MITIGATING SECURITY RISKS ON VIRTUAL PERSONAL ASSISTANT SYSTEMS

TEAM MEMBERS

Mareena Fernandes	8669
Fatima Pereira	8690
Hanita Rego	8693
Ashally Tuscano	8710

> “*Arguing that you don’t care about the right to privacy because you have nothing to hide is no different than saying you don’t care about free speech because you have nothing to say.*
>
> *- Edward Snowden*

VIRTUAL PERSONAL ASSISTANT

- Voice assistant device are getting very popular especially echo devices and google home devices.
- These devices allow the user to command the system with voice only

PayPal

Alexa, ask PayPal to send 1000Rs to Mohit

Alexa, ask medical assistant to give my reports

Alexa, play Today's hits on prime

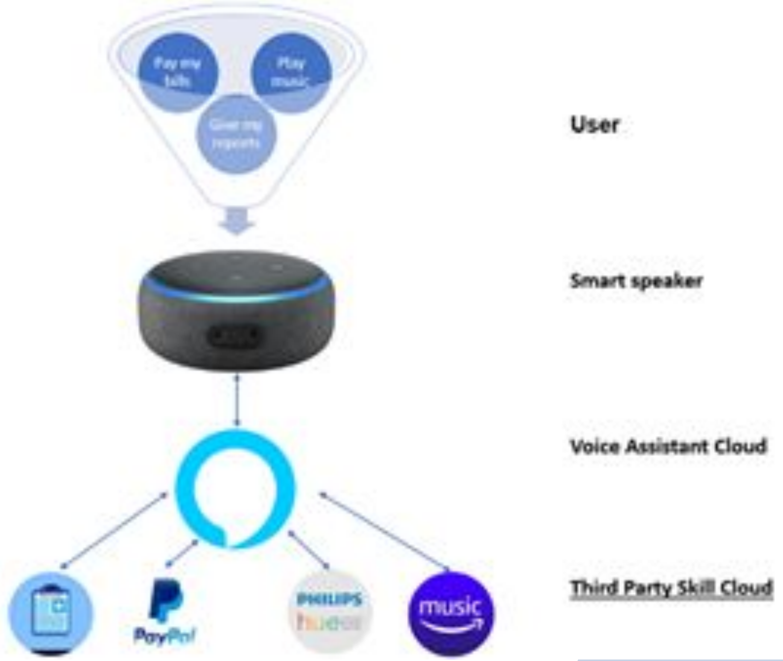
music

Alexa, turn on living room lights

PHILIPS hue

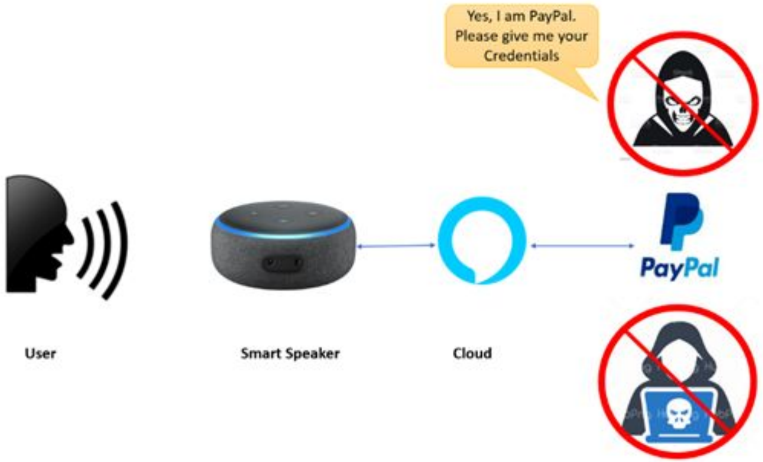
HOW IT WORKS?

