Union internationale des télécommunications

UIT-T

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

X.1235

(01/2022)

SÉRIE X: RÉSEAUX DE DONNÉES, COMMUNICATION ENTRE SYSTÈMES OUVERTS ET SÉCURITÉ

Sécurité du cyberespace – Lutte contre le spam

Technologies de lutte contre l'usurpation de site web pour les organisations de télécommunication

Recommandation UIT-T X.1235

RECOMMANDATIONS UIT-T DE LA SÉRIE X

RÉSEAUX DE DONNÉES, COMMUNICATION ENTRE SYSTÈMES OUVERTS ET SÉCURITÉ

RÉSEAUX PUBLICS DE DONNÉES	X.1–X.199
INTERCONNEXION DES SYSTÈMES OUVERTS	X.200–X.299
INTERFONCTIONNEMENT DES RÉSEAUX	X.300–X.399
SYSTÈMES DE MESSAGERIE	X.400–X.499
ANNUAIRE	X.500–X.599
RÉSEAUTAGE OSI ET ASPECTS SYSTÈMES	X.600–X.699
GESTION OSI	X.700–X.799
SÉCURITÉ	X.800–X.849
APPLICATIONS OSI	X.850–X.899
TRAITEMENT RÉPARTI OUVERT	X.900–X.999
SÉCURITÉ DE L'INFORMATION ET DES RÉSEAUX	
Aspects généraux de la sécurité	X.1000–X.1029
Sécurité des réseaux	X.1030–X.1049
Gestion de la sécurité	X.1050–X.1069
Télébiométrie	X.1080–X.1099
APPLICATIONS ET SERVICES SÉCURISÉS (1)	
Sécurité en multidiffusion	X.1100–X.1109
Sécurité des réseaux domestiques	X.1110–X.1119
Sécurité des télécommunications mobiles	X.1120–X.1139
Sécurité de la toile (1)	X.1140–X.1149
Sécurité des applications (1)	X.1150–X.1159
Sécurité d'homologue à homologue	X.1160–X.1169
Sécurité des identificateurs en réseau	X.1170–X.1179
Sécurité de la télévision par réseau IP	X.1180–X.1199
SÉCURITÉ DU CYBERESPACE	
Cybersécurité	X.1200–X.1229
Lutte contre le spam	**X.1230–X.1249**
Gestion des identités	X.1250–X.1279
APPLICATIONS ET SERVICES SÉCURISÉS (2)	
Communications d'urgence	X.1300–X.1309
Sécurité des réseaux de capteurs ubiquitaires	X.1310–X.1319
Sécurité des réseaux électriques intelligents	X.1330–X.1339
Courrier certifié	X.1340–X.1349
Sécurité de l'Internet des objets (IoT)	X.1350–X.1369
Sécurité des systèmes de transport intelligents	X.1370–X.1399
Sécurité de la technologie des registres distribués (DLT)	X.1400–X.1429
Sécurité des applications (2)	X.1450–X.1459
Sécurité de la toile (2)	X.1470–X.1489
ÉCHANGE D'INFORMATIONS SUR LA CYBERSÉCURITÉ	
Aperçu général de la cybersécurité	X.1500–X.1519
Échange concernant les vulnérabilités/les états	X.1520–X.1539
Échange concernant les événements/les incidents/l'heuristique	X.1540–X.1549
Échange de politiques	X.1550–X.1559
Heuristique et demande d'informations	X.1560–X.1569
Identification et découverte	X.1570–X.1579
Échange garanti	X.1580–X.1589
Cyberdéfense	X.1590–X.1599
SÉCURITÉ DE L'INFORMATIQUE EN NUAGE	
Aperçu de la sécurité de l'informatique en nuage	X.1600–X.1601
Conception de la sécurité de l'informatique en nuage	X.1602–X.1639
Bonnes pratiques et lignes directrices concernant la sécurité de l'informatique en nuage	X.1640–X.1659
Mise en œuvre de la sécurité de l'informatique en nuage	X.1660–X.1679
Sécurité de l'informatique en nuage (autres)	X.1680–X.1699
COMMUNICATIONS QUANTIQUES	
Terminologie	X.1700–X.1701
Générateur quantique de nombres aléatoires	X.1702–X.1709
Cadre de sécurité pour les réseaux QKDN	X.1710–X.1711
Conception de la sécurité pour les réseaux QKDN	X.1712–X.1719
Techniques de sécurité pour les réseaux QKDN	X.1720–X.1729
SÉCURITÉ DES DONNÉES	
Sécurité des mégadonnées	X.1750–X.1759
Protection des données	X.1770–X.1789
SÉCURITÉ DES IMT-2020	X.1800–X.1819

Pour plus de détails, voir la Liste des Recommandations de l'UIT-T.

Recommandation UIT-T X.1235

Technologies de lutte contre l'usurpation de site web pour les organisations de télécommunication

Résumé

L'usurpation de site web constitue une menace majeure pour les organisations de télécommunication, en particulier les opérateurs. Il est recommandé que ces derniers mettent en œuvre des technologies de lutte contre l'usurpation des sites web afin de protéger leurs clients et de préserver leur réputation et leurs recettes. La Recommandation UIT-T X.1235 analyse les principales techniques d'usurpation des sites web et recommande des technologies permettant d'identifier les sites contrefaits. Ces mesures peuvent être considérées comme des lignes directrices destinées aux organisations de télécommunication afin de protéger les sites web contre l'usurpation. Une approche similaire pourra être adoptée contre l'usurpation de tous les sites web, y compris ceux des banques, des compagnies d'assurance, des boutiques en ligne, etc.

Historique

Édition	Recommandation	Approbation	Commission d'études	Identifiant unique*
1.0	UIT-T X.1235	07-01-2022	17	11.1002/1000/14797

Mots clés

Contre-mesures, usurpation de site web.

* Pour accéder à la Recommandation, reporter cet URL http://handle.itu.int/ dans votre navigateur Web, suivi de l'identifiant unique, par exemple http://handle.itu.int/11.1002/1000/11830-en.

AVANT-PROPOS

L'Union internationale des télécommunications (UIT) est une institution spécialisée des Nations Unies dans le domaine des télécommunications et des technologies de l'information et de la communication (ICT). Le Secteur de la normalisation des télécommunications (UIT-T) est un organe permanent de l'UIT. Il est chargé de l'étude des questions techniques, d'exploitation et de tarification, et émet à ce sujet des Recommandations en vue de la normalisation des télécommunications à l'échelle mondiale.

