International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

X.1235

(01/2022)

SERIES X: DATA NETWORKS, OPEN SYSTEM COMMUNICATIONS AND SECURITY

Cyberspace security – Countering spam

Technologies to counter website spoofing for telecommunication organizations

Recommendation ITU-T X.1235

ITU-T X-SERIES RECOMMENDATIONS

DATA NETWORKS, OPEN SYSTEM COMMUNICATIONS AND SECURITY

PUBLIC DATA NETWORKS	X.1–X.199
OPEN SYSTEMS INTERCONNECTION	X.200–X.299
INTERWORKING BETWEEN NETWORKS	X.300–X.399
MESSAGE HANDLING SYSTEMS	X.400–X.499
DIRECTORY	X.500–X.599
OSI NETWORKING AND SYSTEM ASPECTS	X.600–X.699
OSI MANAGEMENT	X.700–X.799
SECURITY	X.800–X.849
OSI APPLICATIONS	X.850–X.899
OPEN DISTRIBUTED PROCESSING	X.900–X.999
INFORMATION AND NETWORK SECURITY	
General security aspects	X.1000–X.1029
Network security	X.1030–X.1049
Security management	X.1050–X.1069
Telebiometrics	X.1080–X.1099
SECURE APPLICATIONS AND SERVICES (1)	
Multicast security	X.1100–X.1109
Home network security	X.1110–X.1119
Mobile security	X.1120–X.1139
Web security (1)	X.1140–X.1149
Application Security (1)	X.1150–X.1159
Peer-to-peer security	X.1160–X.1169
Networked ID security	X.1170–X.1179
IPTV security	X.1180–X.1199
CYBERSPACE SECURITY	
Cybersecurity	X.1200–X.1229
Countering spam	**X.1230–X.1249**
Identity management	X.1250–X.1279
SECURE APPLICATIONS AND SERVICES (2)	
Emergency communications	X.1300–X.1309
Ubiquitous sensor network security	X.1310–X.1319
Smart grid security	X.1330–X.1339
Certified mail	X.1340–X.1349
Internet of things (IoT) security	X.1350–X.1369
Intelligent transportation system (ITS) security	X.1370–X.1399
Distributed ledger technology (DLT) security	X.1400–X.1429
Application Security (2)	X.1450–X.1459
Web security (2)	X.1470–X.1489
CYBERSECURITY INFORMATION EXCHANGE	
Overview of cybersecurity	X.1500–X.1519
Vulnerability/state exchange	X.1520–X.1539
Event/incident/heuristics exchange	X.1540–X.1549
Exchange of policies	X.1550–X.1559
Heuristics and information request	X.1560–X.1569
Identification and discovery	X.1570–X.1579
Assured exchange	X.1580–X.1589
Cyber Defence	X.1590–X.1599
CLOUD COMPUTING SECURITY	
Overview of cloud computing security	X.1600–X.1601
Cloud computing security design	X.1602–X.1639
Cloud computing security best practices and guidelines	X.1640–X.1659
Cloud computing security implementation	X.1660–X.1679
Other cloud computing security	X.1680–X.1699
QUANTUM COMMUNICATION	
Terminologies	X.1700–X.1701
Quantum random number generator	X.1702–X.1709
Framework of QKDN security	X.1710–X.1711
Security design for QKDN	X.1712–X.1719
Security techniques for QKDN	X.1720–X.1729
DATA SECURITY	
Big Data Security	X.1750–X.1759
Data protection	X.1770–X.1789
IMT-2020 SECURITY	X.1800–X.1819

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T X.1235

Technologies to counter website spoofing for telecommunication organizations

Summary

Website spoofing is a major threat for telecommunication organizations, especially for operators. It is recommended for telecommunication operators to adopt technologies to counter website spoofing to protect their customers and guard their reputation and revenue. Recommendation ITU-T X.1235 analyses the main measures to spoof a website and recommends technologies to identify spoofed websites. These measures can be regarded as guidelines for telecommunication organizations to protect websites from being spoofed. A similar approach may be implemented against spoofing of any website, including those of banks, insurance companies, Internet shops, etc.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T X.1235	2022-01-07	17	11.1002/1000/14797

Keywords

Countermeasures, website-spoofing.

