Université Abou Bekr Belkaid
Tlemcen Algérie
جامعة أبي بكر بلقايد
تلمسان الجزائر

République Algérienne Démocratique et Populaire
Université Abou Bakr Belkaid– Tlemcen
Faculté desSciences
Département d'Informatique

Mémoire de fin d'études

pour l'obtention du diplôme de Master en Informatique

Option: *Modèle D'intèligent Et Decision(M.I.D)*

Thème

UNE APPROCHE A BASE D'URL POUR LA DETECTION DES SITES PHISHING

Réalisé par :

- **Boursali Djamel**

Présenté le 16 octobre 2014 devant le jury composé de MM.

- *Boudefla Amine* *(Président)*
- *Belabed Amine* *(Encadreur)*
- *Marzoug Mohamed* *(Examinateur)*
- *Smahi Ismail* *(Examinateur)*

Année universitaire :2013-2014

Remerciement

Avant tout, je rends grâce à DIEU tout puissant de m’avoir accordé la volonté et le courage pour réaliser ce mémoire.

Au terme de ce travail je tiens tout d’abord à exprimer ma profonde gratitude à mon encadreur Mr Belabed Amine.

Qui ma guidé tout au long de ce travail.

Mes remerciements s’adressent à tous les membres du jury

pour l’honneur qu’ils m'ont fait en acceptant

Dédicace

Je dédie ce modeste travail.

À mes très chers parents

Que Dieu les garde

À toute ma famille et mes amis

À tous ceux qui sont proches de mon coeur

et dont je n’ai pas cité les noms

Table Des Matiéres

Chapitre 3 : Implemention

Table Des Matiéres

Le phishing (contraction des mots anglais « fishing », en français pêche, et « phreaking », désignant le piratage de lignes téléphoniques), traduit parfois en « hameçonnage », est une technique frauduleuse utilisée par les pirates informatiques pour récupérer des informations. Le phishing repose sur les techniques d'ingénierie sociale en usurpant des identités d'autres personnes (physiques ou entreprise) via l'envoi de mails qui incitent les utilisateurs à fournir des informations sensibles au pirate.

Le but de notre mémoire est de comprendre les concepts de phishing, ces techniques, et construire une application qui testé si les URLs sont ligimes au sont phishing.

Le présent mémoire est organisé comme suite :

Dans le premier chapitre qui constitue une introduction a notre travail nous exposons les différentes définitions de concepts, le jargon de ce monde.

Dans le deuxième chapitre nous exposons les différentes techniques du phishing, ainsi que les méthodes utilisées pour contrer ces attaques.

Dans le troisième et dernier chapitre nous adoptons une approche de lutte contre le phishing et nous essayons de la mettre en œuvre, avec la création d'un système capable de différencier des adresses légitimes et des adresses de phishing .

1. INTRODUCTION :

Depuis des siècles, la circulation des informations dans le monde a étéembryonnaire .Mais avec l'avènement des progrès scientifiques et techniques, l'on a connu des appareils et machines tels que les radios, les téléphones, les ordinateurs et autres permettant la circulation facile des informations .Ainsi, Internet qui fait l'objet de notre étude voit le jour dans les années 60 en Amérique. Il obéit à des logiques bien précises du genre Plus rapide, plus loin et fiable. Celui –ci de part ses multiples avantages, présente aussi des inconvénients.

2. Définition d'internet : [1]

Internet est un réseau international d'ordinateurs qui communiquent entre eux grâce à des protocoles d'échanges de données standard. Cette communication en réseaux se fait indépendamment des types d'ordinateurs utilisés (Mac, PC, Unix ou autres).

Internet est un outil de communication qui utilise les fils téléphoniques, les fibres optiques, les câbles intercontinentaux et les communications par satellite. Il rend accessibles au public des services comme le courrier électronique et le World Wide Web.

3. Une histoire de l'Internet : [2]

L'ambition d'Internet s'exprime en une phrase : relier entre eux tous les ordinateurs du monde à l'image du téléphone qui permet de converser avec toute personne dont on connaît le numéro, Internet est un système mondial d'échange de documents électroniques : textes, fichiers, images, sons et séquences audiovisuelles. C'est l'alliance de l'informatique et des télécommunications la télématique au véritable sens du terme, selon un mot français peu usité. Les utilisateurs d'Internet sont désignés par le terme d'internautes, synonyme de cybernaute, de surfer ou de netsurfer. Quant aux informations du réseau, elles sont accessibles à partir de « lieux » que l'on appelle les sites Internet.

4. Danger d'internet : [3]

- **Les jeunes : un public ciblé et mal protégé :**

Les jeunes forment la population la plus exposée et la plus ciblée par les prédateurs de toutes sortes sur le Net. C'est en effet un media de référence : en 2003 plus de 87% des 12/17 ans se sont connectés à Internet et le phénomène s'amplifie avec le

développement de l'Internet mobile. La toile est un espace de tentation, un espace ludique mais aussi un espace à risques dans lequel ils pénètrent d'un simple clic ! Voyons les dangers qui les guettent : Des dangers variés Invasion de la vie privée.

La capacité d'interagir et de communiquer avec les autres est un des grands attraits qu'Internet exerce sur les jeunes. Ils aiment discuter dans les messageries instantanées, jouer en ligne et remplir des formulaires pour participer à des concours et sondages. La vie privée des jeunes peut être envahie de différentes manières. C'est le cas, par exemple, quand ils :

- remplissent des formulaires pour participer aux concours des sites Web commerciaux
- donnent des informations les concernant à des inconnus rencontrés dans un chat ou une messagerie instantanée.
- donnent des informations personnelles lors de leur inscription à divers services Internet ou logiciels (messageries instantanées, chat, partage de fichiers, etc.) ;
- fournissent leur profil personnel lors de leur inscription à des comptes de mail ou messageries instantanées gratuits ;

➢ **Cyberintimidation :**

Il existe différentes formes de cyberintimidation parfois, il s'agit d'insultes ou de menaces directement envoyées à la victime par maill ou messagerie instantanée. Les jeunes peuvent aussi faire circuler des commentaires haineux visant une personne, par le biais du maill et des messageries instantanées, ou en les affichant sur des sites Web. Ils le font souvent sous une fausse identité. Par ailleurs, de plus en plus de jeunes sont victimes d'intimidation par le biais de messages textes envoyés sur leur cellulaire. Ce type de téléphone échappe en effet complètement à la surveillance des adultes.

➢ **Spam ou pourriels :**

On entend par pourriels tous les messages non sollicités, publicitaires ou non, qui envahissent de plus en plus nos boîtes à lettres électroniques. Ce fléau s'est développé de manière spectaculaire. en 2002, en Amérique du Nord seulement, le nombre de pourriels a décuplé, et on estime qu'il représente désormais près de la moitié de tout le mail.

- **Contenus violents ou haineux :**

On retrouve sur Internet un univers de violence qui va de pages Web où règne un humour cruel typiquement adolescent à des sites qui n'hésitent pas à diffuser des images de torture et de sadisme. D'un simple clic de souris, les jeunes peuvent télécharger de la musique aux paroles très violentes (parfois censurées dans les disques vendus en magasin) et des images, vidéoclips et jeux en ligne tout aussi inquiétants. Les jeunes aiment également les sites « sanglants » qui montrent des images réelles d'accidents, de tortures ou de mutilations. Beaucoup d'adolescents considèrent ces sites comme inoffensifs, l'équivalent en ligne des films d'horreur, mais on y trouve une inquiétante combinaison de violence et de sexualité.

- **Jeux de hasard :**

La prolifération des jeux de hasard et des sites de paris sur Internet n'a fait qu'augmenter le nombre impressionnant de jeunes qui s'adonnent au jeu. C'est devenu chez les adolescents une addiction plus importante que la cigarette, l'alcool ou les drogues. Les jeunes qui maîtrisent bien les nouvelles technologies se tournent de plus en plus vers les sites Internet de jeux de hasard parce qu'ils sont faciles d'accès, pratiques et anonymes.

5. Attaque d'internet : [4]

- Les hameçonnages (phishing) et les numéroteurs (dialers) peuvent entraîner des achats non autorisés et facturés sur votre carte de crédit ou causer des frais de téléphone pour plusieurs centaines de francs. Il est souvent difficile de convaincre sa banque ou l'opérateur de téléphonie que les transactions effectuées étaient illicites.
- Envoi de messages non sollicités (SPAM) :

Un spam est un courrier non sollicité, envoyé en masse et sans ciblage particulier par l'expéditeur. C'est le volume qui importe, avec l'idée que sur la quantité de destinataires, certains auront le réflexe de répondre. Sur Internet, plus de 90% des e-mails sont du spam ! Un véritable déluge, qui pollue les boîtes électroniques. D'où l'appellation de « pourriel » ou « courrier indésirable ».

- **D'oùviennent-ils :**

- Le spam utilise votre adresse électronique sans votre consentement. Elle peut avoir été récupérée sur une liste d'e-mailing que se revendent des sociétés spécialisées ; elle peut avoir été trouvée sur Internet par un logiciel conçu exprès pour aller « à la pêche » aux e-mails. D'autres logiciels fabriquent automatiquement des adresses en combinant au hasard des noms, des prénoms et des noms de service d'e-mail (Gmail, Hotmail, Yahoo).

➢ Les virus, une nuisance :

On emploie très souvent le terme de virus informatique mais le véritable terme technique est malware. Un malware, ou logiciel malveillant (ou encore maliciel), est un programme qui s'installe à votre insu sur votre ordinateur pour, selon les cas, en perturber le fonctionnement, détruire ou voler des données, permettre à quelqu'un d'en prendre le contrôle à distance voire d'injecter un virus informatique. Les malwares de type « cheval de Troie » ouvrent un accès à votre ordinateur.

6. Solution d'attaque : [5]

- ne pas ouvrir des pièces jointes d'un expéditeur inconnu.
- ne pas relayer n'importe quel message d'alerte (hoax).
- ne pas croire au Père Noël venant par spam.
- être attentif à l'adresse du site, surtout si de l'argent est en jeu.
- ne pas répondre à des demandes de codes par mail (phising).
- sur les réseaux sociaux, savoir distinguer ce qu'on veut rendre public de ce qu'on réserve à un cercle restreint.
- ne pas oublier que ce qui est en ligne un jour est en ligne toujours.