L'Assemblée mondiale de normalisation des télécommunications (AMNT), qui se réunit tous les quatre ans, détermine les thèmes d'étude à traiter par les Commissions d'études de l'UIT-T, lesquelles élaborent en retour des Recommandations sur ces thèmes.

L'approbation des Recommandations par les Membres de l'UIT-T s'effectue selon la procédure définie dans la Résolution 1 de l'AMNT.

Dans certains secteurs des technologies de l'information qui correspondent à la sphère de compétence de l'UIT-T, les normes nécessaires se préparent en collaboration avec l'ISO et la CEI.

NOTE

Dans la présente Recommandation, l'expression "Administration" est utilisée pour désigner de façon abrégée aussi bien une administration de télécommunications qu'une exploitation reconnue.

Le respect de cette Recommandation se fait à titre volontaire. Cependant, il se peut que la Recommandation contienne certaines dispositions obligatoires (pour assurer, par exemple, l'interopérabilité et l'applicabilité) et on considère que la Recommandation est respectée lorsque toutes ces dispositions sont observées. Le futur d'obligation et les autres moyens d'expression de l'obligation comme le verbe "devoir" ainsi que leurs formes négatives servent à énoncer des prescriptions. L'utilisation de ces formes ne signifie pas qu'il est obligatoire de respecter la Recommandation.

DROITS DE PROPRIÉTÉ INTELLECTUELLE

L'UIT attire l'attention sur la possibilité que l'application ou la mise en œuvre de la présente Recommandation puisse donner lieu à l'utilisation d'un droit de propriété intellectuelle. L'UIT ne prend pas position en ce qui concerne l'existence, la validité ou l'applicabilité des droits de propriété intellectuelle, qu'ils soient revendiqués par un membre de l'UIT ou par une tierce partie étrangère à la procédure d'élaboration des Recommandations.

À la date d'approbation de la présente Recommandation, l'UIT n'avait pas été avisée de l'existence d'une propriété intellectuelle protégée par des brevets ou par des droits d'auteur afférents à des logiciels, et dont l'acquisition pourrait être requise pour mettre en œuvre la présente Recommandation. Toutefois, comme il ne s'agit peut-être pas de renseignements les plus récents, il est vivement recommandé aux développeurs de consulter les bases de données appropriées de l'UIT-T disponibles sur le site web de l'UIT-T à l'adresse http://www.itu.int/ITU-T/ipr/.

TABLE DES MATIÈRES

Introduction

L'usurpation de site web joue un rôle majeur dans les escroqueries sur Internet depuis quelques années. Les escrocs ciblent généralement des sites web d'entreprises ou d'organisations reconnues et collectent les informations d'identification des visiteurs ou propagent des programmes malveillants. Il s'ensuit des pertes financières pour les visiteurs et les opérateurs de télécommunication, et une dégradation de la réputation de ces derniers.

Les sites web des opérateurs de télécommunication étant devenus l'une des voies d'entrée principales permettant aux clients de s'informer sur toutes sortes de services et de s'y abonner, il a été observé, dans le monde entier, que les escrocs tentaient continuellement de contrefaire les sites web des opérateurs de télécommunication en vue de tromper les clients. La présente Recommandation propose une analyse poussée de l'usurpation des sites web et recommande une série de contre-mesures coordonnées.

Recommandation UIT-T X.1235

Technologies de lutte contre l'usurpation de site web pour les organisations de télécommunication

1 Domaine d'application

La présente Recommandation recommande des technologies permettant aux organisations de télécommunication d'identifier rapidement les cas d'usurpation de site web et de protéger leurs sites contre les tentatives d'usurpation. Elle présente une analyse systématique des caractéristiques et des techniques d'usurpation et préconise des bonnes pratiques consistant à déployer des technologies côté réseau, avec l'assistance de technologies côté utilisateur afin de lutter contre l'usurpation des sites web.

La conformité à la présente Recommandation ne doit pas être considérée comme une preuve permettant de déclarer la conformité à une législation, une réglementation ou une politique, nationale ou régionale. Les moyens techniques et ceux relatifs à l'organisation et aux procédures décrits dans la présente Recommandation ne garantissent en aucune façon de parvenir à un niveau de sécurité susceptible d'être imposé pour certaines correspondances par une législation, une réglementation ou une politique nationale ou régionale spécifique.

Une approche similaire pourra être adoptée contre l'usurpation de tous les sites web, y compris ceux des banques, des compagnies d'assurance, des boutiques en ligne, etc.

2 Références

La présente Recommandation se réfère à certaines dispositions des Recommandations UIT-T et textes suivants qui, de ce fait, en sont partie intégrante. Les versions indiquées étaient en vigueur au moment de la publication de la présente Recommandation. Toute Recommandation ou tout texte étant sujet à révision, les utilisateurs de la présente Recommandation sont invités à se reporter, si possible, aux versions les plus récentes des références normatives suivantes. Une liste des Recommandations de l'UIT-T en vigueur est publiée périodiquement. La référence à un document figurant dans la présente Recommandation ne donne pas à ce document, en tant que tel, le statut de Recommandation.

Aucune.

3 Définitions

3.1 Termes définis ailleurs

Aucun.

3.2 Termes définis dans la présente Recommandation

La présente Recommandation définit les termes suivants:

3.2.1 réseau neuronal convolutif: algorithme d'apprentissage profond qui prend une image en entrée, attribue une importance à divers aspects ou objets de l'image et est capable de les différencier les uns des autres.

3.2.2 réseau à mémoire à court terme et long terme: type de réseau de neurones récurrents capable d'apprendre une dépendance d'ordre dans les problèmes de prédiction de séquence.

3.2.3 réseau de neurones récurrents: catégorie de réseaux de neurones qui permettent d'utiliser des données de sorties précédentes en tant que données d'entrée, et qui possède des états cachés.