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Introduction

Web spoofing has played a key role in Internet fraud in recent years. Scammers usually choose websites of well-known organizations or companies to collect visitors' credentials or to spread malicious software. This results in financial loss for both visitors and telecommunication operators, and in reputation loss for telecommunication operators.

As the websites of telecommunication operators have become one of the most important portals for their customers to enquire about and subscribe to all kinds of services, it has been observed worldwide that scammers continually try to counterfeit the websites of telecommunication operators with the objective of cheating customers. This Recommendation makes a thorough analysis of website spoofing and recommends a series of organized countermeasures.

Recommendation ITU-T X.1235

Technologies to counter website spoofing for telecommunication organizations

1 Scope

This Recommendation recommends technologies for telecommunication organizations to identify website spoofing in time and protect their websites from being spoofed. After a systematic analysis of spoofing measures and features, it recommends best practices for applying network-side technologies with the assistance of user-side technologies to counter website spoofing.

Conformance with this Recommendation is not to be taken as any proof of evidence for claiming compliance with any national or regional law, regulation or policy. The technical, organizational and procedural means described in this Recommendation do not in any way guarantee the constitution of any level of security that may be put upon certain correspondence by specific national or regional law, regulation or policy.

A similar approach may be implemented against spoofing of any website, including those of banks, insurance companies, Internet shops, etc.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

None.

3 Definitions

3.1 Terms defined elsewhere

None.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 convolutional neural network: A deep learning algorithm which takes in an input image, assigns importance to various aspects/objects in the image, and is able to differentiate one from the other.

3.2.2 long short-term memory networks: A type of recurrent neural network capable of learning order dependence in sequence prediction problems.

3.2.3 recurrent neural network: A class of neural networks that allow previous outputs to be used as inputs while having hidden states.

3.2.4 scale-invariant feature transform (SIFT): A feature detection algorithm in computer vision to detect and describe local features in images, which composes a descriptive identifier invariant to translations, rotations and scaling transformations in the image domain and robust to moderate perspective transformations and illumination variations. Experimentally, the SIFT descriptive

identifier has been proven to be very useful in practice for image matching and object recognition in real-world conditions.

3.2.5 speeded up robust features: A local feature detector and descriptive identifier in computer vision to detect and describe local features in images, which is several times faster than scale-invariant feature transform (SIFT) and is more robust against different image transformations than SIFT.

3.2.6 spoofed website: A website created by website spoofing (see clause 3.2.8).

3.2.7 support vector machine: A supervised machine learning model that solves two-group classification problems with the given sets of labelled training data for each category; they are able to categorize new sets.

3.2.8 website spoofing: This is a set of malicious behaviours to mimic a website that is well known to the public or a group of people; the fake website is then used to abuse the trust of visitors to achieve malicious /illegal goals such as fraud, invasion of privacy, etc.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AI	Artificial Intelligence
DNS	Domain Name Service
GUI	Graphic User Interface
LSTM	Long Short-Term Memory
OCR	Optical Character Recognition
PII	Personally Identifiable Information
QR	Quick Response
SIFT	Scale-Invariant Feature Transform
SURF	Speeded Up Robust Features
SVM	Support Vector Machine
URL	Uniform Resource Locator

5 Conventions

This Recommendation uses the following conventions:

The keywords "**should**" indicate a requirement which is recommended but which is not absolutely required. Thus, this requirement need not be present to claim conformance.

The keyword "**may**" indicate an optional requirement which is permissible, without implying any sense of being recommended.

In the body of this Recommendation, the word **"can"** sometimes appear, in which case they are to be interpreted as **"is able to"**.

6 Analysis of website spoofing

Website spoofing is a set of malicious behaviours to mimic a website known to the public or a group of people. The spoofed websites collect visitors' credentials or spread malware to achieve malicious/illegal goals such as fraud, privacy invasion, etc.

Website spoofing could cause serious damage in several ways. For example, it could lead to financial loss for both customers and operators, damage operators' reputation, etc. Furthermore, the leaked user

information could be misused for a long time even if the spoofed website disappears. Therefore, it is important for telecommunication organizations, especially operators to adopt useful technologies to counter website spoofing.