- Voice assistants act like a relay, proxying and translating conversations between user and skills

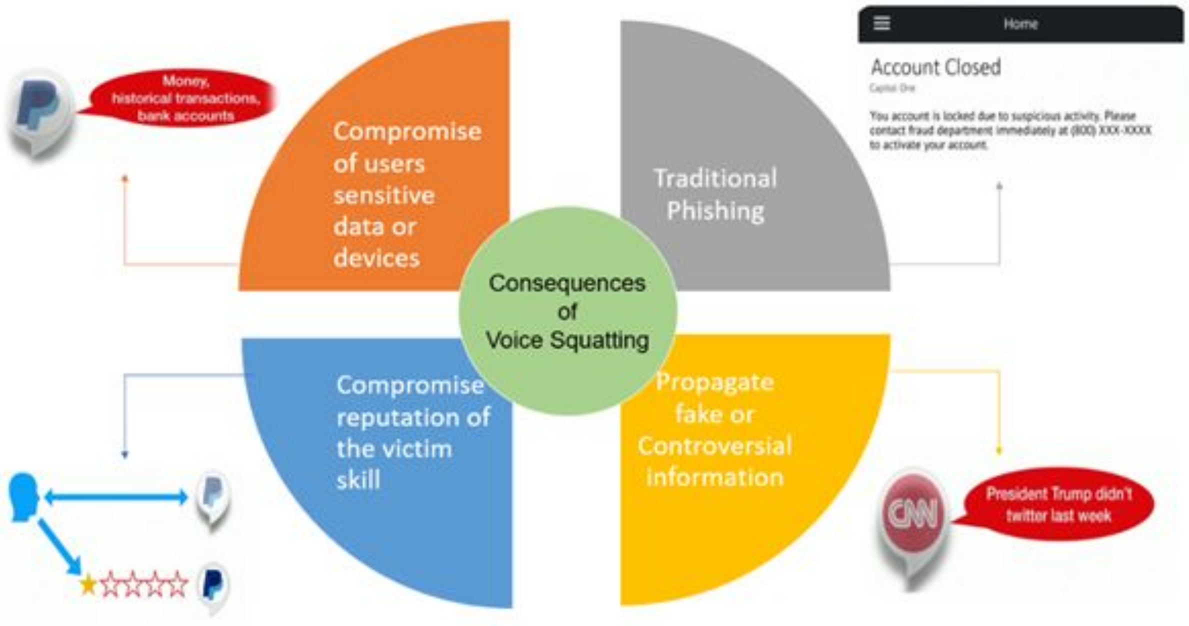


WHAT IS VOICE SQUATTING?

- To voice-squat, an adversary can create a new, malicious skill that is specifically built to open when the user says certain phrases. Those phrases are designed to be similar, if not nearly identical, to phrases used to open legitimate apps.



CONSEQUENCES OF VOICE SQUATTING



Consequences of Voice Squatting

Compromise of users sensitive data or devices

Traditional Phishing

Compromise reputation of the victim skill

Propagate fake or Controversial information

Home

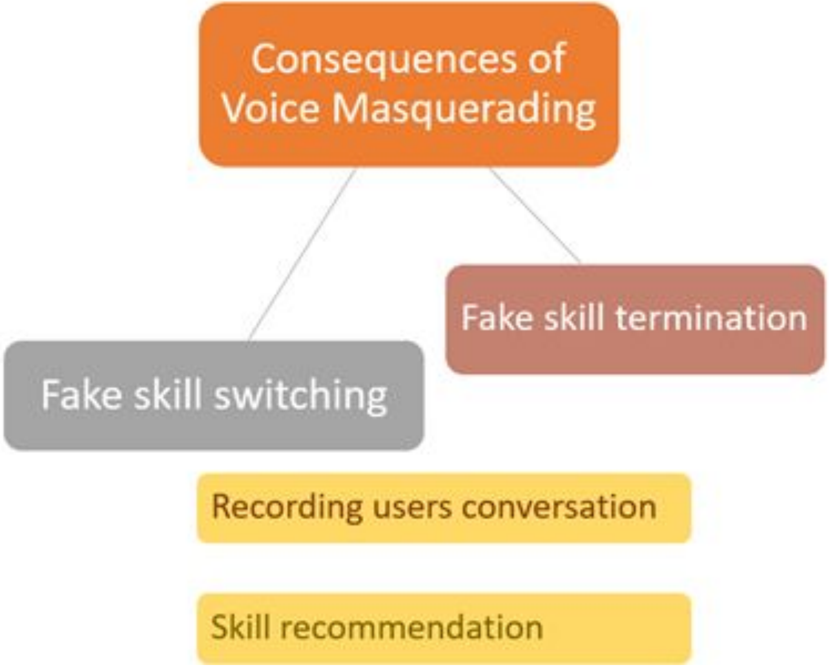
Account Closed

Capital One

You account is locked due to suspicious activity. Please contact fraud department immediately at (800) XXX-XXXX to activate your account.

