"Cocaine Noodles: Exploiting the Gap between Human and Machine Speech Recognition"

1) Target system and service (contributed by 박다원, 안형철)

- Speaker-independent speech recognition systems which are implemented on smartphones, tablets, and wearable devices
- Especially, based on Google Recognition engine API (Google Now application)
- A human operator is not using his device at the time of the attack

2) Vulnerability (contributed by 이은규, 박상욱)

- The vast majority of current generation smart devices do not attempt to authenticate voice commands.
- Machines can recognize voice commands that are not recognizable as human voice by humans because of the differences in how humans and machines understand spoken speech.

3) Exploitation (contributed by 오정석)

- The attacker plays sound that is interpreted as a noise to humans, but as a command to the target devices. Thus, the attacker can maliciously control the devices without the owner's permissions.
- By playing voice commands that are unrecognizable to humans, the attacker may achieve: drive-by-download, making payments based on SMS, enumerating devices, making expensive phone calls, or performing denial-of-service attack.

4) Evaluation and experimental method (contributed by 박다원, 최윤선, 이은규)

- The authors mainly evaluated how humans can accurately recognize mangled voice commands.
- For the evaluation, four types of commands including making a phone call and accessing two websites (both mangled and normal) are tested against one Samsung Galaxy S4 smartphone with Android 4.4.2.
- The sound to attack is generated using a pair of speakers placed about 30 cm from phone, and the experiments are carried out in a quiet room which has 50 dB of background noise.
- For audio mangling, an audio mangler is implemented using MATLAB, which can create mangled commands by adjusting the MFCC parameters of the original voice commands.
- Using a service from Amazon Mechanical Turk, the result of this evaluation shows that it is hard to understand the mangled commands for humans.

5) Defense (contributed by 안형철, 강희도)

- Disabling Google Now application
- Generating audible feedbacks (such as messages) for voice commands
- Adding a human authentication mechanism based on biometric information such as fingerprints

6) Future work (contributed by 이은규, 박철준, 오정석)

- This attack can be extended to other speech recognition systems.
- Using high-frequency sound that cannot heard by humans, but is recognizable by the target speech recognition systems (elder people vs. younger people)
- The mangled voice commands can be used against a lot of victims at once in environments with loud sound noise like club.

7) Questions to presenter (contributed by TA)

- Targeted Google Voice Assistant is a deep learning voice recognition model that continuously learns using the user's voice data. Will this attack still work for Google Voice Assistant in 2021, which is much more sophisticated and complex than it was in 2015 when the paper was published?
- The advantage of the sound-mangling attack method is that the victim does not know the exact meaning of the voice attack. However, as in the case of the demo video, if the sound is too mangled, it is enough to make the average person feel strange even if they cannot clearly understand the meaning. Is it possible to keep the victim from noticing this mangled voice attack attempt?
- The authors' attack method is to mangle the sound by touching the MFCC parameter until the desired result is obtained. However, without a sufficient understanding of the target model, simply crushing the sound makes the success rate and cost of the attack inefficient. How can attackers overcome this?