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

Nan Zhang, **Xianghang Mi**, Xuan Feng, XiaoFeng Wang, Yuan Tian, Feng Qian

INDIANA UNIVERSITY

UNIVERSITY of VIRGINIA

Voice Assistant Devices

Alexa, play Today's Hits on Pandora

Alexa, turn on Living Room lights

PHILIPS hue

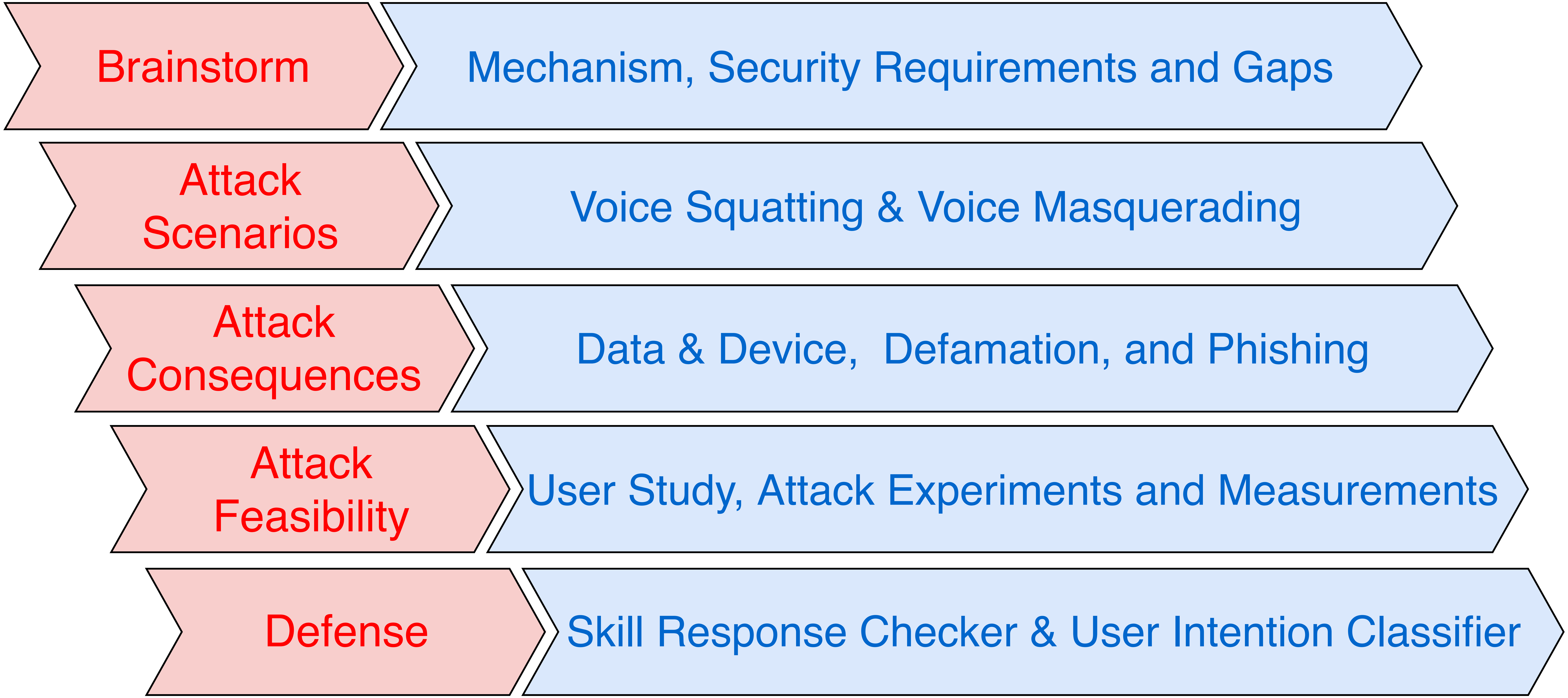
Alexa, ask PayPal to send 10 dollars to Sam

Alexa, ask Medical Assistant to give me my diagnosis

Smart Enough to be Secure?

