire visible pour celui-ci.
- **Authentication-Results :** Cet en-tête est ajouté par les serveurs de messagerie pour indiquer les résultats des vérifications d'authentification, comme SPF, DKIM et DMARC. Il fournit des informations sur la validité du message en identifiant si les contrôles de sécurité ont été passés ou échoués.

2.2 Mesures techniques

Les moyens actuels contribuent à renforcer la sécurité de la chaîne. Chaque protection agit à son niveau et ne s'applique pas à tous les maillons de la chaîne. Il est donc nécessaire de les combiner afin d'obtenir une solution capable de limiter l'impact du phishing. Dans la pratique, nous constatons que l'ensemble n'est que partiellement efficace. De plus, le déploiement complet de ces moyens n'est de loin pas appliqué partout, malheureusement. L'architecture de cette chaîne, héritée des années 80 et conçue dans un esprit ouvert, complique cette tâche car elle ne permet pas de mettre en place une communication sécurisée de bout en bout aussi facilement que le font les applications de messagerie instantanée telles que Signal, WhatsApp ou Telegram, beaucoup plus fermées et propriétaires.

On peut noter différentes tentatives permettant de limiter les attaques de phishing avec des mesures telles que des correctifs du protocole SMTP, la mise en place du Sender Policy Framework (SPF), ou encore l'utilisation de filtres anti-spam. Aucune mesure n'est parfaite et les résultats du terrain le démontrent clairement.

L'authenticité et l'intégrité, deux points essentiels de la sécurité informatique, permettent de réduire considérablement les risques de phishing. Ils consistent à vérifier que l'expéditeur de l'e-mail est bien celui qu'il prétend être et que son message n'a pas été altéré lors de son transit.

2.2.1 Chiffrement et authentification

2.2.1.1 Passkeys

Les clés d'accès (*passkeys*) sont une alternative émergente aux mots de passe traditionnels. Même si l'authentification à deux facteurs (2FA) ou à facteurs multiples (MFA) est ajoutée à ces derniers, ils restent vulnérables à l'hameçonnage. De plus, la gestion des mots de passe et de l'authentification à deux facteurs/facteurs multiples est souvent pénible pour les utilisateur·rice·s et n'est souvent pas faite d'une façon sécurisée.

La *FIDO Alliance*, qui regroupe des acteurs industriels comme 1Password, Amazon, Google et Microsoft, et des acteurs publiques comme le *Bundesamt für Sicherheit in der Informationstechnik* (BSI) allemand et le *National Institute of Standards and Technology* (NIST), est à l'origine des clés d'accès. Complétant la spécification de la *Web Authentication* (WebAuthn) du *World Wide Web Consortium* (W3C), l'idée est de remplacer les mots de passe par la cryptographie asymétrique [fA25].

Dans la cryptographie asymétrique, au lieu d'une seule clé qui doit impérativement être gardée secrète, il existe une paire de clés mathématiquement liées. Comme son nom l'indique, la clé privée est confidentielle, mais la clé publique peut être librement distribuée. Bien que chaque paire est étroitement liée - une clé publique n'appartient qu'à une clé privée et vice versa - il n'est pas possible de déduire la clé privée à partir de la clé publique.

On peut utiliser ces caractéristiques des clés d'accès pour implémenter un schéma d'authentification : quand l'utilisateur·rice crée un compte chez un service en ligne, son appareil génère une nouvelle paire de clés qui sera associé à ce service, et à ce service seul. La clé publique est transférée au service en ligne, tandis que la clé privée reste sur l'appareil. Il est important de noter que cette clé ne quitte jamais l'appareil.

La prochaine fois que l'utilisateur·rice tente de se connecter au service, le service utilise la clé publique pour créer un défi mathématique qui a pour but de prouver que l'utilisateur·rice est en possession de la clé privée associée. Puis l'appareil renvoie la réponse au service, qui à son tour donne accès à l'utilisateur·rice.

Les aspects importants à noter sont :

- c'est l'appareil de l'utilisateur·rice qui lui demande de s'authentifier pour utiliser la clé privée
- l'utilisateur·rice ne s'authentifie jamais directement auprès du service, mais seulement

auprès de son appareil
- l'appareil détermine quelle clé est à utiliser pour s'authentifier auprès du service
- chaque paire de clés est associé à un seul service

Les seuls identifiants dont l'utilisateur·rice doit s'en souvenir, ceux utilisés pour s'authentifier auprès de l'appareil, restent sur l'appareil sous son contrôle. De plus, l'utilisateur·rice ne peut pas se connecter à un site frauduleux avec une paire de clés liées à un service légitime, comme c'est l'appareil qui choisit la paire à utiliser. Finalement, le fait que chaque paire n'est liée qu'à un seul service met fin à la réutilisation de mots de passe [Mic24].

2.2.1.2 SMTP-Auth

Extension du protocole SMTP demandant à l'expéditeur de se connecter au serveur d'envoi en utilisant un mécanisme d'authentification qui est souvent un nom d'utilisateur et un mot de passe. Ainsi, cela signifie que seuls les utilisateurs autorisés peuvent envoyer des e-mails via celui-ci. Ce moyen est indispensable pour éviter que ce dernier devienne un « relais ouvert » ce qui signifierait que le serveur est accessible sans authentification, utilisé abusivement pour envoyer du spam. Cependant, cela ne protège pas contre l'interception et la modification des e-mails en transit entre les serveurs. C'est pourquoi le point suivant est essentiel pour résoudre ce problème, en apportant de la confidentialité.

2.2.1.3 SMTP over TLS

Mesure utilisant le protocole Transport Layer Security (TLS) pour chiffrer et authentifier les e-mails pendant leur transport. Il s'applique entre l'expéditeur et son serveur d'envoi, mais aussi entre les serveurs relais qui acheminent le message sur Internet. Des variantes comme IMAP over TLS et POP over TLS sécurisent également la transmission entre le serveur de réception et le destinataire.

Ce chiffrement empêche un attaquant placé entre deux maillons de la chaîne de lire ou modifier le contenu du message, ou de se faire passer pour l'un des serveurs.

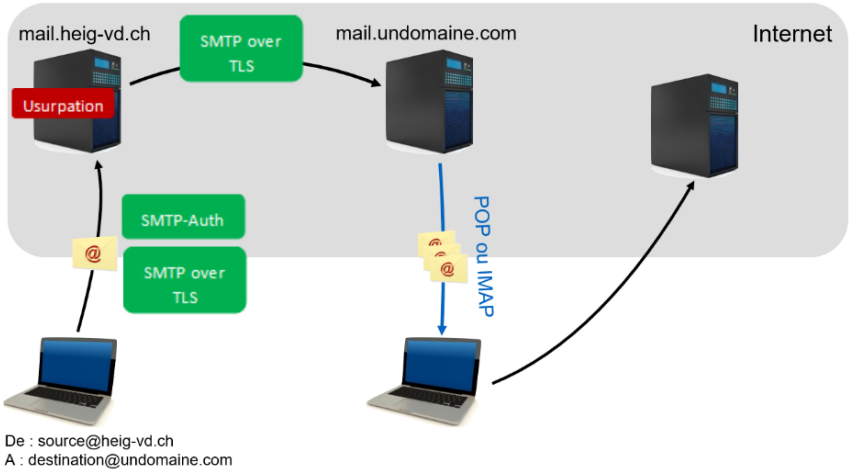

FIGURE 2.3 – Application de SMTP over TLS et SMTP-Auth

Cependant, cette protection ne s'applique que pendant le transit. À chaque étape, le serveur qui reçoit l'e-mail le déchiffre avant de le chiffrer à nouveau pour l'envoyer au maillon suivant. Si un serveur intermédiaire est compromis, il aura donc accès aux messages en clair. Pour garantir la confidentialité de bout en bout, il faut recourir à un chiffrement de type End-to-End Encryption (E2EE).

2.2.1.4 End-to-End Encryption (E2EE)

L'*End-to-End Encryption* (E2EE) est la méthode la plus sûre pour protéger un e-mail : il est chiffré depuis l'expéditeur jusqu'au destinataire, de sorte qu'aucun intermédiaire — serveur, fournisseur de messagerie ou attaquant — ne puisse le lire ou le modifier.

En plus de garantir la confidentialité, l'E2EE assure aussi l'authenticité (l'expéditeur est bien celui qu'il prétend être) et l'intégrité (le contenu n'a pas été altéré).

Deux protocoles majeurs permettent ce chiffrement de bout en bout :

- **S/MIME** (*Secure/Multipurpose Internet Mail Extensions*) : repose sur une infrastructure à clés publiques centralisée (PKI) gérée par des autorités de certification.
- **PGP** (*Pretty Good Privacy*) : s'appuie sur un réseau de confiance distribué où les utilisateurs valident mutuellement leurs clés.

La mise en place de l'E2EE reste cependant complexe pour le grand public, car elle impose la gestion des clés cryptographiques : création, stockage sécurisé, partage avec les correspondants et renouvellement régulier.

2.2.2 Mesures utilisant les enregistrements DNS

2.2.2.1 SPF (Sender Policy Framework)

Le SPF [Kit14, Pro25g] est une méthode d'authentification des e-mails visant à prévenir le spoofing, une technique où un attaquant usurpe l'identité d'un domaine pour envoyer des e-mails frauduleux. Le SPF fonctionne en vérifiant si l'adresse IP de l'expéditeur de l'e-mail correspond à une liste d'adresses autorisées spécifiée dans l'enregistrement DNS du domaine de l'expéditeur. Cet enregistrement SPF contient des règles définissant les serveurs ou adresses IP autorisés à envoyer des e-mails pour ce domaine.

Lorsqu'un serveur de réception reçoit un e-mail, il effectue une requête DNS pour récupérer l'enregistrement SPF (stocké dans un enregistrement TXT) associé au domaine de l'expéditeur. Par exemple, un domaine pourrait publier l'enregistrement suivant :

```
v=spf1 ip4:190.200.210.1 include:mail.example.com -all
```

- `ip4:190.200.210.1` : Une adresse IP qui est autorisée à envoyer des e-mails pour ce domaine.
- `include:mail.example.com` : Les serveurs listés dans l'enregistrement SPF de `mail.example.com` sont aussi autorisés.
- `-all` : Tous les autres serveurs non spécifiés doivent être rejetés.

À noter qu'il est possible de configurer SPF de manière beaucoup plus fine en utilisant d'autres modificateurs ou mécanismes. Par exemple, le modificateur `mx` autorise automatiquement les serveurs configurés comme enregistrements MX (Mail Exchanger) du domaine. De plus, le mécanisme `all` offre plusieurs options : `-all` pour rejeter strictement les serveurs non autorisés, `all` pour les marquer comme suspects sans les rejeter, ou encore `?all` pour adopter une politique neutre.

Cependant, SPF a des limites. Il ne protège pas contre le spoofing de l'adresse `From` visible dans les clients de messagerie, car il vérifie uniquement le domaine utilisé au niveau SMTP, spécifiquement dans l'en-tête `Return-Path` (adresse où les erreurs de livraison doivent être renvoyées). Cet en-tête est souvent invisible pour l'utilisateur final, ce qui permet à un attaquant de falsifier l'adresse `From` affichée dans le client de messagerie, rendant l'usurpation moins évidente. De plus, SPF échoue dans les cas de transfert d'e-mails, car le serveur intermédiaire n'est pas toujours inclus dans l'enregistrement SPF du domaine d'origine. Ces limitations font que le SPF, bien qu'utile, ne constitue pas une solution complète contre le phishing. Pour maximiser son efficacité, SPF doit être utilisé en complément de mesures telles que DKIM et DMARC.

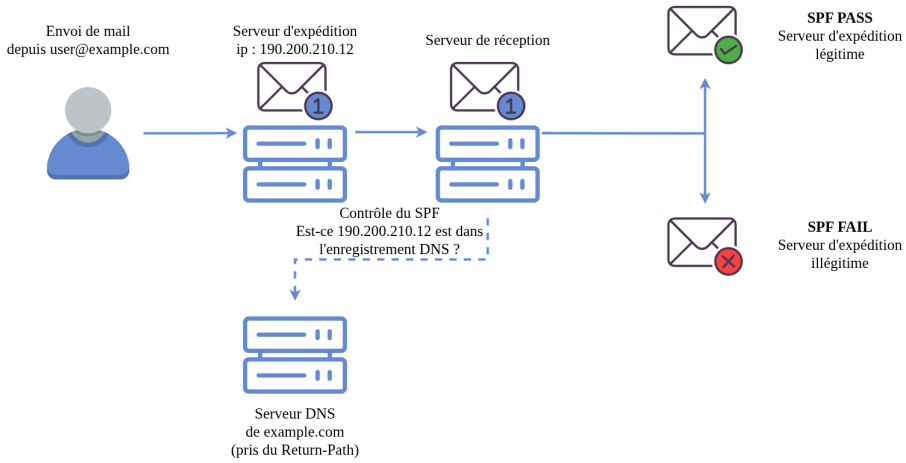

FIGURE 2.4 – Fonctionnement de SPF

2.2.2.2 DKIM (Domain Keys Identified Mail)

DKIM [KCH11, Pro25e] permet au serveur mail de réception de vérifier cryptographiquement que les informations dans l'en-tête de l' e-mail telles que l'adresse de l'expéditeur, ainsi que le contenu du message n'ont pas été modifiées. L'ensemble de ces informations est transformé en une empreinte numérique (hachage), que le serveur d'expédition signe ensuite avec sa clé privée. La signature obtenue est ajoutée dans l'en-tête du message, dans un champ DKIM dédié. Le serveur de réception utilise alors la clé publique du domaine de l'expéditeur (publiée dans l'enregistrement DNS) pour vérifier que la signature est valide et que le message n'a pas été altéré.

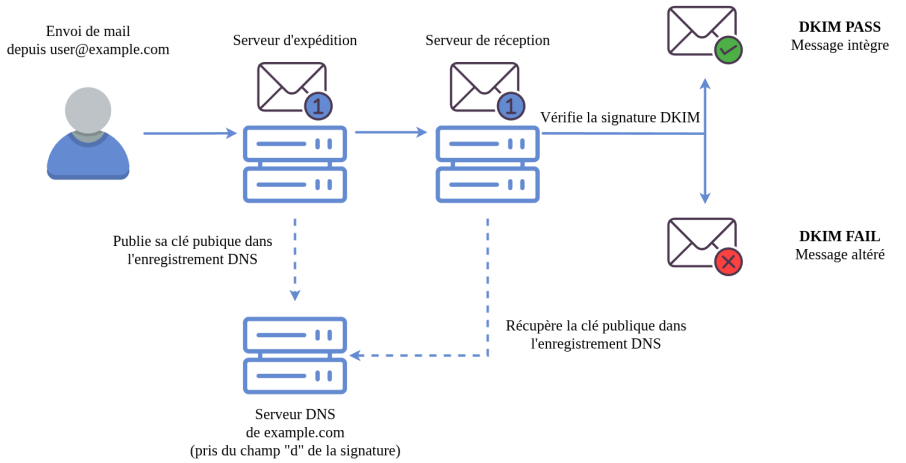

FIGURE 2.5 – Fonctionnement de DKIM

Une signature DKIM possède la forme suivante lorsque l'on inspecte l'en-tête d'un e-mail reçu.

```
DKIM-Signature: v=1; a=rsa-sha256; c=relaxed/relaxed; d=protonuser.me;
    s=protonmail2; t=1681815224; x=1682074424;
    bh=HopzPOvdVOuA7Ue9G14AKilZ4jQVak/25iqXCEH/n34=;
    h=Date:To:From:Subject:Message-ID:Feedback-ID:From:To:Cc:Date:
     Subject:Reply-To:Feedback-ID:Message-ID:BIMI-Selector;
    b=CEhO1fpnW/bhDNaKr5/RCewtAeCnZavAuxuVPd82FWLR/oSXhLF3b60kLyXXs+t8v
     Vy6VR1MAox1m235D+G3I9+uRonzz3z5SvHJcbPQFk68pn1ojd7MuV9f1Gf+ZCcv+v9
     7zjTT4Xe1ujtFhR6Wqrz0BCQpQ1cH8oUZQ5+ahEHb6IKmRjm6R5yWS8MCR/wn/XY7q
     /St9q8LjabgKdbuFgHj6XFOoB4teXzJVKLeJ3QjFflOnE4zItZTfor3tu7ss/bPT4J
     MRIsw7a+CSxQYh1OSMap9GgZOHcQ/027ljlQ+6nbonmzTAa09cpDmo+1XuLF8zqssq
     bLSt/+tkeNafg==
```

FIGURE 2.6 – En-tête DKIM

On y retrouve les champs suivants :

- `v=1` : Version de DKIM (doit toujours être `v=1`).
- `a` : Algorithme utilisé pour créer la signature numérique.
- `d` : Nom de domaine utilisé avec le sélecteur DKIM pour localiser la clé publique.
- `s` : Sélecteur DKIM utilisé pour trouver la clé publique correspondante.
- `t` : Heure à laquelle la signature a été générée (facultatif).
- `x` : Heure d'expiration de la signature (facultatif).
- `bh` : Hachage du corps du message.
- `h` : Liste des champs d'en-tête inclus dans la signature (séparés par des deux-points).
- `b` : Signature numérique, générée à partir des champs `bh` et `h`, signée avec la clé privée.

Quant à l'enregistrement `TXT` contenant les informations DKIM du domaine, on y retrouve les algorithmes utilisés par DKIM, ainsi que la clé publique utilisée pour vérifier l'en-tête DKIM.

```
v=DKIM1; h=sha256; k=rsa; p=MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBiQKBgQD...
```

Comme pour le SPF, DKIM nécessite de configurer un enregistrement DNS sur le serveur de l'expéditeur. Cet enregistrement va contenir le sélecteur DKIM (unique), l'adresse e-mail du domaine, ainsi que la clé publique à utiliser pour vérifier la signature.

2.2.2.3 DMARC (Domain-based Message Authentication, Reporting, and Conformance)

DMARC [KZ15, Pro25f] vérifie que le domaine indiqué dans le champ `From` de l'e-mail correspond bien au domaine validé par SPF ou DKIM. Pour la vérification avec SPF, il

va comparer le champ `From` avec le domaine présent dans le `Return-Path`. Pour DKIM, il compare le domaine du champ `From` avec le domaine indiqué dans la signature DKIM. Si l'allignement avec SPF ou DKIM est validé, il passe ensuite à la vérification SPF ou DKIM comme vu précédemment.

Dans le cas où une vérification échoue, le serveur de réception va quant à lui appliquer la politique DMARC qui est indiquée dans l'enregistrement DNS du serveur d'expédition comme on peut le voir ci-dessous.

```
v=DMARC1; p=quarantine; rua=admin@example.com
```

La politique DMARC peut prendre trois valeurs principales. La première est `none` qui indique qu'aucune action ne doit être entreprise et que l'on délivre tout de même le message, la seconde est `quarantine` indiquant que le message doit être envoyé dans le dossier des e-mails indésirables et finalement la valeur `reject` indiquant que l'e-mail n'est pas délivré. Il est également possible d'ajouter des options comme `rua` (rapports agrégés) ou `ruf` (rapports forensiques) pour définir à quelles adresses e-mail seront envoyés les rapports produits par DMARC.

À noter que pour qu'un e-mail passe la vérification DMARC, il suffit qu'il réussisse l'alignement avec SPF ou DKIM (l'un des deux seulement, pas nécessairement les deux).

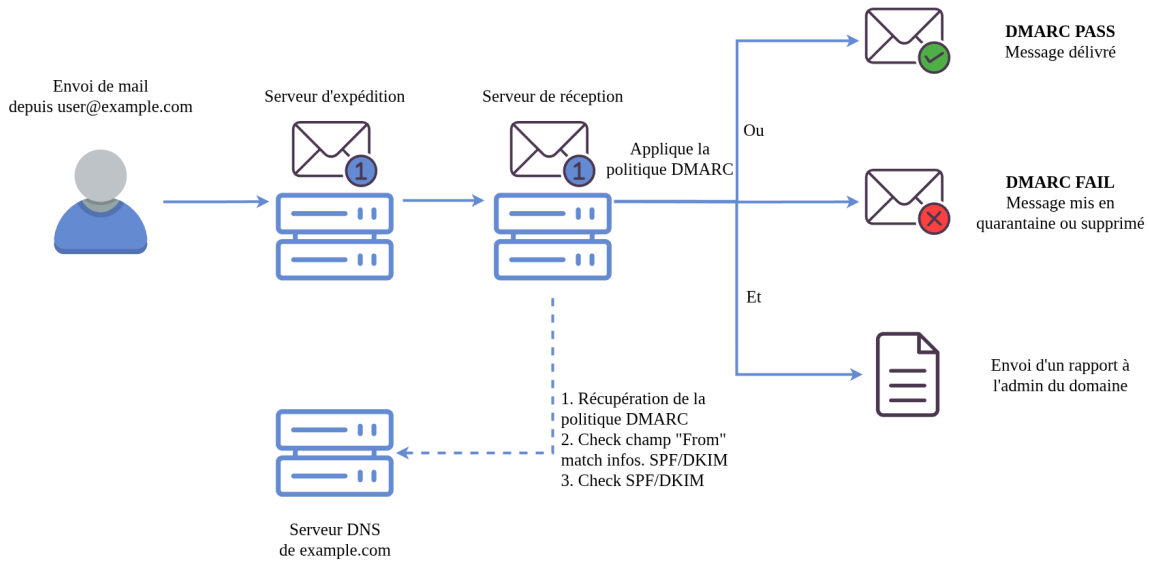

FIGURE 2.7 – Fonctionnement de DMARC

2.2.2.4 ARC (Authenticated Received Chain)

Ces trois mesures, sous forme d'enregistrements DNS, doivent toujours être mises en place ensemble afin d'être réellement efficaces contre les attaques de phishing. Elles protègent uniquement la transmission directe entre le serveur mail d'envoi et celui de réception, sans garantie si le message passe par des serveurs intermédiaires. Cependant, elles ne fonctionnent plus lorsque les e-mails transitent via des intermédiaires, bien que légitimes. En effet, un serveur mail intermédiaire peut tout à fait relayer un e-mail et ainsi modifier son adresse IP source voire modifier son contenu, cassant au passage les vérifications SPF et DKIM. Cela implique donc des problèmes de délivrabilité d'e-mail potentiellement légitimes comme le montre le schéma ci-dessous.

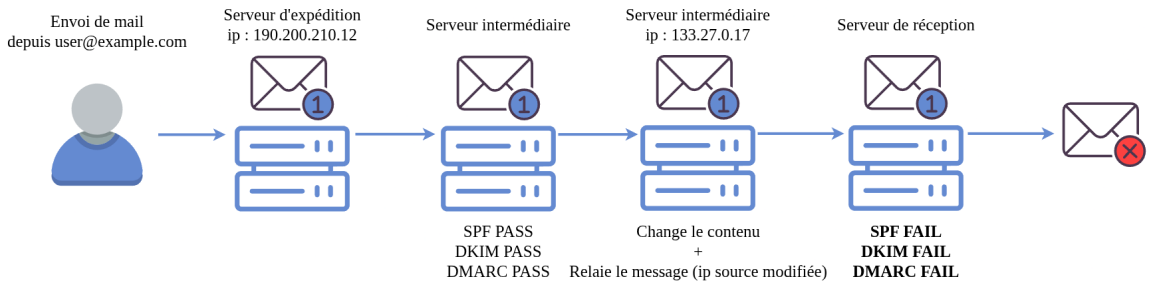

FIGURE 2.8 – Échec de l'envoi d'un e-mail lorsque ARC est désactivé

C'est pourquoi ARC [ALBK19, Pro25d] a été développé afin de préserver les résultats d'authentification du premier serveur intermédiaire par lequel transite un e-mail puis de vérifier l'identité de chaque serveur intermédiaire en cours de route. Pour ce faire, chaque serveur par lequel transite l'e-mail va se charger d'ajouter une signature ARC à l'en-tête du message.

Pour être plus précis, les en-têtes ARC sont composés de trois parties :

- **ARC-Authentication-Results :** Une copie des résultats des vérifications d'authentification de l'e-mail (SPF, DKIM, DMARC) effectuées par le premier serveur intermédiaire.
- **ARC-Message-Signature :** Une signature numérique similaire à une signature DKIM, couvrant l'ensemble du message et des en-têtes (sauf l'en-tête `ARC-Seal`).
- **ARC-Seal :** Une signature semblable à DKIM, couvrant les en-têtes ARC générés par chaque serveur intermédiaire.

FIGURE 2.9 – Exemple d'en-tête ARC (source : Proton)

A chaque serveur intermédiaire par lequel le message passe, le serveur va copier l'en-tête `Authentication-Results` et le mettre dans un `ARC-Autehentication-Results` avec un numéro de séquence pour indiquer l'ordre des serveurs. Il va aussi générer un nouvel ARC-Message-Signature avec encore une fois un numéro de séquence. Finalement, il crée un nouvel `ARC-Seal` qui a pour but de valider l'authenticité de chaque serveur faisant partie de la chaîne ARC.

Cela signifie que le serveur de réception n'a besoin que de vérifier la chaîne du `ARC-Seal` générée, ainsi que le dernier `ARC-Message-Signature`. Les vérifications SPF, DKIM et DMARC ne sont effectuées qu'au niveau du premier serveur intermédiaire rencontré. Ensuite, les serveurs suivants s'appuient uniquement sur la chaîne ARC. Ainsi, le serveur de réception final ne refait pas les vérifications SPF, DKIM ou DMARC du serveur initial, il se base sur la chaîne ARC transmise.

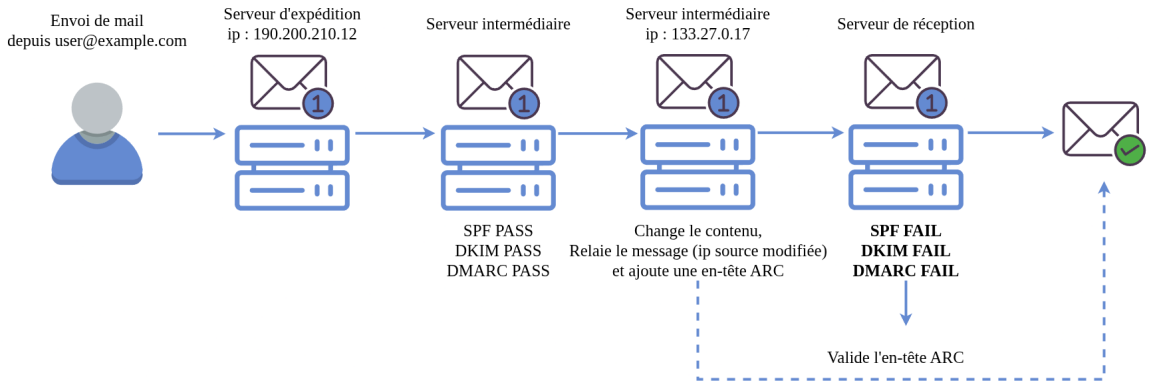

FIGURE 2.10 – Succès de l'envoi d'un e-mail avec ARC

2.2.2.5 BIMI (Brand Indicators for Message Identification)

Cette nouvelle norme, bien que secondaire dans la lutte contre le phishing, permet aux entreprises d'afficher leurs logos, validés par un tiers de confiance, directement dans les boîtes mail de leurs clients à côté du nom de l'expéditeur. L'objectif est de permettre aux destinataires d'identifier plus facilement un e-mail légitime en se basant sur la présence du logo. DMARC doit obligatoirement être configuré pour que BIMI fonctionne. BIMI n'est toutefois pas destiné aux particuliers utilisant des services de messagerie publics et reste encore peu supporté par les fournisseurs de messagerie.