Devoirs:

- respecter la propriété intellectuelle.

- ne pas diffuser de message à caractère illégal.

7. Sécurité de l'internet : [6]

Internet permet à des millions de d'ordinateurs de communiquer librement. C'est une grande force, mais c'est aussi une grande faiblesse car cette liberté a une contrepartie. Chaque utilisateur doit mettre en place les procédures de sécurité adaptées à son cas particulier. La réglementation française était très restrictive. L'annonce récente au mois de Janvier 1999 par le Premier Ministre de la libéralisation totale des procédés de chiffrement va permettre à tous d'utiliser les meilleures technologies disponibles sur Internet.

Dun point de vue technique, la sécurité recouvre un vaste ensemble de moyens qui contribuent à la sécurité. On nous concentrerons sur les problèmes posés par la sécurité des informations lors des échanges au travers de réseaux publics comme Internet. La technologie utilisée (TCP/IP) a permis de simplifier la mise en place des réseaux, donc de réduire le coût des communications. En revanche, les fonctions de sécurité ne sont pas traitées directement par ce protocole.

Sécuriser les données, c'est garantir :

- L'authentification réciproque des correspondants pour être sûr de son interlocuteur

- L'intégrité des données transmises pour être sûr qu'elles n'ont pas été modifiées accidentellement ou intentionnellement.

- La confidentialité pour éviter que les données soient lues par des systèmes ou des personnes non autorisées

- Le non répudiation pour éviter la contestation par l'émetteur de l'envoi de données

8. Solution pour protéger vos infos et votre vie privée sur Internet :

[7]

8.1. Appliquez en permanence une protection de base sur votre ordinateur : installez un logiciel de protection contre les programmes malveillants et mettez-le régulièrement à jour. Les programmes malveillants sont vos plus grands ennemis. Si la plupart des ordinateurs sont aujourd'hui équipés d'un programme antivirus, nombreux sont les utilisateurs qui n'effectuent pas les mises à jour régulières .

Tout comme pour la vaccination à réitérer chaque année pour nous défendre contre les nouvelles souches de la grippe, il faut absolument mettre à jour sa protection antivirus pour protéger correctement son ordinateur des nouveaux virus, vers, chevaux de Troie, logiciels espions et autres programmes néfastes. Cette protection de base constitue la première ligne de défense contre les programmes malveillants, sans cesse plus sophistiqués.

8.2. Choisissez bien ceux avec qui vous échangez des informations sur le Web. si le site Web que vous visitez est la vitrine d'une entreprise avec qui vous n'hésiteriez pas à traiter en personne, vous ne risquez pas grand chose. Si l'entreprise en question n'existe qu'en ligne, vérifiez sa réputation et sa notoriété. Assurez-vous également que l'adresse du site débute par https:// quand vous arrivez à l'étape de saisir des données personnelles ou coordonnées bancaires. Lorsque vous surfez sur des boutiques en ligne ou des sites Web spécialisés, vérifiez qu'il existe un numéro de téléphone pour contacter le service client et testez-le avant d'effectuer vos achats.

8.3. Dans la mesure du possible, n'autorisez pas les boutiques en ligne à stocker vos coordonnées bancaires : .

résistez à l'appel de la facilité et évitez de confier vos coordonnées bancaires à un site d'e-commerce. Nombreux sont ceux qui vous diront que cette procédure accélérera vos prochaines transactions. Sachez toutefois que les informations que vous divulguerez seront stockées sur .

8.4. les serveurs de ladite entreprise, ceux de leur fournisseur de services Cloud ou masquées sur votre propre système. Si le risque existe dès lors que vous divulguez des données, celui-ci sera moindre si vous saisissez les infos à chaque fois que si vous acceptez que vos données soient stockées quelque part.

8.5. N'utilisez pas les mêmes identifiants et mots de passe sur toutes les applicationsetsitesWeb.

utilisez un identifiant et/ou mot de passe différent pour chaque application. La plupart des applications et services d'entreprises et d'administrations sont protégés par un mot de passe. Or les mots de passe peuvent se révéler être les maillons les plus faibles de la chaîne de sécurité. La plupart des gens n'utilisent souvent qu'un seul mot de passe, toutes applications confondues, de crainte de l'oublier. Aussi, dès qu'un pirate parvient à subtiliser un mot de passe, il y a de grandes chances qu'il en devine un bon nombre d'autres, si ce n'est l'ensemble, de vos combinaisons d'accès.

8.6. Méfiez vous des attaques d'ingénierie sociale.

ne divulguez jamais d'informations personnelles si vous avez le moindre doute. Les attaques d'ingénierie sociale sont souvent perpétrées par e-mail. Ces e-mails, légitimes en apparence, peuvent parfois piéger même les plus méfiants d'entre nous. Ne répondez jamais par e-mail à une demande de données confidentielles, comme votre numéro de sécurité sociale, vos coordonnées bancaires ou d'autres informations sensibles. Souvent, les e-mails de « phishing » comportent un lien à suivre pour obtenir des instructions complémentaires. Ce lien aboutit souvent à un site qui semble fiable. Mais ce n'est généralement qu'un stratagème de plus inventé par les cybercriminels pour dérober vos données.

9. Conclusion :

Dance ce chapitre on a vue la vie privé d'internet commençant par définition et histoire ainsi que les risque et ses attaque, ensuite on a parlé et centrer sur la sécurité d'internet et problèmes posés, enfin on a donné quelque solution possible pour protéger nos infos sur internet.

1. Introduction :

Les sites Web sont les outils les plus courants pour l'e-commerce et ils sont au cœur des affaires sur Internet. Cependant, malgré la vaste utilisation de sites Web par les utilisateurs et les industries quotidienne et le développement basé sur les besoins de ceux utilisateurs, les failles de sécurité des sites sont encore apportent divers dangers et des charges financières à leurs sociétés. Ainsi, la sécurité web comme l'une des questions fondamentales de l'internet d'aujourd'hui est celui qui est très important et devrait être traitée en conséquence. [8]

Au cours des dernières années, le nombre d'attaques de phishing a augmenté de façon spectaculaire. Le but de ces attaques sont à exploiter les informations sensibles de l'utilisateur telles que les numéros de cartes de crédit ou de compte bancaire de l'utilisateur de voler de l'argent de leur part. Le manque de solutions robustes contre le phishing est évident. Ainsi, les chercheurs explorent des solutions qui peuvent répondre efficacement phishing, compte tenu du fait que la fourniture de solutions efficaces contre les sites de phishing est important de nos jours.

2. Définition du Phishing :

Le sens de l'hameçonnage en Webopedia [9] est "Le fait d'envoyer un e-mail à un utilisateur prétendant faussement être un enterprize légitime établie dans une tentative d'escroquerie à l'utilisateur en livrant des informations privées qui seront utilisées pour le vol d'identité. "lephishing est une manière de voler les informations des utilisateurs avec l'utilisation des astuces et des techniques d'ingénierie sociale. Les informations volées comprend des cartes de crédit, noms d'utilisateur et mots de passe qui pourraient être utilisés afin d'accéder au compte bancaire de l'utilisateur de voler de l'argent

L'objectif principal d'attaques de phishing est de voler des personnels de valeur ou informations liées à l'identité des utilisateurs. Une fois que les attaquants gagnent des données personnelles telles que mots de passe, date de naissance et les numéros de compte bancaire, ils utilisent l'information à leur avantage en créant des identités fausses ou prendre le contrôle de comptes en ligne .

Les attaques de phishing sont lancées dans de nombreuses façons différentes. La façon la plus courante consiste à envoyer un courrier électronique aux utilisateurs et convaincre les utilisateurs à cliquer sur un lien forgé dans l'e-mail. Lorsque l'utilisateur clique sur le lien forgé, l'utilisateur seront envoyés à un site contrôlé par l'attaquant qui ressemble à un site de confiance (par exemple, la banque en ligne de l'utilisateur). Depuis que le site est sous le contrôle de l'attaquant, toute l'information révélée par l'utilisateur sur le site tel que noms d'utilisateur et mots de passe sont obtenus par l'attaquant. Un utilisateur vigilance peut détecter la différence entre le faux site créé par l'attaquant et le site légitime. Cependant, une population non négligeable des utilisateurs ne sont en proie à ces attaques. Par exemple, les sites Web légitimes (selon la complexité de l'institution) peuvent avoir des messages de bienvenue qui incluent le nom de l'utilisateur à partir de connexions réussies précédentes. Ces informations de connexion n'est pas disponible pour le faux site et par conséquent il ne peut pas personnaliser le message d'accueil.

2.1.La prévalence des attaques de phishing

Le nombre d'attaques de phishing est en augmentation rapide [10]. En raison de leur valeur, les institutions financières sont les cibles favorites des pirates de phishing. Récemment, cette forme d'attaque s'est diversifiée et a commencé à cibler les sites de réseautage social ainsi [10]. En outre, les technologies adoptées par les attaquants deviennent plus sophistiqués chaque jour.

La figure 2.1 présente les faits saillants des statistiques obtenues par une étude récente sur les attaques de phishing. Selon des études récentes, la disponibilité moyenne pour un site de phishing est de 4,5 jours [11] .

Statistical Highlights for Q2 2008	April	May	Jun
Number of unique phishing email reports received by APWG from consumers	24,924	23,762	28,151
Number of unique phishing websites detected	20,410	20,317	18,509
Number of brands hijacked by phishing websites	276	294	227
Country hosting the most phishing websites	China	Turkey	US
Certain some form of target name in URL	28.3%	23.2%	26.1%
No hostname; just IP address	5.5%	13.2%	4%
Percentage of sites not using port 80	.81%	.45%	.49%
Longest time online for website	30 days	31 days	30 days

Fig. 2.1 Faits saillants d'un rapport du Groupe de travail de l'Anti Phishing [11].

Les figures 2.2, 2.3, 2.4 montrent les statistiques relatives aux attaques de phishing qui ont été recueillis par le Groupe de travail de l'Anti Phishing (AWPG) [11].