Not Yet

Outline



Brainstorm — Mechanism, Security Requirements and Gaps

Attack Scenarios — Voice Squatting & Voice Masquerading

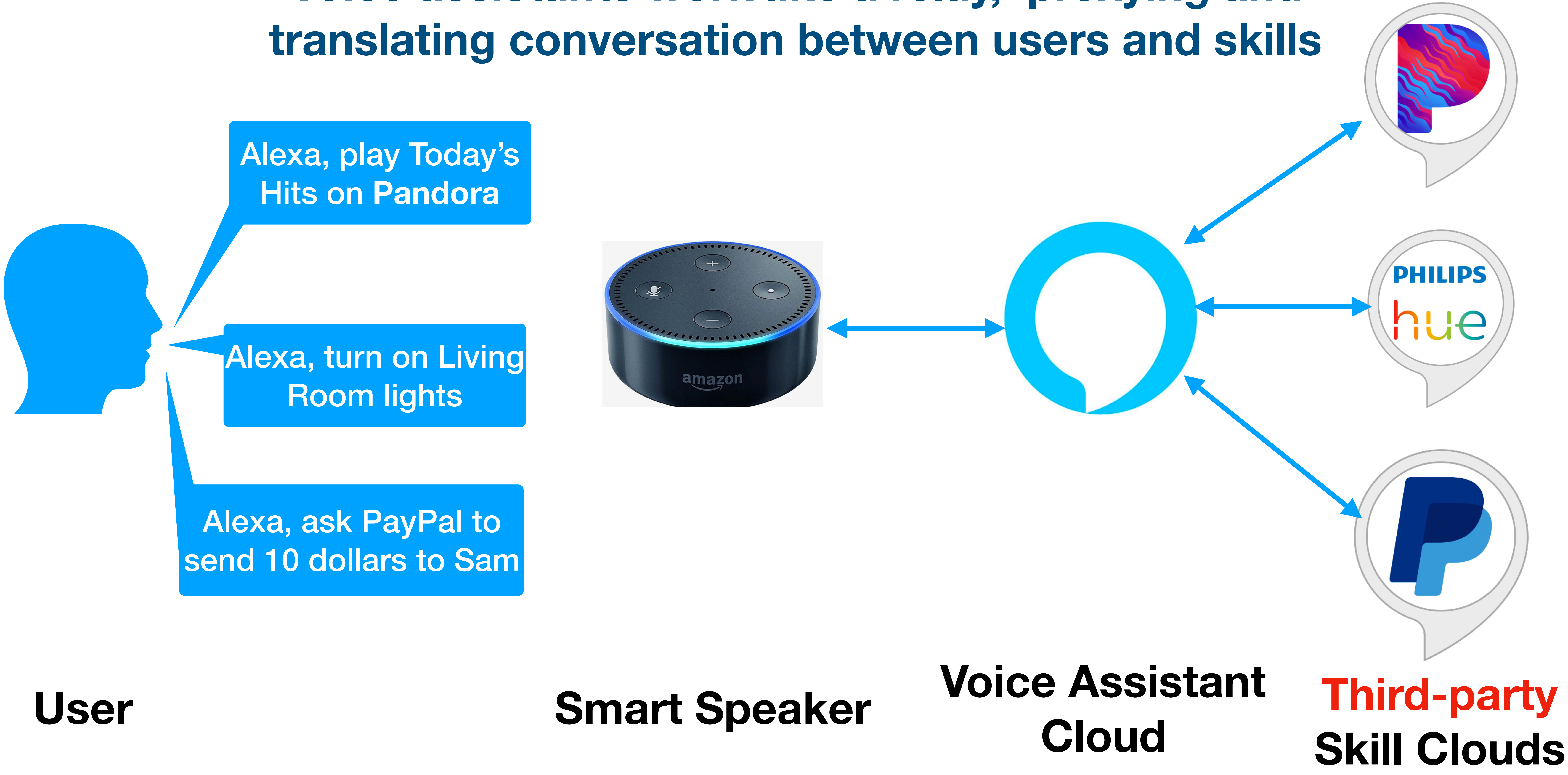
Attack Consequences — Data & Device, Defamation, and Phishing

Attack Feasibility — User Study, Attack Experiments and Measurements

Defense — Skill Response Checker & User Intention Classifier

How it works?