Les mesures vues jusqu'à maintenant évitent qu'un e-mail soit modifié ou lu en cours de route et que l'identité d'un des maillons de la chaîne entre l'expéditeur et le destinataire soit usurpée. Toutefois, elles ne sont pas conçues pour lutter contre la compromission de l'expéditeur lui-même. C'est pourquoi il est essentiel de dresser les remparts suivants.

2.2.2.6 Filtre anti-spam

Les filtres anti-spam sont des systèmes logiciels ou matériels, placés en amont d'un serveur de messagerie ou intégrés directement à celui-ci. Ils utilisent différentes techniques pour identifier et bloquer les e-mails indésirables avant qu'ils n'atteignent la boîte de réception. Voici les principaux types de filtres :

- **Filtrage basé sur le contenu :** Analyse les mots-clés ou expressions spécifiques dans les e-mails. Les spammeurs contournent souvent cette méthode en modifiant légèrement les termes ou en utilisant des images.
- **Filtrage heuristique :** Attribue un score à chaque élément suspect dans un e-mail (liens, pièces jointes, format inhabituel). Si le score dépasse un seuil prédéfini, l'e-mail est classé comme spam.

- **Filtrage bayésien :** Utilise des algorithmes d'apprentissage automatique pour analyser des exemples d'e-mails légitimes et indésirables. Ce filtre calcule la probabilité qu'un e-mail soit un spam en fonction de ses caractéristiques.
- **Listes noires et blanches :** Les listes noires bloquent automatiquement les e-mails provenant de domaines ou adresses IP connus pour envoyer des spams, tandis que les listes blanches permettent explicitement à certains expéditeurs de contourner le filtrage.
- **Greylisting :** Rejette temporairement les e-mails provenant d'expéditeurs inconnus. Les serveurs légitimes réessayent, tandis que les serveurs malveillants abandonnent souvent l'envoi.

Cependant, ces filtres sont limités lorsqu'ils sont confrontés à des e-mails chiffrés de bout en bout. Dans ce cas, le contenu et les en-têtes ne peuvent être analysés par les filtres réseau, et l'analyse doit être déléguée au client mail du destinataire, qui dispose des clés nécessaires pour déchiffrer et inspecter le message. Cela permet de maintenir un certain niveau de protection, notamment contre les liens malveillants ou les pièces jointes infectées.

2.2.2.7 Extensions de détection d'anomalie

Comme vu dans la section sur le filtre anti-spam, il est compliqué de modérer le trafic lorsque celui-ci est chiffré. Cette tâche doit être déléguée au client mail, l'une des solutions consiste à utiliser des extensions spécialisées dans la détection d'anomalies. Celles-ci existent sous différents formats que ça soit pour les clients sur navigateur web (Giant Sentinel [1], ...) ou les applications de messagerie (SPAMfighter [2], ...).

Ces extensions fonctionnent de manière similaire aux filtres anti-spam : elles vérifient la présence et la validité des protections classiques (SPF, DKIM, DMARC, etc.) et peuvent utiliser des algorithmes d'intelligence artificielle pour détecter des contenus suspects.

2.2.3 Solutions de cybersécurité pour les e-mails

Les solutions vues jusqu'à présent représentent des solutions génériques ou des concepts permettant de se protéger face aux attaques. Il existe cependant des entreprises proposant des solutions prêtes à l'emploi qui vont se charger d'analyser les différents marqueurs listés précédemment, le contenu des e-mails ou encore d'exécuter les pièces jointes dans un environnement virtualisé dédidé.

Le but de cette section est de lister quelques acteurs notables du marché et si possible de comprendre comment fonctionnent les technologies qu'ils proposent.

1. `https://giantsentinel.com`
2. `https://www.spamfighter.com`

2.2.3.1 Solutions DNS

2.2.3.2 Protection serveur - Cisco Umbrella

Cisco Umbrella[Cis25] est une couche de sécurité réseau déployée au niveau du pare-feu ou de la passerelle, qui analyse toutes les requêtes DNS pour détecter et bloquer les domaines malveillants. Cette technologie va aussi permettre de bloquer les sites répertoriés comme malicieux par Cisco. L'un de ses avantages est de protéger l'ensemble du réseau et tous les utilisateurs via un seul point de contrôle par lequel transite tout le trafic.

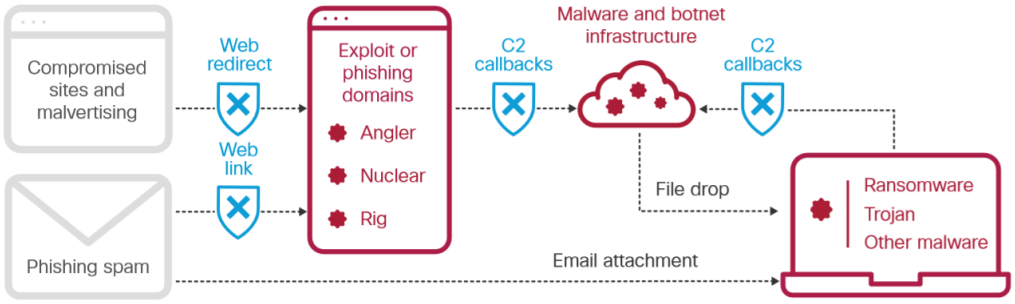

FIGURE 2.11 – Défense proposée par une solution telle que Cisco Umbrella (source : Cisco)

Cisco Umbrella a été pris comme exemple ici, mais il est existe d'autres solutions similaires proposées par des entreprises telles que Fortinet [3] ou Switch [4].

2.2.3.3 Protection client - NextDNS

NextDNS[Nex25] est un service de résolution DNS qui filtre et bloque les domaines indésirables avant qu'ils ne soient chargés par l'utilisateur. Configurable directement sur un appareil, il permet de bloquer les sites malveillants, de limiter le suivi publicitaire et d'améliorer la confidentialité des utilisateurs. Son principal avantage est de fonctionner sur tous les appareils sans nécessiter l'installation d'un logiciel dédié, puisqu'il suffit de configurer le DNS personnalisé proposé par le service.

Il existe d'autres services similaires, tels que DNSFilter [5] ou MullvadDNS [6].

3. `https://www.fortiguard.com/services/sdns`
4. `https://www.switch.ch/en/dns-firewall`
5. `https://www.dnsfilter.com/`
6. `https://mullvad.net/en/help/dns-over-https-and-dns-over-tls`

2.2.3.4 Solutions pour serveur mail

2.2.3.5 Microsoft EOP (Exchange Online Protection) et Defender for Office 365

Microsoft Exchange Online Protection (EOP) [Mic25a] est un service de filtrage cloud conçu pour protéger les organisations contre le spam, les logiciels malveillants, le phishing et autres menaces liées aux e-mails. Inclus par défaut dans les abonnements Microsoft 365 avec des boîtes aux lettres Exchange Online, EOP peut également être déployé pour protéger des environnements de messagerie sur site ou hybrides. Son fonctionnement repose sur plusieurs couches de filtrage, notamment la vérification de la réputation de l'expéditeur, l'analyse des logiciels malveillants et l'application de règles de flux de messagerie personnalisées.

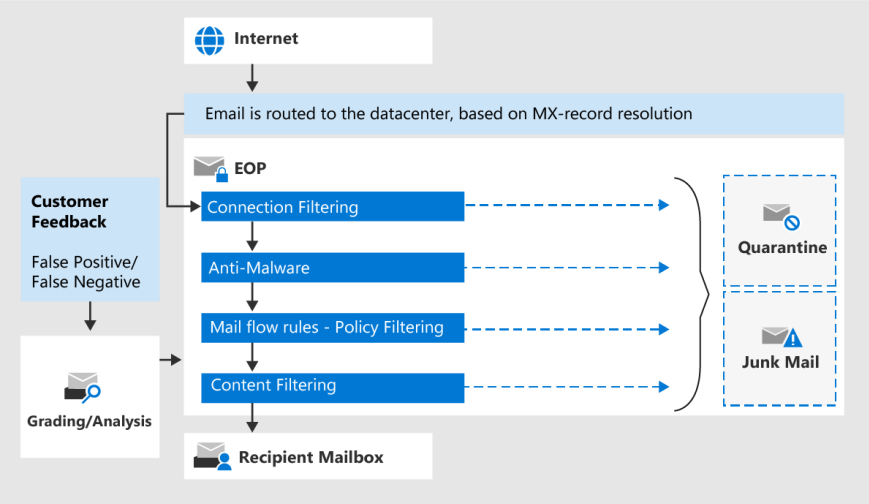

FIGURE 2.12 – Chaîne d'exécution de Microsoft EOP (source : Microsoft)

Microsoft Defender pour Office 365 [Mic25c] complète les protections de base d'Exchange Online Protection en offrant des fonctionnalités avancées contre le phishing. Par exemple, la protection contre l'usurpation d'identité qui détecte et bloque les tentatives d'imitation de domaines ou d'utilisateurs de confiance, réduisant ainsi le risque de phishing ciblé. De plus, la fonctionnalité de `Campaign Views` analyse et suit les campagnes de phishing ciblant l'organisation, offrant une meilleure compréhension des menaces en cours. Enfin, la formation par simulation d'attaques permet de sensibiliser les utilisateurs via de fausses campagnes de phishing, renforçant ainsi leur vigilance face aux menaces réelles. EOP et Defender incluent également des protections comme `Safe Links` et `Safe Attachments`, qui analysent les liens et pièces jointes en temps réel dans un environnement isolé afin de détecter tout contenu malveillant avant qu'il n'atteigne l'utilisateur.

2.2.3.6 Proofpoint

Proofpoint [Pro25a] adopte une approche de sécurité centrée sur les personnes, reconnaissant que les individus constituent souvent le maillon le plus vulnérable face aux menaces. Cette stratégie vise à protéger les employés en tant que cibles principales des attaques, en mettant l'accent sur la compréhension et la mitigation des risques humains. En 2024, Gartner a désigné Proofpoint comme leader du marché dans la sécurité des e-mails. La plateforme est structurée autour de deux composantes principales : Nexus et Zen.

Nexus [Pro25b] est une plateforme de threat intelligence alimentée par l'IA et des informations sur les menaces en temps réel qui permet d'analyser l'ensemble de la chaîne d'un envoi d'un e-mail.

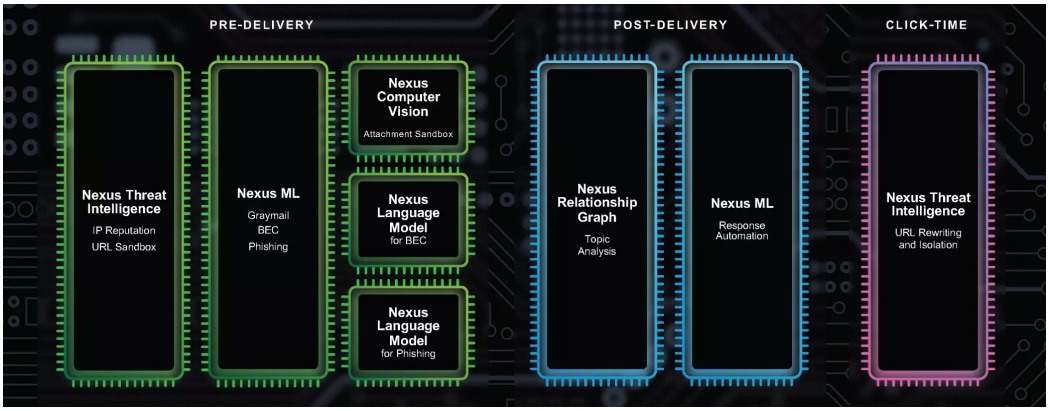

FIGURE 2.13 – Chaîne d'exécution de Proofpoint Nexus (source : Proofpoint)

Les analyses par intelligence artificielle permettent d'identifier des comportements utilisateurs inhabituels et des e-mails au contenu suspect qui vont par exemple utiliser un langage transactionnel ou un champ lexical de l'urgence. Leur plateforme possède aussi une IA reconnaissant les éléments visuels présents dans les e-mails ou sur les sites qui sont envoyés afin de reconnaître des éléments qui pourraient être liés à du phishing. La plateforme maintient également une veille de threat intelligence, c'est-à-dire la collecte et l'analyse d'informations sur les acteurs et techniques de menaces connus.

Zen [Pro25c], quant à elle, est une suite de technologies conçue pour protéger et engager les utilisateurs. Elle comprend des outils tels que ZenWeb, une extension de navigateur qui sécurise la navigation sur tous les appareils, et Zen pour Outlook, qui défend contre les menaces par e-mail et la perte de données. Ces solutions visent à guider les utilisateurs vers des comportements plus sûrs et à réduire les risques associés aux interactions humaines.

Proofpoint est une solution commerciale dont les tarifs ne sont pas publiquement disponibles.

2.2.3.7 Rspamd

Rspamd est une solution open source de traitement de courriels complémentaire au Message Transfer Agent (MTA). Les messages sont évalués à l'aide de plusieurs méthodes de détection de courriel indésirable et reçoivent un score qui, en fonction des stratégies définies, détermine la recommandation de traitement du message transmis au MTA : s'il faut le passer, le rejeter, le mettre en quarantaine ou le modifier.

Rspamd offre les fonctionnalités suivantes :

- filtrage de courriel indésirable appuyé sur l'analyse statistique, les expressions régulières, et l'apprentissage automatique, ainsi que sur des listes de refus d'URL et d'autres modules complémentaires
- contrôle et application des stratégies
- DKIM
- extensibilité via des modules en Lua [Rsp25]

2.2.3.8 Apache SpamAssassin

Apache SpamAssassin est une plateforme anti-spam open source. Elle s'appuie sur une large gamme de tests d'analyse heuristique et statistique pour identifier les courriels indésirables. Le score que Apache SpamAssassin attribue aux messages peut être utilisé par un programme tiers, tel qu'un MTA, pour déterminer leur traitement.

Apache SpamAssassin fournit les fonctionnalités suivantes :

- tests heuristiques
- entraînement d'un filtre bayésien
- intégration avec des bases de données de filtrage collaboratif
- annotation des courriels pour traitement par le client de messagerie des utilisateur·rice·s
- modules d'intégration dans un MTA [Apa25]

2.2.3.9 Solutions pour client mail

2.2.3.10 Ironscales

Ironscales est une solution de protection contre le phishing destinée aux entreprises, qui s'intègre directement dans les messageries cloud comme Microsoft 365 et Google Workspace. Fonctionnant via des connexions API, elle analyse les e-mails en temps réel, y compris après leur livraison, et combine l'intelligence artificielle avec la collaboration entre utilisateurs (crowdsourcing) pour identifier rapidement les nouvelles menaces. Les e-mails suspects sont signalés par des bandeaux d'alerte dans la boîte de réception, et un bouton dédié permet aux employés de remonter facilement les messages douteux. En cas de menace confirmée, Ironscales peut automatiquement retirer les e-mails malveillants des boîtes concernées.

2.2.3.11 Thunderbird

Thunderbird, le client de messagerie de la fondation Mozilla, intègre des protections de base contre l'hameçonnage, comme par exemple un avertissement si les liens visibles dans un message ne correspondent pas aux véritables liens. Le client sollicite également une confirmation de l'utilisateur·rice avant d'ouvrir de tels liens, ce qui évite de les visiter accidentellement ou par inattention [Moz25b].

2.2.3.12 Solutions pour navigateur

2.2.3.13 NetCraft

NetCraft est une extension de navigateur qui offre une protection proactive contre le phishing et les arnaques en ligne. Compatible avec Chrome, Firefox, Edge et d'autres navigateurs, elle s'appuie sur une base de données maintenue par Netcraft pour bloquer automatiquement les sites malveillants connus. Lorsqu'un utilisateur tente d'accéder à une page frauduleuse, un avertissement s'affiche immédiatement et l'accès est bloqué. L'extension permet également de signaler en un clic les sites suspects, enrichissant ainsi la base de données communautaire. Elle complète les protections intégrées aux navigateurs en ajoutant une couche supplémentaire de détection et de blocage.

2.2.3.14 PIXM

PIXM est une extension de navigateur qui utilise l'intelligence artificielle pour détecter les tentatives de phishing en temps réel. Contrairement aux solutions classiques basées sur des listes noires, PIXM analyse visuellement les pages web au moment de leur chargement pour détecter les sites imitant l'apparence de services connus (par ex. Facebook, Google, Office 365). Si une imitation est détectée, l'extension bloque immédiatement la page et affiche une alerte à l'utilisateur. Cette approche basée sur la reconnaissance visuelle permet de détecter même les attaques de phishing récentes, absentes des bases de données traditionnelles.

2.2.3.15 Firefox

Tout comme Thunderbird, Firefox, le navigateur de la fondation Mozilla, intègre des protections de base contre l'hameçonnage. Il offre une vérification préliminaire des sites web que l'utilisateur·rice tente de visiter, ainsi que l'interception des potentiels logiciels malveillants lors des téléchargements [Moz25a].

2.2.3.16 Solutions éducatives

2.2.3.17 Brightside

Brightside [Brs25] est une plateforme visant à réduire l'efficacité des attaques de phishing en diminuant la quantité d'informations personnelles exploitables par les attaquants.

Ils permettent aux employés d'une entreprise ou aux particuliers de scanner le web afin de trouver toutes les informations qui leurs sont liées. Brightside identifie les comptes associés aux adresses e-mail des utilisateurs et recense les informations personnelles accessibles publiquement (nom, prénom, date de naissance, localisation, etc.). Si on remarque que certains de ces comptes sont inutiles, Brightside permet même d'envoyer un e-mail à ces entreprises afin de faire supprimer nos données.

Brightside permet aussi de faire l'historique des fuites de données dans lesquelles nos informations apparaissent afin que l'on puisse réagir en conséquences. Par exemple changer un mot de passe qui aurait fuité. En réduisant la quantité d'informations inutiles disponibles publiquement, on diminue la probabilité qu'une attaque de phishing ciblée soit suffisamment crédible pour tromper la victime. Les deux fonctionnalités précédemment décrites sont proposées au tarif de 8 CHF par mois.

La dernière fonctionnalité que présente Brightside est de réaliser des campagnes de phishing sur les employés. Les e-mails envoyés durant ces campagnes sont construits à partir d'informations réelles collectées en ligne sur les cibles. Cette option qui est activable pour les clients business uniquement coûte 4 CHF par utilisateur et par mois.

2.2.3.18 Gophish

Gophish [Gop25] est un framework open source de simulation de phishing. Il permet aux entreprises de concevoir et de mettre en oeuvre des campagnes d'hameçonnage pour sensibiliser les employé·e·s.

2.2.4 Solutions d'analyse de liens et noms de domaine

Il existe plusieurs outils et API permettant d'analyser dynamiquement les liens avant leur ouverture. Cette section présente quelques solutions notables pouvant être intégrées dans le développement d'une application pour prévenir les attaques de phishing.

2.2.4.1 Google Safe Browsing API

Google Safe Browsing est un service de sécurité gratuit proposé par Google, qui permet de vérifier si une URL est associée à des sites malveillants, notamment du phishing, des malwares

ou des attaques par ingénierie sociale. Lorsqu'une URL est soumise, l'API retourne un verdict basé sur ses bases de données de menaces maintenues à jour. Dans une application, son usage permet d'ajouter rapidement un premier filtre de sécurité sans devoir maintenir sa propre base de sites frauduleux.

L'API est gratuite dans une certaine limite d'utilisation quotidienne, au-delà il est nécessaire de passer à une version payante.

2.2.4.2 PhishTank

PhishTank est une base de données collaborative maintenue par OpenDNS qui centralise les signalements de sites de phishing. Chaque site est vérifié manuellement par la communauté ou automatiquement par les systèmes de détection. PhishTank propose une API gratuite permettant de vérifier si une URL est déjà connue comme étant un site de phishing. Dans le cadre du développement d'une application, PhishTank peut servir de deuxième niveau de vérification après un premier filtrage automatisé.

2.2.4.3 URLScan.io

URLScan.io est un service d'analyse dynamique d'URL. Lorsqu'une URL est soumise, le service explore la page, capture son apparence visuelle, liste ses ressources (scripts, liens externes) et détecte les éléments de phishing visuels ou techniques. L'intégration d'URLScan dans une application permet non seulement d'obtenir des verdicts de sécurité, mais aussi de disposer d'informations détaillées permettant d'analyser l'apparence ou le comportement suspect d'une page.

La version gratuite propose un nombre limité d'analyses par jour, mais une version API commerciale permet un usage étendu.

2.2.4.4 Whois et analyse de l'âge du domaine

La vérification Whois consiste à obtenir des informations sur le propriétaire et la date de création du domaine associé à l'URL. Un domaine créé très récemment est souvent un indicateur de phishing. Des services comme WhoisXML API permettent de faire ces vérifications automatiquement via API. Dans une application, l'analyse de l'âge du domaine peut être intégrée pour augmenter le niveau de suspicion attribué à un lien, notamment si celui-ci copie une marque connue mais a été enregistré récemment.

2.2.4.5 OpenPhish

OpenPhish est une plateforme commerciale proposant un flux automatisé de liens de phishing détectés en temps réel. Contrairement à PhishTank qui repose partiellement sur des signalements manuels, OpenPhish utilise un moteur d'analyse automatique capable de détecter des campagnes émergentes. Son API peut être intégrée dans des applications pour une détection proactive des nouvelles menaces sans attendre les mises à jour communautaires.

2.2.4.6 Flairsafe.ch

Un outil notable apparu à la fin du projet est Flair, une plateforme en ligne lancée en septembre 2025 permettant d'analyser rapidement des contenus suspects. Développé comme initiative personnelle par Sandy Lavorel, spécialiste actif dans la lutte contre la fraude, Flair offre une interface simple où l'utilisateur peut soumettre un message, un lien ou une capture d'écran afin d'obtenir un verdict immédiat sur le risque potentiel d'arnaque. L'outil combine plusieurs techniques d'analyse pour détecter des indices caractéristiques du phishing, du smishing ou d'autres formes d'ingénierie sociale. Bien que Flair soit sorti après la clôture de notre projet et n'ait donc pas pu être intégré dans notre démarche, il représente un exemple intéressant d'initiative visant à fournir une assistance directe et accessible aux utilisateurs. Son apparition confirme la pertinence des besoins identifiés tout au long de ce travail.

2.3 Mesures organisationnelles et légales

2.3.1 Introduction

La présente section fait état des méthodes organisationnelles et légales pour la lutte en matière de phishing. L'objectif est de comprendre l'ensemble de la chaîne sur laquelle il est possible de s'appuyer pour réduire l'impact du phishing. À la base de celle-ci se trouve les lois, principes immuables auxquels l'entièreté des entités doit se soumettre. Ensuite vient l'ensemble des normes, politiques, directives et standards instaurés par les entreprises ayant pour but de garantir la conformité et la sécurité de leurs opérations. Ces éléments définissent un cadre de référence permettant d'orienter les pratiques internes et de répondre aux exigences réglementaires en vigueur. Tout ceci peut s'accompagner de formations au niveau utilisateur·rice. Pour finir, des organismes et des dispositifs suisses mettent à disposition des plateformes numériques publiques dispensant du matériel qui a pour vocation d'aider à identifier et gérer les menaces.

2.3.2 Cadre légal

2.3.2.1 Répression

D'après le site de la Prévention Suisse de la cybercriminalité : « L'hameçonnage ne fait pas l'objet d'une norme pénale particulière en Suisse. » Il peut être sanctionné en vertu de multiples articles du Code pénal suisse [dlC21] and [Con25].

- Art. 143 CP : Soustraction de données
- Art. 143bis CP : Accès indu à un système informatique
- Art. 144 CP : Dommages à la propriété
- Art. 147 CP : Utilisation frauduleuse d'un ordinateur
- Art. 251 CP : Faux dans les titres
- Art. 305bis CP : Blanchiment d'argent

D'autres articles sont mentionnés sur le site Guide Social Romand et le site International Comparative Legal Guides [GSR24, Glo24].

- Art. 144bis CP : Détérioration de données
- Art. 146 CP : Escroquerie
- Art. 179^{novies} CP : Soustraction de données personnelles

Bien que listés sur ces différentes plateformes, leur applicabilité fait l'objet de débats depuis plus de vingt ans [Amm06, Mon09]. La controverse porte principalement sur la question de savoir si les éléments constitutifs objectifs et subjectifs requis par chacun de ces articles sont effectivement remplis. En droit pénal suisse, les éléments constitutifs objectifs (ECO) désignent les faits extérieurs et observables d'une infraction, tels que les actes, le résultat ou encore le lien de causalité. Quant aux éléments constitutifs subjectifs (ECS), ils concernent l'état d'esprit de l'auteur, notamment l'intention ou la négligence. Ces deux éléments doivent être réunis pour qu'un comportement soit pénalement punissable [Leg17]. Les articles en question ont été retenus pour traiter les cas de phishing, dans la mesure où ils s'appliquent déjà à des infractions plus larges liées au piratage informatique [Mon09].

2.3.2.2 Art. 143 CP : Soustraction de données et Art. 143bis CP : Accès indu à un système informatique

Pour les articles 143 et 143bis du Code pénal suisse, une incertitude subsiste autour de la formulation suivante : « ...qui sont spécialement protégées contre tout accès indu de sa part,... ». Selon certain·e·s spécialistes, cette disposition implique nécessairement le contournement d'un dispositif de protection technique, tel qu'un système logiciel, au moyen de procédés techniques.

Dans le cas de l'hameçonnage, aucun mécanisme de protection n'est surmonté, car la victime communique volontairement ses données à la suite de manœuvres manipulatoires [Amm06].

Cependant, une publication plus récente, notamment reprise dans le Commentaire Romand du Code pénal II, a admis que l'utilisation de l'ingénierie sociale pouvait être comparable à l'ingénierie technique et était donc suffisante pour considérer que l'élément constitutif objectif était rempli [Mon09].