6.1 Website spoofing scenario

A typical scenario for creating a spoofed website to accomplish an improper objective has three phases as shown in Figure 6-1.

– **Phase one**: Building the website
  - Step one: Develop and operate a spoofed website which is identical to or similar to some well-known websites.

– **Phase two**: Spread uniform resource locator (URL)
  - Step two: Spread the spoofed website through various methods such as email, short message service and instant messaging services, etc.

– **Phase three**: Collect credentials or spread malware
  - Step three: Trick visitors into inputting credentials or downloading malicious software. Once victims have accessed a spoofed website, they might mistakenly believe that the website is the real one and log in with their credentials or download malicious software.
  - Step four: Retrieve visitors' credentials and prepare for the next scam or other malicious activity.

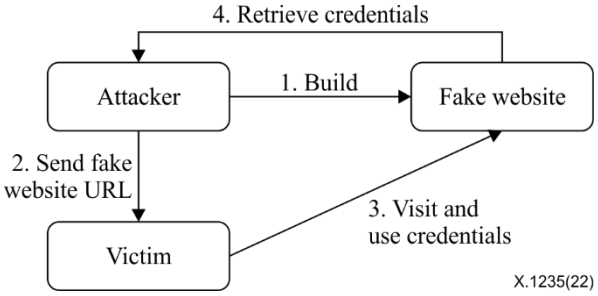

Figure 6-1 – Four phases for using a spoofed website to collect user credentials

6.2 Characteristics of website spoofing

6.2.1 Characteristics of spoofed websites

The spoofed websites use similar techniques to trick visitors into thinking they are visiting the real ones. They have several common characteristics:

– Visual similarity: A spoofed website might be a complete copy of the real one, including its visual content and interactive logic.

– Borrow visual elements: A spoofed website might borrow some important visual elements from the real website.

– Link well-known websites: A spoofed website might include links to other well-known websites to mimic an authorized partner website, a related website, or a different version of the website.

– Similar domain name: The domain name of a spoofed website would be very similar to the real one.

– Obscure domain name: Sometimes attackers use URL shortening services or quick response (QR) codes to obscure the real domain name of a website as sharing a link.

– Duplicate domain name: A set of domain names might be reserved for a spoofed website and to extend the life of the spoofed website. Several of them will be published once a time.

– Obfuscate source code: The source code of a spoofed website might be obfuscated to confuse security scanning programs.
– Credentials input: Spoofed websites might lure visitors to input credentials.
– Redirect to known website: A spoofed website may redirect to known websites (e.g., google.com) during its idle time to avoid automatic online-fraud scanning.
– Fake address bar: Some spoofed websites might be specifically designed for mobile web browsers. The website might place a fake copy of a mobile browser's address bar at the top of the mobile phone's screen and keep the copy locked in by abusing the mobile browser's graphic user interface (GUI) feature.

6.2.2 Characteristics of related activities

To lure or cheat people to visit and interact with spoofed websites, many activities could be taken. The characteristics or tactics of these activities could be one or more of those listed here:

– Spread messages to let as many people as possible know of the existence of the spoofed website.
– Hijack or poison a domain name service (DNS). So that, any visits to the genuine website will be redirected to the spoofed one.
– Use benefits (e.g., coupon and gifts, etc.) to induce users to click on fake website links.
– Install malicious software on the victim's phone to replace the real website request by the fake one.
– Contact users by imitating a customer service to provide them with fake website links.

6.3 Consequences

A spoofed website could have many negative effects on users and website owners. The possible consequences are listed below:

– User property loss: User credentials and personally identifiable information (PII) might be leaked. This information may be used by scammers to go on with further fraud. Users might be tricked into downloading Trojans, crypto-mining codes and other malware. User terminals might be controlled to steal financial accounts, spread spam and infect other users. These activities would result in loss of battery or malfunction of terminals, and would also result in direct and huge financial loss.
– Operators' property loss: The normal transactions between users and operators would get lost and even be restrained for a long term.
– Reputation loss: The reputation of the real websites and even the operator's public trust would be damaged, there would be a large number of customer complaints, and a long-term negative evaluation.