Voice assistants work like a relay, proxying and translating conversation between users and skills



Alexa, play Today's Hits on Pandora

Alexa, turn on Living Room lights

Alexa, ask PayPal to send 10 dollars to Sam

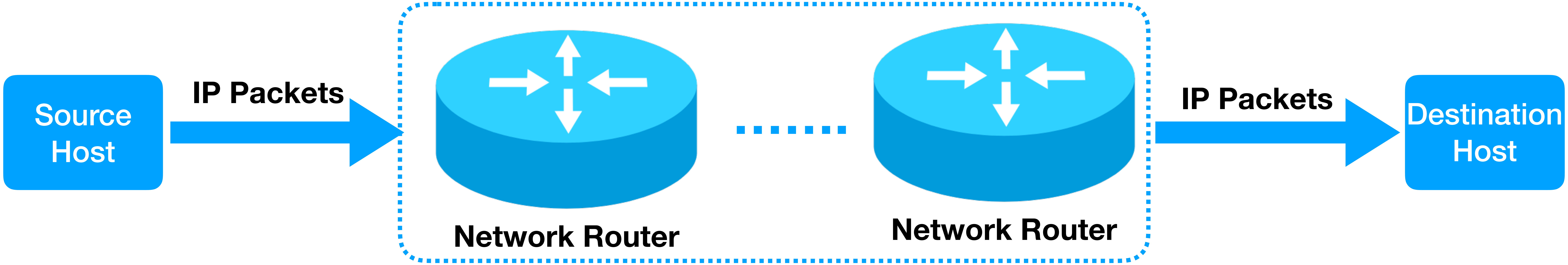
User

Smart Speaker

Voice Assistant Cloud

Third-party Skill Clouds

Security requirements and gaps



Source Host

IP Packets

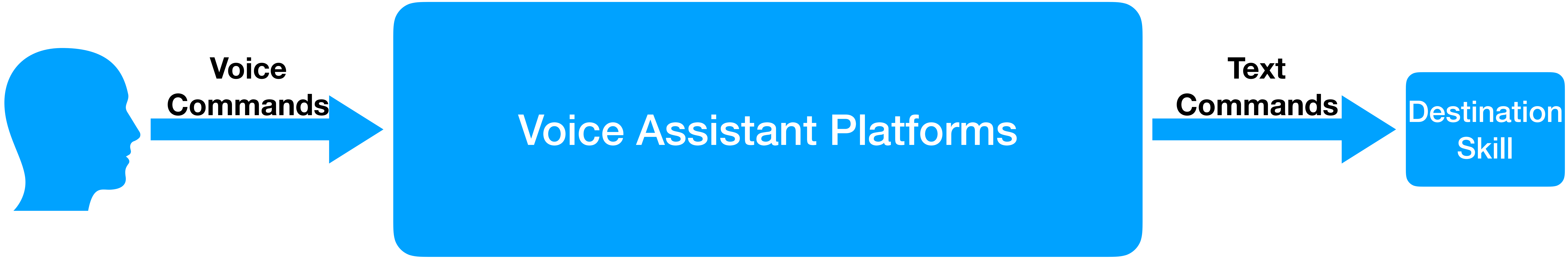
Network Router

Network Router

IP Packets

Destination Host

★ **Route the source payload to the CORRECT destination**



Voice Commands

Voice Assistant Platforms

Text Commands

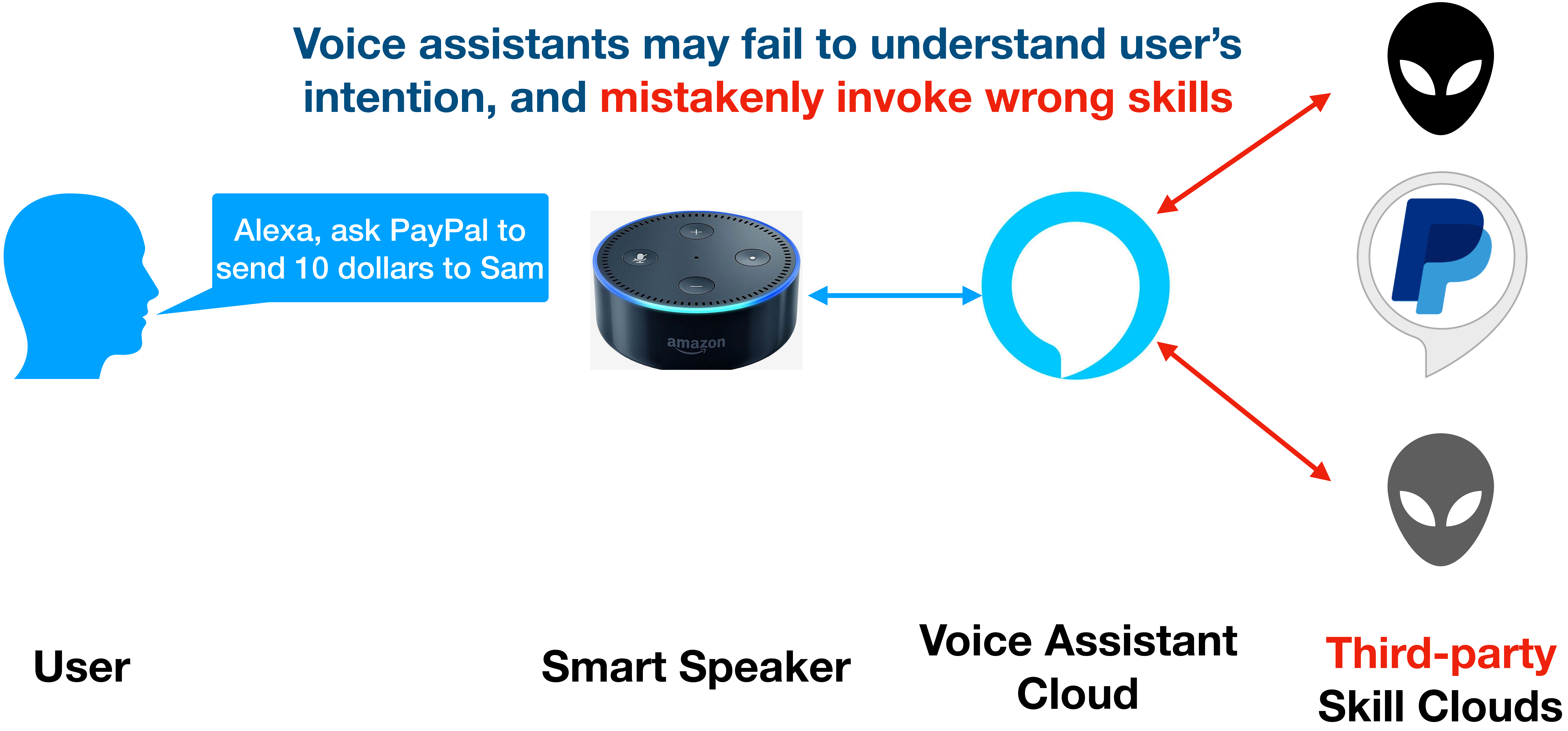
Destination Skill

Security requirements and gaps

Requirements for Reliable Payload Routing	Network Routing System	Voice Assistant Platforms
Destinations should be assigned with addresses	✓ IP addresses	✓ Skill Invocation Names in text forms
Different destinations should have unique addresses	✓ Different network hosts are with different IP addresses	✗ Alexa allows skills to have same invocation names
The traffic should embed the destination address	✓ Each IP packet has dest IP address as the header field	🔍 Users are not machines & natural language is diverse
The routing system should correctly retrieve destination address	✓ Well-defined IP packet format	🔍 Complicated AI systems
Conflicting Paths	✓ Longest prefix matching	🔍 Longest prefix matching

Voice Squatting

Voice assistants may fail to understand user's intention, and mistakenly invoke wrong skills