De plus, un arrêt du Tribunal fédéral datant du 17 mai 2019 montre que le recours d'une femme condamnée selon l'art. 143bis CP a été rejeté, car il a été jugé que la manière dont le mot de passe avait été obtenu n'avait pas d'importance dans la mesure où l'accès au compte mail s'était fait sans autorisation et qu'elle avait outrepassé le système de sécurité censé lui en empêcher l'accès [TRE19].

Ainsi, même en l'absence de contournement technique, le fait d'obtenir un mot de passe par tromperie et d'accéder sans droit à un système protégé suffit à remplir les conditions de l'infraction. L'auteur·rice peut donc être condamné·e dès lors qu'iel accède sans autorisation à un système informatique sécurisé, indépendamment du moyen utilisé pour obtenir les données d'accès.

2.3.2.3 Art. 144 CP : Dommages à la propriété et Art. 144bis CP : Détérioration de données

La distinction entre ces deux articles se fait principalement sur la base de la notion de donnée informatique. Celle-ci est au cœur de l'art. 144bis CP, qui vise spécifiquement les atteintes portées aux données informatiques, tandis que l'art. 144 CP, plus général, sanctionne les dommages à la propriété de nature physique [Per21]. Ainsi, l'article 144bis CP se démarque de l'article 144 CP par son champ d'application limité aux données stockées ou transmises par des moyens informatiques, incluant également les supports externes tels que les clés USB ou les disques durs externes [Mé17].

Dans le cadre de l'hameçonnage, l'article 144bis CP intervient notamment dans les cas de mails de phishing faisant usage d'un malware dissimulé dans une pièce jointe ou se téléchargeant automatiquement par le biais d'un lien. L'altération des données survient à la suite de la propagation du virus [Spa14]. Il en est également question dans les cas d'in-session phishing où la manipulation de flux ou de contenus numériques pour tromper l'utilisateur·rice est caractéristique et constitue une altération du système informatique ou des données [Spa14]. Cette technique, qui repose souvent sur une faille de type cross-site scripting, permet à l'attaquant d'ouvrir une fausse fenêtre ou de rediriger l'utilisateur vers un site frauduleux, modifiant ainsi de manière non autorisée le déroulement d'un traitement informatique [Wik24]. Contrairement au phishing ordinaire, l'in-session phishing est condamnable selon l'article 144bis, car l'attaquant·e manipule le flux en redirigeant les visiteur·euse·s d'un site vers sa version frauduleuse, ce qui constitue de ce fait une modification non autorisée des données ou du déroulement d'un traitement informatique [Spa14].

Concernant l'art. 144 CP, il peut entrer en concours avec l'art. 144bis CP, mais seulement

lorsque des dommages physiques sont causés. Il ne pourrait donc être utilisé que dans des cas où un malware causerait des dégâts physiques au système informatique de la victime ou lorsqu'il faut engager des moyens importants pour réparer les dégâts subis par le système [Mé17].

L'article 144bis CP se concentre donc sur l'intégrité et l'utilisation non perturbée des données informatiques, constituant une infraction de détérioration spécifique informatique, par opposition à l'art. 144 CP qui concerne les dommages aux biens matériels [Per21]. Une distinction importante à noter est que, contrairement aux articles 143 et 143bis CP, les données visées par l'article 144bis CP n'ont pas besoin d'être spécialement protégées. La protection s'étend à toutes les données stockées sur des supports informatiques ou transférées par des moyens informatiques, qu'elles soient ou non protégées [Mé17].

2.3.2.4 Art. 146 CP : Escroquerie

En règle générale, l'obtention des données de connexion n'est pas considérée comme une infraction pénalement sanctionnée. Le phishing est ainsi souvent la prémisse d'une escroquerie, sauf s'il y a soustraction de données ou utilisation frauduleuse. Ici, l'escroquerie suppose qu'une ruse ou une tromperie soit utilisée pour induire une erreur chez la victime afin de ellui soutirer un avantage illicite comme de l'argent ou des biens. Le phishing devient donc punissable dès qu'il franchit ce cap : utilisation des données pour tromper une victime et en tirer un profit indu [Mé17].

Toutefois, une incertitude subsiste quant au lien de causalité entre l'erreur de la victime et l'atteinte à son patrimoine. En effet, les données obtenues par phishing n'ont pas, en elles-mêmes, une valeur économique directe et leur collecte ne cause pas immédiatement un préjudice financier. Il faut généralement qu'une ou plusieurs étapes supplémentaires soient franchies pour que l'infraction d'escroquerie soit pleinement constituée [TG21].

Le phishing seul, sans préjudice direct, n'est donc généralement pas puni selon cet article [Spa14]. Il s'applique dès lors que l'utilisation frauduleuse des données cause un dommage effectif à la victime [Amm06]. Il est donc essentiel d'établir un lien clair entre la tromperie, l'erreur de la victime et la perte subie pour que la responsabilité pénale puisse être engagée [Spa14].

2.3.2.5 Art. 147 CP : Utilisation frauduleuse d'un ordinateur

Le Commentaire Romand du Code pénal II explique « qu'il existe une utilisation indue lorsque l'auteur obtient des données par le système de hameçonnage » [Leg17]. Au sens de l'article 147 CP, l'envoi des mails de phishing n'est qu'un acte préparatoire et n'est donc pas punissable. Ce n'est que lors de l'utilisation des données obtenues, dans le cas d'un transfert d'actifs par exemple, que cela le devient [Amm06].

Cette utilisation des données mène à un processus de traitement de données qui aboutit à un résultat inexact (l'autorisation pour l'auteur·rice non autorisé·e) et provoque un transfert d'actifs au préjudice de la victime [TG21].

2.3.2.6 Art. 179novies CP : Soustraction de données personnelles

La principale différence avec l'art. 143 CP réside dans le type de données. L'art. 179^{novies} CP protège les données personnelles sensibles comme définies dans la LPD. Pour [Amm06], les données d'accès, comme des identifiants, ne constituent pas des données sensibles et il est primordial qu'elles permettent l'obtention ou l'accès à des données plus sensibles, comme des informations liées à la sphère privée ou à la personnalité de la victime [TG21].

En somme, l'art. 179^{novies} CP ne trouve application dans le contexte du phishing que si celui-ci permet d'accéder à des données relevant de la sphère privée ou personnelle sensible de la victime, au-delà des simples identifiants de connexion.

2.3.2.7 Art. 251 CP : Faux dans les titres

Selon l'interprétation du Tribunal fédéral, un courrier de phishing doit être considéré comme un faux matériel car lae véritable auteur·rice du titre ne correspond pas à l'auteur·rice apparent·e [Spa14]. Un site internet falsifié constitue également un faux matériel lorsqu'il reproduit fidèlement l'apparence du site officiel d'une institution bancaire et induit ainsi la victime en erreur, par exemple en ellui permettant d'ouvrir une session de e-banking semblable à celle du véritable site [Spa14].

De plus, il est nécessaire que l'auteur·rice agisse dans le but d'obtenir des renseignements provenant de sa victime et qu'iel les utilise ou les revende, pour que toutes les conditions soient remplies [TG21].

2.3.2.8 Art. 305bis CP : Blanchiment d'argent

Le phishing, en tant que méthode de collecte frauduleuse de données personnelles ou bancaires, peut entretenir un lien indirect avec l'art. 305 CP [Spa14]. En effet, si les données obtenues par phishing, notamment les identifiants bancaires utilisés pour transférer, dissimuler ou réinjecter des fonds provenant d'une infraction préalable, comme l'escroquerie ou la fraude, alors ces actes peuvent constituer du blanchiment [Gro15, Gos20].

Dans les faits, le phishing n'est pas en lui-même un acte de blanchiment, mais il peut servir d'étape préparatoire à l'introduction d'argent illicite dans le système financier, en se servant donc de comptes subtilisés [Gos20, Spa14]. Ainsi, le lien entre le phishing repose sur l'utilisation des éléments obtenus par ce procédé [Gro15].

2.3.2.9 Bases légales

À la base de la lutte contre le phishing se trouvent les lois. Aucune action et aucune décision ne peuvent être prises sans le cadre légal approprié, qui garantit la protection des données et la conformité aux régulations. Les lois imposent donc un certain nombre de règles qui doivent être respectées par toutes et tous. Des outils à la gestion des données, en passant par les protocoles de sécurité, chaque aspect de la prévention et de la réponse au phishing doit être en adéquation avec les exigences légales afin de minimiser les risques et protéger les utilisateurs.

Les institutions suisses doivent donc se conformer à un certain nombre de règles fédérales et européennes [Gro15] : Loi sur la Protection des Données (LPD), Loi sur les Télécommunications (LTC), Règlement Général de Protection des Données (RGPD) dans le cas d'échange avec des résidents de l'Union Européenne, etc. [Gro15, TG21, MMQ17]. Le Préposé fédéral à la protection des données et à la transparence (PFPDT) se charge de veiller à ce que les entreprises helvétiques respectent bien la LPD et qu'elles répondent aux exigences en matière de phishing [TG21]. Afin de garantir cette conformité, plusieurs règles clés s'imposent en matière de traitement des données [sui25] :

1. **Sécurité des données personnelles (Art. 8 LPD)** : Il est nécessaire de mettre en place des mesures techniques et organisationnelles appropriées pour garantir la sécurité des données personnelles [falpddealtP24b]. Cela inclut la protection contre tout traitement non autorisé ou illicite, la perte, l'altération, l'accès ou la divulgation non autorisée de ces données. Les mesures doivent être adaptées au risque et à la sensibilité des données traitées.
2. **Analyse des e-mails et respect de la proportionnalité (Art. 6 al. 2 à 4 LPD)** : Lors de l'analyse des e-mails pour détecter des tentatives de phishing, les entreprises doivent respecter trois principes : se limiter à ce qui est nécessaire (principe de proportionnalité), l'effectuer dans un but clair et légitime (principe de finalité) et les personnes concernées doivent être informées de cette surveillance (principe de transparence) [falpddealtP23].
3. **Formation des collaborateurs (Art. 5 al. 1 et Art. 8 LPD)** : Bien que la LPD ne mentionne pas explicitement la formation, l'obligation de sécurité implique indirectement que les collaborateurs soient formés pour éviter des atteintes à la sécurité des données [falpddealtP24c].
4. **Documentation et traçabilité (Art. 12 à 14 LPD)** : Les entreprises doivent tenir un registre des activités de traitement des données personnelles [falpddealtP24a].
5. **Notification en cas de violation de données (Art. 24 LPD)** : En cas de violation de la sécurité des données, les entreprises doivent notifier le PFPDT sans délai [falpddealtP25].

2.3.3 Méthodes organisationnelles

2.3.3.1 Préambule

Face à la prolifération des attaques par phishing, il est essentiel pour les organisations de mettre en place des stratégies efficaces afin de se protéger contre cette menace en constante évolution. Bien que les solutions techniques, telles que les filtres anti-spam, jouent un rôle clé, les méthodes organisationnelles restent tout autant cruciales. Elles englobent un ensemble de pratiques et de processus internes qui visent à sensibiliser les employés et créer un environnement de travail où la sécurité informatique devient une priorité partagée par tous.

Cette rubrique explore les différentes approches organisationnelles pour contrer le phishing, en mettant l'accent sur les normes, politiques et standards, sur la formation des utilisateurs ainsi que sur le matériel dispensé.

2.3.3.2 Normes, politiques et standards

Les normes, les politiques et les standards sont des contre-mesures procédurales élaborées par les organismes, les entreprises ou les gouvernements pour répondre à leurs besoins spécifiques. Contrairement aux lois, elles ne sont pas imposées et les organisations sont donc libres d'instaurer les règles qu'elles souhaitent afin de renforcer la protection contre l'hameçonnage. Ces règles s'inspirent et s'appuient souvent sur des référentiels reconnus internationalement, tels que les normes ISO/IEC (respectivement International Organization for Standardization et International Electrotechnical Commission) ainsi que sur les bonnes pratiques émises par des organismes comme le National Institute of Standards and Technology (NIST).

L'un des exemples les plus communs est le standard relatif aux mots de passe recommandé par l'ISO/IEC 27001 [Dat22]. Il impose ou suggère aux individus au sein de l'organisme divers critères pour leur création et leur gestion : taille, caractères spéciaux, renouvellement fréquent, limite de réutilisation, etc. [DFM+22].

Cette même norme ISO recommande l'emploi de l'authentification multi-facteurs (MFA) [Dat22]. Renforçant la sécurité des comptes par l'adoption de deux moyens d'authentification indépendants, elle se décline sous différentes formes : SMS contenant un code temporaire, application de génération de mots de passe à usage unique (ou authentificateur), jeton d'authentification physique, etc. [BAJ25].

La mise en place d'une culture cybersécurité permettrait également de consolider le système défensif [Gur23]. Elle est mentionnée et conseillée dans de nombreuses normes ISO : ISO/IEC 27001, ISO/IEC 27002 :2022 et ISO/IEC 27035, pour ne citer que celles-ci [Dat22, CLU14, PEC25].

2.3.3.3 Éducation des utilisateurs

Considérés comme le dernier rempart dans les organisations [CDS+24], les utilisateurs représentent tout de même la plus grande faille dans le système [Bil20]. C'est à eux qu'incombe la décision d'interagir avec le lien ou la pièce jointe du courriel, une fois que celui-ci a traversé toutes les solutions techniques de filtrage déployées. Leur éducation est donc essentielle pour les sensibiliser aux techniques de phishing et leur apprendre à identifier les menaces.

2.3.3.4 Phishing Awareness Training (PAT)

C'est notamment pour cela que de nombreux programmes de formation nommés Phishing Awareness Training ou simplement Awareness Training (PAT) ont vu le jour. L'origine de cette pratique n'est pas connue, mais elle serait issue de multiples recherches en cybersécurité [KHN+22, MGD17].

L'objectif du PAT est de réduire la menace que représente le phishing en créant une stratégie de défense proactive humaine complémentaire aux solutions technologiques déjà en place [SHJL23]. Il vise principalement à entraîner les employés en les sensibilisant et en leur faisant adopter un comportement adapté. La forme que peuvent prendre ces PAT varie souvent selon le temps et le budget à disposition [GDZ+25], mais ils s'articulent très souvent de la manière suivante :

1. Les organisateurs créent une campagne de phishing envoyant un ou plusieurs mails de phishing factices afin d'évaluer la capacité de détection des employés.
2. Du matériel et des entraînements sont dispensés à ceux ayant interagi de façon inappropriée : ouverture de la pièce jointe, accès au site via le lien, renseignement d'informations confidentielles, etc.
3. De nouveaux mails sont envoyés afin d'évaluer les progrès faits.

De nombreux services disponibles sur internet offrent la possibilité de créer et diffuser des campagnes de phishing. Ils permettent, d'une part, de créer de faux mails et de concevoir un site de phishing factice afin de simuler un environnement semblable à ceux déployés par les vrais attaquants [SLLK24]. Ces programmes offrent généralement un tableau de bord fournissant des résultats statistiques et analytiques sur les interactions des usagers avec l'e-mail, le site et la formation dispensée si ceux-ci se sont fait avoir [PAZ20]. Il existe des solutions open-source comme GoPhish [7] et des solutions commerciales telles que KnowBe4 [8] et HoxHunt [9].

La mise en place d'un PAT dans une institution nécessite l'évaluation et la prise en compte de plusieurs facteurs pour garantir le succès de son adoption : les facteurs organisationnels,

7. `https://getgophish.com`
8. `https://www.knowbe4.com`
9. `https://hoxhunt.com`

techniques et environnementaux [GDZ+25]. Cette classification, basée sur le cadre TOE (Technology-Organization-Environment), présente l'avantage que la majorité des éléments abordés dans la littérature peuvent être regroupés sous ces trois catégories.

Facteurs organisationnels

Ces facteurs concernent les attributs internes de l'organisation : sa taille, sa structure, ses processus, ses ressources (financières et humaines), sa culture organisationnelle [NLV22, Bak12, SCG+24].

Traits personnels et comportement

Un grand nombre d'études se sont penchées sur la question de savoir si les caractéristiques démographiques influencent la vulnérabilité des employés à l'hameçonnage [JGST20]. Bien que les conclusions diffèrent selon les recherches, l'âge et le genre sont des attributs très souvent analysés dans les études portant sur le processus de décision [LKv22].

La maîtrise de l'informatique tout comme l'affinité générale des employés avec celui-ci jouerait un rôle déterminant dans la capacité à gérer les mails de phishing. Toutefois, une réserve est émise quant à l'affinité informatique du staff IT, celui-ci ne réduisant pas nécessairement leur vulnérabilité aux attaques [GDZ+25].

D'autres facteurs humains pourraient impacter le traitement des informations, comme le stress, l'attention, les distractions ou encore la fatigue [SHJL23]. Certains chercheurs ont également mis en avant la curiosité des individus à cliquer sur des liens [KPM24] ou à scanner des codes QR d'origine inconnue [SDPJ22]. La lumière a aussi été mise sur la prise de risque : les employés plus enclins à prendre des risques seraient plus susceptibles de cliquer sur des liens malicieux, tandis que ceux plus soucieux de la sécurité auraient tendance à ne pas cliquer, même sur des mails bénins [ADPL21, GDZ+25].

Un autre aspect est l'influence de la surcharge d'informations dans l'environnement de travail [BAAZ23] et la pression liée au temps [SHJL23]. La fatigue, le stress et le manque d'attention qui en résultent font souvent passer la sécurité au second plan chez l'utilisateur [GDZ+25].

Les mesures techniques instaurées par l'entreprise peuvent également créer un faux sentiment de sécurité et amener les employés à prendre plus de risques. Cet a priori résulte d'une compréhension partielle ou erronée des moyens déployés [GDZ+25]. L'auto-évaluation excessive de leur capacité à détecter les menaces, avant ou après la formation, peut porter préjudice à l'entreprise en induisant des comportements non sécurisés et non conformes aux bonnes pratiques [JDWT17].

Normes, politiques internes et éthique

Comme introduit dans la section 3.2, la mise en place de certaines normes et politiques au sein de l'entreprise peut aider à lutter contre l'hameçonnage en influençant le comportement des employés sans nécessiter de longues sessions d'entraînement [SHJL23].

La création d'une « culture cybersécurité » est largement étudiée. Elle aurait pour effet de

rendre les employés plus proactifs face aux menaces [PG24]. L'instauration de programmes de sensibilisation (PAT) aurait pour intérêt d'établir cette culture en communiquant efficacement les normes et les politiques de l'entreprise durant les sessions de formation [Als20].

Pour accroître la résilience à long terme, une formation de sensibilisation régulière est préconisée [GP23]. Elle permet de consolider les connaissances et les capacités acquises tout en tenant les employés à jour [GDZ+25].

Concernant la dimension éthique, il est important d'avertir le personnel avant le lancement d'une campagne de phishing afin d'éviter un climat de méfiance et prévenir l'inconfort [SHJL23]. Informer les régulateurs et le service informatique (IT) s'inscrit dans la même démarche pour éviter des réactions imprévues qui pourraient nuire au bon déroulement de l'exercice [GDZ+25]. L'obtention d'un consentement éclairé est donc essentielle pour garantir la compréhension des objectifs et des modalités [PG24]. Cela respecte les principes éthiques de transparence et de respect de la vie privée. De plus, les simulations de phishing constituent une zone grise en raison de la nature trompeuse des mails envoyés. Il est donc recommandé de faire examiner la formation par un comité éthique [GDZ+25].

Globalement, renforcer les compétences des employés en matière de cybersécurité est aussi un enjeu important pour l'éthique organisationnelle, car cela démontre un engagement envers la protection des données et la sécurité du personnel [ASD22].

Signalement des tentatives

Selon [GDZ+25], un système de signalement efficace constitue un élément crucial pour instaurer une culture « cybersécurité ». Il est donc essentiel d'enseigner aux employés comment traiter les mails suspects en instaurant des procédures de signalement claires.

Les outils mis en place incluent souvent une plateforme dédiée pour signaler les mails et recevoir des informations sur les menaces, ainsi qu'un bouton de signalement directement intégré aux programmes de messagerie [CDS+24, LKv22].

Tous ces éléments encouragent les employés à participer activement à la protection de l'entreprise en matière de cybersécurité et à accroître l'efficacité globale du système [CDS+24].

Facteurs techniques

La dimension technique traite des outils, des systèmes et des technologies mis en œuvre pour la création, l'établissement et la pérennité du PAT [NLV22, Bak12, SCG+24].

Personnalisation des formations

Les raisons évoquées pour l'adaptation des formations selon les individus sont diverses et variées. L'une d'entre elles porte sur le fait que chaque utilisateur n'a pas la même relation avec l'informatique, n'a pas le même rôle au sein de l'entreprise et n'est donc pas exposé de la même façon et à la même fréquence au phishing [SHJL23, JGST20]. Pour cela, les PAT proposent notamment diverses méthodes avec des scénarios à la difficulté progressive et auto-adaptative [SHJL23]. Une autre raison concerne les attaquants utilisant des informations

personnelles pour concevoir des e-mails de phishing ciblés (spear-phishing). Le but est de personnaliser le programme afin de préparer les utilisateurs à ce type de scénario [GDZ+25].

Mises à jour régulières

Dû à l'évolution constante des menaces, les méthodes, les outils et le matériel doivent pouvoir être mis à jour régulièrement. Les attaquants font évoluer leurs techniques afin d'outrepasser les défenses actuelles. Une sensibilisation aux nouvelles méthodes de phishing telles que le Smishing, le Vishing ou le Qishing devrait pouvoir être faite [BAJ25]. Ces techniques étant très récentes, trop peu de stratégies d'atténuation ont été développées, d'où la nécessité de tenir les utilisateurs informés [NPF+23].

Alertes et notifications

Les alertes et notifications sont des indicateurs visuels qui soutiennent l'utilisateur dans sa prise de décision et ajoutent une couche de protection supplémentaire. Toutefois, leur fréquence d'apparition doit être prise en compte, car une surabondance de ces signaux pourrait créer un phénomène d'habituation chez l'utilisateur et en diminuer l'utilité [DFM+22].

Facteurs environnementaux

Ce sont les facteurs externes à l'organisation elle-même qui peuvent influencer la décision et la mise en œuvre du PAT. Cela peut inclure les pressions réglementaires, le soutien des partenaires et la disponibilité des infrastructures externes [NLV22, Bak12, SCG+24].

Régulations

Évoquées dans la section 2.2, le PAT est soumis à des obligations légales. Les développeurs et les praticiens doivent donc prendre en compte les réglementations lors de la planification et de l'exécution des formations [SHJL23].

Par exemple, l'utilisation de marques pour la conception des mails factices est monnaie courante dans ces programmes. Le problème étant que celles-ci sont soumises à des droits d'auteurs et leur emploi peut entraîner des conséquences juridiques [GDZ+25].

Le droit du travail est également à prendre en considération selon les dispositions de l'entreprise en la matière. Suivant le type de données collectées, celles-ci doivent parfois passer par un processus d'anonymisation ou de pseudonymisation avant la phase d'analyse [GDZ+25].

Partenaires externes

Certaines entreprises choisissent de faire appel à des partenaires externes pour le développement et la diffusion du PAT [SLLK24, SHJL23]. Cela permet de réduire la charge de travail interne ; leur spécialisation dans le domaine apporte une expertise qui peut améliorer l'évaluation des connaissances ainsi que la qualité du matériel de sensibilisation [SHJL23].

Des problèmes ont toutefois été évoqués. Le partage d'informations sur les processus et procédures internes, les risques de fuites de données, ou encore les différences juridiques si le prestataire est basé dans un autre pays apportent d'autres risques à prendre en compte [GDZ+25].

2.3.3.5 Méthodes de partage des bonnes pratiques

Les organisateurs de PAT peuvent enseigner les bonnes pratiques de multiples manières différentes. Que ce soit au niveau de la transmission, du type de support ou du contenu, chaque méthode a ses propres caractéristiques qui lui confèrent des avantages et des inconvénients [Dar21]. Elles peuvent être catégorisées et hiérarchisées ainsi [Dar21, JDWT17] :

1. **Modes de transmission :**
   - Formation intégrée (Embedded-based)
   - Internet (Web-based)
   - Cours (Lecture-based)
   - Ateliers (Workshop-based)
2. **Types de support :**
   - Textuel et graphique (Text-based)
   - Vidéo (Video-based)
   - Vidéoludique (Game-based)
3. **Types de contenu :**
   - Règles (Rule-based)
   - Expériences passées (Story-based)

[Dar21] présente ces catégories comme des méthodes distinctes toutes équivalentes, mais ici, le choix a été fait de créer trois niveaux distincts où chaque niveau s'intègre à celui du dessus. Par exemple, une formation intégrée pourrait fournir des conseils sous forme de supports textuels et graphiques, avec un contenu organisé sous forme de règles.

Formation intégrée La formation intégrée au flux de travail se démarque de l'approche ponctuelle par son déroulement en continu et en conditions réelles. Cette méthode utilise généralement des e-mails de phishing simulés contenant des liens ou des pièces jointes. Si les destinataires réagissent incorrectement à ces e-mails en cliquant sur le lien ou en ouvrant la pièce jointe, ils sont redirigés vers un module de formation [GDZ+25].

Internet Les formations de sensibilisation basées sur le web (web-based) sont un moyen par lequel les organisateurs diffusent des conseils et des informations pratiques. Elles offrent souvent une flexibilité aux utilisateurs quant au moment et au lieu d'accès, ainsi que la possibilité de diffusion à large échelle facilement [Dar21].

Réunions Une autre manière de diffuser ces programmes est de faire recours à un instructeur. Contrairement aux autres méthodes qui ont la possibilité d'être dispensées virtuellement, les réunions sont effectuées en présentiel. Cette méthode peut se décliner en plusieurs formats [Dar21] :

La variante sous forme de cours (lecture-based) consiste en une leçon où les bonnes pratiques sont enseignées par l'instructeur. La connaissance provient donc de ce dernier.

Le format ateliers (workshop-based) place l'instructeur en tant qu'observateur d'un dialogue entre les participants. Ces derniers échangent sur leurs expériences passées. Ici, la connaissance est donc issue des participants eux-mêmes.

Supports textuels et graphiques Son faible coût et sa flexibilité d'accès, permettant une consultation selon l'emploi du temps de l'employé, font la force du support textuel et graphique (text-based).

Les bonnes pratiques sont présentées sous la forme de texte et/ou d'images et peuvent être transmises par e-mail ou par le biais de listes de recommandations, de fiches d'information, d'encadrés dans des formations en ligne ou des flyers [JGST20]. Un format qui a notamment été étudié est celui du « texte plus graphiques » combinant du contenu textuel avec des éléments visuels comme des images ou des illustrations. Il peut se présenter sous la forme de séquences d'images ou de bande dessinée [JDWT17].

Vidéos Un type de support relativement similaire au contenu textuel et graphique est celui vidéographique. Accessible en ligne comme son équivalent statique, il est également aisément diffusable à faible coût et consultable selon l'emploi du temps de l'utilisateur [CDS+24, JDWT17, GDZ+25].

