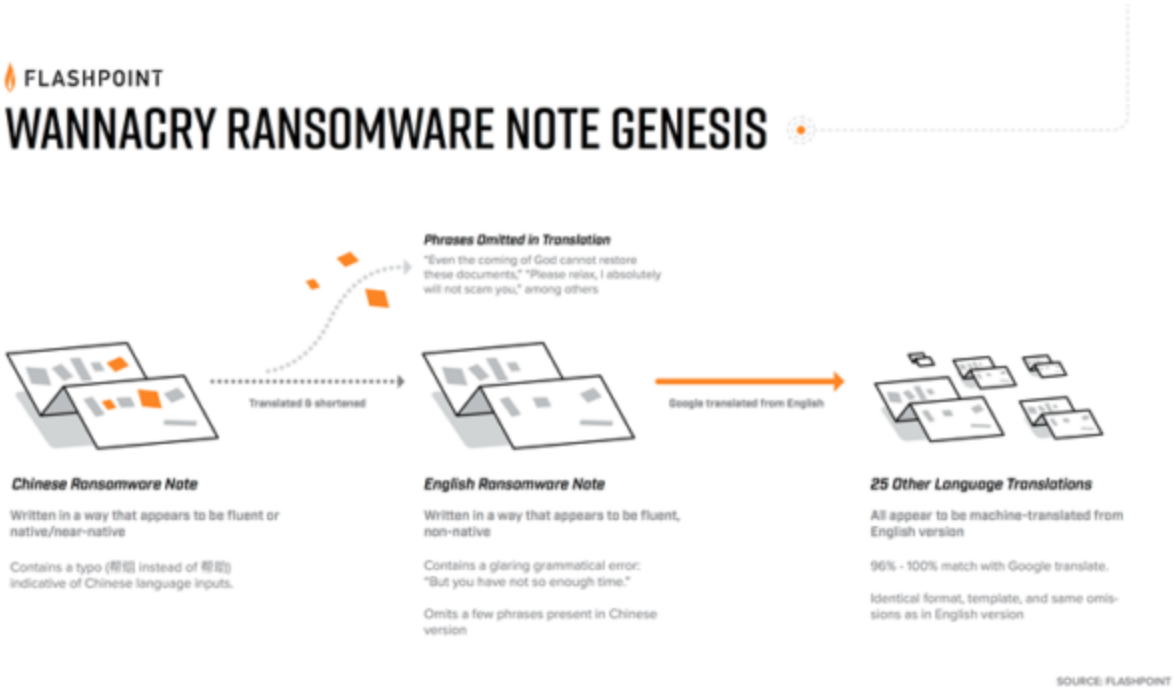
Linguistic Analysis of WannaCry Ransomware Suggests Chinese-Speaking Authors

flashpoint-intel.com/blog/linguistic-analysis-wannacry-ransomware/

May 25, 2017



Blogs

Blog

Linguistic Analysis of WannaCry Ransomware Messages Suggests Chinese-Speaking Authors

Since the May 12, 2017, “WannaCry” ransomware worm attack, researchers have struggled with the question of attribution. As of this writing, a number of researchers have linked the activity to the suspected North Korean-affiliated “Lazarus Group” due to similarities in the code and the infrastructure. Flashpoint analysts conducted similar analyses, but also included a linguistic and cultural review of the 28 ransom notes found within the WannaCry malware to determine the native tongue of the author(s).

Since the May 12, 2017, “WannaCry” ransomware worm attack, researchers have struggled with the question of attribution. As of this writing, a number of researchers have linked the activity to the suspected North Korean-affiliated “Lazarus Group” due to similarities in the code and the infrastructure. Flashpoint analysts conducted similar analyses, but also included a linguistic and cultural review of the 28 ransom notes found within the WannaCry malware to determine the native tongue of the author(s).

Analysis

Flashpoint analyzed each of the notes individually for content, accuracy, and style, and then compared results. Analysts also compared the ransom notes to previous ransom messages associated with other ransomware samples to determine if there was reuse. Unsurprisingly, there are many similarities, but an exact match was not found. The WannaCry samples analyzed by Flashpoint contained language configuration files with translated ransom messages for the following languages:

1. Bulgarian
2. Chinese (Simplified)
3. Chinese (Traditional)
4. Croatian
5. Czech
6. Danish
7. Dutch
8. English
9. Filipino
10. Finnish
11. French
12. German
13. Greek
14. Indonesian
15. Italian
16. Japanese
17. Korean
18. Latvian
19. Norwegian
20. Polish
21. Portuguese
22. Romanian
23. Russian
24. Slovak
25. Spanish
26. Swedish
27. Turkish
28. Vietnamese

 Image 1: WannaCry ransom note in English.***Image 1****: WannaCry ransom note in English.*
Analysis revealed that nearly all of the ransom notes were translated using Google Translate and that only three, the English version and the Chinese versions (Simplified and Traditional), are likely to have been written by a human instead of machine translated.