3.2.4 transformation de caractéristiques invariante à l'échelle (SIFT, *scale-invariant feature transform*): algorithme de détection de caractéristiques dans le domaine de la vision par ordinateur utilisé pour détecter et décrire des caractéristiques locales dans des images, qui compose un descripteur qui ne varie pas en fonction des translations, rotations et transformations d'échelle dans le domaine de l'image, et est robuste vis-à-vis des transformations modérées de perspective et des variations de luminosité. Sur le plan expérimental, le descripteur SIFT s'est avéré très utile en pratique pour l'appariement d'images et la reconnaissance d'objet dans des conditions réelles.

3.2.5 caractéristiques robustes accélérées (SURF, *speeded up robust features*): détecteur de caractéristiques locales et descripteur dans le domaine de la vision par ordinateur, permettant de détecter et de décrire des caractéristiques locales dans des images. Il s'agit d'une technique beaucoup plus rapide que la transformation de caractéristiques invariante à l'échelle (SIFT) et plus robuste que celle-ci vis-à-vis de différentes transformations d'images.

3.2.6 site web contrefait: site web créé par usurpation d'un site web (voir le § 3.2.8).

3.2.7 machine à vecteur de support: modèle d'apprentissage automatique supervisé qui résout les problèmes de classification binaire et est capable de catégoriser de nouveaux ensembles à partir des ensembles définis de données d'entraînement étiquetées pour chaque catégorie.

3.2.8 usurpation de site web: ensemble de comportements malveillants visant à imiter un site web bien connu du public ou d'un groupe de personnes; le faux site est ensuite utilisé pour tromper la confiance des visiteurs à des fins malveillantes ou illicites, par exemple fraude, atteinte à la vie privée, etc.

4 Abréviations et acronymes

La présente Recommandation utilise les abréviations et acronymes suivants:

DNS service de noms de domaine (*domain name service*)

GUI interface graphique utilisateur (*graphic user interface*)

IA intelligence artificielle

LSTM mémoire à court et long terme (*long short-term memory*)

OCR reconnaissance optique de caractères (*optical character recognition*)

PII informations d'identification personnelle (*personally identifiable information*)

QR réponse rapide (*quick response*)

SIFT transformation de caractéristiques invariante à l'échelle (*scale-invariant feature transform*)

SURF caractéristiques robustes accélérées (*speeded up robust features*)

SVM machine à vecteur de support (*support vector machine*)

URL localisateur uniforme de ressources (*uniform resource location*)

5 Conventions

La présente Recommandation utilise les conventions suivantes:

Le terme "**devrait**" indique une exigence qui est recommandée mais qui n'est pas absolument nécessaire. Cette exigence n'est donc pas indispensable pour déclarer la conformité.

Le terme "**pourra**" indique une exigence optionnelle qui est admissible, sans pour autant être en quoi que ce soit recommandée.

Dans le corps de la présente Recommandation, le mot "**peut**" apparaît à quelques occasions. Il doit alors être interprété comme "**est en mesure de**".

6 Analyse de l'usurpation de site web

L'usurpation de site web est un ensemble de comportements malveillants visant à imiter un site web connu du public ou d'un groupe de personnes. Le site contrefait est ensuite utilisé pour collecter les informations d'identification des visiteurs ou propager des programmes malveillants à des fins malveillantes ou illicites: fraude, atteinte à la vie privée, etc.

L'usurpation de site web peut causer des dommages importants de plusieurs façons. Elle peut, par exemple, entraîner des pertes financières pour les clients comme pour les opérateurs, nuire à la réputation des opérateurs, etc. De plus, les informations d'utilisateur dérobées peuvent être employées à mauvais escient longtemps après la disparition du site web contrefait. Il est donc essentiel que les organisations de télécommunication, en particulier les opérateurs, mettent en œuvre des technologies efficaces pour empêcher l'usurpation des sites web.

6.1 Scénario d'usurpation d'un site web

Un scénario typique de création d'un site web contrefait dans un but malveillant comporte trois phases, comme présenté à la Figure 6-1.

- **Phase une**: construction du site web
  - Étape une: développer et exploiter un site web contrefait identique ou similaire à un site web connu.
- **Phase deux**: diffusion du localisateur uniforme de ressources (URL)
  - Étape deux: diffuser le site web contrefait par différents canaux, par exemple par courriel, service de messages courts ou instantanés, etc.
- **Phase trois**: collecte des informations d'identification ou diffusion d'un programme malveillant
  - Étape trois: inciter les visiteurs à saisir des informations d'identification ou à télécharger des logiciels malveillants. Une fois que les victimes ont accédé à un site web contrefait, elles peuvent croire à tort que ce site est le vrai et se connecter avec leurs informations d'identification ou télécharger des logiciels malveillants.
  - Étape quatre: récupérer les informations d'identification des visiteurs et préparer l'escroquerie suivante ou une autre activité malveillante.

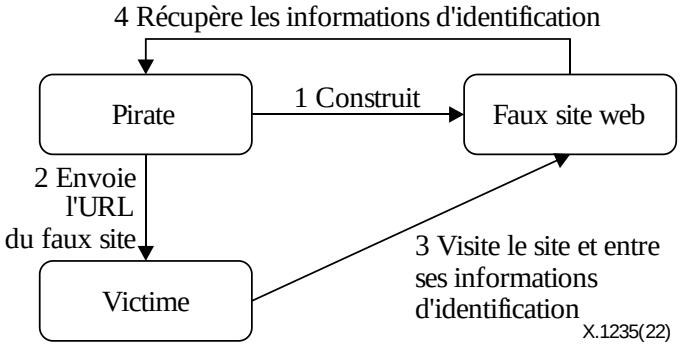


Figure 6-1 – Les quatre étapes d'utilisation d'un site web contrefait pour collecter des informations d'identification

6.2 Caractéristiques de l'usurpation de site web

6.2.1 Caractéristiques des sites web contrefaits

Les sites web contrefaits ont recours à des procédés basés sur la similarité pour faire croire aux visiteurs qu'ils sont sur les vrais sites. Ces sites présentent des caractéristiques communes:

– Similarité visuelle: un site web contrefait peut être une réplique complète du vrai site, y compris en termes de contenu visuel et de logique d'interaction.

– Emprunt d'éléments visuels: un site web contrefait peut emprunter des éléments visuels majeurs du vrai site web.

– Lien vers des sites connus: un site web contrefait peut comporter des liens vers d'autres sites connus, de façon à imiter un site partenaire officiel, un site apparenté ou une version différente du site web.