Les dommages financiers d'attaques de phishing est sévère et choquant. Basé sur les statistiques de Gartner Institute [12] Plus de 56 millions de personnes ont reçu un courriel d'hameçonnage en 2008 et 1,85 million de personnes ont divulgué leurs informations personnelles et financières vers des sites de phishing. Les banques et les sociétés émettrices de cartes de crédit ont perdu plus de 1,3 milliards de dollars en raison d'attaques de phishing [12]. Également 5% des utilisateurs d'Internet sont victimes d'attaques de phishing aux États-Unis chaque année [12] .

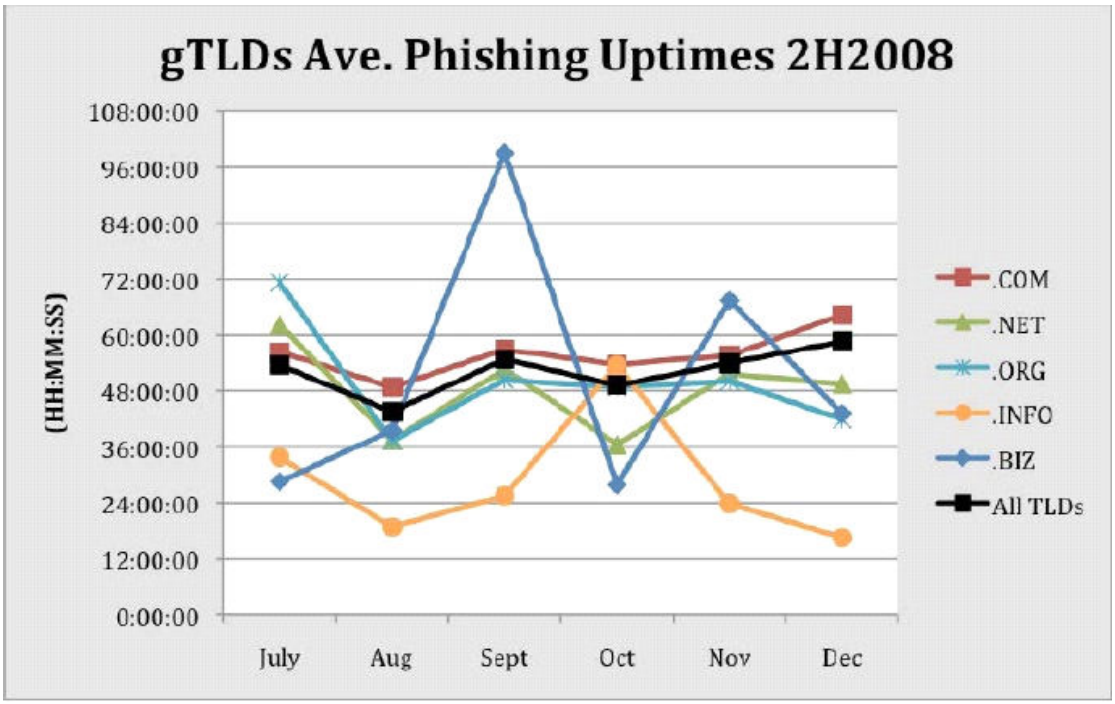


Fig. 2.2 moyenne des temps de bon fonctionnement de phishing pour le second semestre de l'année 2008 [10].

Number of Unique Phishing Web Sites Detected	20,410	20,317	18,509
Unique Domains	6,176	5,849	5,633
Unique Brand-Domain Pairs	7,656	7,267	6,768
Unique Brands	276	294	227
URLs Per Brand	74	69	82

Fig. 2.3 Faits saillants d'un rapport du Groupe de travail de l'Anti Phishing [10].

Fig. 2.4 La plupart des industries ciblées par les attaques de phishing [10].

En raison de la gravité des attaques de phishing, nous avons besoin d'avoir une connaissance très approfondie de ces types d'attaques.

2.2.L'étape d'attaque de phishing:

Nous devons étudier précisément les processus et les étapes d'attaques de phishing afin de trouver une solution globale pour eux. En général, les attaques de phishing sont constituées de 4 étapes [13]. Les attaquants vont suivre les étapes suivantes après la conception d'un faux site très similaire aux sites de confiance des utilisateurs:

- **Distribution des liens malveillants** Dans cette étape, les attaquants essaient d'envoyer à l'utilisateur un des liens malveillants par courrier électronique ou la messagerie instantanée. La désignation des utilisateurs comme des cibles de phishing peut être fait intentionnellement, si les attaquants connaissent les victimes ou il peut être fait au hasard.

- **Visiter les sites de phishing** Dans cette étape, la victime va cliquer sur un lien malveillant qui a été reçu par une sorte de canal de communication. Par conséquent, la victime sera redirigée vers le site de phishing avec une apparence similaire au site réel que l'utilisateur sache.

- **Révéler des informations sensibles** Dans cette étape, en raison de la complète similitude entre le faux site Web et le site réel, l'utilisateur sera persuadé de divulguer leurs renseignements personnels et financiers comme il / elle le fait sur le site réel.

Transfert des informations divulguées à l'attaquant Après la divulgation de l'information au faux site Web par l'utilisateur, l'information sera envoyé aux assaillants et l'identité de l'utilisateur est prise en otage par les assaillants.

Pour une attaque de phishing avec succès toutes les 4 étapes ci-dessus doivent être remplies. Prévention techniques pour une seule étape conduira à l'échec de toute l'attaque de phishing .

Pour cette raison, plusieurs méthodes de prévention ont été proposées pour chacune des étapes ci-dessus et certains d'entre eux seront abordées plus tard.

3. Anatomie d'hameçonnage URL

L'anatomie des URL de phishing a fait l'objet de la recherche anti-phishing pour protéger les utilisateurs contre les nouveaux types d'attaques de phishing et de développer de nouveaux outils pour lutter contre eux. Phishing URL peuvent être classés dans les catégories [14, 15] qui suivent.

• représentation explicite: Par exemple, l'URL réelle est https://onlinebanking.

bankofoklahama.com mais l'URL de phishing affiché dans la barre du navigateur

http: / / www.zglobia.com / Oklahoma / index.php.

• Représentation similaire: Par exemple, l'URL est légitime www.us-bank.com alors que l'URL frauduleuse est www.usbank.com.

• Représentation usurpés: URL présentée dans la barre d'adresse du navigateur de l'utilisateur est le même que légitime. Par exemple, lorsque l'utilisateur accède à un site Web d'hameçonnage, JavaScript rend l'URL visuelle identique à l'URL légitime même si l'URL est différente. Compte tenu de ces approches, il ya beaucoup de possibilités différentes pour la création d'URL de phishing. Voici quelques-unes des techniques d'exemple utilisé pour du phishing sur Internet [15]:

• Ajout d'un suffixe au nom de domaine d'une URL. Par exemple changement www.citybank.com à www.citybank.us.com

.Lien réel qui est définie dans le code source de la page est différente de lien visible. Par exemple , la ligne HTML . <AHREF="http://www.citibank.com.us.ebanking">www.citibank.com</ a> représente un lien visible comme http:/ / www.citibank.com mais le lien réel est http://www.citibank.com.us.e-banking.

• Utilisation d'un bug dans une application Web pour rediriger l'utilisateur vers la page Web d'hameçonnage .

Par exemple, un bug sur le site Web d'eBay peut être utilisé comme ceci: http://cgi.ebay.com/ws/eBayISAPI.dll?fcISAPICommand=RedirectToDomain\&DomainUrl=PHISHINGLINK

pour rediriger les utilisateurs vers un site Web spécifié dans le paramètre . PHISHINGLINK.

• Remplacement de caractères similaires dans le lien réel comme , WWW.CITIBANK.COM à WWW.C1TlBANK.COM que le numéro 1 et le minuscule de la lettre L est utilisé à la place de deux lettres I.

• Encodage le lien pour dissimuler sa vraie valeur à l'aide hexadécimal, DWORD ou octal

coder comme, http://www.visa.com @% 32% 32% 30% 2E% 36% 38% 2E% 32% 31% 34%2E% 32% 31% 33 qui se traduit par http://220.68.214.213 au lieu de

http:/ / www.visa.com.

• L'URL est en fait une partie d'une image, qui utilise les coordonnées de la carte pour définir. la zone de clic. Dans ce cas, une image à l'intérieur d'une page web comporte plus d'un lien qui est défini pour la pression certaine partie d'une image.

• Utilisation d'un service de masquage d'URL comme cjb.net ou tinyurl.com. par exemple

http://jne9rrfj4.CjB.neT/?uudzQYRgY1GNEn peut cacher l'URL réelle et rediriger l'utilisateur vers le site Web d'hameçonnage.

En utilisant le signe "@" dans l'URL. Tout à gauche du "@" est ignorée quand tout à droite est considérée comme une adresse. Par exemple, http://www.usbank.com/update.pl @ 81.109.43.102/usb/upd.pl redirige l'utilisateur à 81.109.43.102/usb/upd.pl au lieu de http://www.usbank.com/ update.pl.

• Utilisation d'une chaîne de texte à la recherche crédible dans l'URL. Par exemple, http://81.109.43.102/ebay/accountupdate/now.php qui utilise le mot clé eBay dans l'URL que l'utilisateur convaincre que c'est une URL légitime .

4. caractéristique de site de phishing :

Un site typique de phishing a les caractéristiques [13] suivantes:

1. Utilisation du logo de la célèbre société et similitudes visuelles de la société de site.

2. L'utilisation de liens réels vrai site de la compagnie dans une partie du faux site.

3. L'utilisation de codes JavaScript pour cacher l'adresse réelle ou des liens réels du faux site.

4. L'utilisation de l'adresse IP dans l'URL.

5. L'utilisation d'un service d'URL de redirection comme URL minuscule.

6. L'utilisation d'un numéro de port différent plutôt que ceux par défaut, tels que 8080 au lieu de 80.

7. L'utilisation du certificat SSL comme, https://www.ebay.com @ 212.45.20.12 /ebay pour indiquer la sécurité en utilisant le protocole HTTPS afin d'attirer l'attention des utilisateurs.

8. Fonctionnalité principale est de rassembler les informations de l'utilisateur seulement, sans fournir d'autres services.

