misc学习笔记1-txt零宽度字符隐写

原创

Amherstieae 于 2020-10-03 13:45:54 发布 3701 收藏 24

分类专栏： misc笔记

本文链接：https://blog.csdn.net/Amherstieae/article/details/108909743

misc笔记 专栏收录该内容

2 篇文章 0 订阅

订阅专栏

写在前面：

这里是β-AS，准备开新坑了，本次笔记记录一下学习misc中遇到的隐写方式，如果有不对的地方希望各位路过的大佬手下留情（逃

第一次见到这种题人都傻了，到比赛结束也没见到下一步的题目内容233333

后来比赛结束后膜了盖乐希师傅的博客（膜

0x01什么是零宽度字符

零宽度字符是一种字节宽度为0的不可打印的Unicode字符, 在浏览器等环境不可见, 但是真是存在, 获取字符串长度时也会占位置, 表示某一种控制功能的字符。

```
零宽空格（zero-width space，ZWSP）用于可能需要换行处。
    Unicode: U+200B  HTML: &#8203;
零宽不连字（zero-width non-joiner，ZWNJ)放在电子文本的两个字符之间，抑制本来会发生的连字，而是以这两个字符原本的字形来绘制。
    Unicode: U+200C  HTML: &#8204;
零宽连字（zero-width joiner，ZWJ）是一个控制字符，放在某些需要复杂排版语言（如阿拉伯语、印地语）的两个字符之间，使得这两个本不会发生连字的字符产生了连字效果。
    Unicode: U+200D  HTML: &#8205;
左至右符号（Left-to-right mark，LRM）是一种控制字符，用于计算机的双向文稿排版中。
    Unicode: U+200E  HTML: &lrm; &#x200E; 或&#8206;
右至左符号（Right-to-left mark，RLM）是一种控制字符，用于计算机的双向文稿排版中。
    Unicode: U+200F  HTML: &rlm; &#x200F; 或&#8207;
字节顺序标记（byte-order mark，BOM）常被用来当做标示文件是以UTF-8、UTF-16或UTF-32编码的标记。
    Unicode: U+FEFF
```

比如在下面这个例子中

```
"大佬早上好呀".length
34
```

我们能看到的只有六个字，但是显示的length有34，零宽度字符就会产生这个效果，他不影响阅读，但是真实存在，也会占长度。

虽然我们看到的样子是上面这个样子，但是实际上他是下面这个样子~

Input

length: 34
lines: 1

大佬早上好呀

Output

start: 34 time: 0ms
end: 34 length: 34
length: 0 lines: 1

大.....佬早上好呀

在在在实际上，他是这个样子

Input

length: 34
lines: 1

大佬早上好呀

Output

start: 204 time: 0ms
end: 204 length: 204
length: 0 lines: 1

\u5927\u200C\u200E\u200B\u200B\u200F\u200C\u200B\u200C\u200D\u200B\u200B\u200C\u200E\u200C\u200C\u200E\u200E\u200F\u200D\u200B\u200C\u200C\u200C\u200D\u200D\u200D\u200F\u200B\u4F6C\u65E9\u4E0A\u597D\u5440