– Nom de domaine similaire: le nom de domaine du site web contrefait est très proche de celui du vrai site.

– Nom de domaine obscur: les escrocs utilisent parfois des services de réduction d'URL ou des codes à réponse rapide (codes QR) pour masquer le véritable nom de domaine lors du partage du lien.

– Noms de domaine multiples: les escrocs peuvent avoir réservé un ensemble de noms de domaine en vue d'usurper un site web. Pour prolonger l'existence du site contrefait, plusieurs d'entre eux seront publiés simultanément.

– Code source brouillé: le code source d'un site web contrefait peut être brouillé pour tromper les analyseurs de sécurité.

– Zone de saisie d'informations d'identification: les sites web contrefaits peuvent inciter les visiteurs à saisir leurs informations d'identification.

– Redirection vers un site connu: un site web contrefait peut rediriger vers des sites web connus (par exemple google.com) pendant les périodes de non-utilisation afin d'éviter d'être détectés par les programmes de recherche automatique de sites frauduleux.

– Fausse barre d'adresse: certains sites web contrefaits peuvent être spécifiquement conçus pour les navigateurs web mobiles. Le site web insère une fausse copie de la barre d'adresse d'un navigateur mobile en haut de l'écran du téléphone mobile et maintient la copie verrouillée en trompant l'interface graphique utilisateur (GUI) du navigateur.

6.2.2 Caractéristiques des activités associées

Il est possible de recourir à de nombreuses activités pour inciter les utilisateurs à visiter un site web contrefait et à interagir avec ce site. Ces activités peuvent s'appuyer sur les caractéristiques ou les tactiques suivantes:

– Diffuser des messages pour informer le plus grand nombre de l'existence du site web contrefait.

– Détourner ou empoisonner un service de noms de domaine (DNS), ainsi les visiteurs se rendant sur le site web officiel sont redirigés vers le site contrefait.

– Utiliser des appâts (bon de réduction, cadeau, etc.)pour inciter les utilisateurs à cliquer sur le lien du site contrefait.

– Installer un logiciel malveillant sur le téléphone de la victime pour afficher le site web contrefait au lieu du site officiel lors de la requête.

– Contacter les utilisateurs en se faisant passer pour un service client afin de leur communiquer des liens vers des sites frauduleux.

6.3 Conséquences

L'usurpation d'un site web peut avoir de nombreuses conséquences négatives pour les utilisateurs et les propriétaires des sites contrefaits. Ces conséquences peuvent être les suivantes:

– Pertes matérielles pour l'utilisateur: les informations d'identification de l'utilisateur et ses informations d'identification personnelle (PII) peuvent être dérobées. Les escrocs peuvent se servir de ces informations pour perpétrer d'autres actes frauduleux. Les utilisateurs peuvent être incités à télécharger un cheval de Troie, du code de cryptominage et d'autres programmes malveillants. Les escrocs peuvent prendre le contrôle du terminal de l'utilisateur pour vider des comptes bancaires, propager du spam et infecter d'autres utilisateurs. Ces activités peuvent réduire la durée de vie de la batterie ou causer un dysfonctionnement du terminal, ce qui peut également entraîner des pertes financières directes de grande ampleur.

– Pertes matérielles pour l'opérateur: les transactions normales entre les utilisateurs et les opérateurs peuvent être perdues et parfois même perturbées à long terme.

– Perte de réputation: l'usurpation peut ternir la réputation du vrai site web et même réduire la confiance du public. Elle peut entraîner un grand nombre de plaintes et des évaluations négatives à long terme.

7 Contre-mesures

Les mesures visant à contrer l'usurpation d'un site web peuvent être divisées en deux groupes: identification et protection. Les mesures d'identification distinguent les sites web contrefaits des autres sites web; voir le § 7.1. Les mesures de protection empêchent que les utilisateurs n'accèdent ou ne soient trompés par des sites web contrefaits identifiés; voir le § 7.2.

Il n'est pas nécessaire pour une organisation de télécommunication de faire appel simultanément à toutes les technologies recommandées dans les paragraphe ci-dessous. En fonction des données acquises, de l'environnement législatif, des exigences des abonnés, etc., l'organisation de télécommunication devrait faire preuve de souplesse dans l'usage des technologies appropriées présentées ici.

7.1 Identification

Les contre-mesures recommandées ci-après se fondent sur les caractéristiques des sites web contrefaits présentées au § 6.2.1.

7.1.1 Comparaison de noms de domaine similaires

Il est recommandé aux opérateurs de tenir à jour une liste des noms de domaine des sites web connus susceptibles d'être contrefaits. Si un nom de domaine est très similaire (mais non identique) à un nom de domaine de la liste, il peut s'agir d'un site web contrefait. Si le nom de domaine a été raccourci au moyen d'un service de réduction d'URL, il devrait être reconverti vers l'URL d'origine avant la comparaison

De nombreuses méthodes permettent de calculer la similarité de deux noms de domaine, par exemple la distance d'édition, la similarité de Jaccard, la plus longue sous-séquence commune, la conversion des similitudes visuelles, etc.

- **Distance d'édition**: soient A et B deux noms de domaine; la méthode calcule le nombre minimum de modifications nécessaires pour passer de A à B. Plus courte est la distance d'édition, plus grande est la similarité.
- **Similarité de Jaccard**: la taille de l'intersection des ensembles de caractères de deux noms de domaine est divisée par la taille de leur union. Plus le rapport est élevé, plus la similarité est grande.

- **Plus longue sous-séquence commune**: on calcule la longueur de la plus longue sous-séquence commune de deux noms de domaine. Plus longue est cette sous-séquence, plus grande est la similarité.

  NOTE – La plus longue sous-séquence commune est la plus longue sous-séquence commune de deux séquences données, à condition que les éléments de la sous-séquence n'aient pas besoin d'occuper des positions consécutives au sein des séquences d'origine. La sous-séquence commune doit être une séquence strictement ascendante des indices des deux séquences données. Par exemple, si "abcde" et "akcve" sont deux séquences données, alors "ace" est la plus longue sous-séquence commune des deux séquences.