7 Countermeasures

The countermeasures for website spoofing can be divided into two aspects: identification and protection. The former identifies spoofed websites from other websites; see clause 7.1. The latter protects users from accessing or being cheated by known spoofed websites; see clause 7.2.

It is not necessary for a telecommunication organization to simultaneously use all the technologies recommended in this clause. Due to acquired data, legislation environment and subscriber requirements, etc., a telecommunication organization should flexibly adopt the appropriate technologies as described in this clause.

7.1 Identification

According to the characteristics of spoofed websites identified as described in clause 6.2.1, the following countermeasures are recommended for consideration:

7.1.1 Similar domain name comparison

It is recommended that operators maintain a list of domain names of well-known websites that are susceptible to be spoofed. If a domain name is very similar (but not identical) to the domain name in the list, it could be the domain name of a spoofed website. If the domain name has been shortened by a URL shortening service, it should be converted back to the original URL before comparison.

There are many methods to calculate the similarity of two domain names, including: edit distance, Jaccard similarity, longest common sub-sequence, visual similarity conversion, etc.

- **Edit distance method**: given two domain names, A and B, the method calculates the minimum number of edits required to convert from A to B. The smaller the edit distance, the higher the similarity.
- **Jaccard similarity method**: the size of the intersection of two domain name characters sets is divided by the size of the union. The bigger the ratio, the higher the similarity.
- **Longest common sub-sequence method**: the length of the longest common sub-sequence of two domain names is calculated. The longer the length, the higher the similarity.

  NOTE – Longest common subsequence means the longest subsequence common to two given sequences, provided that the elements of the subsequence need not occupy consecutive positions within the original sequences. The common subsequence is a strictly ascending sequence of the indices of the two given sequences. For example, if "abcde" and "akcve" are two given sequences, then "ace" is the longest subsequence of the two sequences.
- **Visual similarity conversion method**: visually similar characters are replaced before the comparison. For example, "0" and "o" can be replaced by each other; "1" and "i" can be replaced by each other, etc. For example, the domain name "z00.com" could be converted to "zoo.com". This method could enhance the performance of the similarity. However, under certain circumstances, the visual similarity conversion could decrease the performance. So, it would be prudent to perform the similarity comparison before and after the conversion.

7.1.2 Official logo detection

It is recommended to compare the images or other visual elements in an unknown website with the official logos. Official logos include but are not limited to the company or organization's brand, advertising designs and other iconic elements. The method to detect official logos includes the following steps:

- Before the detection, the official logos should be collected and stored in a logo database.
- The visual elements of an unknown webpage are downloaded or screenshot and compared with the logos in the logo database. The similarity between two images can be calculated by two popular descriptors, which are scale-invariant feature transform (SIFT) and speeded up robust features (SURF).
- If the source code of the detecting webpage has been obfuscated, downloading the visual elements could be hard. And then some kind of browser-based web application automation testing framework can be used to take a screenshot of the web page. Before the logo comparison, the logos can be cropped from the screenshot by using an AI-based logo detection model. This AI-based logo detection model could be well trained using open-source logo databases.

7.1.3 Code obfuscation detection

The spoofed websites might obfuscate the malicious or spoofing code to make its analysis difficult. The obfuscated code is very different from the normal code. There are many methods to detect obfuscated code, including convolutional and recurrent neural network-based classifiers, etc.

- The classifiers based on convolutional neural network can automatically extract the n-gram features of normal and obfuscated source code. An n-gram is a contiguous sequence of n words from a given source code. The number of words can be set to 5 or higher to fully capture the discriminative features for classification.
- The classifiers based on recurrent neural networks are mainly based on long short-term memory (LSTM) networks or transformer networks. These classifiers treat the source code as a character sequence, automatically identify the sequence patterns between the normal code and the obfuscated code, and then determine if it is an obfuscated code.