Money, historical transactions, bank accounts
President Trump didn't twitter last week
CNN

CONSEQUENCES OF VOICE MASQUERADING



Consequences of Voice Masquerading

Fake skill switching

Fake skill termination

Recording users conversation

Skill recommendation

HOW REALISTIC ARE THESE ATTACKS?

Steps taken during the research

STUDY HOW USERS INVOKE SKILLS

Command passed by user	Amazon	Google
Yes, “open Sleep Sounds please”	**64%**	**55%**
Yes, “open Sleep Sounds for me”	**30%**	**25%**
Yes, “open the Sleep Sounds”	**20%**	**14%**

- Users tend to use diverse and natural language utterances
- Longest prefix matching creates attack space for voice squatting

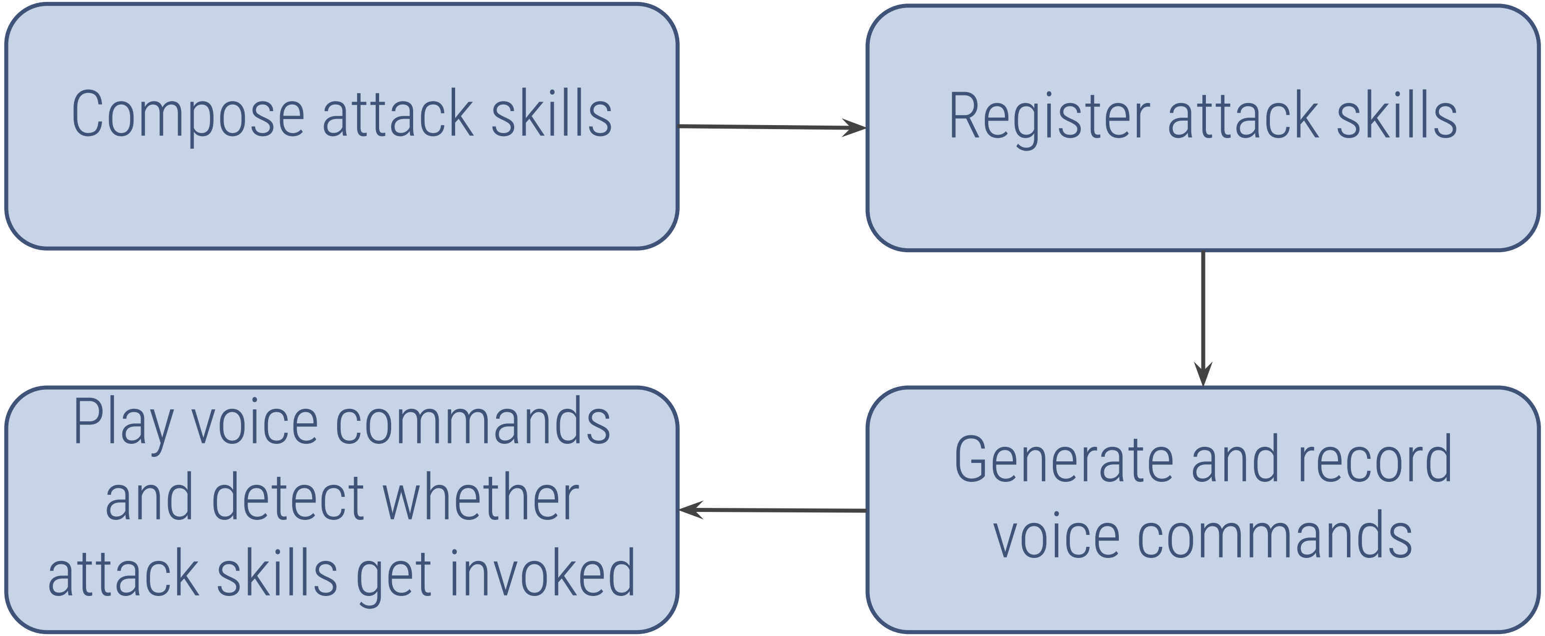
STUDY HOW WELL THE PLATFORM CAN UNDERSTAND VOICE COMMANDS

Platform	TTS Services	Human Subjects
Alexa	**30%**	**57%**
Google	**9%**	**10%**

- The voice assistant platforms are error-prone when recognizing voice commands.