Jeux vidéo Des études se sont penchées sur la « gamification » des méthodes de sensibilisation et certains chercheurs ont mis au point des jeux vidéo ayant pour but d'éduquer les utilisateurs en leur apprenant à détecter les différents indices que contient un mail d'hameçonnage. Voici quelques exemples :

AntiPhishing-Phil Jeu de rôle couvrant la thématique de l'attaque par manipulation d'URL. Le joueur incarne un poisson devant uniquement manger les vers contenant des liens légitimes [WLCA19].

What.Hack Propose d'incarner un employé de banque devant gérer sa boîte mail tout en évitant de se faire phisher. Les caractéristiques du phishing abordées sont les attaques par nom de domaine similaire, la manipulation d'URL et les pièces jointes malicieuses. Le jeu intègre des éléments de gameplay qui se veulent au plus proche du réel comme le flux de travail (workflow), la pression liée au temps, les interactions avec le support informatique et les effets néfastes du phishing [WLCA19].

PickMail Jeu solo en ligne consistant à gérer une boîte mail. Le but est de décider si les mails entrants sont de l'hameçonnage ou non. Contrairement à What.Hack, il faut indiquer les signes d'hameçonnage présents sur le mail lors du signalement [JBL22]. Le jeu évalue donc le processus de pensée du joueur.

PhishDefend Quest Jeu hybride qui utilise des cartes présentant des codes QR à scanner avec son smartphone. Le joueur est engagé par un hôpital pour simuler des attaques de

phishing afin d'éduquer le personnel et il peut faire usage d'informations trouvées grâce à de la recherche en sources ouvertes (OSINT) pour la conception des mails frauduleux. Il améliorerait la compréhension des menaces grâce à son système d'apprentissage interactif [YFJ+24].

UnPhishMe Ce n'est pas un jeu, mais une application mobile qui tire parti des faiblesses propres aux sites d'hameçonnage, tel que le formulaire d'authentification. Il permet aux utilisateurs de se connecter avec de faux identifiants et l'application détermine ensuite si la page de connexion est remplacée par une autre page web après la tentative d'authentification [NKF17].

Règles De nombreux programmes éducatifs utilisent des règles (rule-based) pour enseigner aux utilisateurs les éléments importants auxquels prêter attention [JDWT17, KPM24]. Les experts établissent des règles pour guider la réaction de l'utilisateur. Cette approche cadre la prévention du phishing comme une tâche d'identification d'indices spécifiques [JDWT17, Dar21].

Expériences passées Un autre type de contenu dans les programmes de sensibilisation est celui basé sur les histoires (story-based) [Dar21]. Le but ici est de faire usage de récits d'incidents comme moyen d'éducation pour les utilisateurs [WC18, Dar21].

2.3.4 Organismes et dispositifs

En plus des différents éléments mis en place par les organisations en elles-mêmes, les employés, mais également les particuliers, ont accès à un certain nombre de ressources en ligne.

2.3.4.1 En Suisse

Cette partie a pour objectif de passer en revue les principales plateformes publiques suisses dispensant des conseils, des entraînements, des instructions, des questionnaires ou encore des outils aux utilisateurs afin de les aider à identifier, prévenir et réagir face aux menaces liées au phishing. Elle ne prétend pas être exhaustive, mais elle cherche à fournir un aperçu représentatif des ressources disponibles, en mettant en lumière les initiatives les plus pertinentes et accessibles au grand public. L'objectif est de mieux comprendre comment ces plateformes contribuent à renforcer la sensibilisation et les compétences.

Il est également important de souligner que la plupart des plateformes ne se concentrent pas uniquement dans la lutte contre l'hameçonnage, mais offrent des éléments pour se défendre face à la menace plus globale qu'est le spam et la cybercriminalité en général.

2.3.4.2 ncsc.admin.ch

Sur le site de la confédération suisse, l'Office Fédéral de la Cybersécurité (OFCS) offre une page complète [10] consacrée à l'hameçonnage ainsi qu'à ses principaux dérivés : smishing et vishing. Celle-ci propose une définition ainsi qu'une courte description de chacun de ces phénomènes et fait état de mesures préventives et concrètes ayant pour but d'aider les utilisateurs à se protéger des menaces. Une section relative à l'incidence et aux risques encourus si des actions préjudiciables étaient entreprises est également présente.

Un formulaire d'annonces [11] est disponible et permet de signaler divers incidents, dont ceux relatifs au phishing.

2.3.4.3 antiphishing.ch

Outre les informations mises à disposition par l'OFCS, ces derniers ont créé antiphishing.ch [12] permettant aux particuliers de signaler de potentiels sites de phishing ainsi que d'envoyer des mails d'hameçonnage à une adresse dédiée, `reports@antiphishing.ch`, qui seront traités automatiquement.

2.3.4.4 ibarry.ch

ibarry.ch [13] est une plateforme d'information mise en place et entretenue par la Swiss Internet Security Alliance (SISA). Une rubrique dédiée propose une vidéo explicative ainsi qu'une liste d'éléments caractéristiques au phishing accompagnée d'un mail illustratif. Également, ibarry.ch dispose d'outils [14] permettant entre autres la vérification de sites web, la détection de fuites de données ou encore l'analyse de logiciels malveillants ou obsolètes.

2.3.4.5 ebas.ch (e-Banking en toute sécurité !)

Projet de la Hochschule Luzern (HSLU), e-Banking en toute sécurité ! [15] met à la disposition des visiteurs du site de courtes explications sur l'hameçonnage ainsi qu'un quiz à choix multiples basé sur d'anciens mails et d'anciennes plateformes liées au phishing.

10. `https://www.ncsc.admin.ch/ncsc/fr/home/cyberbedrohungen/phishing.html`
11. `https://www.report.ncsc.admin.ch/fr/`
12. `https://www.antiphishing.ch/fr/`
13. `https://www.ibarry.ch/fr/risques-sur-internet/phishing/`
14. `https://www.ibarry.ch/fr/controles-de-securite/`
15. `https://www.ebas.ch/fr/le-phishing/`

2.3.4.6 cybercrimepolice.ch

cybercrimepolice.ch [16] est une plateforme exploitée par la police cantonale zürichoise et collaborant avec de nombreuses institutions comme l'OFCS, certaines polices cantonales suisses-alémaniques ou encore la Prévention Suisse de la Criminalité. Elle a pour principal intérêt de recenser les cas liés à la criminalité en ligne ainsi que les phénomènes cybercriminels actuels.

2.3.4.7 vd.ch

Au niveau cantonal, vd.ch [17] dispense, en plus de conseils, une courte formation interactive élaborée par le Groupe Latin Sécurité (SIK/CSI). Elle se compose d'une première partie explicative et d'une seconde sous la forme d'un quiz.

2.3.4.8 Organisations victimes d'usurpation (SwissPass, La Poste...)

De nombreuses organisations [18, 19], dont le nom fait l'objet d'usurpation d'identité dans une grande part des mails de phishing visant les Suisses [fdlcO24], mettent à disposition sur leur site une section dédiée à la lutte contre le phishing proposant des conseils et des directives pour aider les utilisateurs à se protéger et à réagir efficacement en cas de fraude.

2.3.4.9 En France

Comme pour la Suisse, chaque pays dispose de ses propres plateformes de prévention et de signalement créées et organisées par le gouvernement ou des organisations tierces. Cette section a pour principal intérêt de mettre en lumière des éléments n'étant pas forcément déjà présents sur les plateformes helvétiques. Le choix de se concentrer sur la France s'explique par sa proximité géographique et linguistique, la disponibilité des sources en français et le fait que les outils français ont déjà été utilisés et testés au préalable.

2.3.4.10 signal-spam.fr

Créée en 2005, signal-spam.fr [20] est une association française à but non lucratif offrant une plateforme collaborative présentant du contenu informatif et éducatif sur les spams ainsi que

16. https://cybercrimepolice.ch/fr
17. https://www.vd.ch/portail-securise-des-prestations-en-ligne/securite-en-ligne/courrier-electroniqueet-phishing
18. https://www.swisspass.ch/phishing?lang=fr
19. https://www.post.ch/fr/notre-profil/responsabilite/la-securite-de-l-information-au-sein-de-la-poste/hameconnage-et-autres-tentatives-d-escroquerie
20. https://www.signal-spam.fr/

permettant aux internautes de signaler de potentiels pourriels. Les signalements sont ensuite pris en charge par les autorités ou des professionnels aptes à mettre en œuvre des actions tangibles.

Une des caractéristiques de signal-spam est le travail en étroite collaboration avec des fournisseurs de services Internet, des entreprises, les autorités publiques et d'autres associations professionnelles. Comme pour bien d'autres plateformes traitant du sujet, la section sur l'hameçonnage n'offre qu'une courte définition de la menace et un moyen de s'y prémunir. Ce dernier est une extension pour les principales boîtes mail disponibles sur le marché. Pour les navigateurs, les boîtes mail installées sur l'ordinateur et les smartphones iOS, ce module détecte, en s'appuyant sur une base de données, les liens frauduleux contenus dans les mails. Lorsqu'il décèle l'une de ces URLs, il affiche une bulle de signalement indiquant que le mail a été marqué comme frauduleux et offre plusieurs possibilités d'action à l'utilisateur :

- Un bouton de signalement permettant de faire remonter le courriel à signal-spam.
- Un bouton faux-positif indiquant à signal-spam que le lien détecté est en réalité légitime.
- Deux liens renseignant sur les contre-mesures déployées ainsi qu'un lien redirigeant vers le tableau de bord de l'utilisateur. Celui-ci comptabilise les mails signalés par l'utilisateur lui-même et affiche diverses statistiques relatives à ces derniers.

Une autre spécificité de ce plugin est que si le mail de spam provient d'un bulletin d'informations (newsletter) auquel l'utilisateur a souscrit, il propose la possibilité de se désabonner. Il est aussi possible de signaler des mails en copiant leur code source dans un formulaire dédié.

2.3.4.11 signal-arnaques.com

Également originaire de France, signal-arnaques.com [21] dispose de nombreuses sections couvrant diverses thématiques relatives à la criminalité en ligne. Il se démarque par ses deux sections analysant respectivement la fiabilité des sites web (SCAMDOC [22]) et des numéros de téléphone (SCAMTEL [23]). Ces outils attribuent un indice de confiance sous forme de pourcentage, sur la base de divers éléments comme l'âge du nom de domaine, le trafic du site, le type de numéro de téléphone ou encore si des signalements relatifs ont été faits.

2.3.4.12 cybermalveillance.gouv.fr

Comme le site de la confédération suisse, le gouvernement français dispose d'un site dédié au crime en ligne sous le nom de cybermalveillance.gouv.fr [24]. Il informe et éduque les utilisateurs sur le phishing, mais propose également une assistance en ligne permettant à ces derniers

21. `https://www.signal-arnaques.com/`
22. `https://fr.scamdoc.com/`
23. `https://fr.scamtel.com/`
24. `https://www.cybermalveillance.gouv.fr/`

d'établir un diagnostic afin de lever le doute. Le résultat de l'examen présente la menace et offre des recommandations sur les actions à entreprendre au niveau technique et au niveau légal. En plus de cela, la plateforme offre des prestations aux entreprises pour protéger leurs systèmes d'information et dispose de modules d'apprentissage enseignant, aux particuliers et aux employés de tous horizons, les bonnes pratiques et les méthodes pour se protéger.

2.3.5 Conclusion

Cette section dresse un état de l'art des mesures organisationnelles et légales existantes en matière de lutte contre le phishing. Sans prétendre proposer de nouvelles solutions, il vise à offrir une vue d'ensemble des méthodes et des dispositifs actuellement en place.

Du côté légal, il met en évidence les articles de loi applicables pour réprimer les actes de phishing ainsi que les bases juridiques servant de cadre et de fondement à la mise en place et le bon déroulement des méthodes de formation. Sur le plan organisationnel, il recense les bonnes pratiques, standards, politiques internes et outils de sensibilisation déjà employés et testés dans différents contextes professionnels, notamment à travers les programmes de Phishing Awareness Training (PAT), qui visent à renforcer les capacités des utilisateurs à reconnaître et réagir face aux tentatives de phishing.

Enfin, une attention particulière est portée aux ressources publiques disponibles en Suisse, qui jouent un rôle essentiel dans la diffusion de l'information, la prévention et le soutien aux victimes. Des ressources françaises sont également présentées à titre informatif, dans une optique comparative, afin de mettre en lumière certaines pratiques ou dispositifs existants ailleurs et qui pourraient inspirer des initiatives similaires en Suisse.

2.4 Retours de terrains

Dans le cadre de ce rapport, le *Center for Digital Trust* (C4DT) a conçu un questionnaire destiné aux partenaires du projet "Combattre le phishing : quelles innovations apporter ?". Compte tenu de leur diversité — Navixia et les *Transports publics de la région lausannoise* (T-L) du secteur privé, la *Police cantonale vaudoise* (PCV) et la *Direction générale du numérique et des systèmes d'information* (DGNSI) du secteur public, et l'EPFL ainsi que l'HEIG-VD des institutions académiques — nous avons estimé qu'ils donneraient un premier aperçu de l'ampleur du problème, et comment il est abordé par les différents acteurs de différents secteurs de la société suisse.

Suite à l'évaluation des réponses, nous avons élaboré quatre études de cas. La première se fonde sur les réponses de la PCV et montre l'impact de l'hameçonnage sur la société suisse. La deuxième étude présente, à partir du témoignage de l'EPFL, la situation dans le milieu académique. La troisième étude montre comment la DGNSI, responsable de la sécurité informatique des services publics du canton de Vaud, met en œuvre des mesures

de protection. Finalement, la quatrième étude explique comment Navixia met en place une défense proactive à travers des formations.

2.4.1 Étude de cas : Police cantonale vaudoise

2.4.1.1 Situation sur le terrain

L'hameçonnage représente 10,3 % des 5 013 cas de criminalité numérique enregistrés dans le canton de Vaud en 2024, avec une augmentation de 41 % par rapport à l'année précédente [cv24]. D'ailleurs, la PCV précise que ces pourcentages sont très probablement sous-estimés car l'hameçonnage est souvent à l'origine d'autres infractions signalées sans que les victimes s'en rendent compte, et ne sont donc pas toujours systématiquement enregistrés.

La PCV différencie entre uniquement l'hameçonnage et l'hameçonnage lié à d'autres phénomènes. L'hameçonnage par e-mail, SMS ou téléphone constitue la plus grande partie des escroqueries dans la première catégorie, tandis que dans la deuxième catégorie, ce sont les cas "faux acheteur" et "faux support technique" qui dominent.

Dans la catégorie "uniquement hameçonnage", nous retrouvons l'hameçonnage via e-mail ou SMS, ainsi que l'utilisation de l'ingénierie sociale par téléphone, qui ont pour but de récupérer des données sensibles — comme p.ex. des numéros de cartes de crédit — des victimes.

L'escroquerie "faux acheteur" repose sur une annonce de vente publiée par le·la plaignant·e sur une plateforme dédiée. Sous prétexte d'être intéressé·e, le·la fraudeur·euse contacte la victime et l'amène à divulguer ses coordonnées bancaires, que ce soit par un formulaire d'hameçonnage ou par téléphone. Des variantes existent, notamment l'utilisation des comptes Twint à la place des comptes bancaires traditionnels.

À l'inverse, l'arnaque "faux support technique" manipule la victime pour qu'elle prenne elle-même contact avec l'auteur·rice. Dans ce but, le navigateur de la victime est bloqué par une fenêtre de dialogue qui l'incite à appeler un faux numéro de support pour résoudre une supposée infection. Pour payer cette "prestation", elle est alors dirigée vers un formulaire de phishing ou poussée à se connecter à son e-banking.

2.4.1.2 Prévention

La sagesse populaire affirme que la prévention est la meilleure protection. À ce titre, la police vaudoise propose des conseils de prévention sur leur site [Pol25b]. Notamment les guides "Les mécanismes psychologiques chez les cybercriminelles et cybercriminels" [cv25b] et "Démarchage téléphonique ou escroquerie téléphonique ?" [cv25a] abordent les arnaques numériques et le phishing. De plus, une newsletter [Pol25a] et un magazine trimestriel [cv25c] traitent aussi des sujets de criminalité numérique.

2.4.1.3 Enquête et élucidation

L'enquête sur les cas d'hameçonnage est entravée par plusieurs facteurs. Il y a un sous-rapportage des cas, notamment parce que les victimes ne signalent souvent que l'abus de leurs coordonnées bancaires, sans faire le lien avec le phishing initial. Les statistiques ne reflètent donc pas la réalité du terrain.

Étant donné que l'hameçonnage cible souvent les finances, la PCV souligne aussi des difficultés de collaboration avec les banques, en particulier en dehors des heures de bureau, ainsi que l'absence de procédures standardisées.

2.4.2 Étude de cas : EPFL

2.4.2.1 Situation sur le terrain

L'EPFL accueille une population diversifiée de près de 18 000 étudiant·e·s et collaborateur·rice·s pour lesquel·le·s l'école doit assurer la sécurité numérique. Cette hétérogénéité, caractérisée par des différences culturelles, un large éventail de compétences techniques et de compréhension des lois suisses pertinentes, ainsi qu'une large gamme de systèmes à protéger, rend la tâche particulièrement complexe [EPF25]. De plus, une tension supplémentaire est le besoin d'équilibrer les exigences de sécurité avec la liberté académique.

Les types d'hameçonnage les plus fréquemment rencontrés par la population d'EPFL sont le "bulk phishing" et le "spear phishing". Le "bulk phishing" n'a pas de cible précise ; une grande partie de la population ou même toute la population est ciblée. Par conséquent, même si les messages donnent l'impression d'être légitimes, par exemple en utilisant l'identité d'entreprise d'une grande entreprise connue, ils ne sont pas individualisés. Le "spear phishing" de l'autre côté cible des personnes précises, et les messages sont en général sur mesure, ce qui les rend très difficile à détecter [IBM25b].

2.4.2.2 Prévention

Au-delà des mesures techniques de détection et de filtrage des e-mails frauduleux, l'EPFL sensibilise ses utilisateur·rice·s au sujet de l'hameçonnage afin de minimiser le risque de succès de ces attaques. Cette sensibilisation est actuellement assurée par une formation obligatoire pour les nouveaux·elles arrivant·e·s. Elle prend la forme de vidéos d'information sur la sécurité numérique, y inclut l'hameçonnage, suivie d'un quiz pour déterminer la compréhension du matériel. Pour compléter cette formation de base, l'EPFL prévoit de mettre en place des campagnes de formation. Ces simulations d'attaques d'hameçonnage permettent d'évaluer les compétences des utilisateur·rice·s dans l'identification des e-mails frauduleux et à renforcer leur vigilance.

2.4.3 Étude de cas : DGNSI

2.4.3.1 Situation sur le terrain

La DGNSI est responsable pour la sécurité numérique de l'infrastructure du canton de Vaud et les 13 500 utilisateur·rice·s associés à l'État. Notamment, leur responsabilité inclut les services d'urgence, les données des différents services de l'État, et des données personnelles des habitant·e·s du canton.

En outre, la DGNSI protège les institutions du canton de Vaud de l'hameçonnage. Des exemples incluent des e-mails imitant des documents Office 365 qui redirigent les utilisateur·rice·s vers une frauduleuse page de connexion Microsoft, des e-mails usurpant l'identité d'un service de livraison de colis, tel que La Poste ou DPD pour récupérer les données de cartes de crédit, ou même, plus récemment, des e-mails présumés de CFF ou SwissPass qui tentent d'obtenir les identifiants SwissPass.

2.4.3.2 Prévention

Chez la DGNSI, la prévention des cyberattaques, y compris l'hameçonnage, repose sur quatre piliers : la conformité avec les lois et les normes pertinentes, la collaboration avec les autres parties prenantes, des solutions techniques et enfin la formation des utilisateur·rice·s.

En Suisse, la Loi sur la sécurité de l'information (LSI) a pour but de sécuriser les traitements d'informations sous la responsabilité de la Confédération ainsi que ses moyens informatiques, et vise à renforcer la résilience de la Suisse face au cybermenaces [Cha23]. La série de normes ISO/IEC 27001 quant à elle fournit des lignes directrices pour la mise en place des systèmes de gestion de la sécurité de l'information et de la gestion des informations confidentielles ainsi que des bonnes pratiques à adopter [IBM25a]. En se mettant en conformité avec la LSI et l'ISO/IEC 27001, la DGNSI établit les fondations d'une prévention efficace des cyberattaques.

Le deuxième pilier est l'étroite collaboration avec les autres parties prenantes, c'est-à-dire la Confédération et les autres cantons latins.

Le *Security Operations Center* (SOC) de la DGNSI monitore l'infrastructure sous sa responsabilité vingt-quatre heures sur vingt-quatre, sept jours sur sept. À cette fin, les différents composants de leur pile sécurité envoient leurs journaux au système *Security Information and Event Management* (SIEM), qui les analyse et les corrèle pour identifier des anomalies. Pour faire face à l'hameçonnage en particulier, la DGNSI s'appuie de plus sur un système de détection automatique de pourriel qui tente aussi de repérer des tentatives d'hameçonnage. Les solutions techniques que la DGNSI déploie pour assurer la sécurité informatique de ses collaborateur·rice·s comprennent un VPN pour les télétravailleur·euse·s, des mots de passe renforcés par l'authentification à facteurs multiples (MFA), le chiffrage automatique

des disques durs, et l'interdiction de Bring Your Own Device (BYOD).

Le dernier pilier sur lequel la DGNSI s'appuie pour la prévention des cyberattaques est la formation des utilisateur·rice·s. La formation se déroule en deux parties : la première partie consiste en des modules de formation sur la protection des données, le secret de fonction et de cybersécurité qui sont suivis lors de l'arrivée d'un·e nouveau·elle collaborateur·rice. La deuxième partie est constituée d'exercices d'hameçonnage annuels, complétés par une phase pédagogique pendant laquelle les participant·e·s recoivent des matériaux e-learning. Les résultats des exercices d'hameçonnage sont ensuite publiés.

2.4.3.3 Remédiation

Une attaque d'hameçonnage est soit détectée par le SOC, soit signalé directement par un·e utilisateur·rice auprès du helpdesk. Dans le dernier cas, le helpdesk informe le SOC à son tour, qui évalue la gravité de l'incident. En général, les attaques de l'hameçonnage sont traités comme des simples incidents.

Une fois qu'une attaque d'hameçonnage est confirmée, une analyse de l'attaque est faite pour déterminer son origine, son étendue et les utilisateur·rice·s touché·e·s. Des mesures de remédiation sont alors mises en œuvre. Une des mesures est l'éradication de la menace. À cette fin, l'expéditeur·rice et les domaines malveillants liés à l'attaque sont bloqués. De même, les messages liés à l'attaque sont supprimés des messageries des utlisateur·rice·s.

En cas d'attaque d'hameçonnage réussie ou de doute, le mot de passe de l'utilisateur·rice concerné·e est réinitialisé, des messages ainsi que des règles suspects sont supprimés de sa messagerie, et les partenaires sont avertis des potentiels messages frauduleux. Un retour d'expérience complète la remédiation.

2.4.4 Étude de cas : Navixia

Navixia est une société de conseil en sécurité numérique. Une de ses offres est la formation de sensibilisation à l'hameçonnage. Contrairement aux études de cas précédentes, cette étude se concentre spécifiquement sur ce sujet.

2.4.4.1 Déroulement

Navixia propose la plateforme DiagnoPhish pour les formations de sensibilisation à l'hameçonnage. Cette plateforme offre à la fois des formations théoriques et la possibilité de simuler des campagnes d'hameçonnage et permet ainsi une formation continue des employé·e·s [Nav25]. Navixia utilise DiagnoPhish aussi bien pour les formations auprès de sa clientèle, que pour ses propres formations internes, qui se déroulent sous forme de campagnes d'hameçonnage mensuelles. À la fin d'une campagne d'hameçonnage, le pourcentage d'utilisateur·rice·s

ayant cliqué sur les liens “frauduleux“ est évalué, ainsi que le pourcentage d'utilisateur·rice·s ayant signalé la vraisemblable tentative d'hameçonnage.

2.4.4.2 Lessons learned

Au fil des formations effectuées auprès de sa clientèle, Navixia a tiré des nombreuses leçons sur ce qui constitue une défense durable contre l'hameçonnage. Une des plus importantes de ces leçons est que la culture de l'entreprise joue un rôle essentiel : si les employé·e·s se sentent en sécurité de signaler leurs inquiétudes, y compris leurs “fautes“, la sécurité est renforcée. Par contre, si les employé·e·s ont peur de le faire, la sécurité est compromise. La deuxième leçon la plus importante est que les formations à long terme, par la répétition et l'automatisation, sont le meilleur moyen de maintenir un niveau de sensibilisation élevé et une vigilance constante. Finalement, l'engagement du management est un facteur essentiel au succès de la formation des employés.

2.4.5 Étude de cas : vishing

Cet étude de cas concerne l'hameçonnage par téléphone ou message vocal, appelé *vishing* (contraction de *voice* et *phishing*). Cette forme d'hameçonnage a connu une croissance inquiétante de 442 % dans la deuxième moitié de 2024 par rapport à la première moitié de l'année [Cro25]. Il est donc pertinent d'inclure une analyse plus détaillée dans ce rapport.

2.4.5.1 Mode opératoire

Le mode opératoire du *vishing* est très similaire à celui du *phishing* via e-mail. Une fois que l'attaquant·e a acquis les coordonnées de la victime, celle-ci est contactée sous un faux prétexte pour l'inciter à effectuer des actions bénéficiant à l'attaquant. La différence principale est que l'interaction avec la victime peut avoir lieu en temps réel [fdlc22].

2.4.5.2 Aspects techniques

La présentation du numéro (*caller ID*) se compose de deux numéros de téléphone de l'appelant (celui qui est affiché au destinataire, et celui qui est partagé avec le fournisseur du service), le *calling line ID* (CLID ou CLI), et, si la région ou le pays le supporte, le *calling party name* (CNAM), le nom associé à ce numéro de téléphone [Mic25b]. Comme pour les e-mails, avoir un numéro de téléphone qui semble fiable à la victime est le premier obstacle à surmonter. L'attaquant·e a donc intérêt à falsifier ces informations.

La présentation du numéro a été mise sur le marché dans les années quatre-vingt, et l'usurpation de la présentation du numéro (*caller ID spoofing*) à grande échelle a suivi dans les

années 2000. Aujourd'hui, les services VoIP ou "voix sur IP", c'est-à-dire la transmission de la voix sur les technologies Internet, ont rendu l'usurpation de la présentation du numéro triviale puisqu'il est souvent possible de configurer le numéro à afficher [aut25].

2.4.5.3 Défenses

Dans leur évaluation de l'authentification du CLI - que nous allons discuter de plus près - l'Ofcom, l'autorité régulatrice des télécommunications au Royaume-Uni, décrit d'autres mesures de défense typiques, dont la majorité consistent à bloquer les numéros usurpés du côté des fournisseurs.