9. Exiger l'installation de certains logiciels pour accéder au site.

10. Barre d'adresse fausse qui montre l'URL de la véritable compagnie pendant que vous êtes sur le site phishing.

11. Utilisation d'un menu Pop-Up et redirection immédiate sur le site réel de l'entreprise après l'achèvement du vol d'informations. Dans un tel cas, l'attaquant informe l'utilisateur que des actions telles que l'activation d'un certain service auront besoin de plus de temps pour être en vigueur puisque l'utilisateur peut ne pas remarquer immédiatement les changements sur le site réel.

12. La désactivation du clic droit de la souris ou d'autres fonctionnalités habituelles de clavier.

13. Traitement réel telles que la validation de carte de crédit données pour être sûr que les chiffres ou les informations qui a été saisi par l'utilisateur sont correctes.

14. L'utilisation de texte caché qui peut être trouvé dans le code source de la page Web afin d'éviter des solutions de filtrage à base de texte.

Les caractéristiques ci-dessus sont les plus courantes tandis que des fonctionnalités supplémentaires sont révélée par de nouvelles attaques de phishing [11].

4.1.EXEPMLE DE PHISHING :

4.1.1. PHISHING PAR PIECE JOINTE :

C’est une méthode très utilisée dans le domaine

le phisheur envoi un email contenant une pièce jointe contenant un formulaire à remplir par la victime, comme l’a subi la SNCF (Société Nationale de Chemins de Fer)(société française)

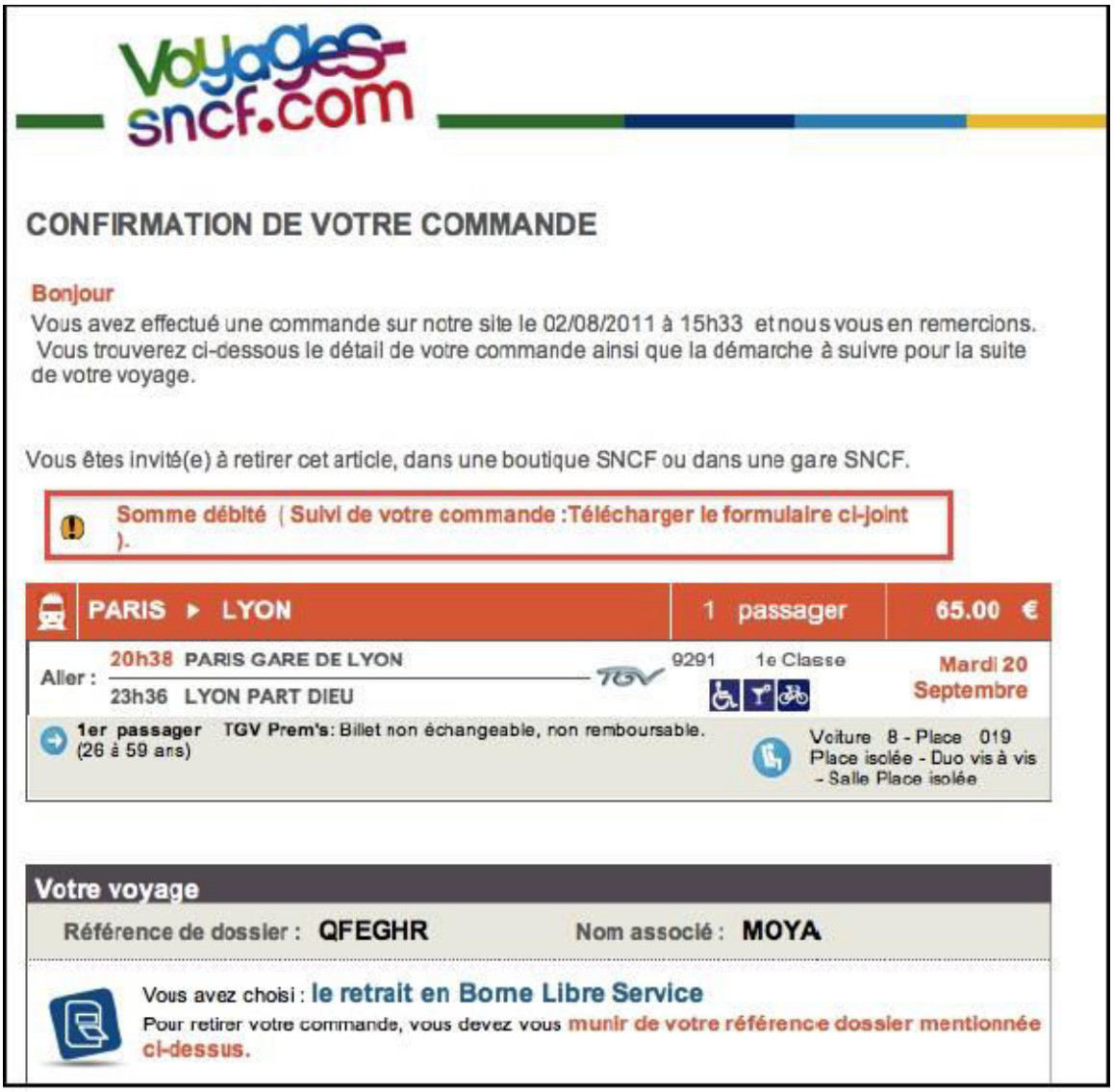
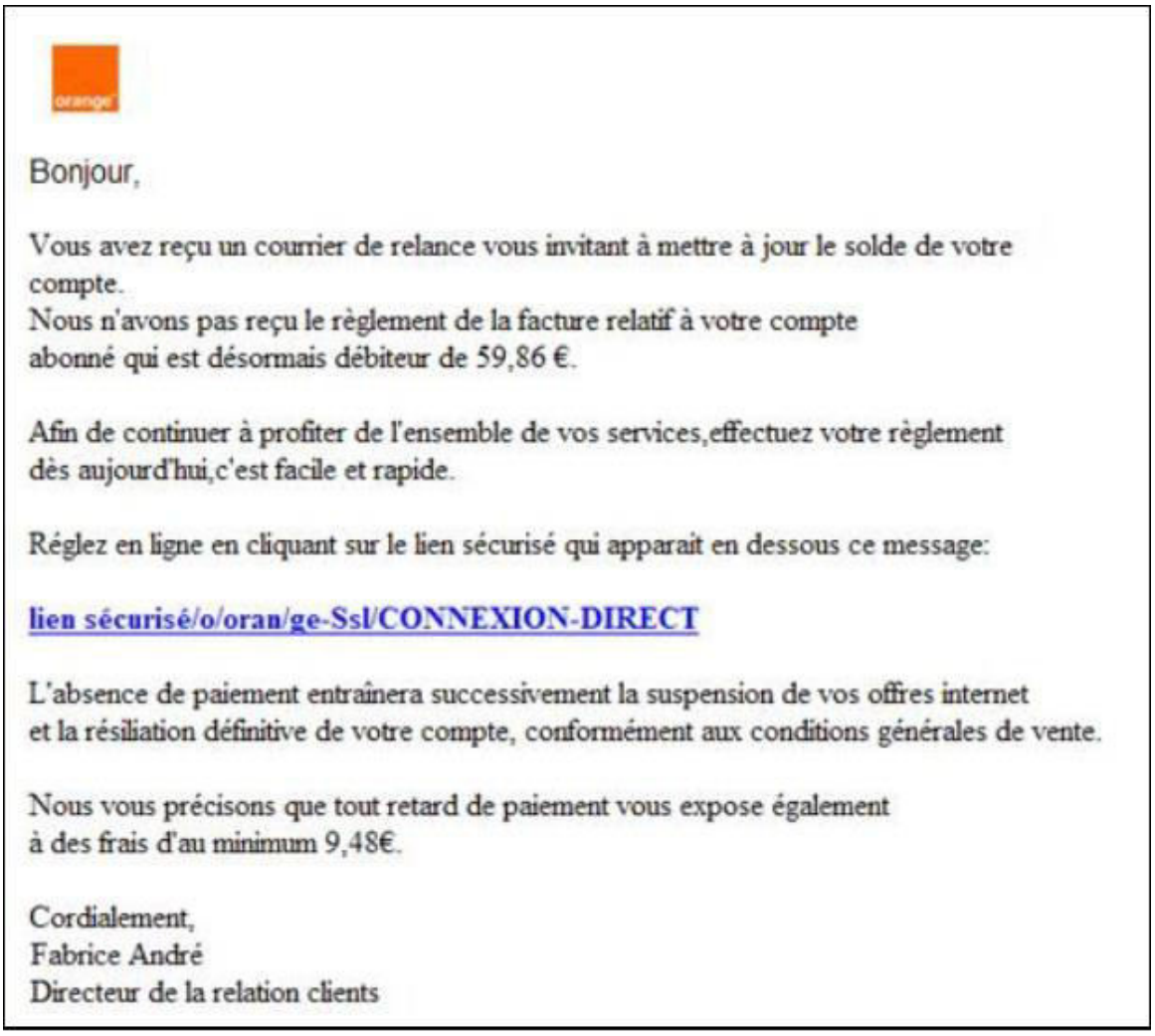


Figure 4.1.1 : Phishing par pièce jointe.

4.1.2. PHISHING PAR REDIRECTION VERS DES SITES MALICIEUX

Dans ce type d'attaques les phishers redirection leurs victimes à de faux sites qui sont identique aux légitimes. Ces sites utilisent souvent JavaScript pour récupérer les informations.



Bonjour,

Vous avez reçu un courrier de relance vous invitant à mettre à jour le solde de votre compte.
Nous n'avons pas reçu le règlement de la facture relatif à votre compte abonné qui est désormais débiteur de 59,86 €.

Afin de continuer à profiter de l'ensemble de vos services,effectuez votre règlement dès aujourd'hui,c'est facile et rapide.

Réglez en ligne en cliquant sur le lien sécurisé qui apparait en dessous ce message:

lien sécurisé/o/oran/ge-Ssl/CONNEXION-DIRECT

L'absence de paiement entrainera successivement la suspension de vos offres internet et la résiliation définitive de votre compte, conformément aux conditions générales de vente.

Nous vous précisons que tout retard de paiement vous expose également à des frais d'au minimum 9,48€.