  Example: Florida state quiz ⟶ Florida snake quiz

EXPERIMENT PROOF OF CONCEPT ATTACK SKILLS



Compose attack skills

Register attack skills

Generate and record voice commands

Play voice commands and detect whether attack skills get invoked

TWO METHODS

- Voice squatting through invocation name extending

  Example: Capital one → "My capital one" OR "Capital one please"

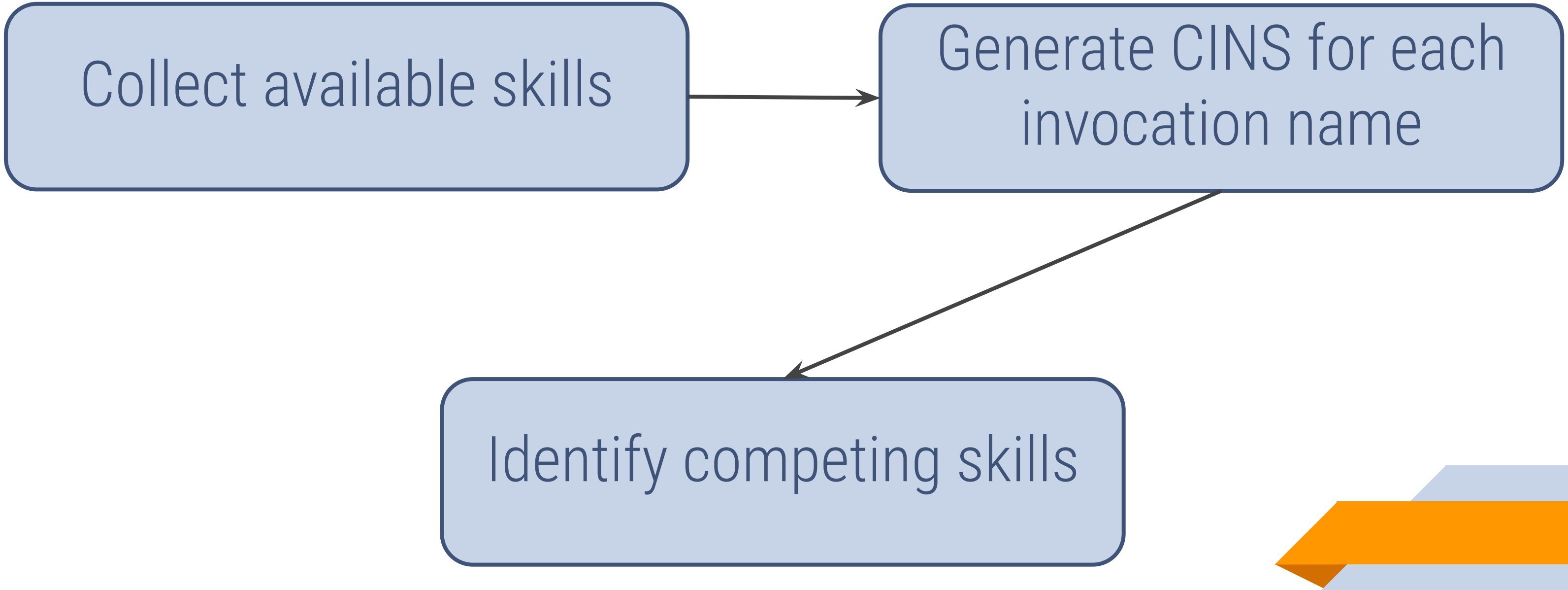
- Voice squatting through similar pronunciation

