Hiding Shellcodes in Korean Texts*

Ji-Hyeon Yoon, *Seoul Women's University* Hae Young Lee†

Abstract

This poster presents two approaches to hiding shellcodes in Korean texts encoded in UTF-16 [1], in which Chinese characters transformed from a shellcode through simple operations are embedded in a Korean text. Another *possible* approach to pseudo-Korean shellcodes (i.e., a Korean version of [2]) is also discussed. Although these approaches address shellcodes encoded in Korean texts, issues posed by them might not be confined to Korea; Korean texts might be considered to be a part of multilingual support. Moreover, they might be, with some modifications, applied to East Asian texts.

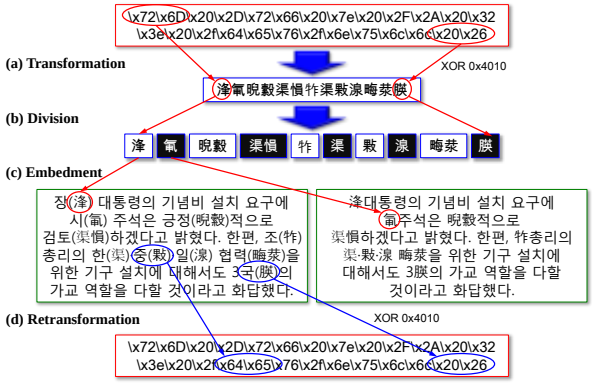

Fig. 1. Hiding shellcodes in Korean texts.

The first two approaches are based on that, in South Korea, Chinese characters are often used to clarify meaning of Sino-Korean words, which mostly originate from Chinese words. Each 2-byte code of a shellcode is first transformed into a Chinese character in UTF-16, through simple operations, such as a bitwise XOR. These characters are then divided into pseudo-Chinese words, with the consideration of the operations in the transformation, and/or of Sino-Korean words in a given Korean text. In Fig. 1, (a) 13 Chinese characters are transformed from a shellcode, and then divided into (b) 10 pseudo- words based on the operations.

Approach 1: In this approach, as shown in the left of Fig. 1(c), each pseudo-Chinese word is placed in the parenthesis followed by a Sino-Korean word within a text. To reconstruct the shellcode from the text can be easily achieved; a simple decoder retransforms each word through the operation hinted by the *very last* Korean character in front of the word, as shown in the left of Fig. 1(d). Thanks to the simplicity of the decoder, it might be also encoded in Korean or Chinese characters.

In order to make the text difficult to be distinguished from 'real' ones, we can insert 'real' Chinese words into it; each of these 'real' words could be ignored by a hint taken from the very last Korean character. The approaches could be simple yet effective against inline payload-based inspection; many Korean people in their 20s and 30s could be not much familiar with Chinese characters, so that they may miss mismatches between Sino-Korean and Chinese words.

Approach 2: In this approach, as shown in the right of Fig. 1(c), Sino-Korean words are replaced with pseudo-Chinese ones. Use of Chinese characters in Korean texts, while old-style, still appears in newspapers, academic literature, judicial texts, and so on. The shellcode can be also reconstructed with ease; each of the words is retransformed through the operation hinted by the *very next* Korean character, as shown in the right of Fig. 1(c). The decoder, due to its simplicity, might be also encoded in Korean or Chinese characters. Other Sino-Korean words can be also replaced with 'real' Chinese ones for more difficulty in the distinction. The approach might be applied to Japanese texts since they also use Chinese characters mixed with Japanese ones.

Approach 3: Another possible approach is, similar to [1], to transform a shellcode into a pseudo-Korean text. A Korean text may include Korean, Chinese, alphanumeric characters and symbols, which make up a large portion of Unicode characters. Thus, we expect that most instructions, including those of decoders, could be encoded within these characters. The approach might be applied to Chinese texts.

Our future work includes: 1) automation of the first two approaches, along with encoding of decoders in Korean or Chinese words, 2) concretization of the last approach, and 3) detection of Korean shellcodes.

References

[1] ISO/IEC, *Information Technology – Universal Coded Character Set (UCS)*, ISO/IEC 10646:2012, Jun. 2012.

[2] J. Mason, S. Small, F. Monrose, G. Mac Manus, "English Shellcode," *Proc. of CCS '09*, Nov. 2009.

* This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2013R1A1A1006542).

† haelee@swu.ac.kr