Cordialement,
Fabrice André
Directeur de la relation clients

Figure 4.1.2 : Phishing par redirection vers un site frauduleux.

5. LES TYPE DES ATTAQUE :

5.1. EMPOISONNEMENT DU CACHE DNS (Cache poisoning) :

Lorsqu'un serveur DNS est obligé d'interroger un autre serveur DNS pour obtenir l'adresse IP d'un nom de domaine faisant l'objet d'une requête, ce qui est le cas le plus général, il stocke temporairement (2 jours en moyenne) le résultat dans sa mémoire cache.

Ceci lui permet de pouvoir fournir immédiatement ce numéro en cas de nouvelle requête. Puisque ce cache est conçu pour recevoir des informations en provenance de l'extérieur, on conçoit que, s'il existe une faille de sécurité sur ce serveur, il soit possible à un pirate d'insérer un nom de domaine connu (par exemple www.google.fr) et lui faire correspondre le numéro IP d'un autre site (site piégé envoyant des programmes malveillants). Un visiteur utilisant ce serveur DNS sera donc redirigé vers le site piégé au lieu d'atteindre le site demandé (Google dans l'exemple choisi). C'est l'attaque par empoisonnement du cache.

5.2. INJECTION DE CODE :

Les attaques de type **Cross-Site Scripting** (notée parfois *XSS* ou *CSS*) sont des attaques visant les sites web affichant dynamiquement du contenu utilisateur sans effectuer de contrôle et d'encodage des informations saisies par les utilisateurs. Les attaques Cross-Site Scripting consistent ainsi à forcer un site web à afficher du code HTML ou des scripts saisis par les utilisateurs. Le code ainsi inclus (le terme « injecté » est habituellement utilisé) dans un site web vulnérable est dit « malicieux ». Il est courant que les sites affichent des messages d'information reprenant directement un paramètre entré par l'utilisateur. L'exemple le plus classique est celui des « pages d'erreur 404 ». Certains sites web modifient le comportement du site web, afin d'afficher un message d'erreur personnalisée lorsque la page demandée par le visiteur n'existe pas. Parfois la page générée dynamiquement affiche le nom de la page demandée. Appelons *http://site.vulnerable* un site possédant une telle faille. L'appel de l'URL *http://site.vulnerable/page-inexistante* correspondant à une page n'existant pas provoquera l'affichage d'un message d'erreur indiquant que la page « page-inexistante » n'existe pas. Il est ainsi possible de faire afficher ce que l'on souhaite au site web en remplaçant « page-inexistante » par toute autre chaîne de caractère. Ainsi, si aucun contrôle n'est effectué sur le contenu fourni par l'utilisateur, il est possible d'afficher du code HTML arbitraire sur une page web, afin d'en changer l'aspect, le contenu ou bien le comportement. De plus, la plupart des navigateurs sont dotés de la capacité d'interpréter des scripts contenus dans les pages web, écrits dans différents langages, tel que JavaScript, VB Script, Java, ActiveX ou Flash. Les balises HTML suivantes

permettent ainsi d'incorporer des scripts exécutables dans une page web : <SCRIPT>, <OBJECT>, <APPLET>, and <EMBED>.

Il est ainsi possible à un pirate d'injecter du code arbitraire dans la page web, afin que celui-ci soit exécuté sur le poste de l'utilisateur dans le contexte de sécurité du site vulnérable. Pour ce faire, il lui suffit de remplacer la valeur du texte destiné à être affiché par un script, afin que celui s'affiche dans la page web. Pour peu que le navigateur de l'utilisateur soit configuré pour exécuter de tels scripts, le code malicieux a accès à l'ensemble des données partagées par la page web de l'utilisateur et le serveur (cookies, champs de formulaires, etc.).

5.3.Attaque man in the middle :

L'attaque « **man in the middle** » (littéralement « attaque de l'homme au milieu » ou «attaques de l'intercepteur »), parfois notée MITM, est un scénario d'attaque dans lequel un pirate écoute une communication entre deux interlocuteurs et falsifie les échanges afin de se faire passer pour l'une des parties. La plupart des attaques de type « man in the middle » consistent à écouter le réseau à l'aide d'un outil appelé sniffer.

5.4.phishing avec pièces jointes :

Une tendance lourde est actuellement en cours chez les phisheurs. De plus en plus d'arnaques mettent en jeu un message frauduleux accompagné d'une pièce jointe comprenant le formulaire de vol d'informations. La SNCF a été victime récemment d'une attaque de phishing de ce type, avec un message potentiellement à même de tromper un certain nombre internautes. Evidemment, le piège demeure plus aisément identifiable si vous n'avez pas commandé récemment de billets de train en ligne ! Des indices peuvent également permettre à un individu vigilant de déceler la tentative d'arnaque. Quelques fautes se sont en effet glissées dans les légères modifications textuelles effectuées par le pirate.

5.5.Phishing par virus (malwares) :

Les keyloggers sont assez rare, les phisheurs installent un programme sur l‟ordinateur de la victime qui enregistre tous ce qui est tapé au clavier y compris les mots de passes.

Pour lutter contre ce type de piratage il faut avoir un antivirus de qualité, et surtout un PC mis à jours!

5.6.Phishing par fenêtres pop-up :

Durant l'utilisation d'un site de confiance sécurisé, une fenêtre pop-up s'affiche invitant l'internaute à réinscrire son identifiant et son mot de passe. Une fois les informations validées, l'instigateur de l'attaque peut les réutiliser. Ce type d'attaque utilise, généralement, un script javascript. Il est techniquement possible pour un script Javascript de déclencher une action si le site prédéterminé est visité en même temps que le site contenant le script. Si c'est le cas, le script javascript se déclenche et ouvre le pop-up. Cette attaque est notamment basée sur la faille de Cross Site Scripting.

5.7. Utilisation de bar d'adresses :

De nombreux sites web d'hameçonnage désactivent la "barre d'adresse" du navigateur, ce qui signifie que vous ne pouvez pas voir l'adresse du site web que vous visitez. Ceci est délibéré, afin que vous ne remarquiez pas que le site que vous consultez est contrefait et n'a pas la bonne adresse.

6. La solution proposée :

6.1.solutions à base de détection :

Les solutions proposées dans cette catégorie sont utilisés après le phishing s'est produite. En d'autres termes, ces solutions ne fonctionnent que quand ils rencontrent des

preuves d'une attaque de phishing. Les mécanismes de détection sont classés en 3 catégories principales [13]

1. Restreindre la réception de spams et les escroqueries.

2. Détection de phishing sites similaires à des sites les plus connus.

3. Filtrage de contenu.

6.1.1. Restriction de la réception de spams et les escroqueries:

Le spam est l'abus de systèmes de messagerie électronique, pour envoyer des messages non sollicités sans discrimination [16]. Escroqueries [8] sont le sous-ensemble de spams qui ne sont pas considérés comme aussi importants que les spams. Bien qu'il existe des solutions qui permettent de détecter les spams, il ya quelques problèmes qui empêchent l'utilisation de solutions de filtrage de spam pour détecter les fraudes de manière efficace. Les escroqueries sont envoyés à un petit nombre de personnes par rapport à spams. Ce facteur empêche la détection des fraudes automatiquement sur la base de la tête de l'email. Les assaillants ont également envoyer des escroqueries par différentes voies dans une grande période de temps afin d'éviter la détection de fraudes à l'aide de filtres anti-spam.

6.1.2. **Détection des sites d'hameçonnage similaires aux sites les plus connus**

Ces solutions tentent d'identifier si le contenu des sites web de phishing est similaire à la contenu hébergé sur les sites célèbres. Observateur du site [18] est un exemple des solutions qui détecte le contenu des similitudes avec ceux à sites célèbres et informe les utilisateurs quand ils veulent visiter ces sites Web. Dans ces solutions, les utilisateurs doivent présenter les sites qui sont tenus d'être protégés contre le phishing au logiciel. Ensuite, toutes les caractéristiques des sites désignés en fonction de leur contenu, tels que des textes, des images, des styles de page et des modèles utilisés, sont introduits dans le système automatiquement par le logiciel. Le système accède ensuite à travers tout le courrier électronique de l'utilisateur et fait enquête sur toutes les pages que les URL dans les e-mails font référence. Si le système trouve des similitudes avec les sites célèbres qui ont été introduites dans le système de protection, il la considère comme un site phishing et élimine l'e-mail correspondant à partir de la boîte de réception de l'utilisateur et informe l'utilisateur .

Les similitudes de contenu textuel que ces solution utilise s'appuie sur des similitudes de mots clés entre les deux pages.

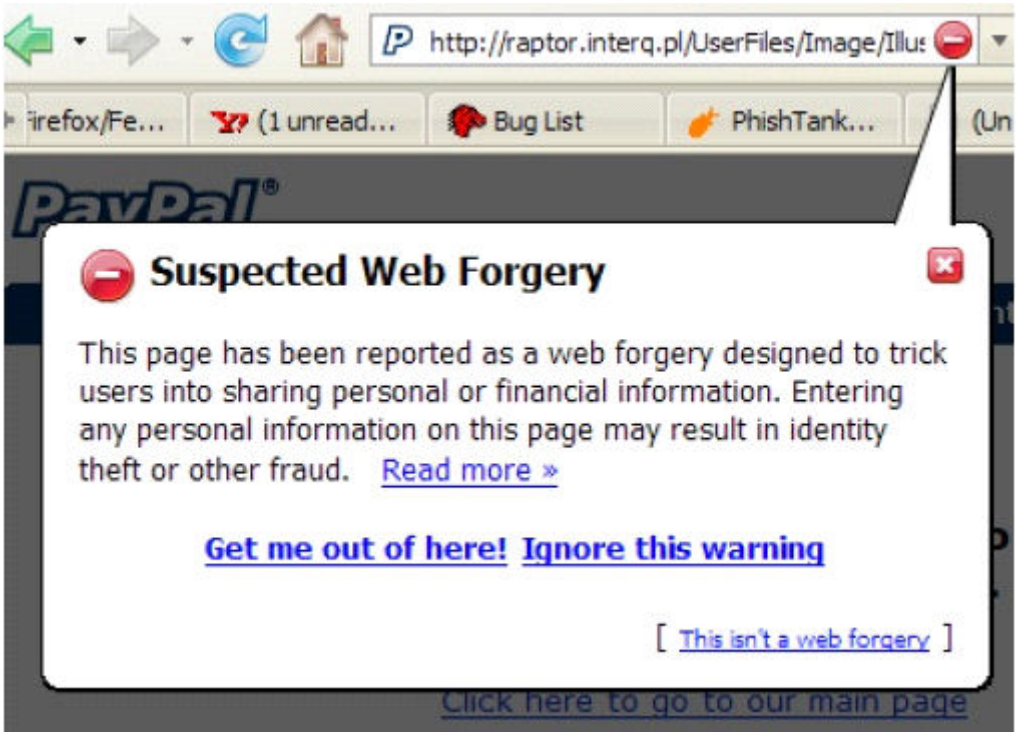


Fig 6.1.2 : Google utilisation de la navigation sécurisée.