7.1.4 Credentials input detection

Credential inputs usually exist in the form of "input forms" in the source code of webpages. There is usually a "password" attribute in the source code of the input form (this attribute guarantees that the content of the password is not displayed when the user enters it). The attribute information can be quickly located using regular expressions or webpage analysis tools. If the website source code is obfuscated, the optical character recognition (OCR) method can be used to detect credential inputs in the screenshots of the website. OCR method will locate the text areas in a screenshot of a web page first, and then convert the text areas into texts. If the converted texts contain some keywords such as "password", then credential inputs are detected in the webpage.

7.1.5 Third-party security service

It is recommended that operators use the security services provided by third-party security vendors to obtain the attributes and the statistics of a website, including website traffic, reputation, plug-ins, domain name registration, certification and other security information. This information could be used as additional information to identify a spoofed website.

7.1.6 Link analysis

The spoofed website may copy the hyperlinks in the official website to reuse the visual elements of the official website in its spoofed webpages. This behaviour could be detected by link analysis. Link analysis collects the links of the visual elements on the official website and compares them to the links on the website to be detected. If the two websites have the same hyperlinks, or many of them, it may be a potential spoofed website.

7.1.7 Final identification based on reputation scoring

It is recommended that the operators use an artificial intelligence (AI) decision model to combine all the analysis or detection results of the countermeasures recommended above, and make a comprehensive decision on whether a website is a spoofed website or not. In the AI decision model, the weight of each countering measure has been determined automatically after the training phase.

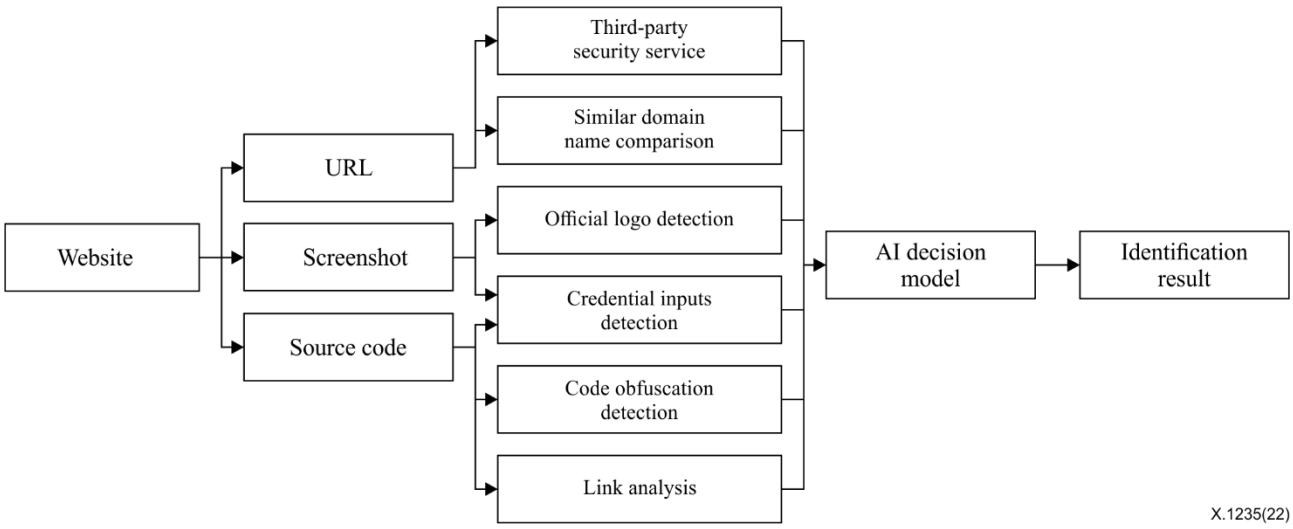

Figure 7-1 – Final identification based on reputation scoring

As shown in Figure 7-1, the website is analysed based on its URL, its screenshots and its source code. For the webpage URL, domain name similarity score is computed, and the result of third-party security services is checked. For the screenshots, official logos and credential inputs are detected and analysed. For source code, credential inputs is detected, and link analysis is performed. To determine if the website is spoofed, the AI decision model should consider all these aspects.