- **Conversion des similitudes visuelles**: les caractères qui sont visuellement similaires sont remplacés avant de procéder à la comparaison. Par exemple, "0" et "o" peuvent être remplacés l'un par l'autre; de même pour "1" et "i", etc. Par exemple, le nom de domaine "z00.com" pourrait être converti en "zoo.com". Cette méthode peut accroître les performances de la similarité. Toutefois, dans certaines circonstances, la conversion des similitudes visuelles pourrait également réduire les performances. Il est donc prudent d'effectuer la comparaison de similarité avant et après conversion.

7.1.2 Détection des logos officiels

Il est recommandé de comparer les images et autres éléments visuels d'un site web inconnu avec les logos officiels. Par logo officiel, on entend notamment le visuel de la marque de l'entreprise ou de l'organisation, les dessins publicitaires et les autres éléments iconiques. La méthode de détection des logos officiels comprend les étapes suivantes:

- Avant la détection, collecte des logos officiels et stockage dans une base de données de logos.
- Téléchargement ou capture d'écran des éléments visuels d'une page web inconnue et comparaison avec les logos stockés dans la base de données. La similarité entre deux images peut être calculée au moyen de deux descripteurs populaires, le descripteur SIFT (transformation de caractéristiques invariante à l'échelle) et le descripteur SURF (caractéristiques robustes accélérées).
- Si le code source de la page web testée a été brouillé, il peut être difficile de télécharger les éléments visuels. Dans ce cas, un cadre de test automatique des applications web basé sur le navigateur peut être utilisé pour prendre une capture d'écran de la page web. Avant d'être comparés, les logos peuvent être détourés sur la capture d'écran par un modèle de détection de logo utilisant l'intelligence artificielle. Il est possible d'entraîner correctement un tel modèle de détection de logo utilisant l'intelligence artificielle au moyen de bases de données de logos à code source ouvert.

7.1.3 Détection du brouillage de code

Un site web contrefait peut volontairement brouiller le code malveillant ou d'usurpation pour rendre son analyse plus difficile. Le code brouillé est très différent du code normal. De nombreuses méthodes permettent de détecter un code brouillé, notamment les outils de classification fondés sur des réseaux neuronaux convolutifs et récurrents, etc.

- Les outils de classification fondés sur un réseau neuronal convolutif peuvent extraire automatiquement des n-grammes caractéristiques du code source normal et du code brouillé. Un n-gramme est une séquence contigüe de n mots d'un code source donné. Le nombre de mots peut être fixé à 5 ou plus pour capturer entièrement les caractéristiques discriminantes en vue de la classification.

- Les outils de classification fondés sur les réseaux de neurones récurrents utilisent principalement les réseaux LSTM ou de type transformeur. Ces outils traitent le code source comme une séquence de caractères, identifient automatiquement les motifs de séquence entre le code normal et le code brouillé et déterminent ensuite si un code est brouillé.

7.1.4 Détection des zones de saisie d'informations d'identification

Les zones de saisie d'informations d'identification se présentent habituellement sous la forme de "formulaires de saisie" dans le code source des pages web. Le code source du formulaire de saisie comporte habituellement un attribut "mot de passe" (qui garantit que le contenu du mot de passe est masqué lors de sa saisie). Les informations d'attribut peuvent être rapidement localisées au moyen d'expressions régulières ou d'outils d'analyse des pages web. Si le code source d'un site web a été brouillé, la méthode de reconnaissance optique de caractères (OCR) peut permettre de détecter les zones de saisie d'informations d'identification sur les captures d'écran du site suspect. La méthode OCR localise les zones de texte sur la capture d'écran de la page web, puis les convertit en textes. Si les textes obtenus après conversion contiennent certains mots-clés tels que "mot de passe", il est possible d'en déduire la présence d'une zone de saisie d'informations d'identification sur la page web.

7.1.5 Service de sécurité tiers

Il est recommandé aux opérateurs d'utiliser les services de sécurité proposés par des fournisseurs de sécurité tiers afin d'obtenir les attributs et les statistiques d'un site web, y compris le trafic du site, sa réputation, les modules d'extension, l'enregistrement du nom de domaine, la certification et d'autres informations de sécurité. Ces informations peuvent apporter des éléments complémentaires pour repérer un site web contrefait.

7.1.6 Analyse des liens

Il se peut que le site web contrefait copie les liens hypertexte du site web officiel pour réutiliser les éléments visuels de celui-ci dans les pages web frauduleuses. L'analyse des liens permet de détecter cette tactique en collectant les liens des éléments visuels du site web officiel et en les comparant aux liens figurant sur le site suspect. Si les deux sites comportent les mêmes liens hypertextes, ou s'ils en ont beaucoup en commun, il peut s'agir d'une usurpation.

7.1.7 Identification finale par évaluation de la réputation

Il est recommandé aux opérateurs de recourir à un modèle décisionnel fondé sur l'intelligence artificielle pour combiner tous les résultats d'analyse ou de détection des contre-mesures recommandées ci-dessus, et de prendre une décision globale sur la question de savoir si un site web est un site contrefait ou non. Dans le modèle décisionnel fondé sur l'intelligence artificielle, le poids de chaque contre-mesure est déterminé automatiquement après la phase d'apprentissage.

