

# A Potential IRI based Phishing Strategy

Yingjie Fu, Xiaotie Deng, Liu Wenyin

Dept of Computer Science, City University of Hong Kong, 83 Tat Chee Ave.,  
Hong Kong SAR., China  
{anthony@cs., csdeng@, csliuwy@}cityu.edu.hk

**Abstract.** We anticipate a potential phishing strategy by obfuscation of Web links using Internationalized Resource Identifier (IRI). In the IRI scheme, the glyphs of many characters look very similar while their Unicodes are different. Hence, certain different IRIs may show high similarity. The potential phishing attacks based on this strategy are very likely to happen in the near future with the boosting utilization of IRI. We report this potential phishing strategy to provoke much further dissections of related counter measures.

**Keywords:** Internet security, Anti-phishing, Internationalized Resource Identifier (IRI)

## 1. Introduction

Phishing webpages are webpages forged to mimic the webpages of certain real companies offering Internet transactions in order to spoof end users to leak their private information. It is a popular way for phishers to use visually and semantically similar URIs and similar webpages to spoof people. Unwary Internet users who are induced to access phishing webpages could be deceived to expose their bank accounts, passwords, credit card numbers, or other important information to the phishers.

In this paper, we report a potential phishing attack that depends on the utilization and popularization of Internationalized Resource Identifier (IRI) [2], as could be a quite severe problem. We also anticipate provoking a series of related further dissections on the potential flaw of IRI on the anti-phishing aspect.

## 2. IRI based Phishing Strategy

Rapid evolution of the Internet requires advanced security developments to make it safe. It is not an over claim if we say it is unimaginable if the Internet is not available. However, the Internet is still not mature, and its development is an endless process. In the very beginning, people use IP address directly to access a webpage or other Internet resources. Later, uniform resource identifiers (URI) [1], which are ASCII based, are used to locate or access a webpage or other Internet resources. With the popularization of the Internet, people speaking languages other than English are

demanding to use a unified representation of internet resource identifiers to locate the information resources, while URI has its inborn deficiency for this purpose. It uses ASCII and can represent a set of very limited number of readable characters. IRI is such a standard proposed as a complement of URI. An IRI is a sequence of characters chosen from the Unicode [4], which could be used instead of URIs to identify resources. However, the utilization of IRI could bring in severe, potential phishing attacks, since the Universal Character Set (UCS)[4] covers almost all characters in the word for information exchanges, in which a lot of visually similar characters and semantically similar characters co-exist. Figure 1 shows some of the similar characters to “a”.



**Figure 1.** Similar characters of “a” in Arial Unicode MS Font (adapted from UC-SimList [3], and the codes for under each character are in hexadecimal form)

It is very easy for phishers to spoof users by replacing characters in a target IRI with similar ones. Although a phishing IRI looks very similar or exactly the same with a target one, they are definitely different in coding level, and people could be victims of this kind of scams without knowing what is happening. Hence, this kind of phishing attacks is very likely to happen in the near future.

## Conclusion and Future Work

An IRI based potential phishing strategy is report in this paper. Phishers may use visually and semantically similar characters in UCS to mimic to IRIs of real webpages. Various methods can be used to detect this kind of phishing attacks. As a following up research, we consider nondeterministic finite automaton (NFA) to be a possible counter measure of this kind of phishing attacks.

## Acknowledgement

We thank Dr. Felix Sasaki from University of Bielefeld for his useful discussion and suggestions.

## Reference

1. Berners-Lee T., Fielding R., Masinter L., *RFC 3986: Uniform Resource Identifier (URI): Generic Syntax*, The Internet Society (2005), Jan. 2005.
2. Duerst M.,Suignard M., *RFC 3987: Internationalized Resource Identifiers (IRIs)*, The Internet Society (2005), Jan. 2005.
3. Fu Y., <http://www.cs.cityu.edu.hk/~anthony/AntiPhishing/IRI>, Jun. 2005
4. The Unicode Consortium, <http://www.unicode.org/>