  Example: Capital one → "Capital **Won**" OR "**Captain** one" OR "**Capitol** one"

- Attack skills were not published to the skills market

IDENTIFY REAL WORLD ATTACKS

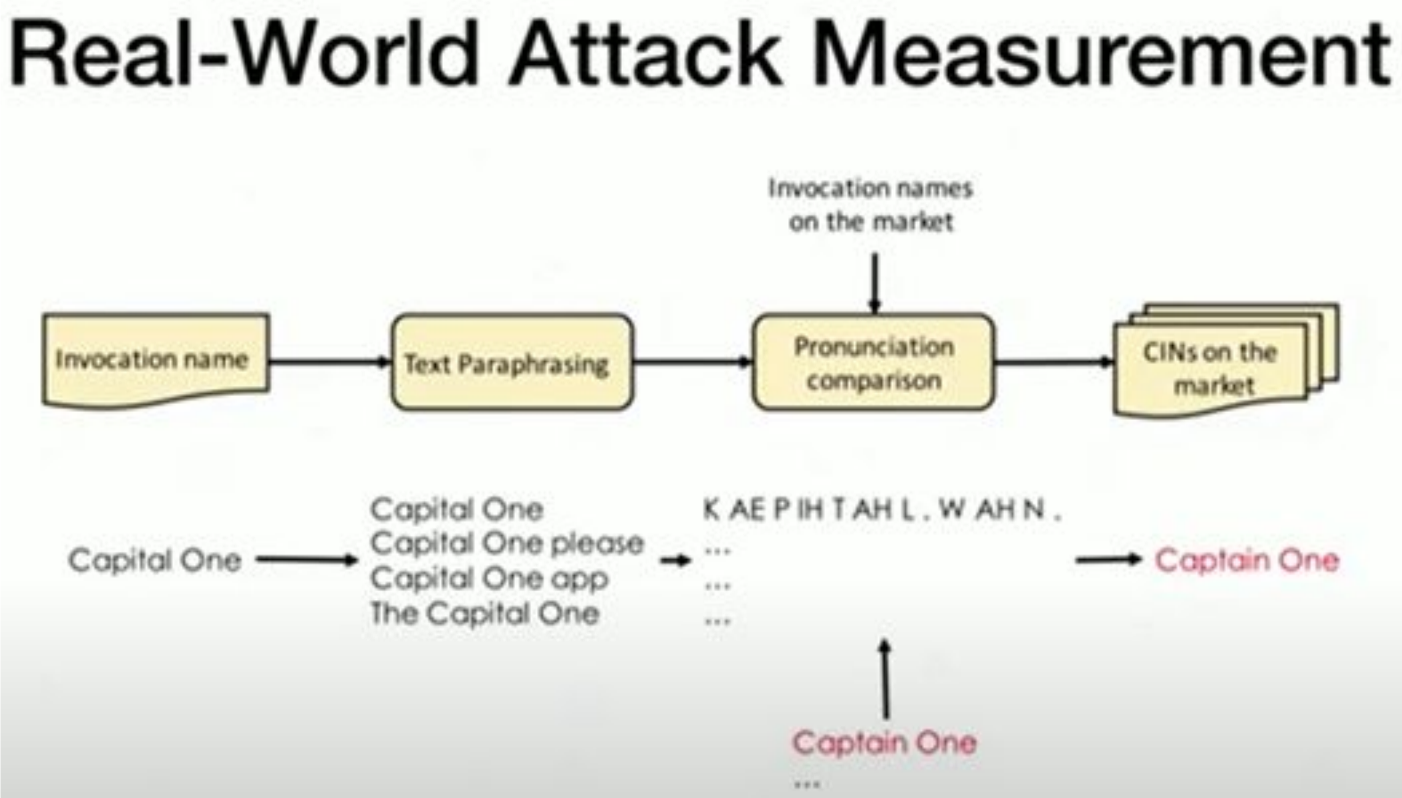
Finding voice squatting skills: Identify skills with competitive invocation names(CINS)



Collect available skills

Generate CINS for each invocation name

Identify competing skills



Real-World Attack Measurement

Invocation names on the market

Invocation name → Text Paraphrasing → Pronunciation comparison → CINs on the market

Capital One → Capital One
Capital One please
Capital One app
The Capital One → K AE P IH T AH L . W AH N .
...
...
... → Captain One

Captain One
...