Une défense répandue consiste en des listes appelées "Do not originate". Elles contiennent des numéros fréquemment usurpés, comme des numéros publics de services publics ou de banques qui, en général, ne reçoivent que des appels, mais ne passent pas d'appels eux-mêmes. Ces listes sont fournies aux fournisseurs de services qui bloquent les appels sortants semblant provenir de ces numéros.

L'authentification du CLI est implémentée par exemple par les standards STIR/SHAKEN. Ces standards reposent sur la cryptographie asymétrique et les certificats électroniques pour ajouter des informations sur l'origine d'un appel. Le fournisseur de départ vérifie le numéro de téléphone et sa fiabilité et crée une en-tête avec ces informations. Cette en-tête est chiffrée - l'appel est dit signé - et est envoyé avec l'appel au fournisseur de destination. Cet fournisseur obtient alors la clé publique du fournisseur de départ ainsi que le certificat associé de l'autorité de certification, pour lui permettre de vérifier à la fois que les informations sur l'appel qu'il reçoit n'ont pas été manipulées en transit ainsi que le taux de fiabilité [Trale].

Cependant, l'Ofcom note que cette solution implique des coûts considérables, surtout en termes de charge supplémentaire sur les réseaux non-IP, tout en n'étant pas une solution universelle, et a donc recommandé de ne pas poursuivre STIR/SHAKEN. Des réserves similaires sont probablement la raison pour laquelle STIR/SHAKEN est très peu adopté en dehors des États-Unis et du Canada [Ofc24].

2.4.6 Étude de cas : smishing

Nous allons présenter une étude de cas sur l'hameçonnage via SMS, ou *smishing*. Par rapport au deuxième trimestre de 2024, les attaques de ce type ont augmenté de 22 % [APW24]. Nous allons examiner de plus près ce que sont les aspects techniques du *smishing*, ses caractéristiques, ainsi que les raisons pour lesquelles ces attaques sont en croissance.

2.4.6.1 Mode opératoire

Tout comme avec le *vishing*, l'attaquant·e a intérêt à fournir des informations plausibles sur l'origine du message.

2.4.6.2 Aspects techniques

SMS, ou *Short Message Service*, fait partie des standards GSM (*Global System for Mobile Communications*). Un SMS peut être envoyé entre téléphones mobiles ou entre une application et un téléphone mobile, dans les deux sens. Un exemple courant de cette deuxième situation est l'authentification à deux facteurs via SMS [SMP19].

Dans le cas de l'envoi entre deux téléphones mobiles, c'est le fournisseur qui détermine cette information en fonction de l'IMSI (*international mobile subscriber identity*) associée à la carte SIM du téléphone. Les possibilités d'usurpation sont par conséquence limitées pour l'attaquant·e [Tsu24].

Le protocole standard pour l'envoi de SMS entre applications et téléphones mobiles est le *Short Message Peer-to-Peer Protocol* (SMPP). Ce protocole permet à l'expéditeur·rice de renseigner librement l'information sur l'origine du message, et si le fournisseur n'a pas de moyens de détection d'usurpation adéquats en place, même des numéros existants peuvent être utilisés sans que cela empêche l'envoi du message [Tsu24] (voir figure 2.14).

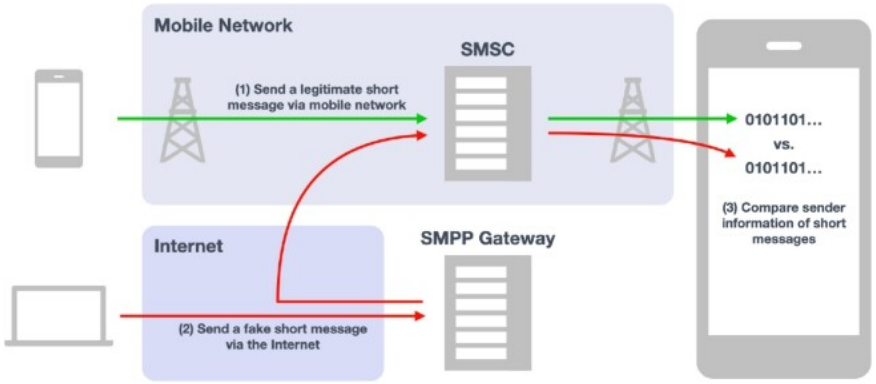

FIGURE 2.14 – Verification of spoofability of the originating number : (1) a legitimate short message sent via mobile network ; (2) a fake short message sent via the Internet ; (3) comparison of the sender information. [Tsu24]

2.4.6.3 Défenses

Il existe plusieurs défenses contre le *smishing* qui peuvent être mises en place par les fournisseurs. Par exemple, le filtrage des SMS tente d'identifier et de supprimer les SMS contenant un contenu douteux. D'autres moyens de défense peuvent inclure la limitation du nombre de SMS qu'un seul numéro peut envoyer en un temps donné, ou le nombre de cartes SIM pouvant être achetées par transaction [Ofc22].

2.4.7 L'IA, un défi émergent

Dans leur Digital Defense Report 2025, Microsoft dévoile des détails alarmants sur l'impact de l'IA sur la cybersécurité qui grandit à mesure que la société, et les acteur·rice·s de menace deviennent compétent·e·s dans l'utilisation de cette technologie émergente. Concernant l'hameçonnage, la situation est particulièrement préoccupante : il ne se limite pas à la génération d'e-mails plus convaincants, mais touche chaque étape d'une attaque d'hameçonnage.

L'hameçonnage dans sa forme étendue qui est sujet de notre rapport se déroule à plusieurs niveaux. Par exemple, le contenu d'un e-mail ou SMS, un appel téléphonique ou le site web vers lequel la victime est dirigée. L'IA, et notamment l'IA générative, touche à tous ces aspects.

Un premier exemple d'abus de cette technologie est la création et l'envoi des e-mails ou de SMS d'hameçonnage. Ici, l'impact de l'IA est particulièrement dévastateur : le taux de clic, qui mesure combien d'utilisateur·rice·s ayant reçu le lien ont également cliqué dessus, passe de 12

Les liens qui sont envoyés aux victimes doivent paraître légitimes, et mener vers des sites qui ressemblent le plus possible aux sites ciblés. L'IA est utilisée dans ces deux cas.

D'abord, l'IA aide à améliorer le cybersquattage. Cette technique permet à l'attaquant·e d'acquérir des noms de domaines qui sont très proches du nom de domaine du site légitime. Cela est fait soit en exploitant des fautes de frappes, caractères similaires, ou en ajoutant des mots ou sous-domaines. L'IA permet d'automatiser ce processus et de tester les résultats contre des défenses standards. L'IA permet aussi de créer facilement et à grande échelle des sites web contrefaits qui sont difficiles à distinguer de leur contrepartie légitime.

Finalement, l'IA est utilisée pour créer des hypertrucages des voix ou même des images de personnes pour, par exemple, mener à bien un hameçonnage reposant sur un faux support technique.

En conclusion, le rapport de Microsoft montre que l'IA a un impact non négligeable sur l'hameçonnage. Le passage à l'échelle, l'impressionnante amélioration du taux de clic et les hypertrucages posent des défits inédits dans une situation qui étaient déjà tendue avant l'arrivée de cette technologie [Mic25d].

Chapitre 3

Ateliers d'échange

Table des matières du chapitre

Les ateliers d'échange avaient pour objectif de rassembler l'ensemble des partenaires impliqué·e·s dans le projet afin de réfléchir collectivement à des solutions innovantes contre le phishing. Ces ateliers ont permis de croiser les regards et de bénéficier de l'expertise variée de chaque participant·e, issu·e·s de domaines complémentaires. Cette démarche collaborative visait à identifier non seulement les mesures actuellement efficaces, mais aussi à comprendre les limites des approches existantes et à en analyser les causes.

Chaque atelier était structuré autour d'un objectif précis, permettant d'avancer étape par étape : d'abord établir un état des lieux partagé, puis générer des idées, les affiner, et enfin en sélectionner certaines pour les développer. Ce format a permis d'ancrer les réflexions dans des situations concrètes, tout en laissant la place à la créativité.

3.1 Synthèses des ateliers

D'une manière générale, les ateliers ont été une expérience très enrichissante pour chacun·e des participant·e·s, représentant les différents partenaires. Les échanges étaient extrêmement agréables et constructifs. Les partenaires sont entrés très rapidement dans une dynamique de partage et d'échanges qui a été très bénéfique. Chacun·e a pu échanger sa vision, ses expériences, ses besoins, tout en confrontant ses idées aux regards complémentaires.

3.1.1 Premier atelier : état des lieux et identification des failles

Le premier atelier s'est tenu le 5 décembre 2024. Même si certaines tâches ont été réalisées au préalable, il a également été l'occasion de faire le kick-off officiel du projet. Les différents partenaires ont pu tous se rencontrer et démarrer les échanges.

3.1.2 Ordre du jour prévu initialement

1. Accueil et présentation des personnes / partenaires (10min)
2. Introduction au projet par Sylvain Pasini (20min)
3. Présentation des travaux sur l'état de l'art (45min)
   - Présentation des mesures techniques
   - Présentation des mesures organisationnelles et légales
4. État de la situation actuelle (45min)
   - Retours de terrain
   - Points défaillants
5. Discussion des prochains ateliers, synthèse et documents
6. Divers

3.1.2.1 Déroulement de l'atelier

L'atelier a été ouvert par le porteur de projet. Il a débuté par un rappel des objectifs du projet, de la coordination globale des partenaires, reprenant les points importants sur le déroulement du projet, la planification et autres aspects.

L'atelier a ensuite permis de partager l'état de l'art réalisé au préalable. Les présentations ont permis d'expliquer aux autres partenaire les mesures actuelles de lutte contre le phishing, tant sur le plan technique (HEIG-VD) qu'organisationnel et légal (Unil). Voir les détails aux Sections 2.2 et 2.3 respectivement. Ces exposés ont été complétés par des retours de terrain (EPFL, DGNSI, PCV), mettant en lumière la réalité des incidents, modes opératoires et les réactions observées. Voir les détails à la la Section 2.4.

Une discussion collective a ensuite permis d'identifier des failles récurrentes, notamment liées à des erreurs de configuration, à la faible responsabilisation des utilisateur·rice·s, à des situations de peur ou de honte d'annonce, ainsi qu'à l'adaptation rapide des attaquant·e·s.

Les participant·e·s ont été invité·e·s à réfléchir en groupes à partir de leur propre expérience sur des thèmes ciblés : d'autres mesures techniques, la responsabilité des utilisateur·rice·s, des exemples concrets d'attaques et les limites des formations actuelles.

3.1.2.2 Points clés

- Les mesures techniques (SPF, DKIM, etc.) sont utiles et efficaces pour protéger un domaine. Elles ne sont en général pas contrournées directement.
- Ces mesures sont cependant insuffisantes, car facilement contournées d'un point de vue humain, sur un autre domaine (ex. : typosquatting avec un domaine bien configuré).
- La sensibilisation reste essentielle, mais elle est souvent mal calibrée ou peu engageante.
- La PCV indique que de manière générale les incidents (annoncés) concernent principalement les particulier·ère·s.
- La peur du blâme limite les signalements d'incidents, notamment lorsqu'une erreur a été commise.
- Il existe un besoin clair d'impliquer davantage les utilisateur·rice·s dans la sécurité, en les responsabilisant sans les culpabiliser.
- Il peut être intéressant d'encourager les victimes à annoncer, et les cibles à signaler et rapporter les données.
- Il peut être délicat de développer un système d'annonce, d'aide ou d'analyse, car ce dernier nécessite de récupérer les données de l'email, ce qui peut poser des problèmes de confidentialité ou de sphère privée. Une remontée de données hashée pourrait être une idée à creuser.

3.1.3 Deuxième atelier : génération d'idées innovantes

Ce deuxième atelier s'est déroulé le 12 février 2025. Il avait pour objectif principal de produire des idées concrètes pour améliorer la lutte contre le phishing.

3.1.4 Ordre du jour prévu initialement

1. Brainstorming sur des idées (en groupe de 3) (30min)
2. Présentations, retours, discussions sur les idées (30min)
3. Mise en commun et structuration des idées (20min)
4. Votation et sélection des 3 meilleures idées (10min)
5. Approfondir 3 idées (en 3 groupes) : détailler le fonctionnement, UX, besoins externes, évaluation du temps de dev + tests (30min)

3.1.4.1 Déroulement de l'atelier

Après un bref rappel des mesures existantes et de leurs limites, les participant·e·s ont été réparti·e·s en quatre groupes de travail. Chaque groupe devait produire autant d'idées que possible, les noter, structurer.

Par la suite, une restitution a été faite et les idées partagées et discutées collectivement. Ci-dessous, on retrouve une brève descriptions des idées principales. Une description plus longue, ainsi que des illustrations se situent à la Section 3.2

Plugin Outlook d'aide et de signalement. Un plugin Outlook permettant de signaler facilement des e-mails suspects. Le plugin pourrait également permettre en cas de doute d'obtenir un retour sur leur dangerosité, avec un code couleur clair.

Outils de vérification de configuration des serveurs mail. Construire ou recommander des outils permettant de vérifier la configuration des serveurs de messagerie, visant surtout les PME et leurs administrateur·rice·s à détecter les faiblesses de configuration et les corriger.

Application mobile/desktop de prévisualisation de liens. L'idée serait d'attraper les cliques sur des liens (email, réseaux, SMS, etc.). Un système de prévisualisation des liens permettrait à l'utilisateur·rice de voir le véritable domaine derrière le lien. Cela lui permet par exemple de savoir après avoir cliqué sur `ubs.support.com` qu'il se connecte sur `support.com` et non pas sur `ubs.com`.

Système TLP de coloration des liens. Un système TLP [1] (code couleur) pour marquer la fiabilité des liens, directement dans les e-mails. Par exemple, rouge indiquerait que le lien a déjà été répertorié comme phishing avéré, orange comme suspect et bleu comme n'ayant rien trouvé de suspect (mais n'indique pas pour autant que c'est sûr).

1. `https://en.wikipedia.org/wiki/Traffic_Light_Protocol`

Sensibilisation. Des modules d'apprentissage et de sensibilisation pour identifier les liens suspects.

Protections intégrées aux gestionnaires de mots de passe. Un exemple d'application serait le suivant : si l'utilisateur pense être sur `ubs.com` mais se trouve en réalité sur `ubs.net`, l'auto-complétion du gestionnaire ne s'active pas. Pour prévenir le risque qu'il colle tout de même son mot de passe manuellement, il faudrait bloquer ou alerter explicitement lors de cette action. L'idée de ce projet serait que le password managers puisse proposer une protection intégrée pour alerter l'utilisateur·rice en cas de tentative de saisie sur un site de phishing.

Plateforme régionale. Cette idée rassemble différentes propositions sous une seule et unique : une plateforme de sensibilisation. Elle rassemblerait toutes les informations et service en un seul endroit, ce que nous n'avons pas en Suisse.

Elle devrait permettre de : *former, informer, annoncer, aider et assister (levée de doute), hoax, supporter en cas d'incident, etc.* Elle contiendrait des contenus ludiques (serious games, quiz, actualités locales sur le phishing) pour aider le public à mieux repérer les arnaques.

Un autre aspect de cette plateforme était de proposer aux utilisateur·rice·s un moyen de soumettre l'e-mail sur lequel ils auraient un doute afin d'avoir un retour sur la dangerosité de celui-ci. Un dernier point était de proposer des guides sur les outils à adopter et comment les utiliser pour guider l'utilisateur·rice à travers la multitude de moyens proposés sur Internet.

Autres canaux (SMS, voix et QR) Il ressort que la problématique des emails est déjà bien couverte en sensibilisation. L'idée serait d'élargir et d'avoir un volet sur le smishing, vishing et le quishing, pour élargir la sensibilisation vers les SMS, appels téléphoniques ou QR codes.

Il était initialement prévu de voter et d'élire les idées les plus prometteuses. Cependant, suite à la discussion autour des idées innovantes, il est rapidement apparu qu'il serait nécessaire de trouver des critères afin de départager les idées.

Un ensemble de critères a donc été défini pour anticiper et affiner leur évaluation future, dans l'objectif de pouvoir éliminer ou retenir entre 1 et 3 idées prioritaires. Les critères discutés étaient les suivants :

- Degré d'innovation
- Faisabilité technique
- Faisabilité vis-à-vis des dépendances externes (données, logiciels, etc.)
- Impact potentiel (social, privé, industriel)
- Facilité d'adoption
- Faisabilité budgétaire
- Pérennité et maintenance

3.1.4.2 Points clés

- **8 idées principales** ont été générées, avec une forte diversité (technique, UX, sensibilisation).
- **Aucune idée n'a été rejetée à ce stade** : elles seront toutes évaluées objectivement lors de l'atelier suivant.
- Des **critères d'évaluation clairs** ont été définis : innovation, faisabilité, impact, adoption, dépendances, maintenance. Ils aideront à analyser la pertinence des idées et sélectionner les plus adéquates pour la suite du projet.

3.1.5 Troisième atelier : analyse et pré-sélection des idées

Ce troisième atelier s'est déroulé le 10 mars 2025. L'objectif était de pouvoir analyser et ne retenir que les idées les plus adéquates pour la suite du projet. L'objectif derrière était de pouvoir lancer le développement des démonstrateurs (PoCs) et de planifier la suite du projet.

3.1.6 Ordre du jour prévu initialement

1. Présentation et rappel des solutions envisagées (15min)
2. Discussion des critères d'évaluation (15min)
3. Évaluation des idées (30min)
4. Elimination d'idées (non réaliste, hors budget, sans impact, existant déjà) (15min)
5. Approfondissement des idées (fonctionnement, UX, besoins externes, temps de dev, faisabilité PoC) (30min)
6. Sélection des meilleures idées (30min)
7. Planification et discussion concernant la suite (30min)

3.1.6.1 Déroulement de l'atelier

Lors de ce troisième atelier, les idées issues du brainstorming ont été évaluées à l'aide des critères établis. Les partenaires ont été répartis en groupes pour évaluer chaque idée selon les critères. Un tableau de scores a permis de guider la discussion. Chaque idée a ensuite été discutée collectivement pour décider de la suite.

Le fait de travailler en groupe rendait l'analyse plus objective, sans que certains a priori se répercutent de manière trop forte sur une idée. Ensuite, les notes de chaque groupe ont été mises ensemble pour former une évaluation finale.

Chaque critère faisait l'objet d'une notation de 1 à 5. Par exemple, le degré d'innovation était représenté par 1 très faible, 2 faible, 3 moyen, 4 élevé, 5 très élevé. De même, la

faisabilité technique était évaluée par 1 non faisable, 2 probablement non faisable, 3 à évaluer, 4 réalisable, 5 réalisable facilement.

Il est important de noter que les critères ne peuvent pas être pondérés. En effet, même si une idée à une note très élevée mais n'est pas faisable, elle devra être éliminée. D'autres scénarios de ce type sont possibles et l'évaluation finale nécessitera donc une analyse détaillée des critères et non de la note globale.

	Plateforme régionale	Plugin Outlook d'aide et de signalement	Canaux SMS/Voice/QR	Outils de vérification de configuration des serveurs mail	Application de prévisualisation de liens.	Système TLP de coloration des liens	Sensibilisation liens	Protections intégrées aux gestionnaires de mots de passe
État de l'art / innovation	3	3	5	2	4	4	2	4
Faisabilité technique	5	3	3	5	3	3	5	3
Dépendances externes	3	3	3	5	4	4	5	2
Faisabilité budgétaire	5	4	4	5	4	4	5	3
Impact social	4	3	4	2	4	5	3	3
Facteur d'acceptation	4	4	3	3	3	4	3	3
Maintenance et futur	3	3	4	4	4	4	4	2
Total	27	24	27	26	26	28	27	20

FIGURE 3.1 – Résumé des notes attribuées par l'ensemble des partenaires

A la suite de cette évaluation, certaines propositions se démarquaient nettement, en vert et rouge. En revanche, les autres ont été longuement débattues. En effet, pour certaines, il était difficile de trancher car nous avions des doutes sur la faisabilité technique, mettant la

proposition directement en élimination, ou non.

Au final, les idées ont été classées en trois catégories :

- **Idées rejetées** :
  - Protection dans les gestionnaires de mots de passe : jugée irréaliste à implémenter, il est plus judicieux de laisser les entreprises développant ces applications s'en occuper.
  - Outils de vérification de serveurs mail : existent déjà, intégrable via les guides réalisés durant la valorisation et les périodes de sensibilisation.
- **Idée retenue directement** :
  - Sensibilisation aux liens : ce concept est déjà pris en compte initialement lors de la déposition du projet. Cette sensibilisation fera partie de la valorisation.
- **Idées en attente d'étude de faisabilité** :
  - Prévisualisation de liens : très bien notée, mais incertitude technique. Il n'était pas sûr de ce qui était possible de faire sur chaque système d'exploitation cible. Est-ce qu'il était réellement possible d'intercepter les clics des utilisateurs au niveau système ? Après avoir capturé ce clic, qu'est-il possible de faire ?
  - Plugin Outlook : nécessite analyse des capacités réelles. Quelles actions sont réalisables via ces plugins ? Peut-on intéragir avec chaque e-mail et si oui quand (à l'ouverture, à la réception) ? Quelles parties du mail sont accessibles et que peut-on en faire ?
  - Coloration TLP des liens : flou sur l'implémentation et la cohérence du concept. Comme pour la prévisualisation de lien, la question était de savoir quelles sont les limites sur chaque sytème. Est-ce qu'il est vraiment possible d'intéragir avec tous les liens au niveau système ? Si oui, quelles interactions (ex : colorer) ?
  - Smishing / vishing : pertinence reconnue, mais manque d'informations techniques. Il faudrait savoir pourquoi ces types de phishing ont lieu et comment les éviter. Pourquoi est-ce qu'on en reçoit autant et qui en est responsable.
  - Plateforme (information, levée de doute, assistance) : pertinente, mais à préciser (public, contenu, structure, maintenance). Il faut aussi se poser la question du but réel que l'on veut donner à cette plateforme. Est-ce qu'il ne serait pas plus judicieux de rejoindre un projet déjà existant, afin de lui apporter notre soutien et nos ressources.

3.1.6.2 Points clés

- Le **tableau de notation partagé** a facilité les discussions et objectivé les décisions.
- **2 idées ont été rejetées** (gestionnaire de mots de passe, outils de vérification).
- **1 idée retenue directement** : sensibilisation aux liens.
- **5 idées jugées intéressantes mais nécessitant des études de faisabilité**, réparties et attribuées aux différents partenaires pour le prochain atelier.

3.1.7 Quatrième atelier : études de faisabilité et choix final des PoCs

Ce quatrième atelier s'est déroulé le 11 avril 2025. L'objectif était donc de retenir les idées les plus prometteuses et de planifier les développements des PoCs.

3.1.7.1 Déroulement de l'atelier

Ce quatrième atelier faisait suite aux études de faisabilité réalisées sur les idées issues des ateliers précédents. Chaque partenaire en charge a présenté les résultats techniques et les contraintes identifiées, en particulier sur les plateformes mobiles. Les résultats ont permis de statuer sur la faisabilité réelle des idées et de trancher pour le choix des PoCs à développer.

Voici un résumé des principaux retours par idée :

- Prévisualisation de lien :
  — Faisable sur Android via une app intermédiaire.
  — Sur iOS, faisable uniquement via une extension Safari ou une app de navigation dédiée.
  — Windows : déjà des solutions existantes (BrowserBox, Menlo).
  Conclusion : jugée faisable et retenue comme PoC.
- Coloration TLP des liens :
  — Android : partiellement possible avec les Accessibility Services.
  — iOS : pas faisable au niveau système.
  — Windows : seulement faisable au sein des clients mails.
  Conclusion : trop limité techniquement et incohérent conceptuellement, abandonné.
- Plugin Outlook :
  — Impossible d'analyser automatiquement les mails à l'ouverture.
  — Interactions possibles uniquement via actions manuelles (ex. clic sur un bouton).
  Conclusion : restrictions trop fortes, abandonné.
- Accès aux SMS et appels :
  — Cette étude est liée aux autres types de phishing, afin de déterminer ce qu'il est possible de faire au niveua de l'appareil de l'utilisateur.
  — Android : accès complet possible via permissions.
  — iOS : accès très limité (uniquement expéditeur, corps du message, code pays).
  Conclusion : faisable partiellement, mais intégré uniquement comme contenu de sensibilisation.
- Plateforme de levée de doute :
  — S'appuie sur l'existant (antiphishing.ch, Hafnova).
  — Possibilité d'offrir un service de soumission/analyse de liens et e-mails suspects.
  Conclusion : faisable et retenu comme PoC.

Ces résultats ont permis de trancher sur les idées à conserver ou à abandonner.

3.1.7.2 Points clés

- **2 idées ont été définitivement abandonnées** :
  — Coloration TLP des liens : trop limité techniquement, incohérence conceptuelle.
  — Plugin Outlook : restrictions fortes de l'API, impact trop faible.
- **2 PoCs ont été retenus pour développement** :
  — **Une application de prévisualisation de liens** avant ouverture sur mobile.
  — **Une plateforme de levée de doute** permettant à un utilisateur de soumettre un lien ou un e-mail suspect pour obtenir un retour.
- Les autres idées (ex. smishing/vishing) seront intégrées sous forme de contenus dans les supports de sensibilisation.

3.1.8 Cinquième atelier : rencontre avec l'OFCS

Ce dernier atelier s'est déroulé le 13 mai 2025 et avait comme objectif de coordonner les efforts réalisés dans ce projet avec l'OFCS qui est déjà actif dans certains domaines proches, notamment avec leur système d'annonce `antiphishing.ch`.

3.1.8.1 Déroulement de l'atelier

Ce dernier atelier a consisté en une réunion avec l'Office fédéral de la cybersécurité (OFCS). Le projet [seal] y a été présenté, la méthodologie, les idées de PoCs retenues, ainsi que nos conclusions. Nos conclusions portaient sur les besoins de la société en un système centralisé regroupant toutes les informations, permettant d'informer et d'aider les individus. Nous avons également conclu qu'il était préférable que la plateforme soit centralisée, qu'on ne souhaite pas faire quelque chose d'indépendant, et aussi exprimé le souhait de coordination. La discussion a porté sur les possibilités de collaboration, notamment autour de potentielles contributions que nous pourrions offrir dans le cadre de leur plateforme antiphishing.ch.

L'OFCS a précisé qu'il ne souhaite pas faire évoluer les fonctionnalités techniques de la plateforme, mais est ouvert à intégrer du contenu de sensibilisation à destination de trois publics cibles : population, entreprises, communes. L'OFCS a également évoqué sa participation au Cyber Security Month 2025, qui pourrait être un point d'ancrage pour des actions communes. Nous n'avons pas obtenu d'attentes, de contenus ou formats sur lesquels nous pourrions nous baser pour concevoir ces medias ou documents.