0x02零宽度字符有什么用处捏

1.数据防爬

将零宽度字符插入关键词文本中，使得匹配关键字时不能正确匹配，但是不影响用户的正常阅读

2.信息隐藏（类似水印

类似上面例子我们可以将信息隐藏在正常文字中而不影响阅读

0x03零宽度字符怎么进行隐写呢

隐写方法是可逆的，但是需要是相同的方法或者是网站呀~

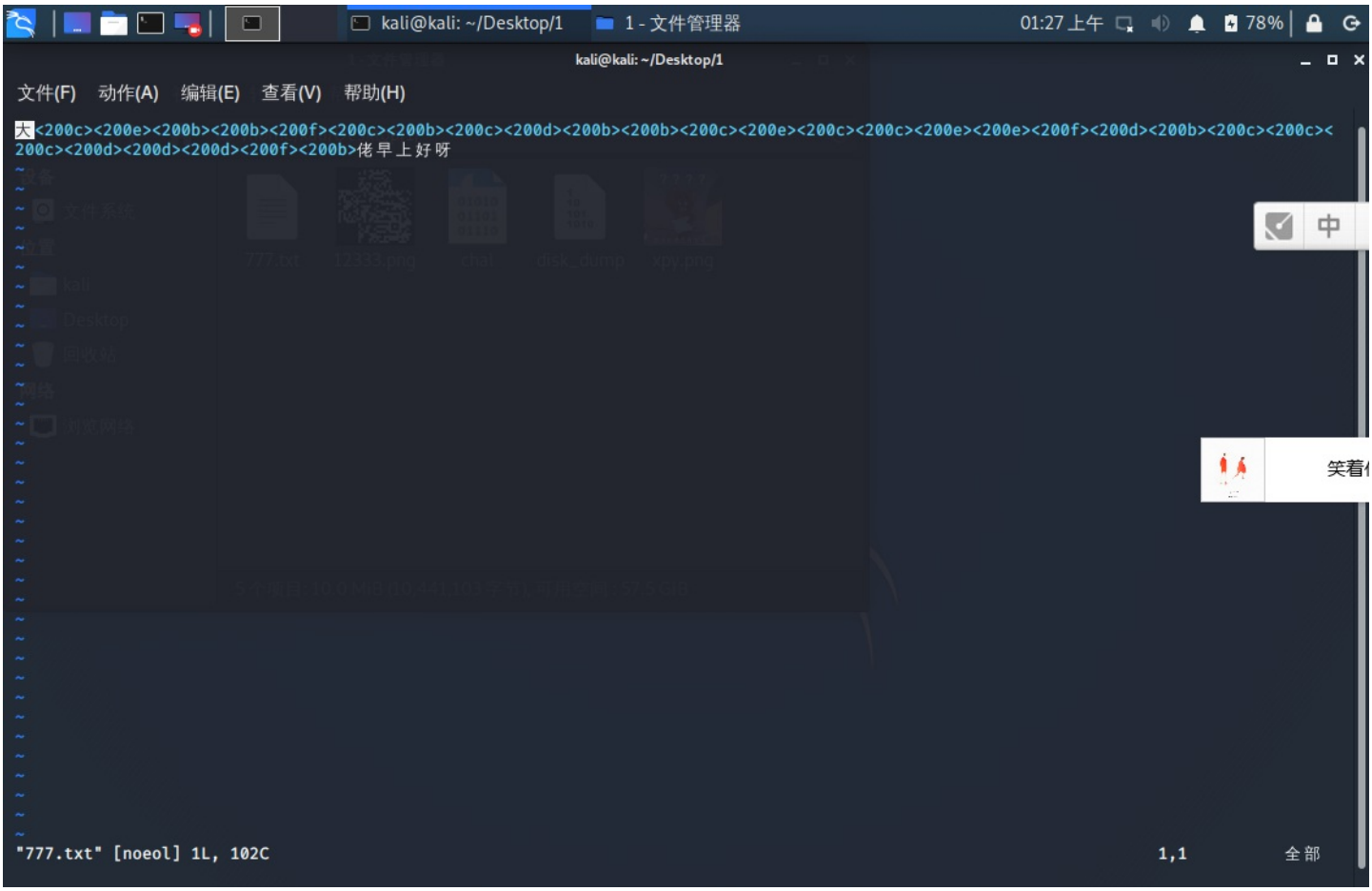
具体可以参考盖乐希师傅的博客2333（日常膜盖乐希

http://www.ga1axy.top/index.php/archives/20/

0x04怎么判断是零宽度字符呢

1.Kali中

vim file.txt



蓝色的就是零宽度字符啦

2.在光标移动中有明显的停顿感

（无图，脑补233333）

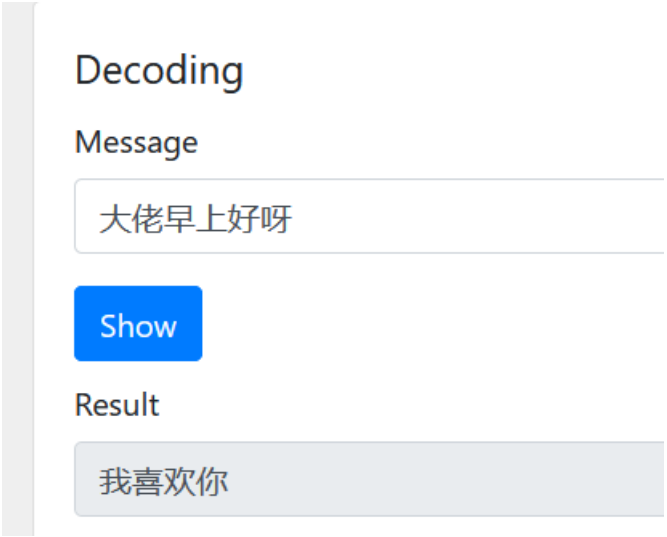
3.在010或者winhex打开

0x05解密方式

1.https://offdev.net/demos/zwsp-steg-js

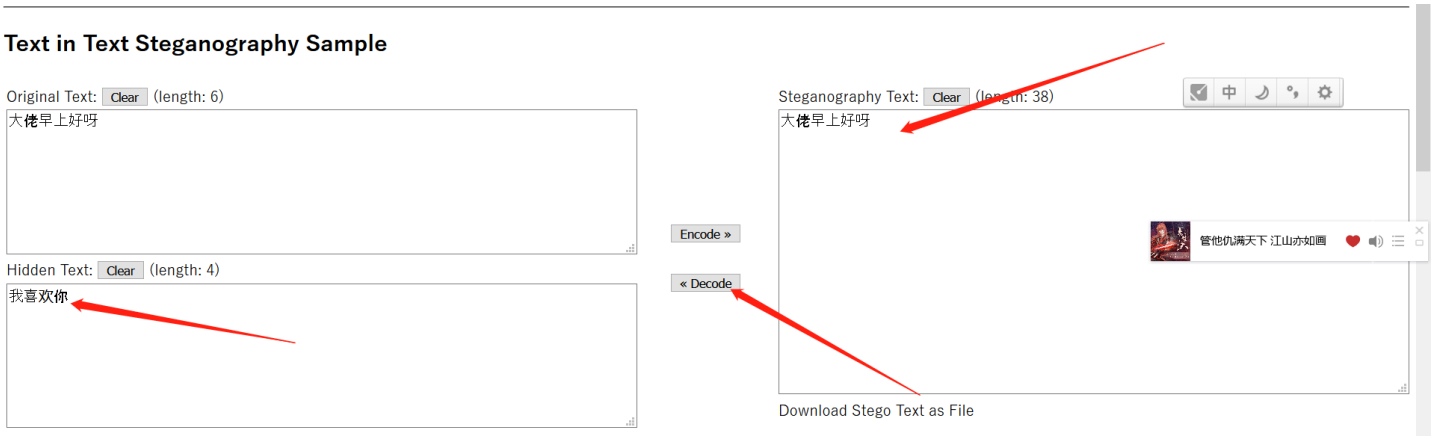
将文本全部复制到解码框中，点击show，下面就是结果啦

如下



2.http://330k.github.io/misc_tools/unicode_steganography.html

将文本复制到右边文本框中，点decode即可



注意此解码网站还可根据零宽度字符的不同选择