6.1.3. **Filtrage de contenu**

Dans ces solutions, le contenu des sites visités et les informations que l' utilisateur essaie d'entrer en entrée à ces sites est analysé. Sur la base des activités suspectes qui peuvent être rencontrées, l'utilisateur sera informé ou son / sa accès à ces sites sera bloqué. L'analyse du contenu peut être fait automatiquement dans les programmes côté client comme SpoofGaurd [17] ou peut être fait par d'autres prestataires de services qui préparent une sorte de dépôt de la liste noire, qui permet aux utilisateurs de savoir si un site est phishing ou non. Ces solutions sont principalement basées sur le rapport des utilisateurs avant de faire n'importe quel type d'analyse de contenu sur un site Web donné.

6.2.des solutions basées sur la prévention : .

Ce sont des solutions qui empêchent phishing. Contrairement aux solutions dans la catégorie de détection qui sont utilisés après le phishing a lieu, ces solutions aident les utilisateurs dans la lutte contre les attaques de phishing.

Les meilleures solutions connues dans cette catégorie sont [1322]:

1. authentification

2. Authentification E-mail

3. gestion des correctifs

4. Sécurité des applications Web

6.2.1. authentification :

L'une des failles fondamentales qui aident les attaques de phishing, n'est pas en mesure d'identifier les utilisateurs corrects. L'authentification est une solution à sens unique où seuls les utilisateurs peuvent donner leur identité aux fournisseurs Internet et les fournisseurs de Web ne peuvent pas identifier les utilisateurs à l'avance.

Les sites Web utilisent généralement des certificats numériques pour prouver leur identité pour les utilisateurs. Les utilisateurs ne sont souvent pas conscients des certificats numériques et ne connaissent pas les propriétés SSL soit. Ainsi, les pirates peuvent facilement tromper en leur fournissant un site visuellement similaires.

Utilisation de personnalisation visuelle [19] est un moyen de protéger les utilisateurs contre le phishing. Dans ces solutions la page de connexion de chaque utilisateur est modifié en fonction de la mise en page que l'utilisateur choisit pour lui-même / elle-même dans une certaine période de temps. Par exemple, l'utilisateur peut mettre son / sa favorite image comme fond de page de connexion. Cela rend difficile pour les attaquants de construire une réplique du site pour un tel utilisateur. Si le modèle choisi n'apparaît pas dans la page de connexion, l'utilisateur devra se méfier et ne divulguera aucune information sur le site présumé.

Cependant, cette solution ne fonctionne pas contre les attaques de phishing basée sur l'empoisonnement DNS d'une attaque en milieu-homme, puisque ces modèles peuvent être détectées et remplacées par Régime de sécurité dynamique [19] est une solution qui a intégré le protocole Secure Remote (SRP) pour résoudre le problème ci-dessus. Dans

cette solution, une barre d'outils est installé sur le navigateur de l'utilisateur. Les fournisseurs de services produisent des images en fonction de mot de passe de l'utilisateur, puis le renvoyer à l'utilisateur comme un fond de leur site actuel ou le fond de log-in la page de l'utilisateur. Ensuite, l'utilisateur génère une image basée sur son propre mot de passe en utilisant la barre d'outils correspondante. Enfin, l'utilisateur compare ces deux images générées pour voir si elles sont identiques ou non. Similitude entre les deux images informe l'utilisateur que le fournisseur de services sait son mot de passe et peut être considéré comme fiable. Toutefois, il ne peut pas protéger les utilisateurs contre les attaques au Moyen-homme complètement.

Les figures suivantes montrent un exemple de l'image générée à la fois par la barre d'outils et le fournisseur de service.

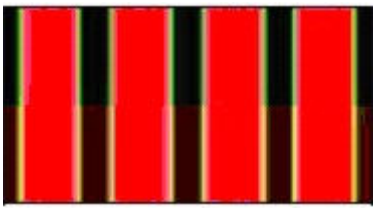


Fig. 6.2.1 : Image généré sur la base du mot de passe de l'utilisateur qui s'affiche

comme l'image d'arrière-plan d'autres formes d'enchaînement ultérieures.

Il ya d'autres solutions qui ont utilisé la même idée, mais de différentes façons pour

aider les utilisateurs contre les attaques de phishing.

Certaines barres d'outils utiliser l'idée de Passpet [20] qui demande à l'utilisateur d'attribuer un label pour les sites de confiance qu'il / elle est prêt à visiter. Chaque fois qu'un utilisateur visite un site Web qui a été déjà attribué une étiquette de la barre d'outils, la barre d'outils affiche l'étiquette, de sorte que l'utilisateur peut identifier les sites de phishing, depuis sites web de phishing ne peuvent pas produire la même étiquette. D'autres solutions utilisent des tables de hachage [21] pour générer un nouveau mot de passe basé sur le mot de passe de l'utilisateur et l'adresse du site en visite. Depuis le phishing

sites ont des adresses différentes à partir des sites Web de confiance, le mot de passe généré serait inutile et les attaquants seulement recevoir un mot de passe qui n'est pas le mot de passe de l'utilisateur. Cependant, aucune de ces solutions ne peut tolérer les

attaques DNS empoisonnement Man in the middle qui permet de basculer l'adresse d'un site Web d'hameçonnage avec l'adresse d'un vrai site web.

Fig. 6.2.2 : L'image générée par le fournisseur de services.

Authentification à deux facteurs [22] est un autre exemple de ces solutions pour protéger

les utilisateurs contre le phishing. Dans cette solution, deux facteurs différents sont utilisés pour désigner l'identité d'un utilisateur. Les utilisateurs ont besoin de leurs mots de passe et leurs appareils mobiles pour accéder aux sites Web. Dans ces solutions, comme [22], les prestataires de services envoient un mot de passe unique pour les appareils mobiles des utilisateurs par SMS. Les fournisseurs de services ont déjà été informés par les utilisateurs au sujet de leurs numéros de téléphone mobile. Depuis attaquants n'ont pas accès au réseau mobile et les téléphones mobiles des utilisateurs, ils ne peuvent pas accéder au mot de passe généré.

Aussi [22] a introduit une nouvelle approche qui utilise un jeton matériel qui génère un nombre différent chiffres six toutes les 60 secondes. Nombres générés sont différents pour chaque utilisateur et peut être utilisé comme un mot de passe qui expire après 60 secondes et est inutile si les attaquants utilisent eux après leur expiration. Cette solution peut également ne pas tolérer Man-in-Moyen-attaque qui permet aux pirates d'utiliser le mot de passe dans les 60 secondes de leur période de validité.

Il ya d'autres solutions qui ne sont pas sensibles aux attaques au Moyen-homme comme [23], ce qui oblige les utilisateurs à installer le matériel sur leurs ordinateurs, ce qui n'est pas rentable .

6.2.2. **Patch Management et la sécurité des applications Web :**

Patch Management et Web Application Security [8] sont les solutions les plus importants dans la lutte contre le phishing. Les attaquants utilisent les avantages des failles dans les applications web dans le but de tromper les gens. Toutefois, ces deux solutions permettent aux utilisateurs d'identifier tout mauvais comportement d'un site web. Ils sont également basés sur une sorte de dépôt de la liste noire qui aide les fournisseurs de connaître les défauts, afin que les fournisseurs puissent préparer un patch ou une mise à jour pour prévenir les activités malveillantes et supprimer les défauts.

6.3. solutions à base de correction :

Ce sont des solutions qui peuvent être utilisés après l'apparition d'une attaque de phishing afin de diminuer les dégâts d'une telle attaque [8]. Il existe deux types de solutions dans cette catégorie:

• PhishingTakedown du site [24].

• Forensics enquête [11].

6.3.1. PhishingTakedown du site

Pour protéger les autres usagers attaques de phishing, certains groupes et organisations ont été fondées comme APWG qui désactive les sites de phishing après leur détection et supprime ces sites à partir d'Internet. Cependant, il n'est pas facile d'abattre les sites de phishing depuis un grand nombre d'entre eux sont hébergés dans différents pays. Ces organisations anti-phishing n'ont pas la capacité de désactiver les sites de phishing car elle nécessite des accords internationaux et exige un effort ardu.

6.3.2. Forensics enquête :

Après une attaque de phishinga lieu, les chercheurs tentent d'analyser et de les comprendre en étudiant les techniques qui ont été utilisées en eux. Cette recherche peut être utilisé dans les nouvelles solutions contre le phishing. Il aide également les chercheurs à connaître les menaces possibles de ces attaques afin de mesurer l'ampleur des dégâts causés par eux et d'informer efficacement les utilisateurs à ce sujet.

7. Conclusion :

Dans ce chapitre, nous allons étudier les différents types de phishing, ainsi que leurs caractéristiques. Nous définissons d'abord le phishing en détail et l'importance des techniques de prévention, selon les statistiques qui montrent l'augmentation des attaques de phishing au cours des dernières années. Ensuite, nous allons montrer les étapes de phishing ainsi que les e-mails de phishing et les sites Web, enfin on a proposée des solution contre ces attaque qui seront les derniers sujets abordés dans ce chapitre.