La question de l'accès aux données brutes de signalement a aussi été abordée. Un accès direct n'est pas envisageable pour des raisons légales, mais l'OFCS est disposé à fournir des statistiques anonymisées à condition d'un besoin clairement formulé.

3.1.8.2 Points clés

- Pas d'intégration possible de PoCs techniques sur antiphishing.ch.
- Ouverture à une collaboration sur les contenus de sensibilisation (guides, vidéos, etc.), mais sans préciser les attentes.
- Possibilité de contribuer au Cyber Security Month 2025 via des contenus adaptés aux publics cibles (actifs, seniors, jeunes).
- Accès aux données brutes exclu, mais possible d'obtenir des statistiques anonymisées si la demande est précise et justifiée.
- Nécessité de coordination avec les autres initiatives suisses (ex. SKP, Pay Attention, E-Banking aber sicher).

3.2 Détails sur les idées innovantes produites lors des ateliers

Les échanges entre partenaires ont permis de faire émerger un ensemble d'idées complémentaires pour renforcer la lutte contre le phishing. Elles couvrent des approches variées, telles que la sensibilisation, l'assistance de l'utilisateur, les outils techniques ou intégrations logicielles et sont présentées ci-dessous avec une description détaillée de leur fonctionnement et de leur valeur ajoutée.

3.2.1 Plugin Outlook d'aide et de signalement

Un groupe a proposé le développement d'un plugin Outlook permettant de signaler en un clic un e-mail suspect directement depuis la boîte de réception. L'outil aurait deux fonctions principales :

- **Faciliter la remontée d'informations** vers une équipe de sécurité ou une plateforme centralisée de confiance, afin de mutualiser les signalements et renforcer la visibilité sur les menaces en circulation.
- **Fournir un retour immédiat à l'utilisateur·rice** sous la forme d'un code couleur clair indiquant le niveau de dangerosité estimé : rouge (lien identifié comme phishing avéré), orange (lien inconnu mais présentant des signes suspects), bleu (lien inconnu et ne présentant pas de signes suspects particuliers).

L'intérêt de ce plugin réside dans sa simplicité d'usage et son intégration directe dans l'environnement quotidien des employés, limitant ainsi la friction lors du signalement d'un e-mail suspect.

Comme pour la plateforme de sensibilisation, ce système aurait vocation à être géré au niveau cantonal ou fédéral par un organisme reconnu (par ex. Confédération, canton, police). Cela garantirait la légitimité et la confiance nécessaires pour encourager les signalements et assurer une gouvernance solide des données collectées.

Un enjeu majeur réside dans la robustesse de l'algorithme de classification. Celui-ci doit être capable d'évaluer la dangerosité des liens soumis tout en résistant aux tentatives de manipulation. Un risque serait par exemple l'empoisonnement des données : un attaquant pourrait soumettre massivement des liens légitimes en les présentant comme suspects, ou au contraire introduire de faux positifs/ négatifs pour biaiser la base de données. Il est donc crucial de concevoir des mécanismes de vérification et de validation humaine afin d'éviter ces dérives et de préserver la fiabilité du système.

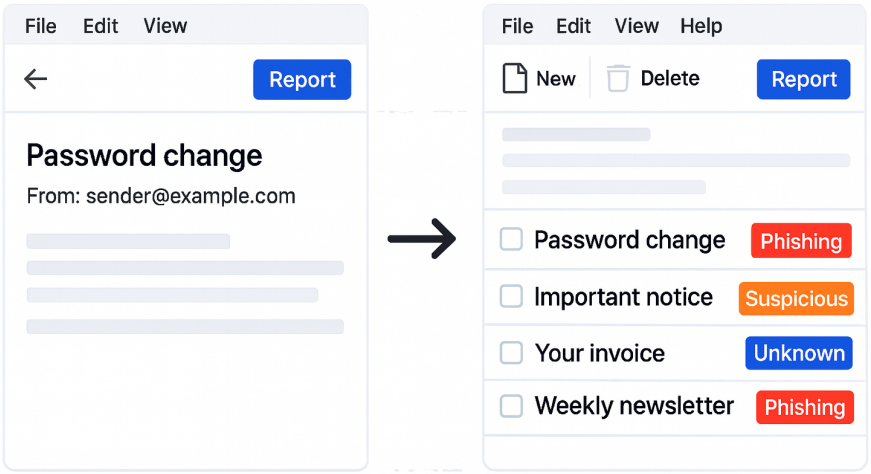

FIGURE 3.2 – Maquette représentant l'idée du plugin Outlook

3.2.2 Outils de vérification de configuration des serveurs mail

Une autre piste identifiée visait les administrateurs système, en particulier dans les PME. L'idée consistait à développer ou à mettre à disposition des outils simples pour vérifier la configuration de leur serveur mail (SPF, DKIM, DMARC, etc.). L'objectif : réduire les erreurs fréquentes de configuration et améliorer la résilience globale face aux attaques de phishing.

3.2.3 Application mobile/desktop de prévisualisation de liens

L'idée consiste en une application mobile servant d'intermédiaire entre le clic sur un lien et son ouverture dans le navigateur. Lorsqu'un utilisateur clique sur un lien reçu par e-mail, SMS ou toute autre application, celui-ci est d'abord redirigé vers l'application. Le lien y est analysé et un retour est fourni à l'utilisateur sous la forme d'un feedback de sécurité.

L'application présente ensuite deux options : poursuivre l'ouverture du lien dans le navigateur ou annuler l'action. Dans cette logique, l'application doit être définie comme navigateur par défaut, afin de s'activer automatiquement à chaque tentative d'ouverture de lien. Ce mécanisme vise ainsi à ajouter une couche de protection générale couvrant l'ensemble des canaux de communication utilisés par les attaquants.

FIGURE 3.3 – Maquette représentant l'idée de l'application faisant de la prévisualisation de lien

3.2.4 Système TLP de coloration des liens

Inspiré du Traffic Light Protocol (TLP), cette idée visait à attribuer une couleur à chaque lien pour en indiquer la fiabilité :

- Rouge : site identifié comme phishing confirmé.
- Orange : site suspect, nécessitant vigilance.
- Bleu : aucun élément suspect détecté.

Le principe s'appliquerait de manière systémique : tous les liens présents sur l'appareil mobile seraient automatiquement analysés et affichés avec un code couleur. L'utilisateur disposerait ainsi d'un repère visuel immédiat et constant sur le niveau de sécurité des liens qu'il consulte, quel que soit le canal de communication par lequel ils sont reçus.

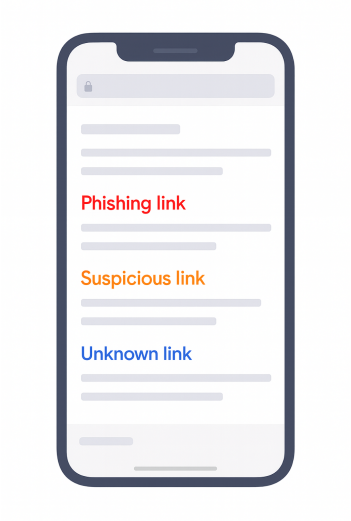

FIGURE 3.4 – Maquette représentant l'idée de la coloration de lien au niveau système

3.2.5 Modules d'apprentissage pour identifier les liens suspects

Certains participants ont proposé des modules d'apprentissage courts et interactifs, intégrés par exemple dans les environnements de travail ou accessibles via la plateforme de sensibilisation. Ces modules viseraient à entraîner les utilisateurs à reconnaître les indices d'un lien frauduleux (nom de domaine altéré, absence de certificat, structure étrange de l'URL, etc.).

3.2.6 Protections intégrées aux gestionnaires de mots de passe

L'idée consistait à renforcer le fonctionnement des gestionnaires de mots de passe existants en y intégrant un module de protection supplémentaire. Actuellement, ces outils n'autocomplètent les identifiants que si le domaine correspond exactement à celui enregistré (par exemple `ubs.com`). Cependant, un utilisateur peut tout de même être tenté de copier-coller ou de saisir manuellement son mot de passe sur un site frauduleux imitant l'original (par ex. `ubs.net`).

Le module proposé viserait donc à :

- Détecter ces tentatives de saisie manuelle ou de copie-collé sur un domaine suspect.
- Bloquer ou alerter explicitement l'utilisateur, afin d'empêcher la compromission des identifiants.

Ce mécanisme transformerait le gestionnaire de mots de passe en un garde-fou actif, en complément de ses fonctions actuelles, et réduirait significativement les risques liés aux attaques de phishing basées sur la confusion entre domaines.

3.2.7 Plateforme régionale

L'idée principale consistait en la création d'une plateforme centralisée dédiée à la sensibilisation du public face au phishing. Elle reposerait sur trois axes complémentaires :

- **Former** : proposer des contenus interactifs et engageants (quiz, serious games, actualités locales sur les arnaques en cours) afin de rendre la sensibilisation plus concrète et adaptée aux menaces réellement rencontrées.
- **Assister** : offrir un espace où l'utilisateur peut soumettre un e-mail ou un lien suspect et obtenir un retour automatisé ou accompagné sur sa dangerosité. Ces signalements alimenteraient également une base de données utile pour mieux comprendre les campagnes actives.
- **Guider** : mettre à disposition une documentation claire sur les outils de protection existants et les bonnes pratiques à adopter au quotidien, afin d'aider les utilisateurs à s'orienter dans la diversité des solutions disponibles.

La portée de la plateforme serait volontairement cantonal ou fédéral, afin de cibler les menaces réellement actives en Suisse et de renforcer la pertinence des contenus proposés. Ce positionnement local permettrait aussi d'exploiter les remontées des utilisateurs pour analyser les phénomènes régionaux et adapter en continu les protections et messages de sensibilisation.

Une inconnue importante concerne toutefois la gouvernance et la pérennité de la plateforme. Son succès dépendrait de sa capacité à être hébergée et maintenue par un acteur reconnu et légitime — par exemple un organisme public comme la Confédération, un canton (ex. État de Vaud) ou encore la police. Ce choix conditionnerait à la fois la confiance des utilisateurs et la durabilité de l'initiative.

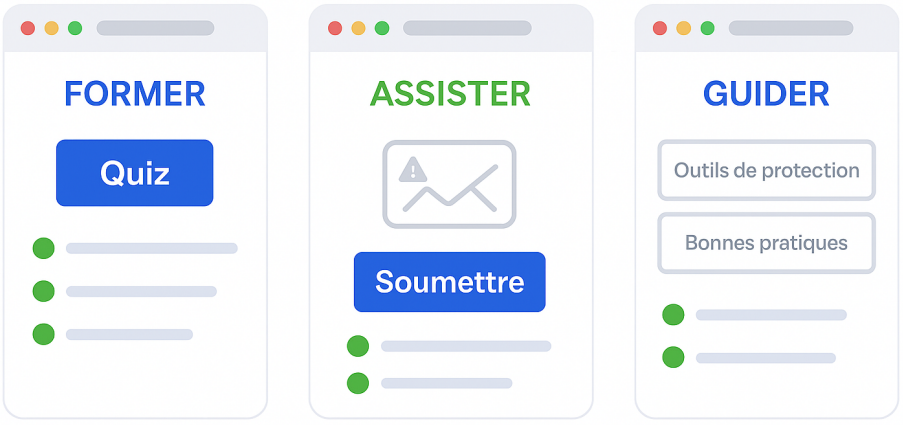

FIGURE 3.5 – Maquette représentant l'idée de la plateforme

3.2.8 Autres canaux (SMS, voix et QR)

Plusieurs participants ont souligné l'importance d'élargir la sensibilisation à d'autres vecteurs que l'e-mail, de plus en plus exploités par les attaquants :

- Les SMS frauduleux (smishing), qui redirigent souvent vers des pages de phishing.
- Les appels téléphoniques trompeurs (vishing), exploitant la confiance ou l'urgence pour obtenir des informations sensibles.
- Les QR codes malveillants, insérés dans des messages ou sur des supports physiques, et qui redirigent vers des sites piégés.

Contrairement à l'e-mail, dont les modes opératoires sont bien connus et documentés, ces vecteurs restent encore relativement opaques. Avant même d'élaborer des mesures de protection ou des campagnes de sensibilisation, il est essentiel de mieux comprendre pourquoi ces attaques fonctionnent et quels moyens techniques elles exploitent.

Par exemple, dans le cas du vishing, la pratique du spoofing rend l'identification des appels difficile, puisqu'un numéro peut en usurper un autre. De même, dans le smishing, les informations accessibles aux opérateurs sont limitées, ce qui complexifie la détection et le suivi. Ces zones d'ombre limitent également les enquêtes, comme l'ont relevé certains représentants des forces de l'ordre (p. ex. Police cantonale vaudoise), qui peinent actuellement à remonter la chaîne d'un appel ou d'un SMS frauduleux.

Une meilleure compréhension technique de ces attaques permettrait non seulement d'appuyer les enquêtes, mais aussi de sensibiliser plus efficacement les utilisateurs. Par exemple, la plateforme de sensibilisation proposée par le premier groupe pourrait intégrer des modules pédagogiques spécifiques au smishing, au vishing et aux QR codes, afin d'élargir la vigilance au-delà du seul canal e-mail.

3.3 Conclusion des ateliers

Les cinq ateliers ont permis de dresser un état des lieux des pratiques actuelles, de générer collectivement des idées innovantes, puis de sélectionner deux solutions concrètes à développer.

De manière générale, l'analyse a montré que les mesures existantes, qu'elles soient techniques, organisationnelles ou légales, offrent une efficacité relativement bonne, mais restent insuffisantes face à l'évolution constante des modes opératoires des cybercriminels. Ces derniers exploitent principalement le facteur humain, en trompant les utilisateurs et en les redirigeant vers des domaines frauduleux. Dans ce contexte, les défenses mises en place au niveau des domaines légitimes deviennent souvent inopérantes.

Un constat unanime a émergé : le besoin d'une plateforme centralisée et régionale regroupant les différents aspects de la lutte contre le phishing, à savoir la formation, l'information, l'aide aux utilisateurs et les annonces officielles. Un tel outil permettrait d'unifier les efforts

dispersés et d'améliorer la visibilité des dispositifs existants. Toutefois, son développement sort du cadre de ce projet en raison de la forte nécessité d'une coordination avec les entités déjà en charge de ces thématiques, mais aussi des contraintes budgétaires, des problématiques de maintenance future ainsi que du manque d'accès aux données.

Toutefois, l'idée derrière la plateforme qui permettrait d'analyser les liens pour donner un retour sur leur niveau de danger potentiel reste intéressante et mérite d'être creusée dans le cadre de ce projet. À ce titre, il est intéressant de relever que, peu après la fin du projet, une initiative indépendante nommée Flair (voir Section 2.2.4.6) a vu le jour et propose précisément un service d'analyse accessible aux utilisateurs, illustrant la pertinence du besoin identifié lors des ateliers.

À l'issue de cette démarche, deux PoCs ont finalement été retenus pour développement :

- Un prototype de service d'analyse de liens permettant à un utilisateur ou à une application de soumettre un lien et de recevoir une évaluation de la dangerosité de celui-ci.
- Une application mobile Android permettant d'attraper les clics de l'utilisateur et de prévisualiser le lien, en mettant en évidence le domaine. Cette application pourra aussi être capable de fournir une analyse du lien en se basant sur le service précédemment développé.

Chapitre 4

Expérimentations et PoC

Table des matières du chapitre

Suite aux ateliers d'échange menés avec les différents acteurs du projet, plusieurs pistes de solutions ont été identifiées pour répondre aux enjeux liés au phishing. Afin de valider leur faisabilité et d'en évaluer la pertinence, deux preuves de concept (PoCs) ont été développées :

- une **application Android de prévisualisation de liens**, jouant le rôle de couche intermédiaire avant l'ouverture d'une URL et offrant des informations contextuelles pour aider l'utilisateur à décider ;
- un **prototype de service d'analyse de liens** a également été mise en place, celui-ci s'est révélée nécessaire pour l'application Android de prévisualisation de liens, car une analyse complète des liens était indispensable. Ce service fait donc office de point d'entrée unique, combinant les résultats de différents services et systèmes de détection afin de fournir une réponse consolidée sur le caractère potentiellement frauduleux d'un lien ou d'un message.

Initialement, nous avions commencé à développer notre propre application en Flutter qui accueillerait par la suite le service d'analyse de liens. Mais en cours de travail, nous avons identifié l'application open-source *URLCheck* (`https://github.com/TrianguloY/URLCheck`). Nous avons décidé de réutiliser cette base côté Android en y ajoutant un **module "LinkRisk"** connecté à notre service d'analyse. Le périmètre fonctionnel reste inchangé (prévisualisation et avis de risque). Cela permet de garantir une maintenance plus simple et stable si le projet venait à continuer.

4.1 Application Android de prévisualisation de liens

4.1.1 Objectifs et rôle du PoC

L'idée est d'abord d'aider l'utilisateur à décider avant de cliquer : afficher un avis clair (*Phishing détecté* si une source de confiance répertorie le lien, sinon un niveau *Risque élevé / modéré / faible*) pour éviter les clics à l'aveugle, tout en laissant le choix final à l'utilisateur.

L'application Android adaptée dans ce PoC a pour objectif d'assister les utilisateurs lorsqu'ils interagissent avec des liens potentiellement suspects. Plutôt que d'ouvrir directement le lien dans leur navigateur, l'application intercepte celui-ci et effectue une analyse préalable. Les utilisateurs reçoivent ainsi un retour clair sur la dangerosité potentielle du site avant d'y accéder réellement, leur permettant de décider en connaissance de cause s'ils souhaitent continuer ou non.

Cette approche constitue une première couche de protection supplémentaire, transversale à tous les canaux d'attaque : qu'il s'agisse d'un lien reçu par e-mail, SMS, messagerie instantanée ou tout autre vecteur, l'application traite de manière générique le point commun de ces attaques, à savoir l'ouverture d'un lien.

4.1.2 Architecture et conception

L'architecture de l'application est volontairement simple et repose sur deux éléments principaux :

- un **frontend mobile**, qui gère l'interface utilisateur et l'interception des liens au niveau du système d'exploitation ;
- un **backend distant**, représenté par le service d'analyse de lien développée dans un autre PoC, qui centralise l'analyse des liens et renvoie un verdict sur leur sécurité.

Lorsqu'un utilisateur clique sur un lien, l'application intercepte cet événement, envoie l'URL au service d'analyse de liens, puis présente à l'utilisateur un retour visuel. Cette conception en couches permet de déléguer la logique d'analyse au backend, tout en maintenant un client léger et portable.

La version finale de l'application retenue s'appuie sur *URLCheck* (TrianguloY), dans lequel nous avons intégré un **module "LinkRisk"** qui envoie l'URL au service d'analyse puis affiche *Phishing détecté* (si une source reconnue répertorie le lien) ou un *niveau de risque* (élevé / modéré / faible) selon le score renvoyé par le modèle. L'endpoint est configurable et aucune clé API (GSB/OFCS/URLhaus) n'est embarquée côté application : tout est traité par le backend. L'analyse est également déclenchée automatiquement à l'ouverture du lien, afin d'éviter une action manuelle supplémentaire.

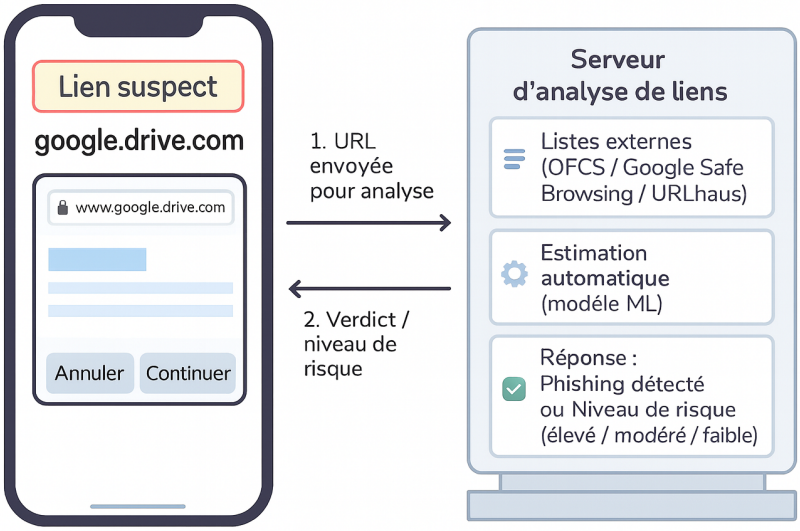

FIGURE 4.1 – Représentation de l'architecture

4.1.3 Exemple d'utilisation

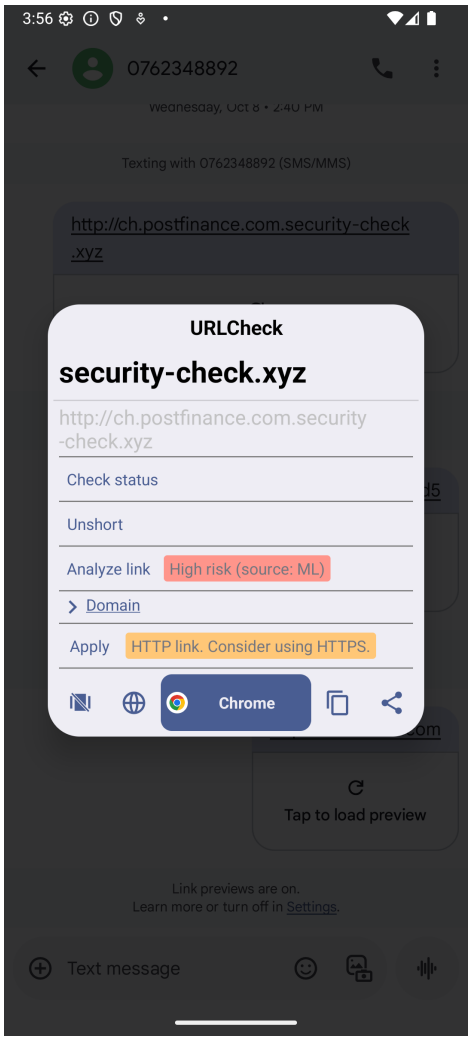

FIGURE 4.2 – Prévisualisation d'un lien avec analyse intégrée et choix du navigateur.

Les captures proviennent de l'application URL-Check accompagnée de notre module "LinkRisk". Le parcours utilisateur est le suivant :

1. l'utilisateur clique sur un lien.
2. l'application affiche une page de prévisualisation.
3. Un verdict laisse le choix d'ouvrir ou non dans le navigateur préféré.

Côté échanges réseau, l'application envoie :

```
POST /check_url
{
    "url": "<lien-cliqué>"
}
```

et interprète la réponse selon des règles simples et explicites :

- **Phishing détecté** si une base externe reconnue (OFCS/PhishDB, Google Safe Browsing, URLhaus) signale l'URL.
- **Risque élevé** si aucune base ne signale l'URL mais que le modèle estime une probabilité élevée.
- **Risque modéré** si le modèle indique des indices suspects.
- **Risque faible** si le score du modèle est bas et qu'aucune base externe ne remonte de signal.

Il est tout de même important de noter que l'analyse est indicative et qu'une attention particulière de l'utilisateur est toujours nécessaire.

Enfin, cette explication peut être représentée à l'aide du schéma séquentiel suivant :

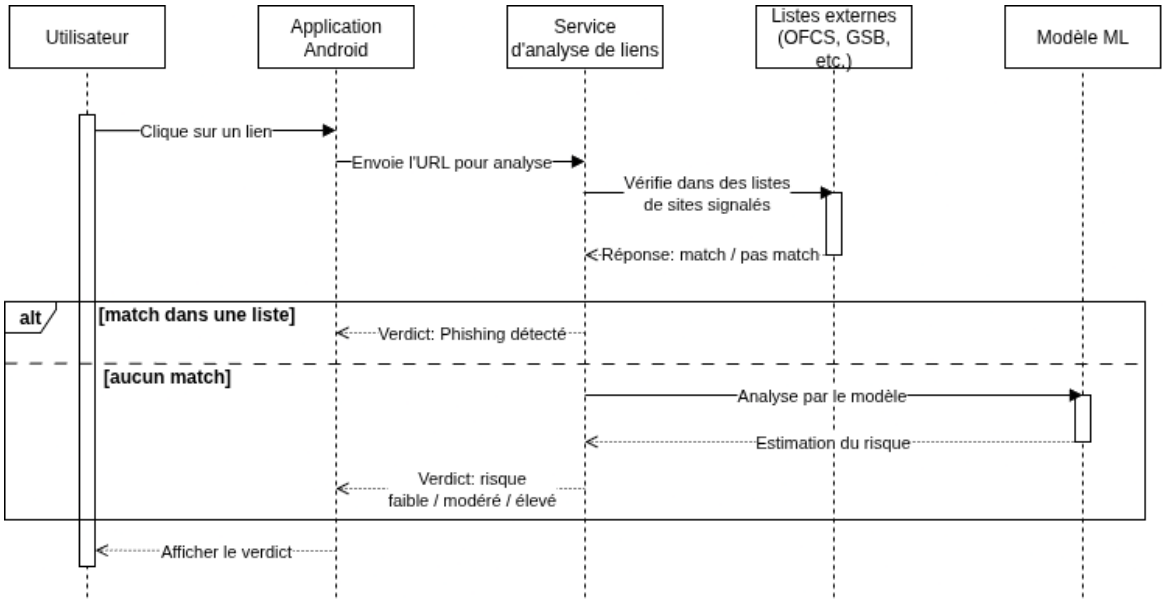

FIGURE 4.3 – Représentation des différentes requêtes de manière séquentielle

4.1.4 Limites et perspectives

La qualité du verdict dépend surtout de la couverture des bases (OFCS/GSB/URLhaus) et du jeu d'entraînement : des campagnes récentes peuvent passer sous le radar. Par choix de sécurité, la prévisualisation WebView tourne sans JavaScript, ce qui réduit les risques mais peut dégrader le rendu de certains sites.

Les prochaines étapes qui pourraient être étudiée seraient le mode hors ligne qui n'utiliserait que le score obtenu grâce au machine learning ainsi que le stockage des liens soumis pour continuer d'alimenter la base de donnée et réentraîner régulièrement le modèle. Cette option permettrait ainsi de rester à jour, d'ajuster les seuils et limiter les faux positifs/négatifs.

4.2 Service d'analyse de liens

4.2.1 Objectifs et rôle du PoC

Avant tout, ce service d'analyse de liens sert de garde-fou : quand un lien arrive (mail, SMS, messagerie), on peut le soumettre et obtenir un retour lisible pour décider s'il est prudent de continuer. Ce service répond avec un **niveau de risque** et la **source** qui justifie ce résultat :

si une base reconnue répertorie l'URL, on affiche *Phishing détecté* ; sinon, on indique *Risque élevé / modéré / faible* selon les indices trouvés par le modèle. L'idée est d'aider l'utilisateur à trancher, sans donner de faux sentiment de sécurité.

