IDN Spoof漏洞自动化挖掘

lyle / 2018-12-09 07:32:00 / 

---

# 0x00 背景

## 国际化域名

国际化域名（Internationalized Domain Name，IDN）又称特殊字符域名，是指部分或完全使用特殊的文字或字母组成的互联网域名，包括中文、法语、阿拉伯语、希伯来语或拉丁字母等非英文字母，这些文字经

## 同形异义攻击

同形异义字是利用IDN中一些非拉丁字符语种的字母与拉丁字符非常相似，字面看很难区分的特性，找到对应的字符来实现钓鱼攻击。例如16□.com(U+10D5)、16□.com (U+1012)、16□.com (U+0499)
都在一定程度上和163.com有相似性，基于一些开放的https证书服务这些域名还能取得相应的证书，进一步增加钓鱼成功的可能性。

## Punycode

Punycode是RFC 3492标准设计的编码系统，用于把Unicode转换为可用的DNS系统的编码，比如16□.com就会被转成xn--16-8tc.com，这在一定程度上可以防止IDN欺骗。

# 0x01 漏洞介绍

在主流浏览器中，Chromium Project对这类漏洞关注度较多，甚至特别在安全类型中设置了对应的idn-spoof标签。

在chromium中维护了一个domain list，内置了一些较有知名度的域名，当有域名被认为和这些相似域名相似时，就会转成punycode显示。

其完整的检测算法可以在这里看到，对其详细的解释可以参考这篇文章。

总的来说，只要找到了一个字符的组合，可以通过Spoof Check，且在浏览器地址栏中显示的字形和top domain相似，就可以认为找到了一个IDN Spoof漏洞。

# 0x02 挖掘方法

Unicode的字符较多，因此笔者考虑一定程度上将漏洞挖掘的过程自动化。最直接的思路是，将域名是否同形的问题转换为图像相似度的问题。

我们可以遍历所有的Unicode字符，使用浏览器地址栏渲染的字体生成其对应的图像，当其图像和域名中允许出现的ascii字符相似度较高时，则认为是可能造成Spoof的字符

在这里笔者使用了感知哈希算法作为图像相似度的计算的方式，其大致步骤如下：

- 将图像缩小至相同尺寸：用于去除图像的细节，保留结构等基本信息
- 简化色彩：将缩小后的图像转为64级灰度，减少颜色带来的影响
- 计算平均值：计算所有像素的灰度平均值
- 比较像素的灰度：将每个像素的灰度，与平均值进行比较，记录结果
- 计算Hash值：将上一步的结果组合在一起构成一个整数作为图片的指纹

在获取到Unicode字符图片对应的Hash值后，使用计算汉明距离的方式计算两个图片的距离，就可以得到较为相似的字符列表了。

当找到符合条件的字符后，则找到包含该字符的域名，替换该字符进行测试，检测其是否能通过Spoof Check。这里直接使用Chrome测试不太方便，这里笔者抽取了其中部分代码形成独立的脚本进行测试。

另外对一些特别的字符，如 / / ? / . / # 等，则构造包含对应字符的URL进行测试。

通过图像相似度和Spoof Check的测试后，最后进行人工的确认，如确实是可能造成Spoof的字符，则认为是漏洞并报告。

# 0x03 挖掘结果

经测试，笔者成功找到了 `crbug.com/904325` 、 `crbug.com/904627` 等尚未修复的IDN Spoof漏洞。

# 0x04 参考资料


- https://en.wikipedia.org/wiki/Internationalized_domain_name
- https://www.unicode.org/faq/idn.html
- https://xlab.tencent.com/en/2018/11/13/cve-2018-4277/

- Gontmakher A . The Homograph Attack[J]. Communications of the Acm, 2002, 45(2):128.
- https://en.wikipedia.org/wiki/IDN_homograph_attack
- https://tw.saowen.com/a/72b7816b29ef30533882a07a4e1040f696b01e7888d60255ab89d37cf2f18f3e
- https://en.wikipedia.org/wiki/Perceptual_hashing





上一篇：macOS上一个模拟鼠标攻击的0day 下一篇：某CMFX_2.2.3漏洞合集


1. 0 条回复
   - 动动手指，沙发就是你的了！