1. Introduction :

Les pages de phishing sont l'un des problèmes majeurs de sécurité sur internet. La majorité des attaques utilisent des méthodes sophistiquées comme les fausses pages pour tromper les utilisateurs afin d'acquérir des informations sensibles. La méthode la plus simple pour éviter la visite des sites frauduleux est l'utilisation des listes noires si ces listes sont maintenues a jour immédiatement âpres qu'un site frauduleux est créé ce qui est pratiquement impossible a mettre en place. Dans ce chapitre nous allons étudier une méthode pratique d'anti-phishing, ce qui consiste à un système de classification automatique.

2. Approche :

Pour l'application nous avons utilisé l'approche d'analyse d'URLs pour dire qu'un site est sans risque sans avoir a examiner son contenu, cette approche cherche a éviter de télécharger le contenu puis faire l'analyse ce qui crée une importante latence qui dérange l'utilisateur.

L'URL tout seul peut contenir un lot important d'informations qui nous donne la possibilité de juger d'une façon proactive son contenu. Nous avons réalisé un système de classification qui peut analyser une base d'URLs étiquetés selon un nombre d'heuristiques qui peuvent détecter les techniques utilisées pour tromper les utilisateurs.

* La présence de formes : Le but d'un site de phishing est d'avoir les informations de la victime ce qui se fait par l'utilisation de formes d'input. Ce paramètre peut être utilisé comme un pré filtre « s'il y a des formes d'input nous faisons l'analyse sinon pas la peine de faire ».

* L'âge du domaine : En général les sites de phishingon une très courte durée de vie avant qu'ils soient fermés par leurs hébergeurs a cause des plaintes d'utilisateurs et des autorités, alors on peut prendre une durée de vie minimum pour dire qu'un site est sur ou pas .

* La longueur d'URL : Les sites de phishing utilisent deux façons pour tromper les utilisateurs, des URLs avec une importante longueur pour cacher la vraie adresse, dans d'autres cas on utilise des URLs réduites (comme le fait TinyURL.com) pour cacher la vraie adresse et passer l'analyse de la longueur. Nous avons prit la longueur comme paramètre entier.

* Adresses avec IP : pour détourner l'attention des utilisateurs on utilise des adresses avec juste un IP, ce qui n'est pas très évident pour une importante compagnie qui n'a pas un nom de domaine. Un paramètre booléen pour designer si l'adresse contient un IP affecter par la valeur 1 sinon 0.

* Points dans l'URL : Les adresses de phishing contiennent généralement un nombre de sous-domaines pour que l'url apparaisse légitime. Nous avons pris la valeur 3 comme paramètre de division entre phishing ou pas car une valeur importante de nombre de point indique un ou plusieurs sous-domaines.

* Niveau de sécurité : L'utilisation de protocoles sécurisés est un facteur important pour dire est ce sur de visiter un lien ou pas malgré qu'il y a des sites de phishing avec https.

Notre système cherche si l'adresse est sécurisée alors il affecte au paramètre la valeur 1 (pour https) sinon il l'affecte un 0 (en cas de http, ftp ou encas d'absence de protocole dans le lien).

* Caractères spéciaux '@','-' ,'_' : On peut faire une redirection directe d'une adresse a une autre a l'aide de '@', et on utilise '-','_' pour faire faire une adresse ressemblante a la légitime.

Nous avons pris la valeur 1 comme paramètre de division (1 si le nombre est supérieur à 1, 0 sinon).

* Modification d'encodage d'URL: Pour détourner l'analyse on modifie les URLs en changeant des caractères par leurs codes html %20 pour un espace par exemple.

Nous avons pris la valeur 1 comme paramètre de division (1 si le nombre est supérieur à 2, 0 sinon).

3. BASE D'URLS UTILISEE :

Pour la base de d'URLs nous avons eu recours à une base publique utilisée par OPENDNS, cette base est publiée par leur site phishtank.com (phishtank.com est créé par OPENDNS) qui est un site ou les internautes peuvent signaler les sites suspects. La base rendue publique par le site est vérifiée par des experts qui disent que ces adresses sont effectivement des adresses de phishing. Nous avons utilisé deux versions avec des dates de sorties différentes pour plus de précision. Pour la base des adresses sures nous avons utilisé, plusieurs sources comme ALEXA.com, et Google ranking spécialisés dans les statistiques du trafique sur internet et qui donnent le classement des sites les plus populaire périodiquement. Pour le nombre d'instances de la base nous avons 4806 : adresses de phishing. 535 : adresses.

ISTRUCTURE DE LA BASE DE PHISHING :

La base de données est présente sous forme d'un fichier XML, avec les informations suivantes (un exemple d'instance) :

```
<url>http://aerospecialties.aerospecialties.com/osc22/mastercard.number/account.php</url>

<phish_id>1278385</phish_id>

<phish_detail_url>http://www.phishtank.com/phish_detail.php?phish_id=1278385</phish_detail_url>

<ip_address>209.161.24.98</ip_address>

<submission_time>2011-05-20T01:05:01+00:00</submission_time>

<verified>yes</verified>

<verification_time>2011-05-20T01:40:58+00:00</verification_time>

<online>yes</online>

<target>Mastercard</target>
```

Figure3.1 : Base de phishing sous forme XML.

phish_id : ID ou référence du site du phishing.

phish_detail_url : details sur phishtank du site en question.

url : L'url du site

submission_time : la date et l'heure de la déposition d'alarme sur le site

Verified : site vérifier ou pas mais comme phishtank n'ajoute que les sites vérifies alors tous on une valeur :yes

verification_time : date et temps de vérification

online : statut du site .

target : la compagnie ou la marque visée par l'attaque (visa, mastercard…)

4. FONCTIONNALITES DE L'APPLICATION :

Notre application se constitue de deux parties majeures :

* La première partie (Partie développeur) : contient tout se qui a un lien avec le développement, ainsi nous trouvons l'interface pour le calcul des différents paramètres que nous avons vu précédemment.

* La deuxième partie (Partie utilisateur) : Une partie utilisée par l'utilisateur final, qui peut décider pour lui si c'est le site est légitime ou c'est un phish, son implémentation est une barre d'un navigateur web qui bloque l'accès aux sites sensibles

5. Evaluation :

Pour la partie d'évaluation nous avons utilisé la suite de logiciel d'apprentissage automatique et d'exploration de données « WEKA », et ce pour avoir plusieurs classifieurs en main, faire la comparaison, et avoir une analyse claire des résultats. Pour cela nous avons transformé le fichier de sortie de notre application en fichier ARFF de ce format

```
@RELATION phishing   % Nom de l'ensemble de données

@ATTRIBUTE longueur                numeric          % déclaration des
@ATTRIBUTE nombredepoints          {0,1}            % attributs et leurs
@ATTRIBUTE securiteduprotocol      {0,1}            % types
@ATTRIBUTE sansnomdudomaine        {0,1}
@ATTRIBUTE caracteresspeciaux      {0,1}
@ATTRIBUTE caracteresunicodes      {0,1}
@ATTRIBUTE class                   {Phishing, SUR}

@DATA                % la partie data
55 , 0 , 0 , 0 , 0 , 0 , Phishing      % les instances avec leurs
55 , 0 , 0 , 0 , 0 , 0 , Phishing      % classes
55 , 1 , 0 , 0 , 0 , 0 , Phishing
55 , 0 , 0 , 0 , 0 , 0 , Phishing
55 , 0 , 0 , 0 , 0 , 0 , Phishing
55 , 0 , 0 , 0 , 0 , 0 , Phishing
31 , 0 , 0 , 0 , 0 , 0 , SUR
31 , 0 , 1 , 0 , 0 , 0 , SUR
31 , 0 , 1 , 0 , 0 , 0 , SUR
31 , 1 , 1 , 0 , 0 , 0 , SUR
32 , 0 , 0 , 0 , 0 , 0 , SUR
32 , 0 , 0 , 0 , 0 , 0 , SUR
32 , 0 , 0 , 0 , 0 , 0 , SUR
```

Figure 3.3 : fichier au format ARFF

6. RESULTATS SOUS WEKA

Nous avons utilisé plusieurs algorithmes d'apprentissage

* Arbres de décision.

* Classification bayesienne probabiliste (naïve bayes, réseaux bayesiens).

Machine à vecteur de support(SVM).

7. Resultat sous WEKA :

Nous avons utilisé plusieurs algorithmes d'apprentissage

* Arbres de décision.

* Classification bayesienne probabiliste (naïve bayes, réseaux bayesiens).

* Machine à vecteur de support(SVM).

Le tableax suivant résume les résultats obtenus avec ces classifieurs

	SVM (SMO)	Arbres (J48)	RandomTree	Réseaux Bayesiens	Naïve Bayes
Instances Correctement Classées	4346 (90.4286%)	4472 (93.0504%)	4459 (92.7799%)	4385 (91.2401%)	3241 (67.4365%)
Instances Mal Classées	460 (5.5714%)	334 (6.9496%)	347 (7.2201%)	421 (8.7599%)	1565 (32.5635%)

Tableau 7.1 : Résultats de classification

Le graphe suivant compare le nombre d'instances correctement classées par rapport au nombre total d'instances.

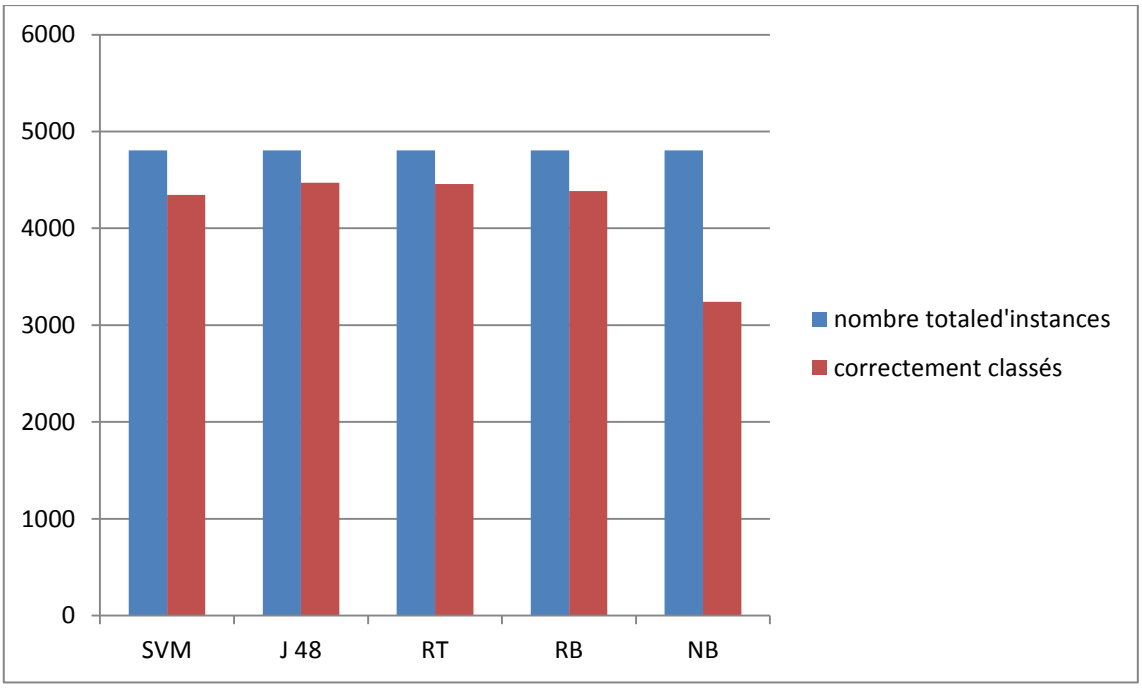


Fig.7.2: nombre d'instances corretement classifiées .