To train this AI decision model, all the official websites to be protected, the known spoofed websites and other normal websites should be collected as samples. All the samples (that is, the collected websites) are scored by the countermeasures recommended above as shown in Figure 7-1 and every sample will have a score vector. All these score vectors and the types (spoofed or not spoofed) of all the samples can be used as training data to train a classifier to form the AI decision model. The classifiers could be support vector machine (SVM)-based or deep neural network-based classifiers.

It would be useful that the spoofed websites identified by the AI decision model be audited by human reviewers to avoid misidentification.

7.2 Protection

It is recommended that the operator protects users by adopting the following countermeasures with the permission of users:

- Warn users: Once a user visits a known spoofed website, the operator is recommended to hold the request, redirect it to a warning page to alert the spoofing risk, and ask the user to confirm the request.
- Create a block list: The operator is recommended to create a block list of spoofed websites to interrupt all requests on the block list. The block list could serve in the gateways in the operator's networks or in the operator's DNS servers or other proper network entities.
- Prevent propagation: The operator is recommended to block the jam messages containing the link to the spoofed websites.
- Guide users to protect themselves from suspicious spoofed websites: Periodically remind users of the risks and the features of the spoofed websites and inform them of the best practices to avoid using spoofed websites. The best practices recommended include:
  - Do not click on strange links in emails or messages.
  - Carefully observe the domain name information in the address bar and compare it with the official website.
  - Use the security function of web browsers to check the authenticity of a website.

- Protect the user by discovering living domain names of the same spoofed websites: A set of domain names of a spoofed website could be used to predict other living domain names of the website. If the content of the websites under new living domain names complies with that of the known spoofed websites, they could be treated as the same spoofed websites directly and take the protection measures recommended above.

8 Mechanism

A systematic mechanism for countering spoofed website is recommended with the combination of all the countermeasures described in clause 7.

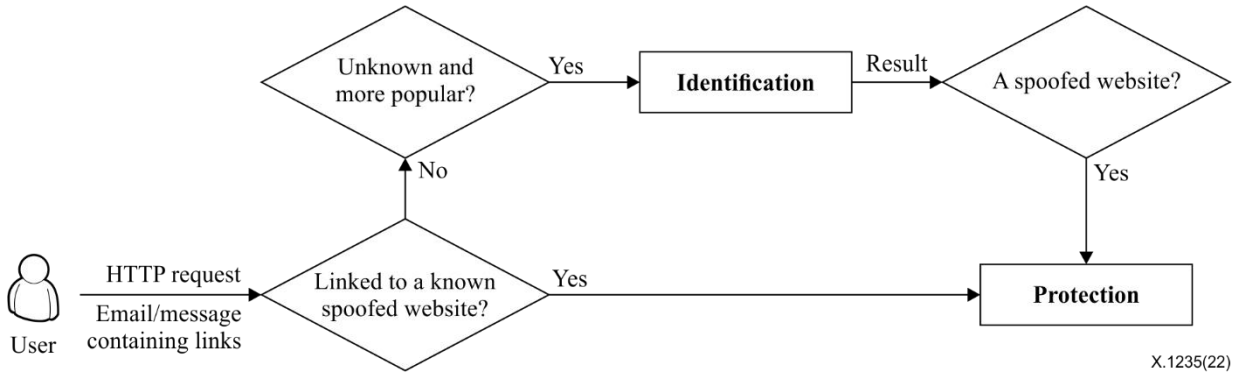

Figure 8-1 – A mechanism for countering spoofed websites

1) As shown in Figure 8-1, URLs may come from user HTTP requests, or user emails or instant messages (shared by a third-party or authorized by users).
2) If a URL points to a known website, protection countermeasures should be taken to handle the request or message.
3) If a URL points to a popular unknown website, identification countermeasures should be taken to identify the type of website.
4) If the result of identification countermeasures indicates a spoofed website, protection countermeasures should be taken to protect the user from accessing the website.

Appendix I

An example of the mechanism for countering spoofed websites

(This appendix does not form an integral part of this Recommendation.)