Though the English note appears to be written by someone with a strong command of English, a glaring grammatical error in the note suggest the speaker is non-native or perhaps poorly educated.

Flashpoint found that the English note was used as the source text for machine translation into the other languages. Comparisons between the Google translated versions of the English ransomware note to the corresponding WannaCry ransom note yielded nearly identical results, producing a 96% or above match.

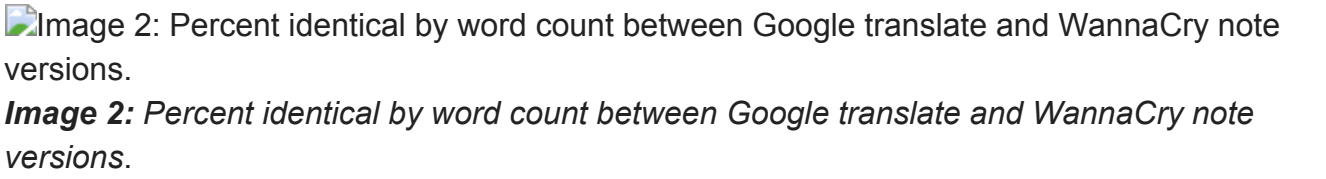


Image 2: Percent identical by word count between Google translate and WannaCry note versions.
Image 2: *Percent identical by word count between Google translate and WannaCry note versions*.

Chinese Ransomware Notes

The two Chinese ransom notes differ substantially from other notes in content, format, and tone. Google Translate fails in both Chinese-English and English-Chinese tests, producing inaccurate results that suggests the Chinese text was likely not have been similarly generated by the English text.

A number of unique characteristics in the note indicate it was written by a fluent Chinese speaker. A typo in the note, “帮组” (bang zu) instead of “帮助” (bang zhu) meaning “help,” strongly indicates the note was written using a Chinese-language input system rather than being translated from a different version. More generally, the note makes use of proper grammar, punctuation, syntax, and character choice, indicating the writer was likely native or at least fluent. There is, however, at least one minor grammatical error which may be explained by autocomplete, or a copy-editing error.

The text uses certain terms that further narrow down a geographic location. One term, “礼拜” for “week,” is more common in South China, Hong Kong, Taiwan, and Singapore; although it is occasionally used in other regions of the country. The other “杀毒软件” for “anti-virus” is more common in the Chinese mainland.

Perhaps most compelling, the Chinese note contains substantial content not present in any other version of the note, is lengthier, and differs slightly in format.

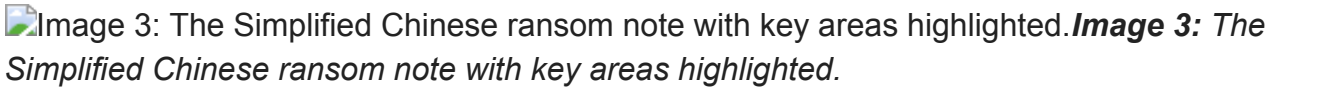


Image 3: The Simplified Chinese ransom note with key areas highlighted.***Image 3:*** *The Simplified Chinese ransom note with key areas highlighted.*

Conclusions

Flashpoint assesses with high confidence that the author(s) of WannaCry's ransomware notes are fluent in Chinese, as the language used is consistent with that of Southern China, Hong Kong, Taiwan, or Singapore. Flashpoint also assesses with high confidence that the author(s) are familiar with the English language, though not native. This alone is not enough to determine the nationality of the author(s).

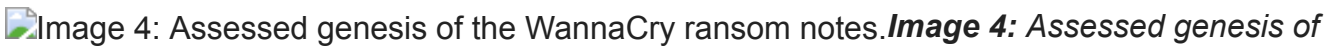


Image 4: Assessed genesis of the WannaCry ransom notes.***Image 4:*** *Assessed genesis of the WannaCry ransom notes.*

Flashpoint assesses with moderate confidence that the Chinese ransom note served as the original source for the English version, which then generated machine translated versions of the other notes. The Chinese version contains content not in any of the others, though no other notes contain content not in the Chinese. The relative familiarity found in the Chinese text compared to the others suggests the authors were fluent in the language—perhaps comfortable enough to use the language to write the initial note.

Given these facts, it is possible that Chinese is the author(s)' native tongue, though other languages cannot be ruled out. It is also possible that the malware author(s)' intentionally used a machine translation of their native tongue to mask their identity. It is worth noting that characteristics marking the Chinese note as authentic are subtle. It is thus possible, though unlikely, that they were intentionally included to mislead.