Le tableau suivant récapitule les valeurs de précision et rappel

	SVM (SMO)		Arbres (J48)		RandomTree		Réseaux Bayesiens		Naïve Bayes	
Classe	Phishing	sur	Phishing	sur	Phishing	sur	Phishing	sur	Phishing	Sur
Précision	0.904	0.818	0.946	0.705	0.947	0.67	0.917	0.697	0.914	0.603
Rappel	0.644	0.963	0.979	0.472	0.975	0.485	0.993	0.15	0.644	0.963

Tableau 7.3 : Tableau récapitulatif des précisions et rappels.

8. CHOIX DU MEILLEUR RESULTATS :

Le but de cette étude est de éliminer voir réduire les faux négatifs pour les deux classes (les adresses phishing classée comme valides et inversement). Pour cela et après l'analyse de données nous avons trouvé que l'arbre de decision (J 48) et la meilleure solution avec la matrice de confusion.

```
==== Confusion Matrix === (Arbre De Decison J 48)
a        b        ←Classified as
4255     91     /  a=  Phishing
243      217    /  b= Sur
```

Fig 7.4 : Matrice de confusion pour L'arbre de decision (J48) .

9. Descution des resultat : .

Apres avoir essayé plusieurs classifieurs nous avons trouvé que l'arbre de decision et Randomtree sont les mieux adaptés, les mieux performants avec une précision avoisinant les 94 %.

En se basant sur la dernière matrice de confusion nous pouvons dire deux choses

1- Le système détecte des faux positifs (fausses alarmes 243 instances), qui n'est pas un vrai problème vis-à-vis l'utilisateur, malgré que c'est un argument de confort pour l'utilisateur.

2- Le problème majeur est qu'il y a un nombre d'instances qui sont passés inaperçus (91 instances confirmées phishing) par le système alors qu'ils constituent un vrai danger pour l'utilisateur inexpérimenté.

10. Conclusion :

Malgré les bons résultats obtenus par notre système, il y a beaucoup d'améliorations à faire comme l'ajout d'heuristiques, et l'intégration d'autres approches par prendre en considération le corps de l'email et le site en considération.

La base que nous avons utilisée est vraiment réduite (4806 instances) et ne reflète pas vraiment tout les cas possibles, les formes imaginables que peut avoir une adresse phishing. Alors pour faire un système fiable et applicable en réalité il faut avoir beaucoup plus d'instances dans la base d'apprentissage.

Comme nous avons vu a travers cette étude le phishing est un sérieux danger en ligne, chaque jours en signale des milliers de nouveaux sites de phishing. Les techniques évoluent aussi très vite.

Une vraie guerre ou les mesures de contre attaque restent un pas en arrière car il y a toujours des adresses de phishing qui échappent des filtres quoi que se soit les performances des systèmes ils n'atteindront jamais le maximum.

Pour notre application qui reste un peut simpliste, il y a beaucoup d'améliorations en perspective, comme l'utilisation des techniques anti-spam avec l'intégration de plus d'heuristiques, et prendre en considération le site en lui-même en utilisant différentes méthodes statistique et probabilistes, et l'intégration du Protocol WHOIS qui peut nous donner des informations sur le domaine et son âge.

Le seul élément clef de ce combat reste l'humain utilisant la machine, il faut réveiller son attention pour assister ces systèmes à identifier ces dangers.

- [1] nojoomcirta.com/upload/attach/92ded5d8ff.pdf vue le 24/10/2012
- [2] http://www.futura-sciences.com/magazines/high-tech/infos/dico/d/internet-internet-3983/ vue le 21/06/2001
- [3] www.college-castillonnes.fr/IMG/pdf/dangers-internet-long.pdf vue le 25/01/2012
- [4] web.univ-pau.fr/~cpham/M2SIR/BIBLIO/DOC01-02/Attaque.doc année 2002
- [5] www.unilim.fr/sci/IMG/pdf/dangers-internet-2.pdf année 13/04/2012
- [6] www.securite-informatique.gouv.fr/gp_rubrique34.html année 2001
- [7] http://www.lemonde.fr/technologies/article/2009/11/12/comment-proteger-sa-vie-privee-sur-internet_1266460_651865.html le 12/11/2009
- [8] M. Jakobsson and S. Myers, Phishing and Countermeasures: Understanding the

Increasing Problem of Electronic Identity Theft.Wiley-Interscience, 2006.

- [9] "What is phishing." http://www.webopedia.com/TERM/P/phishing.html.
- [10] "Phishing attack trends report - second half 2008."

http://www.apwg.org/reports/apwg report H2 2008.pdf, Mar. 2009.

- [11] "APWG." http://www.antiphishing.org, Anti Phishing Working Group.
- [12] "Gartner institute." http://www.gartner.com.
- [13] M. Wu, Fighting phishing at the user interface. PhDthesis, Cambridge, MA, USA, 2006. Adviser-Miller, Robert C.

- [14] D. K. McGrath and M. Gupta, "Behind phishing: an examination of phisher modi operandi," in LEET'08: Proceedings of the 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats, (Berkeley, CA, USA), pp. 1–8, USENIX Association, 2008.

- [15] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Beyond blacklists: learning to

detect malicious web sites from suspicious urls," in KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data

mining, (New York, NY, USA), pp. 1245–1254, ACM, 2009 .

- [16]"Spam wikipedia, the free encyclopedia." http://en.wikipedia.org/wiki/Spam (electronic).
- [17] "Spoofguard." http://crypto.stanford.edu/SpoofGuard/.
- [18] "Sitewatcher - anti-phishing service." http://www.sitewatcher.com/.
- [19] R. Dhamija and J. D. Tygar, "The battle against phishing: Dynamic security skins," in SOUPS '05: Proceedings of the 2005 symposium on Usable privacy and security, (New York, NY, USA), pp. 77–88, ACM, 2005.
- [20] K.-P. Yee and K. Sitaker, "Passpet: convenient password management and phishing protection," in SOUPS '06: Proceedings of the second symposium on Usable privacy and security, (New York, NY, USA), pp. 32–43, ACM, 2006.
- [21] B. Ross, C. Jackson, N. Miyake, D. Boneh, and J. C. Mitchell, "Stronger password authentication using browser extensions," in SSYM'05: Proceedings of the 14th conference on USENIX Security Symposium, (Berkeley, CA, USA), pp. 2–2, USENIX Association, 2005.
- [22] G. Yang, D. S.Wong, H.Wang, and X. Deng, "Two-factor mutual authentication based on smart cards and passwords," J. Comput. Syst. Sci., vol. 74, no. 7, pp. 1160–1172, 2008.
- [23] "RSA. Security-Technology." www.rsa.com/rsalabs/technotes/One-TimePWWP.pdf

[24] T. Moore and R. Clayton, "Examining the impact of website take-down on phishing," in eCrime '07: Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit, (New York, NY, USA), pp. 1–13, ACM, 2007

Résumé

L'hameçonnage est un type de vol d'identité qui tente de voler des données confidentielles et personnelles comme l'information de cartes de crédit ou de comptes bancaires. Plusieurs stratégies ont été proposées pour vaincre l'hameçonnage ; la plupart d'entre elles dépent d'une base de données. Dans cette Mémoire, nous présentons un système au un programme crées en DELPHI pour la detections des sites de phishing ; l'execution de l'application c'est une fenetre de navigateur qui ouvre et saffiche et dite si le lien url est un site phishing au ligitime, Et pour la partie d'évaluation nous avons utilisé la suite de logiciel d'apprentissage automatique et d'exploitation de données « WEKA», et ce pour avoir plusieurs classifieurs en main , faire la comparaison et avoir une analyse claire des resultats.

Abstract

Phishing is a type of identity theft that attempts to steal confidential and personal data such as credit card information or bank account numbers. Several strategies have been proposed to overcome phishing; most of them depent a database. In this memory, we present a system created in Delphi for detections phishing sites program; the execution of the application is a browser window that opens, and said right will be displayed if the url link is a phishing site ligitime, and for the evaluation part we used the following software and machine learning data Mining "WEKA" and for several classifiers in hand, to compare and have a clear analysis of the results.

ملخص

التصيد هو نوع من سرقة الهوية الذي يحاول سرقة البيانات السرية والشخصية مثل معلومات بطاقة الائتمان أو أرقام حسابات البنوك. وقد اقترحت عدة استراتيجيات للتغلب على التصيد. معظمهم يتطلب قاعدة بيانات. في هذه الأطروحة، نقدم نظام إنشاؤها بواسطة المكتشفة دلفي لبرنامج مواقع التصيد. تنفيذ التطبيق هو الذي يفتح نافذة هو موقع التصيد الشرعي، وبالنسبة للجزء تقييم URL المتصفح، وقال الحق سيتم عرض إذا كان الارتباط وعدة مصنفات في متناول اليد، لمقارنة "WEKA" استخدمنا التعلم البرمجيات وآلة التالية تعدين البيانات والحصول على تحليل واضح للنتائج.