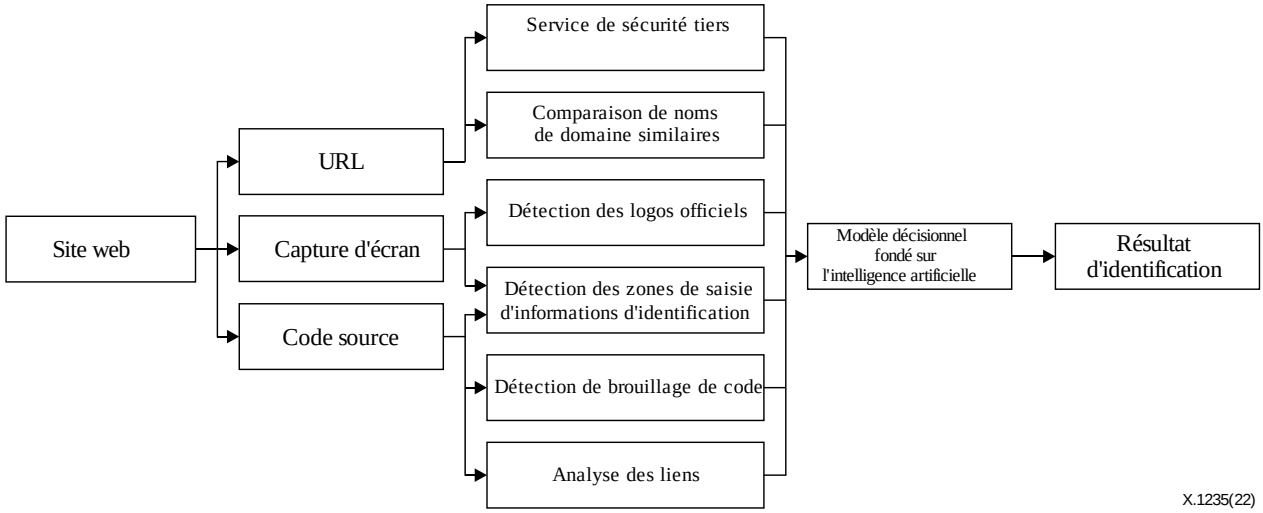


Figure 7-1 – Identification finale par évaluation de la réputation

Comme l'indique la Figure 7-1, le site web est analysé sur la base de son URL, de captures d'écran et de son code source. En ce qui concerne l'URL de la page web, le score de similarité du nom de domaine est calculé et on vérifie le résultat des services de sécurité tiers. Les captures d'écran sont soumises à une détection et une analyse des logos officiels et des zones de saisie d'informations d'identification. Le code source est examiné pour détecter les zones de saisie d'informations d'identification et une analyse des liens est effectuée. Le modèle décisionnel utilisant l'intelligence artificielle devrait prendre en compte tous ces aspects pour déterminer si un site Web est contrefait.

L'entraînement d'un tel modèle décisionnel devrait être effectué en collectant tous les sites web officiels à protéger, les sites contrefaits identifiés comme tels et des sites web normaux en tant qu'échantillons. Tous ces échantillons, c'est-à-dire tous les sites web collectés, sont évalués au moyen des contre-mesures recommandées ci-dessus (voir la Figure 7-1), ce qui permet d'attribuer un vecteur de score à chaque échantillon. Tous ces vecteurs de score et les types (contrefait ou véritable) de tous les échantillons peuvent être utilisés comme données d'apprentissage pour entraîner un outil de classification et former le modèle décisionnel utilisant l'intelligence artificielle. L'outil de classification peut être de type SVM (machine à vecteur de support) ou reposer sur un réseau neuronal profond.

Il serait judicieux que les sites web contrefaits identifiés comme tels par le modèle décisionnel utilisant l'intelligence artificielle soient vérifiés par des examinateurs humains afin d'éviter toute erreur d'identification.

7.2 Protection

Il est recommandé aux opérateurs de protéger les utilisateurs en adoptant les contre-mesures suivantes avec l'autorisation des utilisateurs:

- Avertir les utilisateurs: lorsqu'un utilisateur se rend sur un site dont on sait qu'il a été contrefait, il est recommandé à l'opérateur de mettre la requête en attente, de la rediriger vers une page d'avertissement pour alerter l'utilisateur du risque d'usurpation, et de demander à l'utilisateur de confirmer sa demande.
- Créer une liste de blocage: il est recommandé à l'opérateur de créer une liste de blocage des sites contrefaits afin de bloquer toute requête vers un site figurant sur cette liste. La liste de blocage peut être mise en œuvre au niveau des passerelles des réseaux de l'opérateur ou des serveurs DNS de l'opérateur, ou d'autres entités réseaux appropriées.
- Empêcher la propagation: il est recommandé à l'opérateur de bloquer les messages trompeurs contenant les liens vers les sites web contrefaits.

- Amener les utilisateurs à se protéger contre les sites web contrefaits: en leur rappelant périodiquement les risques et les caractéristiques des sites web contrefaits et en les informant des bonnes pratiques permettant d'éviter d'utiliser des sites web contrefaits. Les bonnes pratiques à recommander peuvent consister:
  - à ne pas cliquer sur un lien suspect dans un courriel ou un message;
  - à examiner soigneusement le nom de domaine dans la barre d'adresse et à le comparer avec celui du site officiel;
  - à utiliser les fonctions de sécurité des navigateurs web pour vérifier l'authenticité d'un site web.
- Protéger l'utilisateur en recherchant les noms de domaine actifs pour un même site web contrefait: un ensemble de noms de domaine d'un site web contrefait peut servir à prédire d'autres noms de domaine actifs du site contrefait. Si le contenu des sites correspondant aux nouveaux noms de domaine actifs ressemble à celui des sites contrefaits connus, ces sites peuvent être traités directement comme les sites contrefaits connus et faire l'objet des mesures de protection recommandées plus haut.

8 Mécanisme

Il est recommandé de mettre en œuvre un mécanisme systématique de lutte contre l'usurpation des sites web en combinant toutes les contre-mesures détaillées au § 7.

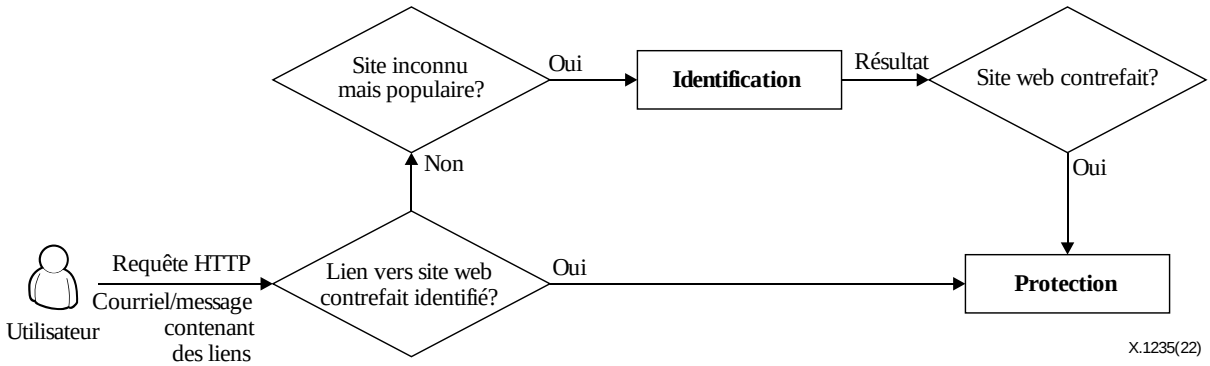


Figure 8-1 – Mécanisme de lutte contre l'usurpation des sites web

1) Comme indiqué à la Figure 8-1, les URL peuvent provenir des requêtes HTTP de l'utilisateur, ou bien de courriels ou de messages instantanés (partagés par une tierce partie ou autorisés par l'utilisateur).
2) Si un URL pointe vers un site web connu, il convient de mettre en œuvre des mesures de protection pour traiter la requête ou le message.
3) Si un URL pointe vers un site web inconnu populaire, il convient de mettre en œuvre des mesures d'identification pour déterminer la nature du site.
4) Si le résultat des mesures d'identification indique que le site web est contrefait, il convient de mettre en œuvre des mesures de protection pour empêcher l'utilisateur d'accéder au site.