Fig: Generation of CINS process

OVERALL ANALYSIS

- 19% (3718 skills) have same pronunciation
- 2.7% (531 skills) same pronunciation but different spellings
- 1.8% (345 skills): longest prefix matching

CASE STUDY

Interesting Case

EXAMPLE

- Skill - “dog fact” was invoked by “me a dog fact”
- Skills - “Scuba Diving Trivia” and “Soccer geek” registered “space geek” as invocation name even though clearly, they have nothing in common.

DEFENCE

Skill Response Checker

SRC

User Intention Classifier

UIC

SKILL RESPONSE CHECKER

- To defend against such attacks, our core idea is to control the avenues that a malicious skill can take to simulate either the VPA system or a different skill, allowing the user to be explicitly notified of VPA system events when a security risk is observed.
- For this purpose, SRC maintains a set of common utterance templates exclusively used by the VPA system to capture the similar utterances generated by a running skill
- A challenge here is how to reliably measure whether a given response is similar enough to one of those templates, as the attacker could morph the target system utterance.

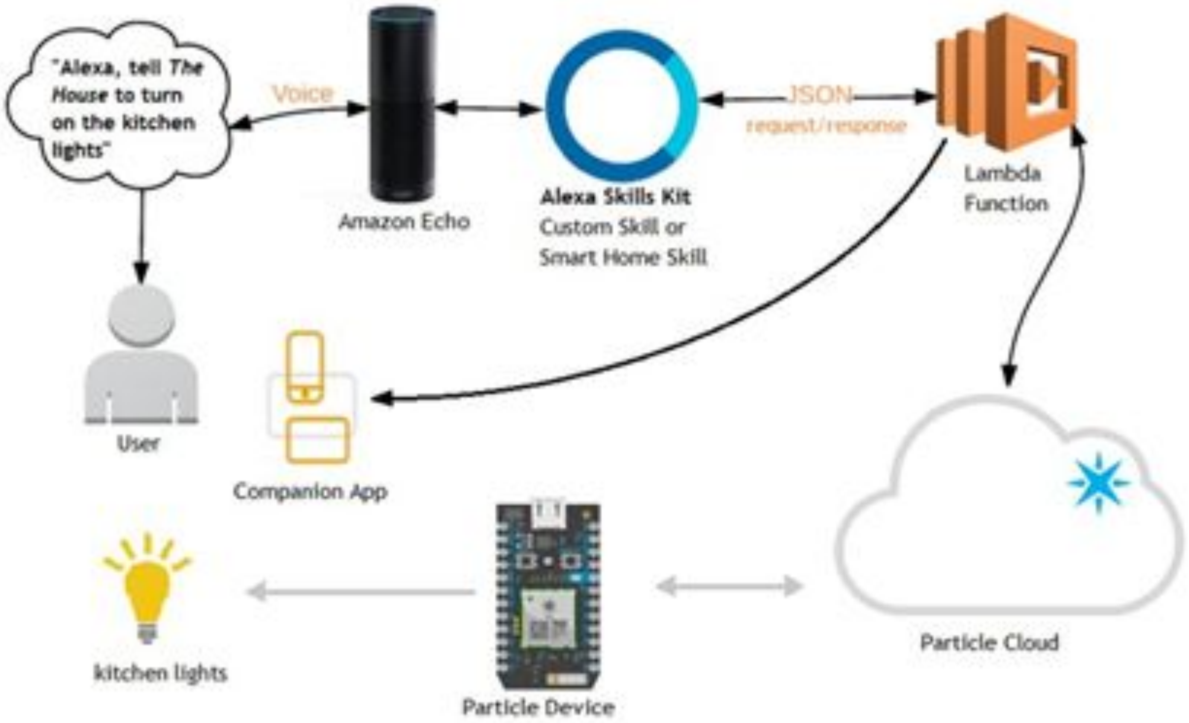
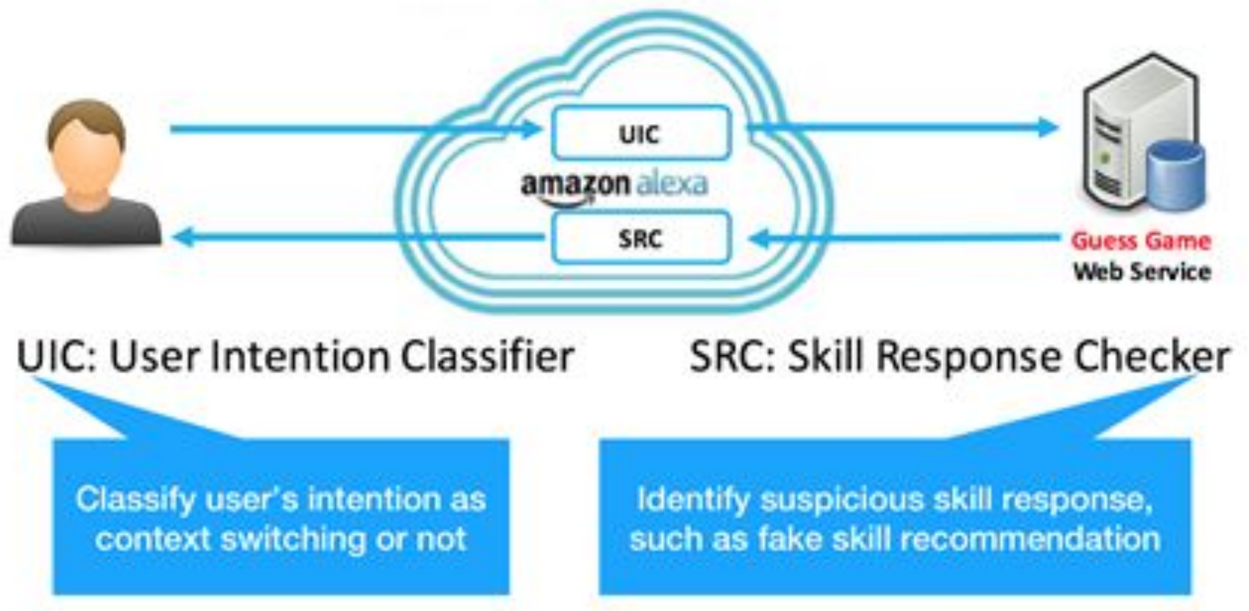