Concrètement, ce PoC fournit une API centralisée capable d'évaluer automatiquement la dangerosité d'un lien URL. Pensée pour être intégrée à d'autres systèmes (comme l'application Android utilisée en parallèle), ce service agit comme un moteur de détection en ligne : à chaque appel avec une URL en paramètre, elle renvoie un verdict indiquant si le lien comporte un risque *élevé*, *modéré*, ou *faible*.

Ce composant vise à externaliser l'intelligence de détection, en encapsulant toute la logique d'analyse dans un service indépendant et interrogeable par n'importe quel client (mobile, web, navigateur, etc.).

4.2.2 Architecture et conception

Le service d'analyse de lien repose sur une architecture simple :

- un **backend Python** développé avec `FastAPI`, qui expose une interface REST pour recevoir les requêtes et répondre avec un verdict.
- un **modèle de machine learning** (`RandomForestClassifier`), entraîné en amont, qui estime la probabilité qu'un lien soit malveillant à partir de ses caractéristiques lexicales et techniques.
- des **appels API à des bases de données externes** pour enrichir la détection, notamment Google Safe Browsing, URLhaus ou encore PhishDB (OFCS), permettant de détecter certains liens connus même sans analyse intelligente.

Cette séparation permet de combiner plusieurs sources de vérification (statique et dynamique) pour une analyse plus robuste.

4.2.3 Implémentation technique

Le service s'appuie sur **FastAPI** (exposition HTTP), **scikit-learn** (modèle de classification) et des appels sortants via **Requests** vers des sources externes (Google Safe Browsing, URLhaus, PhishDB/OFCS).

Pour le modèle, nous utilisons un *RandomForestClassifier* entraîné pour une classification binaire (*phishing* vs *legitimate*). À la réception d'une URL, on en extrait des caractéristiques simples et robustes : longueurs (URL/hôte), comptages de caractères spéciaux, ratio de chiffres, TLD, nombre de sous-domaines, présence de punycode `xn-`, schéma `http/https`, extension de chemin, détection d'IP et usage de raccourcisseurs. Les variables catégorielles sont encodées de façon compacte (par ex. `http`→0, `https`→1, *one-hot* limité pour `tld` et `path_ext`) et l'ordre exact des variables est figé à l'entraînement puis réappliqué à l'inférence

pour éviter tout décalage. Le modèle renvoie une probabilité p que l'URL soit du phishing.

Le flux est le suivant : l'API commence par vérifier l'URL dans des listes externes (OFCS/-PhishDB, Google Safe Browsing, URLhaus). Si l'URL y est référencée, le service renvoie immédiatement « *Phishing détecté* » en citant la source. Sinon, la probabilité p issue du modèle est traduite en niveau de risque côté client selon des seuils fixes (déterminés arbitrairement) :

- $p \geq 0{,}80 \rightarrow$ *Risque élevé*,
- $0{,}60 \leq p < 0{,}80 \rightarrow$ *Risque modéré*,
- $p < 0{,}60 \rightarrow$ *Risque faible*.

Pour l'apprentissage, nous avons constitué un premier jeu de données issu de plusieurs provenances, afin de refléter l'usage réel et les menaces courantes :

- *sites légitimes les plus visités par des utilisateurs suisses*,
- *bases publiques de liens suspects*,
- la base *PhishDB/OFCS*.

Pour équilibrer les classes sans biaiser la distribution, nous conservons tous les exemples de phishing, tous les liens légitimes « `.ch` », puis un échantillon aléatoire de liens légitimes hors « `.ch` » après déduplication.

L'évaluation se fait par validation stratifiée (ou k-fold), avec précision, rappel et F1-score sur la classe *phishing*, ainsi que l'AUC-ROC. Une matrice de confusion sert à ajuster les seuils selon le compromis faux positifs / faux négatifs. En exploitation, nous prévoyons de journaliser, avec consentement et rétention limitée, les URLs soumises et les verdicts afin d'alimenter une base interne et de réentraîner régulièrement le modèle pour rester à jour face aux nouvelles campagnes.

Ce design permet une analyse rapide, y compris en mode dégradé, tout en tirant parti de listes de référence dès qu'elles sont disponibles.

4.2.4 Exemple d'utilisation

Une requête typique se fait via un simple appel HTTP POST, avec un JSON contenant l'URL à analyser :

POST **/api/check-url**

check a URL for phishing

Parameter

no URL parameter, body only

Body application/json

```
{
  "url": "http://phishing.test/login"
}
```

Response application/json

200 successful analysis

```
{
  "verdict": "malicious",
  "source": "phishing-database",
  "proba": 0.92
}
```

400 bad request

```
{
  "error": "missing 'url' parameter"
}
```

Cela permet à toute application cliente (comme l'application mobile) d'obtenir un retour clair et exploitable.

4.2.5 Limites et perspectives

Bien que le modèle de machine learning fournisse des résultats prometteurs, sa précision dépend fortement de la qualité et de l'équilibre du jeu de données. Or, les URL de phishing sont bien plus nombreuses que les URL suisses réellement légitimes, ce qui crée un déséquilibre difficile à corriger.

De plus, certaines attaques très sophistiquées peuvent ne présenter aucun indice évident dans l'URL elle-même. L'ajout d'un module d'analyse dynamique (rendu HTML, comportement JavaScript, etc.) pourrait améliorer la détection à l'avenir.

Enfin, un autre axe d'amélioration envisagé est l'ajout d'un système de feedback utilisateur, permettant de réentraîner le modèle en continu à partir des cas mal classés.

4.3 Évaluation du prototype — deux scénarios complémentaires

L'objectif du PoC n'est pas uniquement de *détecter du phishing*, mais aussi de *ne pas* déclencher d'alerte sur des sites légitimes. Pour valider ces deux aspects, nous avons testé deux scénarios distincts, chacun ciblant un objectif précis :

- **Scénario A - Détection :** un lot de 3'643 URLs fourni par la Police cantonale vaudoise (PCV), composé de sites frauduleux qui leur ont été remontés entre 2019 et 2025. On s'attend ici à de nombreux signalements dans les bases (OFCS, Google Safe Browsing, URLhaus) et à des scores de machine learning élevés.
- **Scénario B - Fiabilité :** un échantillon de 1'000 domaines issus du Top 1'000'000 Tranco (liste du 27.10.2025), représentant les sites les plus visités dans le monde. L'objectif est inverse : vérifier que le PoC reste prudent et n'associe pas à tort un risque élevé à des sites légitimes. A noter que nous sommes parti de l'hypothèse que les sites les plus visités au monde sont aussi des sites sûrs. Cette hypothèse est plausible, car il est peu probable qu'une campagne de phishing ait réussi à avoir un impact supérieur à celui d'un site courrament visité.

Dans les deux cas, chaque URL est traitée par le service d'analyse de liens, qui interroge successivement les trois bases de données (**OFCS**, **Google Safe Browsing** et **URLhaus**), puis réalise une analyse avec le modèle de machine learning. Le résultat indique combien de ces URLs étaient présentes dans chacune des bases de données. Ainsi que la répartition de l'attribution à travers les trois niveaux de dangerosité. Pour les URLs qui se trouvaient déjà dans les bases de données, on a en plus décidé d'afficher le score attribué, ainsi que l'URL affectée.

Dans un seconde temps, ces URLs seront intégrées au système en tant que nouvelle base de données à consulter avant d'interogger le modèle.

Scénario A — Jeu PCV (3'643 URLs orientées phishing)

Ce premier scénario sert à évaluer la capacité du PoC à identifier efficacement des liens frauduleux. On s'attend logiquement à ce que les bases de données reconnaissent une part importante des URLs et que le modèle attribue des scores élevés à la majorité des autres.

DB Phishing (au moins une base)	
Total	556
URLhaus	0
OFCS	373
GSB	183
Machine learning (quand aucune DB ne match)	
Total	3'607
Risque élevé (≥ 0.80)	3'604
Risque modéré $[0.60, 0.80)$	1
Risque faible (< 0.60)	2
Erreurs / timeouts	36

TABLE 4.1 – Résultats sur le lot PCV (phishing connu).

Sur ce jeu, les trois bases signalent 556 URLs, dont une majorité via OFCS, suivie de GSB. Parmi ces URLs, 36 URLs sont détectées simultanément par plusieurs sources (GSB et OFCS), ce qui renforce la confiance dans la détection. Parmi les 3'643 URLs, la grande majorité a été détectée comme ayant un risque élevé de contenir du phishing. Seulement 3 URLs n'ont pas été identifiées comme ayant un risque élevée et une trentaine ont généré une erreur.

Scénario B - Échantillon Tranco (1'000 domaines populaires)

Ce second scénario évalue la stabilité du PoC face à des sites réputés sûrs. L'idée est de vérifier qu'il ne considère pas à tort des plateformes légitimes comme suspectes.

DB Phishing (au moins une base)	
Total	10
URLhaus	0
OFCS	10
GSB	0
Machine learning (quand aucune DB ne match)	
Total	945
Risque élevé (≥ 0.80)	0
Risque modéré $[0.60, 0.80)$	0
Risque faible (< 0.60)	945
Erreurs / timeouts	55

TABLE 4.2 – Résultats sur l'échantillon Tranco (sites légitimes).

Sur 1'000 domaines parmi les plus visités, seuls 10 ont été signalés par OFCS. Il s'agit de plateformes légitimes connues (par exemple des raccourcisseurs d'URL ou services de formulaires), qui peuvent être temporairement exploitées à des fins malveillantes. Le modèle, qui évalue ici le domaine dans son ensemble, les classe néanmoins tous en *risque faible*, ce qui correspond au comportement attendu. Aucun site n'a été faussement évalué comme risqué, confirmant que le PoC reste conservateur sur les hôtes populaires.

Analyse croisée et conclusion

Les deux tests racontent une histoire cohérente : le PoC détecte efficacement les sites frauduleux lorsqu'il le doit, et reste mesuré sur les sites légitimes. Les bases de données fournissent des signaux fiables et ponctuels, tandis que le modèle de machine learning généralise bien le comportement des URLs suspectes sans tomber dans la sur-sensibilité.

Sur le plan opérationnel, les quelques erreurs et timeouts (36 pour PCV, 55 pour Tranco) mettent en évidence la nécessité d'un mécanisme de relance et de mise en cache pour stabiliser le traitement à grande échelle.

4.4 Synthèse et conclusion

Les deux prototypes (PoCs) que les partenaires avaient décidé d'investiguer lors des ateliers de synthèse (voir Section 3.3) ont pu être réalisés comme prévu. Les résultats obtenus sont à la fois pertinents et prometteurs :

L'application Android de prévisualisation de liens se révèle très prometteuse grâce à la mise en évidence du domaine (la partie essentielle de l'URL). Elle aide efficacement l'utilisateur à éviter de se rendre sur des domaines frauduleux et limite ainsi le risque de devenir victime d'une attaque.

Le service d'analyse de liens, bien qu'encore au stade de prototype, est déjà fonctionnel et constitue une réelle avancée pour la sécurité. Il permet notamment d'identifier et de signaler les liens déjà connus comme dangereux.

La combinaison des deux PoCs apporte une valeur ajoutée : l'application Android peut tirer parti du service d'analyse pour offrir un retour immédiat en cas de détection d'un lien malveillant. Même si cette approche n'est pas parfaite, elle représente une amélioration tangible.

Au fil des développements, nous avons finalement intégré notre module à l'application open-source URLCheck (auteur : TrianguloY) plutôt que de poursuivre notre prototype Flutter. Cette intégration nous a permis de nous concentrer sur la logique d'analyse et l'expérience utilisateur, tout en capitalisant sur une base applicative robuste.

De manière générale, il reste nécessaire de réaliser des tests en conditions réelles et de recueillir davantage de retours utilisateurs pour confirmer la pertinence et l'efficacité des prototypes.

En conclusion :

- Les deux idées et prototypes retenus apparaissent pertinents pour réduire l'impact du phishing, même s'ils ne permettent pas de l'éliminer totalement, comme on pouvait s'y attendre.
- La méthodologie adoptée, reposant sur des ateliers collaboratifs et des échanges entre partenaires, s'est avérée un excellent moyen de faire émerger et de sélectionner les idées les plus pertinentes.

Chapitre 5

Sensibilisation et valorisation

Table des matières du chapitre

L'un des principaux objectifs de ce projet est de pouvoir réduire l'impact du phishing. En ce sens, il semblait essentiel de pouvoir communiquer autour de ce projet, de sa réalisation, des réflexions, ainsi que de pouvoir sensibiliser un large public à la problématique et aux bons réflexes.

Lors des ateliers, plusieurs idées ont été discutées et échangées. On peut se poser des questions sur les publics à adresser (particuliers ou professionnels), les types de phishing (email, sms, QR codes, etc.), la manière de les adresser, le type de media ou encore de partenaire.

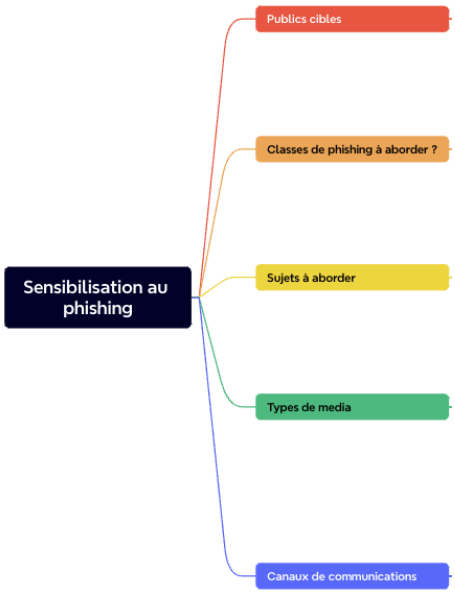

FIGURE 5.1 – Éléments de réflexion autour de la communication et sensibilisation

A l'heure de l'écriture de ce document, le plan d'action n'a pas été finalisé, en revanche, l'approche et les réflexions sont synthétisées.

5.1 Public cible

Deux catégories principales de publics cibles ont pu être identifiées. Au sein de chaque public cible, on retrouve des sous-groupes pour lesquels on explique l'importance de les sensibiliser face aux menaces.

La première catégorie est **la société**, qui est composée de tous les individus qui utilisent les moyens de communication modernes pour des raisons personnelles. Ce catégorie est composée des enfants, des personnes actives et des seniors. Ce public est particulièrement vulnérable, car la plupart de ces personne ne sont pas formés pour réagir correctement en cas de menace.

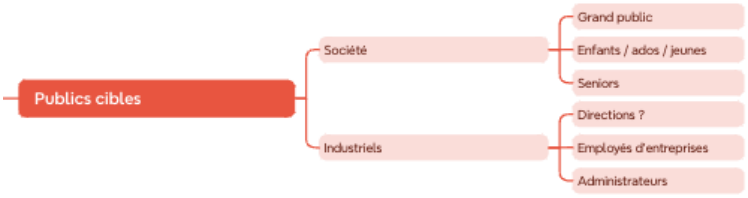

FIGURE 5.2 – Publics cibles

La deuxième catégorie est le **monde industriel**. On retrouve **trois sous-groupes** nécessitant une sensibilisation différente :

- **La direction** nécessite une formation particulière car elle prend la plupart des décisions au sein de sa propre institution. Ces décision sont de plusieurs types, comme le fait de former ou non ces employés aux menaces actuelles, ou simplement d'imposer un authentification multifacteurs à ses employés. De plus, la direction requière une sensibilisation particulièrement plus avancée car elle possède des privilèges plus élevés au sein de son institution, et est donc une cible privilégiée par les attaquants.
- **Les employés d'entreprise** doivent avoir accès à des contenus de sensibilisation de qualité afin de réagir correctement en cas d'attaque.
- **Les administrateurs du système informatique** doivent être sensibilisés aux dernières menaces afin de pouvoir réagir rapidement en cas d'incident et aussi connaître les technologies permettant de défendre son institution et limitant au maximum la surface d'attaque.

5.2 Classes de phishing

Le phishing est souvent associé aux e-mails frauduleux contenant des liens vers de faux sites web où les victimes saisissent des informations sensibles, comme leurs identifiants ou leurs données bancaires. Cependant, ce vecteur d'attaque s'est largement diversifié au fil des années et touche désormais de nombreux autres canaux de communication. Ci-dessous sont présentées les principales formes de phishing, ainsi que les enjeux de sensibilisation qui leur sont associés.

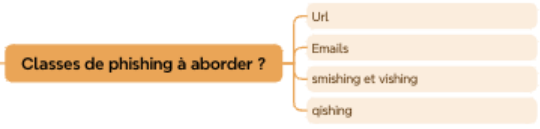

FIGURE 5.3 – Types de phishing

La plupart des attaques de phishing exploitent la confiance – voire la crédulité – que les

utilisateurs accordent aux liens reçus. C'est pourquoi nous commençons par examiner les liens et les URLs, en soulignant que leur compréhension et une sensibilisation adéquate sont essentielles.

5.2.1 Problématique commune autour des URLs

Les attaques basées sur les URLs consistent à rediriger les victimes vers des sites frauduleux imitant des services légitimes (banque, messagerie, plateformes de commerce en ligne, etc.). Les attaquants diffusent ces liens piégés par e-mail, SMS, QR codes ou autres canaux de communication, incitant ensuite l'utilisateur à saisir ses identifiants ou des informations sensibles.

Ce type d'attaque est particulièrement efficace car de nombreux utilisateurs se fient uniquement à l'apparence visuelle du site sans vérifier l'adresse complète. Des techniques comme le typosquatting (ex. `paypa1.com` au lieu de `paypal.com`) ou l'usage d'homographes (caractères ressemblants issus d'autres alphabets, comme le cyrillique) renforcent ce risque.

La sensibilisation doit insister sur l'importance de vérifier attentivement l'URL avant de cliquer et de privilégier la saisie manuelle des adresses des services sensibles plutôt que de passer par des liens reçus.

5.2.2 E-mail

L'e-mail est le vecteur historique du phishing et reste aujourd'hui l'un des plus utilisés. Les attaquants y diffusent des messages imitant l'apparence d'institutions connues afin d'inciter les destinataires à cliquer sur un lien, télécharger une pièce jointe ou transmettre des informations confidentielles.

Malgré l'évolution des filtres anti-spam, les cybercriminels parviennent encore à contourner les protections en exploitant des moyens d'ingénierie sociale, tels que l'urgence (paiement bloqué, compte suspendu) ou l'autorité supposée de l'expéditeur. La sensibilisation doit amener les utilisateurs à identifier les signaux d'alerte les plus fréquents (adresse de l'expéditeur incohérente, fautes, demandes pressantes) et à signaler tout message douteux.

5.2.3 Smishing

Le smishing, ou phishing par SMS, repose sur l'envoi de messages courts contenant généralement un lien frauduleux ou un numéro de téléphone à rappeler. Exploitant la confiance associée au canal SMS, les attaquants s'appuient sur le réflexe des utilisateurs à consulter et à répondre rapidement à ce type de messages.

Ce canal est d'autant plus efficace que les outils de filtrage sont limités et que les SMS

paraissent crédibles par leur concision. Les victimes ne disposent souvent pas du temps nécessaire pour vérifier la légitimité du message. Les campagnes de sensibilisation doivent rappeler qu'aucune organisation sérieuse ne demande d'informations sensibles par SMS et encourager à se méfier des liens raccourcis ou anonymes.

5.2.4 Vishing

Le vishing, ou phishing vocal, se déroule par téléphone. L'attaquant appelle directement la victime en se faisant passer pour une banque, une administration ou un service technique, afin d'obtenir des informations personnelles ou de pousser à réaliser une action (par exemple un virement immédiat).

La force de cette attaque repose sur l'effet de surprise et sur la pression psychologique exercée lors de la conversation. Beaucoup de victimes, déstabilisées, n'osent pas remettre en question l'autorité supposée de l'interlocuteur. La sensibilisation doit insister sur la nécessité de rester vigilant au téléphone et d'interrompre tout appel suspect pour rappeler soi-même via les numéros officiels.

5.2.5 Quishing

Le quishing, ou phishing par QR code, tire parti de la popularité des codes QR. Les attaquants diffusent ou collent de faux codes dans l'espace public par exemple par dessus des codes existants (ex. QR code pour payer le parking), dans des e-mails ou sur des documents, qui redirigent vers des sites malveillants.

Cette technique est redoutable car, contrairement à une URL classique, l'adresse cible n'est pas immédiatement visible avant le scan. De plus, la banalisation des QR codes rend leur utilisation quasi automatique pour de nombreux utilisateurs. Une sensibilisation efficace doit rappeler qu'il est important de vérifier l'URL affichée après le scan avant de cliquer et de ne scanner que des codes provenant de sources fiables.

5.3 Sujets à aborder

Sensibiliser de manière efficace et concise constitue un véritable défi. Pour maximiser son impact, plusieurs thématiques clés doivent être systématiquement abordées afin d'assurer une compréhension claire des risques et des moyens de protection.

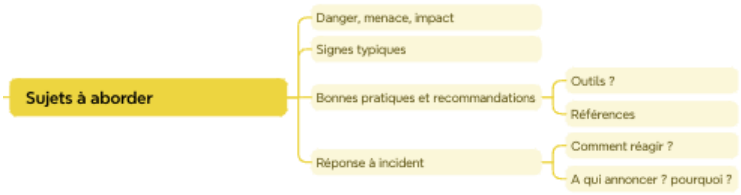

FIGURE 5.4 – Sujets à aborder

5.3.1 Danger, menace et impact

Pour susciter une prise de conscience immédiate, il est crucial de commencer par présenter les dangers potentiels, les menaces concrètes, ainsi que les impacts possibles (financiers, juridiques, réputation, etc.).

5.3.2 Signes typiques

Décrire les signes courants permettant d'identifier une tentative de phishing, une compromission ou une autre forme d'attaque permet aux utilisateurs de reconnaître rapidement les situations à risque.

5.3.3 Bonnes pratiques et recommandations

Fournir des conseils clairs, applicables au quotidien, comme la vérification des liens ou encore la méfiance face aux pièces jointes inattendues.

5.3.4 Outils à disposition

Présenter les outils internes ou externes permettant de se protéger ou de signaler un incident : filtres anti-spam, gestionnaires de mots de passe, VPN, etc.

5.3.5 Références utiles

Inclure des ressources fiables pour approfondir les connaissances : sites officiels, guides de bonnes pratiques (ex. : MELANI, NCSC, ANSSI, etc.).

5.3.6 Réponse à incident

Expliquer clairement la procédure à suivre en cas de suspicion ou d'incident avéré :

- Que faire immédiatement (déconnecter, ne pas interagir davantage, etc.) ;
- À qui remonter l'information (contact IT, RSSI, supérieur hiérarchique) ;
- Pourquoi cette remontée est cruciale (limiter les dégâts, enclencher les mesures correctives, conformité réglementaire).

5.4 Type de médias

Le choix du support de sensibilisation constitue un enjeu important, car il conditionne la portée et l'efficacité du message. Plusieurs formats complémentaires peuvent être envisagés afin de maximiser l'impact et de toucher un large public.

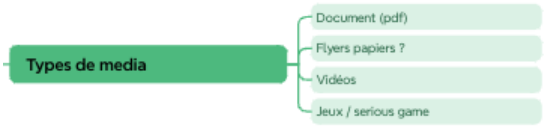

FIGURE 5.5 – Types de media

- L'utilisation d'un **document** partagé au sein des entreprises et des particuliers. Celui-ci permettrait de communiquer plein d'information de manière complète.
- Un **flyer** permettrait de fournir de la sensibilisation en aide-mémoire, ce qui permettrait aux bénéficiaires de réagir correctement en cas de suspicion d'attaque de phishing ou si une attaque aurait réussi, de réagir le plus rapidement et efficacement possible.
- Une **vidéo** permettrait de communiquer les même informations que le document, mais avec un format plus ludique et plus apprécié par la plupart des gens.
- Un **jeux/serious game** serait la manière la plus pertinente de passer l'information, car les bénéficiaires auraient la capacité d'avoir une formation didactique dans laquelle ils pratiquent et donc apprennent mieux.

Chacune de ces méthodes sont utiles, et permettent de fournir facilement des ressources de qualité.

5.5 Canaux de communications

FIGURE 5.6 – Canaux à utiliser

5.5.1 Pages web

Afin d'intégrer les contenus, il ne serait pas très pertinent de créer une page web uniquement pour nos contributions. En effet, il serait alors nécessaire de maintenir ce site Web et de réaliser un travail de communication afin de partager aux mieux les éléments.

Nous trouvons plus pertinent de s'intégrer à des pages et contenus existants. Pour ce faire, nous avons listé certaines opportunités, tout en analysant leur pertinence.

Site seal Cela permettrait de faire de la publicité pour le projet, et aussi de montrer les résultats des projets **[seal]**.

Site de ProSenectute Cela facilite notre intégration et diffusion, mais uniquement auprès des Seniors. Pour rappel, ProSenectute est une association privée d'utilité publique. Son but est de contribuer au bien-être matériel, physique et moral des personnes âgées vaudoises, ainsi que de préserver ou renforcer leur capacité de vivre indépendantes et intégrées à la vie du pays. Nous pourrions alors intégrer notre sensibilisation avec leur service, en intégrant notre contenu sur leur plateforme. Cela aiderait considérablement les personne âgées qui ont des difficultés avec les outils informatiques.

Plateforme antiphishing.ch Cette plateforme est maintenue par l'OFCS. Notre idée initiale était de publier notre contenu sur leur plateforme. Cependant, la collaboration ne semble pas si évidente pour des raisons d'objectifs. En effet, antiphishing.ch se veut être positionné comme une plateforme d'annonce à incidents, et non de sensilisation à la problématique.

Site d'une entité cantonale La dernière idée serait de s'intégrer dans un des programme cantonal. Le canton possède plusieurs plateformes afin de partager des informations, notamment Etat de Vaud, DGNSI, PCV, etc.

5.5.2 Réseaux sociaux

Deux options sont envisageables : sensibiliser au travers des réseaux sociaux, ou communiquer et valoriser le projet en lui-même.

Pour la première option, il serait intéressant de s'appuyer sur des canaux existants tels que l'`e-cop` vaudois [1] ou des institutions telles que la PCV ou la DGNSI.

Pour la seconde option nous pourrions simplement exploiter les réseaux sociaux afin de communiquer autour de notre projet de recherche, notamment ce document, en utilisant les canaux sociaux de différents partenaires tels que la PCV, la DGNSI, ou encore l'EPFL, la HEIG-VD, l'Unil, voire d'autres.

5.5.3 Autres idées

D'autres idées ont également été abordées et sont listées ci-dessous.