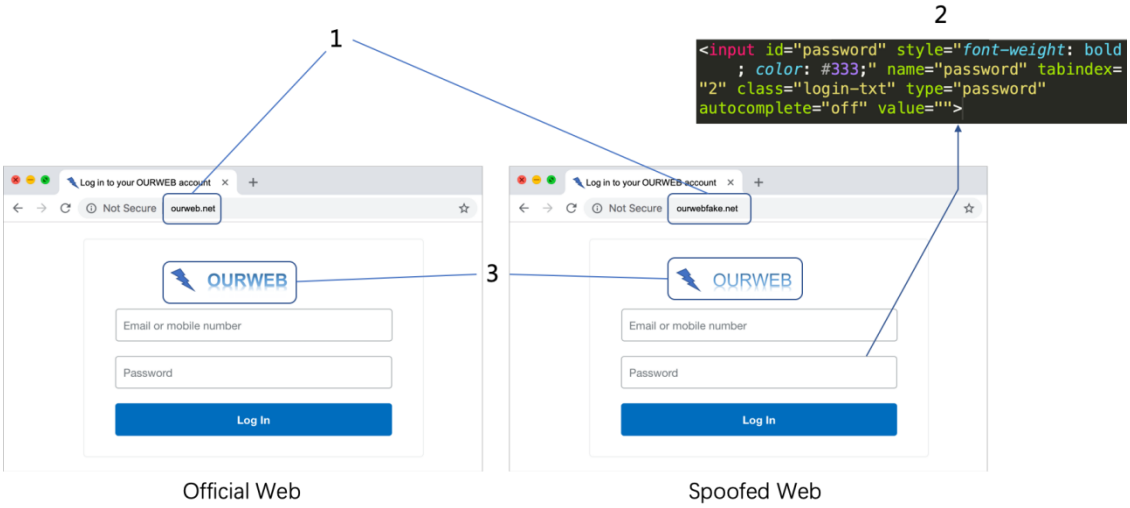

Figure I.1 – An example of a website and a spoofed website

Suppose there is a well-known website called OURWEB whose domain name is "ourweb.net". A hacker tries to create a spoofed website whose domain name is "ourwebfake.net". Suppose the domain name of the spoofed website is unknown to the operator, then when there are many user requests for the spoofed website, the website becomes a popular unknown website. Then the website should be treated by identification countermeasures. As shown in Figure I.1, there are three countermeasures to identify the spoofed website.

1) The domain names of the two websites are similar. Assume the similarity score is 0.8, which corresponds to the proportion of common characters.

2) There is a credential input in the spoofed website which can be identified from the source code. The score of credential inputs is 1.

3) The logo of the spoofed website is similar to the official website. Suppose the similarity score is 0.9, which is the proportion of common visual descriptive identifiers such as SIFT descriptive identifier.

4) Suppose the scores of the other countermeasures are zero.

Then we can compose a score vector [0.8,0.9,0,1,0,0]. The first value in the vector is the similarity score of the domain names. The second value is the similarity score of the logo. The fourth score is credential input score. This vector can be input to a SVM classifier to get the result. If the result indicates a spoofed website, it will be treated by protection countermeasures to protect users from accessing the website. For example, the domain name of the website should be added to the block list in the gateways of the operators' network.

Appendix II

Examples of technical measures

(This appendix does not form an integral part of this Recommendation.)

Example II.1: Similar domain name comparison

Suppose the official domain name is "abc123.cn". A spoofed website uses "abc123cn.org" as its domain name. By calculating the Jaccard similarity (the size of the intersection of two domain name characters sets is divided by the size of the union.) of the two domain names, we get the similarity score 8/11 = 0.73.

Example II.2: Official logo detection

As shown in Figure II.1, the official website of OURWEB is shown on the left, and the corresponding spoofed website is shown on the right. Suppose the official logo of OURWEB is added to the logo database. The logo of the spoofed website on the right is very similar to the official logo of OURWEB. Using the similarity comparing algorithm, a similarity score (e.g., 0.9) can be determined.