Appendice I

Exemple de mécanisme de lutte contre les sites web contrefaits

(Cet Appendice ne fait pas partie intégrante de la présente Recommandation.)

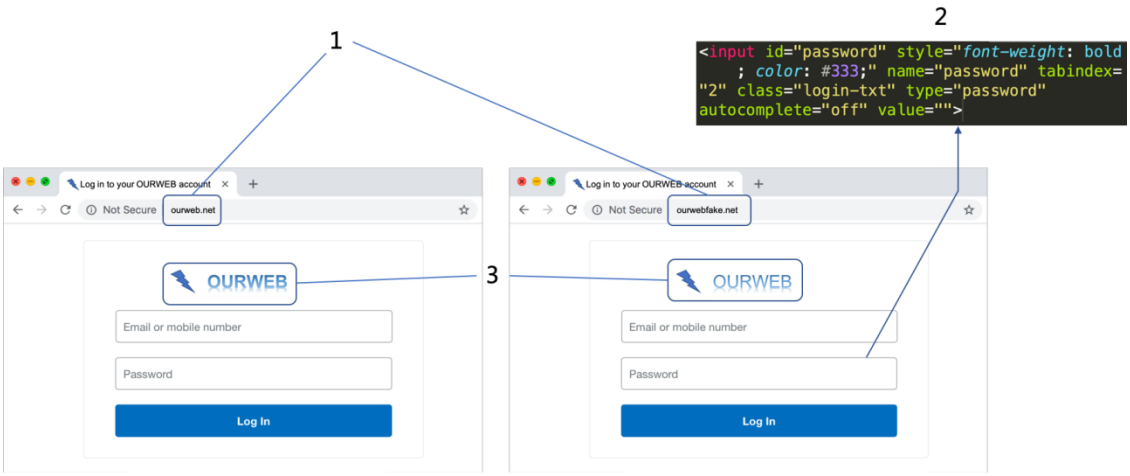


Figure I.1 – Exemple de site web officiel et de site web contrefait

Soit un site web connu appelé OURWEB, ayant pour nom de domaine "ourweb.net". Un pirate tente de créer un site web contrefait ayant pour nom de domaine "ourwebfake.net". En supposant que le nom de domaine du site web contrefait n'est pas connu de l'opérateur; lorsqu'un grand nombre d'utilisateurs demandent à accéder à ce site contrefait, celui-ci devient un site web inconnu populaire. Il convient alors de soumettre ce site à des mesures d'identification. Comme illustré à la Figure I.1, trois mesures permettent d'identifier le site web contrefait.

1) Les noms de domaine des deux sites sont similaires. Le score de similarité est estimé à 0,8, ce qui correspond à la proportion de caractères communs.

2) Le site web contrefait contient une zone de saisie d'informations d'identification, qui peut être identifiée à partir du code source. Le score est de 1 pour les zones de saisie d'informations d'identification.

3) Le logo du site web contrefait est similaire à celui du site web officiel. Le score de similarité est estimé à 0,9, correspondant à la proportion des descripteurs visuels communs tels qu'un descripteur SIFT.

4) Les scores des autres mesures sont estimés comme étant chacun égal à zéro.

Il est possible de construire le vecteur de score suivant: [0.8,0.9,0,1,0,0]. La première valeur du vecteur est le score de similarité des noms de domaine. La deuxième valeur est le score de similarité des logos. La quatrième valeur est le score relatif aux zones de saisie d'informations d'identification. Ce vecteur peut être rentré dans un outil de classification SVM pour obtenir un résultat global. Si le résultat indique que le site web est contrefait, des mesures de protection sont mises en œuvre afin d'empêcher les utilisateurs d'y accéder. Par exemple, le nom de domaine du site web peut être ajouté à la liste de blocage active sur les passerelles du réseau de l'opérateur.

Appendice II

Exemples de mesures techniques

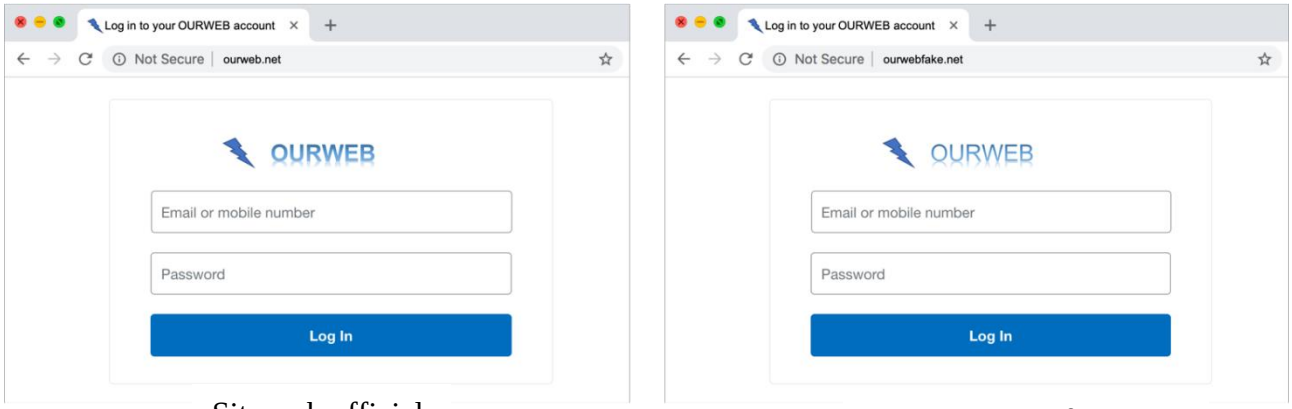
(Cet Appendice ne fait pas partie intégrante de la présente Recommandation.)