Fig: Skill Response Checker (SRC)

USER INTENTION CLASSIFIER

- UIC further protects the user attempting For this purpose, it aims at automatically detecting such erroneous commands from the user, based upon the semantics of the commands and their context in the conversation with the running skill.
- If such attempts can be perfectly identified by the VPA, it can take various actions to protect the user, e.g., reminding her that she is talking to the skill, not the VPA, or following the instructions to terminate the skill, which closes the surface for the impersonation attack.
- The challenges come from not only the variations in natural-language commands (Example:, “open sleep sounds” vs. “sleep sounds please”)

Defense



UIC: User Intention Classifier

Classify user's intention as context switching or not

SRC: Skill Response Checker

Identify suspicious skill response, such as fake skill recommendation

Fig: User Intention Classifier (UIC)

FEATURES

- The SR between the utterance and the skill’s response prior to the utterance.
- the top-k SRs between the utterance and the sentences in the skill’s description (we pick k=5)
- The average SR between the user’s utterance and the description sentences.

Fig: Features

CONCLUSION

- Analysis of popular VPA ecosystems and their vulnerability.
- The attacks are found to pose a realistic threat to VPA IoT systems
- To mitigate the threat, a skill-name scanner and ran it against Amazon and Google skill markets.

REFERENCE

Dangerous Skills: Understanding and Mitigating Security Risks of Voice-Controlled Third-Party Functions on Virtual Personal Assistant Systems:

https://www.computer.org/csdl/proceedings-article/sp/2019/666000a263/19skfw8oHZe

Written by: Nan Zhang, Xianghang Mi, Xuan Feng, XiaoFeng Wang, Yuan Tian and Feng Qian

THANK YOU!