- S'associer à des journaux (papier) comme relais de communication. Une idée pertinente serait d'essayer de s'insérer dans `La Lettre Cyber` du Temps.
- Réaliser une émission à la télévision (p.ex. RTS) afin de communiquer sur le projet, la thématique du phishing, et tenter de sensibiliser le public.
- Réaliser une émission à la radio et imaginer une campagne de sensibilisation.
- S'associer à la Poste, ou autre entité, et imaginer des flyers papiers percutants et instructifs.

1. `https://www.tiktok.com/@ecop.francois/video/7519211610133515542?lang=fr`

Chapitre 6

Conclusion

Ce projet s'est inscrit dans un contexte où le phishing demeure l'une des principales menaces en matière de cybersécurité touchant aussi bien les citoyens que les organisations publiques et privées. Face à cette problématique sociétale, le projet a confirmé un besoin urgent, a cherché à comprendre les mécanismes sous-jacents de cette menace, a tenté d'identifier les leviers d'action possibles, et a finalement proposé différentes approches innovantes pour y faire face.

Porté par une collaboration étroite entre des acteurs académiques, institutionnels et industriels du canton de Vaud, le projet s'est distingué par sa dimension interdisciplinaire et son ancrage concret dans la réalité du terrain. Cette diversité d'expertises a permis de croiser les regards techniques, organisationnels, juridiques et humains afin d'aborder le phishing sous un angle global et cohérent.

L'objectif principal a été d'explorer comment renforcer la résilience collective face à cette menace, en combinant compréhension, innovation et sensibilisation.

6.1 Approche et méthodologie

Les partenaires ont adopté une méthodologie innovante, centrée sur la collaboration et l'échange. Plutôt que de travailler de manière individuelle, les membres du projet ont été réunis à plusieurs reprises au sein d'ateliers d'échanges à thématiques (voir Section 3), permettant de confronter directement les points de vue techniques, organisationnels, juridiques et institutionnels autour d'une même table. Ces ateliers ont constitué un espace privilégié d'interaction et de co-construction, où chacun a pu partager son expérience, ses compétences, ses contraintes et ses attentes, ainsi qu'apprendre de celles des autres.

Cette approche a favorisé une compréhension partagée du phénomène du phishing et a permis de faire émerger des pistes d'action concrètes, parmi lesquelles plusieurs Proof of Concepts

(PoCs) ont été imaginés, puis certains développés afin d'en démontrer la faisabilité.

Au-delà des résultats obtenus, cette méthodologie s'est révélée être un modèle reproductible pour d'autres projets d'innovation, en démontrant qu'une collaboration structurée et ouverte entre acteurs publics, académiques et privés pouvait produire des solutions adaptées aux besoins réels du terrain.

6.2 Résultats sur les mesures actuelles.

L'analyse de l'état de l'art (voir Chapitre 2) et les discussions menées lors des ateliers ont permis de dégager plusieurs constats clés. D'un point de vue technique, les mécanismes existants pour contrer le phishing, tels que SPF, DKIM, DMARC, ou encore ARC et BIMI, se révèlent efficaces lorsqu'ils sont correctement déployés, mais leur adoption demeure inégale. De nombreuses organisations, notamment de petite taille, ne les activent pas ou les configurent partiellement, laissant subsister des vulnérabilités exploitables. Surtout, ces mesures purement techniques ne suffisent pas à elles seules. Les attaques montrent que les cybercriminels contournent les défenses techniques en exploitant le facteur humain, que ce soit par la tromperie, la pression sociale ou la manipulation. L'analyse de l'état de l'art (voir Chapitre 2) et les discussions menées lors des ateliers ont permis de dégager plusieurs constats clés. D'un point de vue technique, les mécanismes existants pour contrer le phishing, tels que SPF, DKIM, DMARC, ou encore ARC et BIMI, se révèlent efficaces lorsqu'ils sont correctement déployés, mais leur adoption demeure inégale. De nombreuses organisations, notamment de petite taille, ne les activent pas ou les configurent partiellement, laissant subsister des vulnérabilités exploitables. Surtout, ces mesures purement techniques ne suffisent pas à elles seules. Les attaques montrent que les cybercriminels contournent les défenses techniques en exploitant le facteur humain, que ce soit par la tromperie, la pression sociale ou la manipulation. Par exemple, plutôt que de tenter de contourner la mesure technique SPF pour forger un courriel depuis le domaine légitime *ebanking.mabanque.ch*, un attaquant va enregistrer et utiliser un domaine visuellement similaire, tel que *mabanque.e-banking.ch*, tirant parti de la ressemblance pour tromper l'utilisateur plutôt que le système de protection.

Ces constats rejoignent les retours des partenaires institutionnels et industriels, qui soulignent la nécessité de renforcer la sensibilisation et la formation des utilisateurs. Si les outils techniques contribuent à limiter l'exposition, la vigilance de l'utilisateur reste un maillon déterminant et aujourd'hui encore trop fragile.

Sur le plan légal, l'analyse menée dans le cadre du projet montre que la Suisse ne dispose pas d'une infraction spécifique au phishing, celui-ci étant réprimé au travers de plusieurs articles du Code pénal relatifs notamment à la fraude, à l'accès indu à un système informatique ou à l'utilisation frauduleuse d'un ordinateur. Si ces dispositions permettent de sanctionner certaines formes d'attaques, leur application reste complexe et dépend souvent de l'interprétation des tribunaux. Cette approche morcelée rend difficile une réponse cohérente face à

un phénomène en constante évolution. En complément, la législation sur la protection des données (LPD) impose aux entreprises des obligations de sécurité et de proportionnalité dans le traitement des informations, mais celles-ci relèvent davantage de la conformité que de la prévention active. Ces constats mettent en évidence la nécessité d'un cadre plus clair et mieux coordonné, articulant répression, prévention et sensibilisation des utilisateurs.

6.3 Résultats directs du projet

De nombreuses idées ont été imaginées pour améliorer la situation globale (voir Section 3.2). Ce projet étant limité en temps et en budget, uniquement une partie de ces idées ont pu être réalisées et testées.

Les ateliers ont mis en évidence différents aspects améliorables, notamment un manque d'assistance directe à l'utilisateur, en particulier dans les phases d'incertitude lorsqu'il reçoit un message suspect ou ne sait pas comment réagir.

Cette lacune est d'autant plus visible en Suisse, où, malgré l'existence d'un système d'annonce performant (anitiphishing.ch via l'OFCS), il manque une plateforme centralisée permettant également d'obtenir un accompagnement concret, des informations, conseils, et analyses.

C'est précisément à cette problématique que les deux Proofs of Concept (PoCs) développés (voir Chapitre 4) ont cherché à répondre :

- Application mobile de prévisualisation de liens : intercepte les liens reçus (mail, SMS, messagerie) et affiche un verdict clair avant ouverture — *Phishing détecté* si une source fiable le répertorie, sinon un niveau de risque (*élevé / modéré / faible*) calculé à partir du score renvoyé par le service d'analyse.
- Service d'analyse de liens : API centrale combinant l'interrogation de bases externes (OFCS/PhishDB, Google Safe Browsing, URLhaus) et un modèle d'apprentissage automatique exploitant différentes caractéristiques d'URL (structure, longueur, date de création, etc.) afin de produire un verdict justifié et réutilisable par d'autres outils.

Ces prototypes visent à renforcer l'autonomie de l'utilisateur tout en lui offrant un appui technique accessible, sans qu'il doive disposer de compétences avancées. Ils illustrent la manière dont les partenaires ont su transformer les constats des ateliers en solutions concrètes et transférables à d'autres contextes.

6.4 Perspectives futures

Le phishing reste une préoccupation sociétale majeure et il est nécessaire d'agir. Au-delà des réalisations de ce projets, plusieurs résultats, pistes de réflexion et d'expérimentation ont émergé au fil des ateliers et mériteraient d'être explorées dans le cadre de futurs travaux.

Certaines idées prometteuses n'ont pas pu être concrétisées en raison de contraintes de temps, de budget, de faisabilité technique, ou encore de collaboration. En revanche, elles conservent toute leur pertinence au regard des besoins identifiés sur le terrain et notamment pour la société.

Une autre idée récurrente, soulevée tant dans les ateliers que dans les échanges avec les acteurs de terrain, concerne la création d'une plateforme centralisée de sensibilisation, d'assistance et d'annonce. Ce projet de long terme pourrait centraliser les signalements, proposer des outils de vérification automatisée et offrir un accompagnement aux utilisateurs confrontés à une tentative d'hameçonnage. Sa mise en oeuvre permettrait de renforcer la coordination entre les dispositifs existants et d'assurer une meilleure diffusion de l'information au niveau national. À titre d'illustration, une initiative indépendante nommée Flair (voir Section 2.2.4.6) a vu le jour peu après la fin du projet, proposant une analyse accessible de contenus suspects. Son apparition confirme la pertinence des besoins identifiés lors des ateliers, notamment en matière d'assistance directe aux utilisateurs.

Parmi ces pistes, la question du smishing (phishing par SMS) et du spoofing téléphonique mérite une attention particulière. Ces formes d'ingénierie sociale connaissent une forte montée en puissance, comme en témoignent les signalements croissants relevés par les partenaires institutionnels. Pourtant, elles demeurent encore peu documentées et mal comprises en Suisse, les contre-mesures relevant surtout du domaine des opérateurs de télécommunications. Faute d'expertise spécifique et de ressources adaptées, le consortium n'a pas pu traiter ces aspects en profondeur, mais ils constituent des prolongements naturels du projet, essentiels pour une approche réellement globale de la lutte contre le phishing.

De manière plus générale, les résultats, échanges et prototypes produits dans le cadre de ce projet constituent une base riche pour de futurs travaux. Les idées explorées et les constats partagés offrent une matière concrète à prolonger, que ce soit sous la forme de nouveaux outils, de campagnes de sensibilisation ou d'approches éducatives adaptées aux différents publics.

En définitive, la lutte contre le phishing repose avant tout sur une action collective, où la technique, la coordination et la sensibilisation avancent ensemble pour réduire le taux de succès des attaques et renforcer la résilience des utilisateurs.

Bibliographie

[ADPL21] H. Abroshan, J. Devos, G. Poels, and E. Laermans. Phishing happens beyond technology : The effects of human behaviors and demographics on each step of a phishing process. *IEEE Access*, 9 :44928–44949, 2021. Consulté en 2025.

[ALBK19] Kurt Andersen, Brandon Long, Seth Blank, and Murray Kucherawy. The Authenticated Received Chain (ARC) Protocol. RFC 8617, July 2019.

[Als20] M. Alshaikh. Developing cybersecurity culture to influence employee behavior : A practice perspective. *Computers & Security*, 98 :102003, 2020. Consulté en 2025.

[Amm06] M. Ammann. Sind phishing-mails strafbar ?, 2006. Consulté en 2025.

[Apa25] Apache Software Foundation. Apache spamassassin, 2025.

[APW24] APWG. Phishing activity trends report 3rd quarter 2024, 2024.

[ASD22] M. F. Ansari, P. K. Sharma, and B. Dash. Prevention of phishing attacks using ai-based cybersecurity awareness training. *International Journal of Smart Sensor and Adhoc Network*, pages 61–72, 2022. Consulté en 2025.

[aut25] Plusieurs auteur·rice·s. Caller id spoofing, 2025.

[BAAZ23] P. Burda, A. M. Altawekji, L. Allodi, and N. Zannone. The peculiar case of tailored phishing against smes : Detection and collective defense mechanisms at a small it company. In *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 232–243, Delft, Netherlands, 2023. IEEE. Consulté en 2025.

[BAJ25] S. K. Birthriya, P. Ahlawat, and A. K. Jain. Detection and prevention of spear phishing attacks : A comprehensive survey. *Computers & Security*, 151 :104317, 2025. Consulté en 2025.

[Bak12] J. Baker. The technology–organization–environment framework. In Y. K. Dwivedi, M. R. Wade, and S. L. Schneberger, editors, *Information Systems Theory*, volume 28 of *Integrated Series in Information Systems*, pages 231–245. Springer New York, New York, NY, 2012. Consulté en 2025.

[Bil20] B. B. Bilet. Security awareness training as a countermeasure to phishing attacks, 2020. Consulté en 2025.

[Brs25] Brside. Brside, 2025.

[CDS+24] X. Chen, S. Doublet, A. Sergeeva, G. Lenzini, V. Koenig, and V. Distler. What motivates and discourages employees in phishing interventions : An exploration of expectancy-value theory, 2024. Consulté en 2025.

[Cha23] Chancellerie fédérale. Loi sur la sécurité de l'information (lsi), 2023.

[Cis25] Cisco. Cisco umbrella, 2025.

[CLU14] CLUSIF. Analyse de la norme iso 27035, 2014. Consulté en 2025.

[Con25] Confédération suisse. Code pénal suisse, 2025. Consulté en 2025.

[Cro25] CrowdStrike. Crowdstrike 2025 global threat report, 2025.

[cv24] Police cantonale vaudoise. Rapport annuel vaud 2024 - spc24, 2024.

[cv25a] Police cantonale vaudoise. Démarchage téléphonique ou escroquerie téléphonique ?, 2025.

[cv25b] Police cantonale vaudoise. Les mécanismes psychologiques chez les auteurs de cybercriminalité, 2025.

[cv25c] Police cantonale vaudoise. Polcant info, 2025.

[Dar21] A. Darem. Anti-phishing awareness delivery methods. *Engineering, Technology & Applied Science Research*, 11(6) :7944–7949, 2021. Consulté en 2025.

[Dat22] DataGuard. Iso 27001 - annex a.9 access control, 2022. Consulté en 2025.

[DFM+22] G. Desolda, L. S. Ferro, A. Marrella, T. Catarci, and M. F. Costabile. Human factors in phishing attacks : A systematic literature review. *ACM Computing Surveys*, 54(8) :1–35, 2022. Consulté en 2025.

[dlC21] Prévention Suisse de la Cybercriminalité. Phishing, 2021. Consulté en 2025.

[EPF25] EPFL. Epfl en chiffres, 2025.

[fA25] fido Alliance. Passkey security, 2025.

[falpddealtP23] Préposé fédéral à la protection des données et à la transparence (PFPDT). Protection des données dans les associations, 2023. Consulté en 2025.

[falpddealtP24a] Préposé fédéral à la protection des données et à la transparence (PFPDT). Déclarer les activités de traitement, 2024. Consulté en 2025.

[falpddealtP24b] Préposé fédéral à la protection des données et à la transparence (PFPDT). Sécurité de l'information, 2024. Consulté en 2025.

[falpddealtP24c] Préposé fédéral à la protection des données et à la transparence (PFPDT). Traitement des données par l'employeur, 2024. Consulté en 2025.

[falpddealtP25] Préposé fédéral à la protection des données et à la transparence (PFPDT). Guide sur la notification des violations de la sécurité des données, 2025. Consulté en 2025.

[fdlc22] Office fédéral de la cybersécurité. Hameçonnage (phishing), vishing, smishing, 2022.

[fdlcO24] Office fédéral de la cybersécurité (OFCS). Rapport anti-phishing 2024, 2024. Consulté en 2025.

[GDZ+25] T. Geppert, T. Dudas, S. Zimmermann, T. Sutter, and N. Ebert. How to successfully implement phishing awareness training in organizations : A technology adoption perspective, 2025. Consulté en 2025.

[Glo24] Global Legal Group. Iclg - cybersecurity laws, 2024. Archive Location : United Kingdom, Publisher : Global Legal Group, Consulté en 2025.

[Gop25] Gophish Project. Gophish, 2025.

[Gos20] B. Gossin. Escroquerie et hameçonnage en vue de l'utilisation frauduleuse d'un ordinateur en tant qu'infractions préalables au blanchiment d'argent, 2020. Consulté en 2025.

[GP23] L. Gamisch and D. Pöhn. A study of different awareness campaigns in a company. In *Proceedings of the 18th International Conference on Availability, Reliability and Security*, pages 1–8, Benevento, Italy, 2023. ACM. Consulté en 2025.

[Gro15] Groupe de coordination en matière de blanchiment d'argent (GCBF). Rapport sur l'évaluation nationale des risques de blanchiment d'argent et de financement du terrorisme en suisse, 2015. Consulté en 2025.

[GSR24] GSR. Abus des moyens de télécommunication et réseaux sociaux, 2024. Consulté en 2025.

[Gur23] S. Gurzhii. Aspect organisationnel et technique de la lutte contre le phishing, 2023. Consulté en 2025.

[IBM25a] IBM. Iso/iec 27001 compliance - ibm cloud, 2025.

[IBM25b] IBM. Phishing - ibm topic, 2025.

[JBL22] G. Jayakrishnan, V. Banahatti, and S. Lodha. Pickmail : A serious game for email phishing awareness training. In *Proceedings 2022 Symposium on Usable Security*, San Diego, CA, 2022. Internet Society. Consulté en 2025.

[JDWT17] M. L. Jensen, M. Dinger, R. T. Wright, and J. B. Thatcher. Training to mitigate phishing attacks using mindfulness techniques. *Journal of Management Information Systems*, 34(2) :597–626, 2017. Consulté en 2025.

[JGST20] D. Jampen, G. Gür, T. Sutter, and B. Tellenbach. Don't click : towards an effective anti-phishing training. a comparative literature review. *Human-centric Computing and Information Sciences*, 10(1) :33, 2020. Consulté en 2025.

[KCH11] Murray Kucherawy, Dave Crocker, and Tony Hansen. DomainKeys Identified Mail (DKIM) Signatures. RFC 6376, September 2011.

[KHN+22] J. Kävrestad, A. Hagberg, M. Nohlberg, J. Rambusch, R. Roos, and S. Furnell. Evaluation of contextual and game-based training for phishing detection. *Future Internet*, 14(4) :104, 2022. Consulté en 2025.

[Kit14] Scott Kitterman. Sender Policy Framework (SPF) for Authorizing Use of Domains in Email, Version 1. RFC 7208, April 2014.

[KPM24] D. Köhler, W. Pünter, and C. Meinel. How users investigate phishing emails that lack traditional phishing cues. In C. Pöpper and L. Batina, editors, *Applied Cryptography and Network Security*, volume 14585 of *Lecture Notes in Computer Science*, pages 381–411. Springer Nature Switzerland, Cham, 2024. Consulté en 2025.

[KZ15] Murray Kucherawy and Elizabeth Zwicky. Domain-based Message Authentication, Reporting, and Conformance (DMARC). RFC 7489, March 2015.

[Leg17] Legalis.ch. Commentaire romand - code pénal ii, 2017. Consulté en 2025.

[LKv22] D. Lain, K. Kostiainen, and S. Čapkun. Phishing in organizations : Findings from a large-scale and long-term study. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 842–859, San Francisco, CA, USA, 2022. IEEE. Consulté en 2025.

[MGD17] G. D. Moody, D. F. Galletta, and B. K. Dunn. Which phish get caught ? an exploratory study of individuals' susceptibility to phishing. *European Journal of Information Systems*, 26(6) :564–584, 2017. Consulté en 2025.

[Mic24] Microsoft. Implement passkeys, 2024.

[Mic25a] Microsoft. Exchange online protection (eop) documentation, 2025.

[Mic25b] Microsoft. Manage caller id for users, 2025.

[Mic25c] Microsoft. Microsoft defender for office 365 documentation, 2025.

[Mic25d] Microsoft. Microsoft digital defense report 2025, 2025.

[MMQ17] Macaluso, Moreillon, and Q. Queloz. *Code pénal. 2 : Art. 111-392 CP*. Helbing Lichtenhahn, Bâle, 2017. Consulté en 2025.

[Mon09] G. Monnier. Le piratage informatique en droit pénal, 2009. Seite 141, Consulté en 2025.

[Moz25a] Mozilla. How does phishing and malware protection work ?, 2025.

[Moz25b] Mozilla. Thunderbird's scam detection, 2025.

[Mé17] S. Métille. Internet et droit, 2017. Consulté en 2025.

[Nav25] Navixia SA. Diagnophish – simulation de phishing, 2025.

[Nex25] NextDNS. Nextdns, 2025.

[NKF17] J. D. Ndibwile, Y. Kadobayashi, and D. Fall. Unphishme : Phishing attack detection by deceptive login simulation through an android mobile app. In *2017 12th Asia Joint Conference on Information Security (AsiaJCIS)*, pages 38–47, Seoul, South Korea, 2017. IEEE. Consulté en 2025.

[NLV22] T. H. Nguyen, X. C. Le, and T. H. L. Vu. An extended technology-organization-environment (toe) framework for online retailing utilization in digital transformation : Empirical evidence from vietnam. *Journal of*

Open Innovation : Technology, Market, and Complexity, 8(4) :200, 2022. Consulté en 2025.

[NPF+23] B. Naqvi, K. Perova, A. Farooq, I. Makhdoom, S. Oyedeji, and J. Porras. Mitigation strategies against the phishing attacks : A systematic literature review. *Computers & Security*, 132 :103387, 2023. Consulté en 2025.

[Ofc22] Ofcom. Tackling scam calls and texts, 2022.

[Ofc24] Ofcom. Calling line identification (cli) authentication assessment and future roadmap, 2024.

[PAZ20] S. Pirocca, L. Allodi, and N. Zannone. A toolkit for security awareness training against targeted phishing. In S. Kanhere, V. T. Patil, S. Sural, and M. S. Gaur, editors, *Information Systems Security*, volume 12553 of *Lecture Notes in Computer Science*, pages 137–159. Springer International Publishing, Cham, 2020. Consulté en 2025.

[PEC25] PECB. Iso/iec 27002 - education and certification for individuals, 2025. Consulté en 2025.

[Per21] B. Perrin. Accès indu à un système informatique, soustraction et détérioration de données : contribution à la résolution de quelques questions interprétatives, 2021. Consulté en 2025.

[PG24] G. Prüzelius and T. Gyllner. Phishing in the workplace : Organizational practices, culture and phishing susceptibility, 2024. Consulté en 2025.

[Pol25a] Police cantonale vaudoise. Newsletter de la police cantonale vaudoise, 2025.

[Pol25b] Police cantonale vaudoise. Votrepolice.ch – prévention numérique, 2025.

[Pro25a] Proofpoint. Proofpoint - site officiel, 2025.

[Pro25b] Proofpoint. Proofpoint nexus platform, 2025.

[Pro25c] Proofpoint. Proofpoint platform zen, 2025.

[Pro25d] Proton. What is authenticated received chain (arc) ?, 2025.

[Pro25e] Proton. What is dkim ?, 2025.

[Pro25f] Proton. What is dmarc ?, 2025.

[Pro25g] Proton. What is sender policy framework (spf) ?, 2025.

[Rsp25] Rspamd Project. Rspamd, 2025.

[SCG+24] W. C. Satyro, J. C. Contador, J. A. Gomes, S. F. D. P. Monken, A. P. Barbosa, F. S. Bizarrias, J. L. Contador, L. S. Silva, and R. G. Prado. Technology-organization-external-sustainability (toes) framework for technology adoption : Critical analysis of models for industry 4.0 implementation projects. *Sustainability*, 16(24) :11064, 2024. Consulté en 2025.

[SDPJ22] F. Sharevski, A. Devine, E. Pieroni, and P. Jachim. Phishing with malicious qr codes. In *Proceedings of the 2022 European Symposium on Usable Security*, pages 160–171, Karlsruhe, Germany, 2022. ACM. Consulté en 2025.

[SHJL23] O. Sarker, S. Haggag, A. Jayatilaka, and C. Liu. Personalized guidelines for design, implementation and evaluation of anti-phishing interventions. arXiv :2311.12827 [cs], 2023. Consulté en 2025.

[SLLK24] J. Scott, Y. Levy, W. Li, and A. Kumar. Comparing phishing training and campaign methods for mitigating malicious emails in organizations. In *Proceedings of the 10th International Conference on Information Systems Security and Privacy*, pages 643–651, Rome, Italy, 2024. SCITEPRESS - Science and Technology Publications. Consulté en 2025.

[SMP19] SMPP.org. Smpp, 2019.

[Spa14] M. Spas. Phénomènes cybercriminels, November 2014. Consulté en 2025.

[sui25] Confédération suisse. Loi fédérale sur la protection des données, 2025. Consulté en 2025.

[TG21] F. Teichmann and L. Gerber. Cybercriminalité en suisse – le phishing, 2021. Consulté en 2025.

[Trale] TransNexus. Understanding stir/shaken, non disponible.

[TRE19] TREX. Document juridique sur swisslex, 2019. Consulté en 2025.

[Tsu24] Akaki Tsunoda. Can serious gaming tactics bolster spear-phishing and phishing resilience ? securing the human hacking in information security. *Digital Threats : Research and Practice, Volume 5, Issue 1*, 5 :1 – 13, 2024.

[WC18] R. Wash and M. M. Cooper. Who provides phishing training ? facts, stories, and people like me. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, Montreal, QC, Canada, 2018. ACM. Consulté en 2025.

[Wik24] Wikipedia contributors. Cross-site scripting, 2024. Consulté en 2025, Page Version ID : 213135057.

[WLCA19] Z. A. Wen, Z. Lin, R. Chen, and E. Andersen. What.hack : Engaging anti-phishing training through a role-playing phishing simulation game. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, Glasgow, Scotland, UK, 2019. ACM. Consulté en 2025.

[YFJ+24] A. Yasin, R. Fatima, Z. JiangBin, W. Afzal, and S. Raza. Can serious gaming tactics bolster spear-phishing and phishing resilience ? securing the human hacking in information security. *Information and Software Technology*, 170 :107426, 2024. Consulté en 2025.

Table des figures

Annexe A

Partenaires et contributeur-trice-s

Pour chaque partenaire, les personnes ayant contribué au projet sont listées ci-dessous. Les contacts principaux sont indiqués en gras.

Haute École d'Ingénierie et de Gestion du Canton de Vaud (HEIG-VD)
Route de Cheseaux 1, 1400 Yverdon-les-Bains
Contributeurs : **Sylvain Pasini**, Alexis Martins, Axel Vallon, Pablo Saez

École Polytechnique Fédérale de Lausanne (EPFL), *1015 Lausanne*
Contributeur-trice-s : **Linus Gasser**, Carine Dengler

Université de Lausanne (UNIL), *1015 Lausanne*
Contributeurs : **Thomas Souvignet**, Johann Polewczyk, Valentin Diaz

Direction Générale du Numérique et des Systèmes d'Information (DGNSI)
Avenue de Longemalle 1, 1020 Renens
Contributeurs : **Guillaume Fumeaux**

Police Cantonale Vaudoise (PCV), *Chemin de la Lanterne 2, 1052 Le Mont-sur-Lausanne*
Contributeur-trice-s : **Adrien Schopfer**, Bérangère Jacquart, Chloé Berthet

Navixia SA, *Route du Bois 1, 1024 Ecublens*
Contributeur : **Tristan Leiter**

Transports Lausannois (T-L), *Chemin du Closel 15, 1020 Renens*
Contributeur : **Serge Miéville**