Security Risk of Malicious Sites Direction Caused by Misinterpretation of Special Characters in URL

Satoki TSUJI†, Hajime SHIMADA†, Yukiko YAMAGUCHI†, Hirokazu HASEGAWA†
†Nagoya University

Summary
In this study, we focus on the automatic linking of URLs from plain text, and describe the behavior of automatic linking that splits and embeds unintended URLs when the text contains special characters. As a result, we confirm that there is a risk of users being led to malicious sites without being aware of it.

Automatic linking
Automatic linking is a feature that determines URLs by regular expressions and links to them automatically.

Plain text
Hi, a.example.com and b.example.com are great!

Linked text
Hi, a.example.com and b.example.com are great!

Diverted to Malicious Sites
URL is split and diverted to malicious sites by automatic linking libraries that cannot handle special characters properly.

Malicious URL
http://example.com?q=ABC...&e=^evil.example.com?q=CDE...

Application A
http://example.com?q=ABC...

Application B
http://example.com?q=ABC...

Benign site
http://example.com?q=ABC...

Application B, which cannot handle special characters, will generate malicious links.

Real World Case Studies
• We confirmed this risk in some applications (e.g. Fig. 1).
• We also confirmed this risk in a library.*
• Countermeasures and development of safe libraries will be required.

* We reported them to developer, but they answered "phishing related risks are not vulnerability."