但是此网站没有U+200F，含有U+200F可以用上面的网站

Zero Width Characters for Steganography:

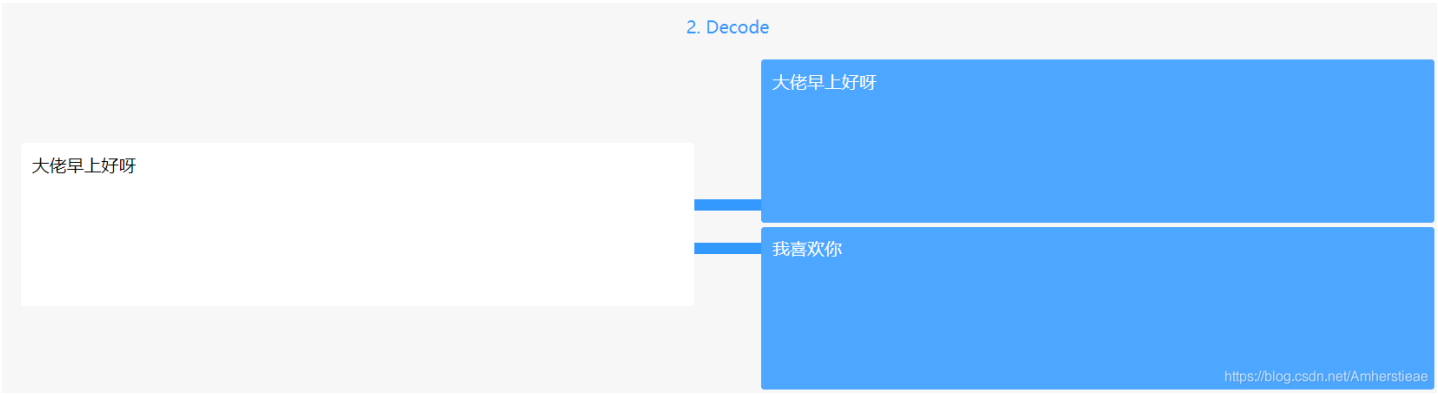
- ☐ U+200B ZERO WIDTH SPACE
- ☑ U+200C ZERO WIDTH NON-JOINER
- ☑ U+200D ZERO WIDTH JOINER
- ☐ U+200E LEFT-TO-RIGHT MARK
- ☐ U+202A LEFT-TO-RIGHT EMBEDDING
- ☑ U+202C POP DIRECTIONAL FORMATTING
- ☐ U+202D LEFT-TO-RIGHT OVERRIDE
- ☐ U+2062 INVISIBLE TIMES
- ☐ U+2063 INVISIBLE SEPARATOR
- ☑ U+FEFF ZERO WIDTH NO-BREAK SPACE

3.将零宽度字符替换成01或者莫斯解码

附两篇文章

- 转化为二进制的加密：https://zhuanlan.zhihu.com/p/87919817
- 转化为Morse编码的加密：https://zhuanlan.zhihu.com/p/75992161

4.https://yuanfux.github.io/zero-width-web/



over完结撒花✿✿ヽ(°▽°)ノ✿