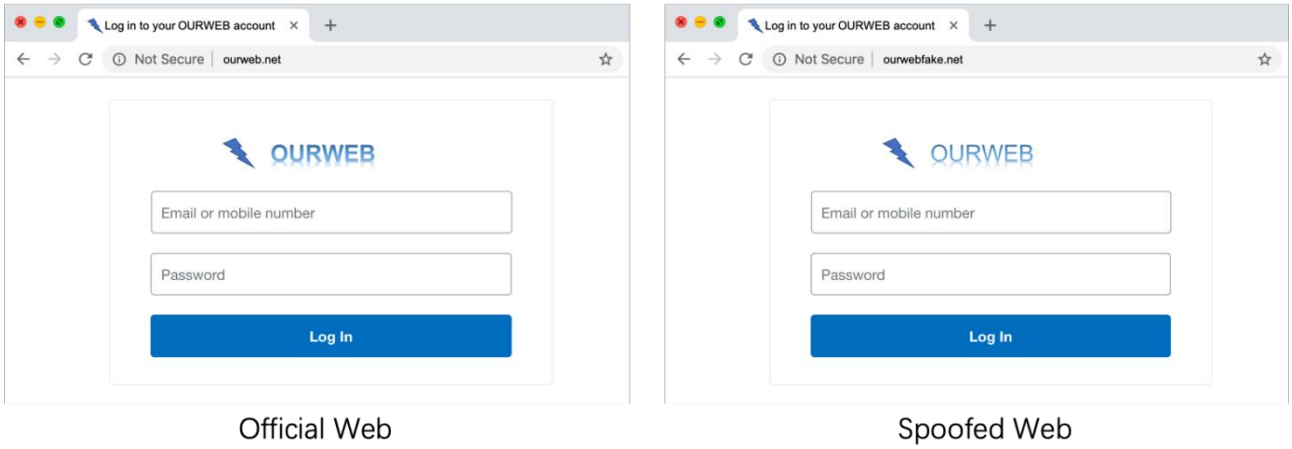

Figure II.1 – A spoofed website with official logo

Example II.3: Code Obfuscation Detection

Some websites use JavaScript to dynamically generate web content, and use an online obfuscation service (e.g., https://obfuscator.io/) to encrypt JavaScript code.

As shown in Figure II.2, the obfuscated code is significantly different from the normal code. An AI-based text classification model can be used to predict whether the code is obfuscated. The algorithm gives the source code a prediction probability value (e.g., 0.7).

Figure II.2 – Examples of JavaScript code obfuscation

Example II.4: Credentials input detection

As shown in Figure II.1, the spoofed website tricks users into entering credentials. The website can be scored by checking whether there is a form on the page for entering credentials. If there is a form for entering credentials, the site is scored 1, otherwise it is scored 0.

Example II.5: Third-party security service

Online URL query services from security vendors (such as https://www.urlvoid.com/) can help determine if a website is a spoofed website. If the detected website is classified as a spoofed website, the score is 1, otherwise it is 0.

Example II.6: Link analysis

In Figure I.1, the logo may refer to the resources of the OURWEB's official website. Link analysis can determine that the image resource is from ourweb.net (not ourwebfake.net). If there are links to official image resources on the website, it can be scored 1, otherwise it can be scored 0.

Example II.7: Combining results

The scores of the different methods of spoofed websites are combined into a vector. Table II.1 is an example of scoring a website. The second column of the table forms a score vector. This vector can be entered into a classifier to determine if the website is a spoofed website.

Table II.1 – An example of scoring a website

Identification Methods	Score
Similar domain name comparison	0.8
Official logo detection	0.9
Code obfuscation detection	0.7
Credentials input detection	1
Third-party security service	0
Link analysis	1

Table II.2 – Mapping between identification methods and characteristics of spoofed websites

Identification methods	Characteristics
Similar domain name comparison	Similar domain name
Official logo detection	Visually similar Borrow visual elements
Code obfuscation detection	Obfuscate source code
Credentials input detection	Credentials input
Third-party security service	
Link analysis	Link well-known websites

Bibliography

[b-ITU-T X.1126] Recommendation ITU-T X.1126 (2017), *Guidelines on mitigating the negative effects of infected terminals in mobile networks*.

[b-ITU-T X.1244] Recommendation ITU-T X.1244 (2008), *Overall aspects of countering spam in IP-based multimedia applications*.

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

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	**Data networks, open system communications and security**
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2022