Exemple II.1: Comparaison de noms de domaines similaires

Soit un site web dont le nom de domaine officiel est "abc123.cn". Un site web contrefait utilise "abc123cn.org" comme nom de domaine. Le calcul de la similarité de Jaccard (obtenue en divisant la taille de l'intersection des ensembles de caractères des deux noms de domaine par la taille de leur union) pour les deux noms de domaines permet d'obtenir un score de similarité de 8/11 = 0,73.

Exemple II.2: Détection des logos officiels

La Figure II.1 présente le site web officiel de l'entité OURWEB sur la gauche, et le site contrefait correspondant sur la droite. On suppose que le logo officiel de OURWEB a été ajouté à la base de données de logos. Le logo du site contrefait sur la droite est très similaire au logo officiel du site OURWEB. L'algorithme de comparaison des similitudes permet de déterminer un score de similarité; par exemple, un score de 0,9 peut être déterminé ici.



Site web officiel

Site web contrefait

Figure II.1 – Site web contrefait comportant un logo officiel

Exemple II.3: Détection de brouillage de code

Certains sites web utilisent JavaScript pour générer du contenu web de façon dynamique, et font appel à un service de brouillage en ligne (par exemple, https://obfuscator.io/) pour brouiller le code JavaScript.

La Figure II.2 permet de constater que le code brouillé est très différent du code normal. Un modèle de classification de texte utilisant l'intelligence artificielle peut être utilisé pour prédire si un code est brouillé. L'algorithme attribue une valeur de probabilité de prédiction au code source (par exemple, 0,7).

Figure II.2 – Exemples de brouillage de code JavaScript

Exemple II.4: Détection des zones de saisie d'informations d'identification

Comme le montre la Figure II.1, le site web contrefait incite les utilisateurs à saisir leurs informations d'identification. Il est possible d'attribuer un score au site web en vérifiant s'il comporte un formulaire de saisie d'informations d'identification. Si c'est le cas, le site obtient le score de 1; dans le cas contraire, il obtient le score de 0.

Exemple II.5: Service de sécurité tiers

Des services de requêtes d'URL en ligne de fournisseurs de services de sécurité (tels que https://www.urlvoid.com/) peuvent aider à déterminer si un site web est contrefait. Si le site web examiné est classé comme contrefait, il obtient le score de 1; dans le cas contraire, il obtient le score de 0.

Exemple II.6: Analyse des liens

Sur la Figure I.1, le logo peut renvoyer aux ressources du site web officiel OURWEB. L'analyse des liens permet de déterminer que la ressource image provient du site ourweb.net (et non du site ourwebfake.net). Si les liens d'un site web renvoient à des ressources images officielles, le site web se voit attribuer le score de 1; dans le cas contraire il obtient le score de 0.

Exemple II.7: Combinaison des résultats

Les scores déterminés par les différentes méthodes appliquées au site suspect sont combinés pour former un vecteur. Le Tableau II.1 présente un exemple d'évaluation d'un site web. La deuxième colonne du tableau forme le vecteur de score. Ce vecteur peut être rentré dans un outil de classification afin de déterminer si le site web en question est un site contrefait.

Tableau II.1 – Exemple d'évaluation d'un site web

Méthodes d'identification	Score
Comparaison de noms de domaines similaires	0,8
Détection des logos officiels	0,9
Détection de brouillage de code	0,7
Détection de zones de saisie d'informations d'identification	1
Service de sécurité tiers	0
Analyse des liens	1

Tableau II.2 – Correspondance entre les méthodes d'identification et les caractéristiques des sites web contrefaits

Méthodes d'identification	Caractéristiques
Comparaison de noms de domaines similaires	Nom de domaine similaire
Détection des logos officiels	Similarité visuelle Emprunt d'éléments visuels
Détection de brouillage de code	Code source brouillé
Détection des zones de saisie d'informations d'identification	Saisie d'informations d'identification
Service de sécurité tiers	
Analyse des liens	Lien vers des sites connus

Bibliographie

[b-UIT-T X.1126] Recommandation UIT-T X.1126 (2017), *Lignes directrices relatives à l'atténuation des effets négatifs des terminaux infectés dans les réseaux mobiles*.

[b-UIT-T X.1244] Recommandation UIT-T X.1244 (2008), *Aspects généraux de la lutte contre le spam dans les applications multimédias IP*.

[b-SVM] Nature biotechnology, 2006, 24(12) 1565-1567*: What is a support vector machine?*

SÉRIES DES RECOMMANDATIONS UIT-T

Série A	Organisation du travail de l'UIT-T
Série D	Principes de tarification et de comptabilité et questions de politique générale et d'économie relatives aux télécommunications internationales/TIC
Série E	Exploitation générale du réseau, service téléphonique, exploitation des services et facteurs humains
Série F	Services de télécommunication non téléphoniques
Série G	Systèmes et supports de transmission, systèmes et réseaux numériques
Série H	Systèmes audiovisuels et multimédias
Série I	Réseau numérique à intégration de services
Série J	Réseaux câblés et transmission des signaux radiophoniques, télévisuels et autres signaux multimédias
Série K	Protection contre les perturbations
Série L	Environnement et TIC, changement climatique, déchets d'équipements électriques et électroniques, efficacité énergétique; construction, installation et protection des câbles et autres éléments des installations extérieures
Série M	Gestion des télécommunications y compris le RGT et maintenance des réseaux
Série N	Maintenance: circuits internationaux de transmission radiophonique et télévisuelle
Série O	Spécifications des appareils de mesure
Série P	Qualité de transmission téléphonique, installations téléphoniques et réseaux locaux
Série Q	Commutation et signalisation et mesures et tests associés
Série R	Transmission télégraphique
Série S	Équipements terminaux de télégraphie
Série T	Terminaux des services télématiques
Série U	Commutation télégraphique
Série V	Communications de données sur le réseau téléphonique
Série X	**Réseaux de données, communication entre systèmes ouverts et sécurité**
Série Y	Infrastructure mondiale de l'information, protocole Internet, réseaux de prochaine génération, Internet des objets et villes intelligentes
Série Z	Langages et aspects généraux logiciels des systèmes de télécommunication

Imprimé en Suisse
Genève, 2022