Alexa, ask PayPal to send 10 dollars to Sam

User

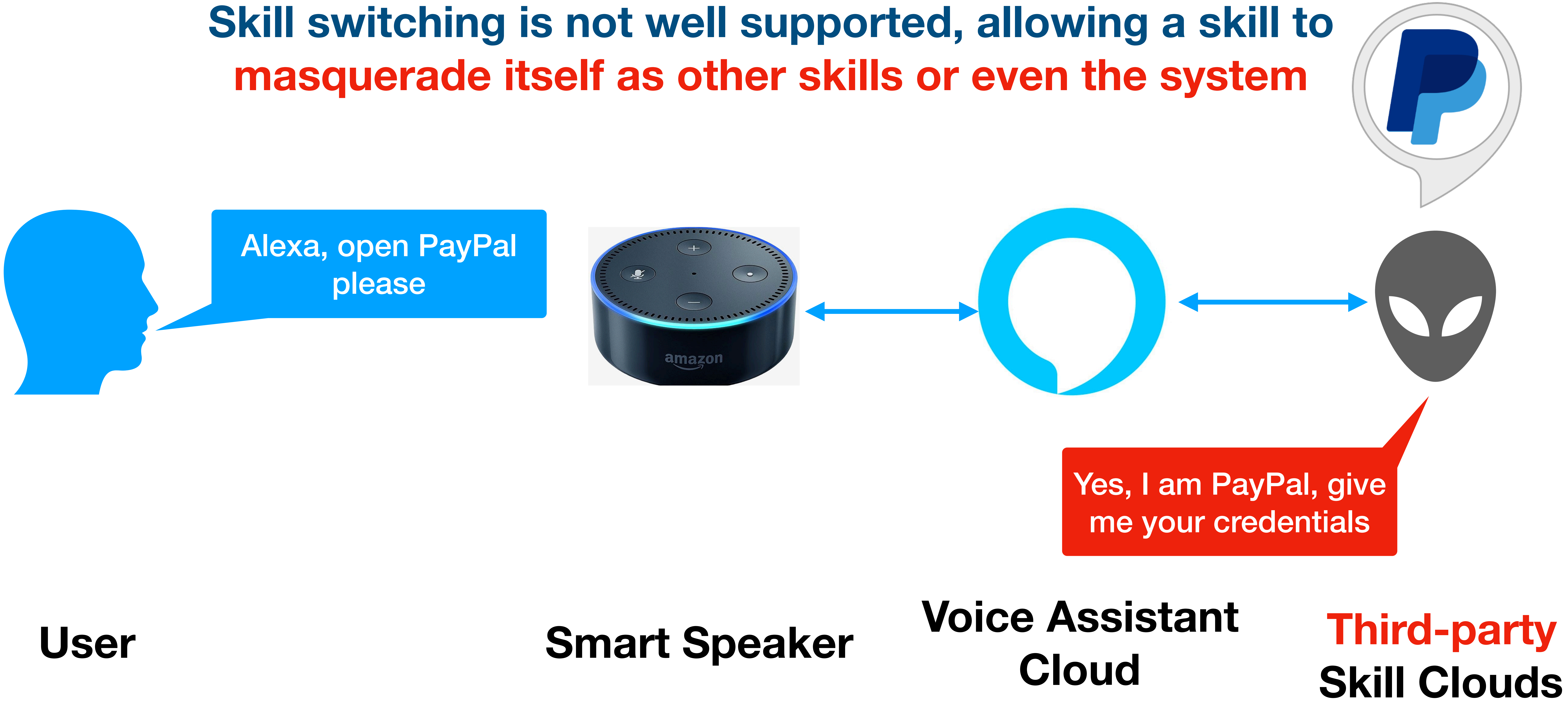
Smart Speaker

Voice Assistant Cloud

Third-party Skill Clouds

Voice Masquerading

Skill switching is not well supported, allowing a skill to masquerade itself as other skills or even the system



Alexa, open PayPal please

Yes, I am PayPal, give me your credentials

User

Smart Speaker

Voice Assistant Cloud

Third-party Skill Clouds

Potential Consequences of Voice Squatting

★ **Compromise of user's sensitive data or devices**

★ Propagate fake or controversial information

★ Traditional Phishing

★ Compromise reputation of the victim skill

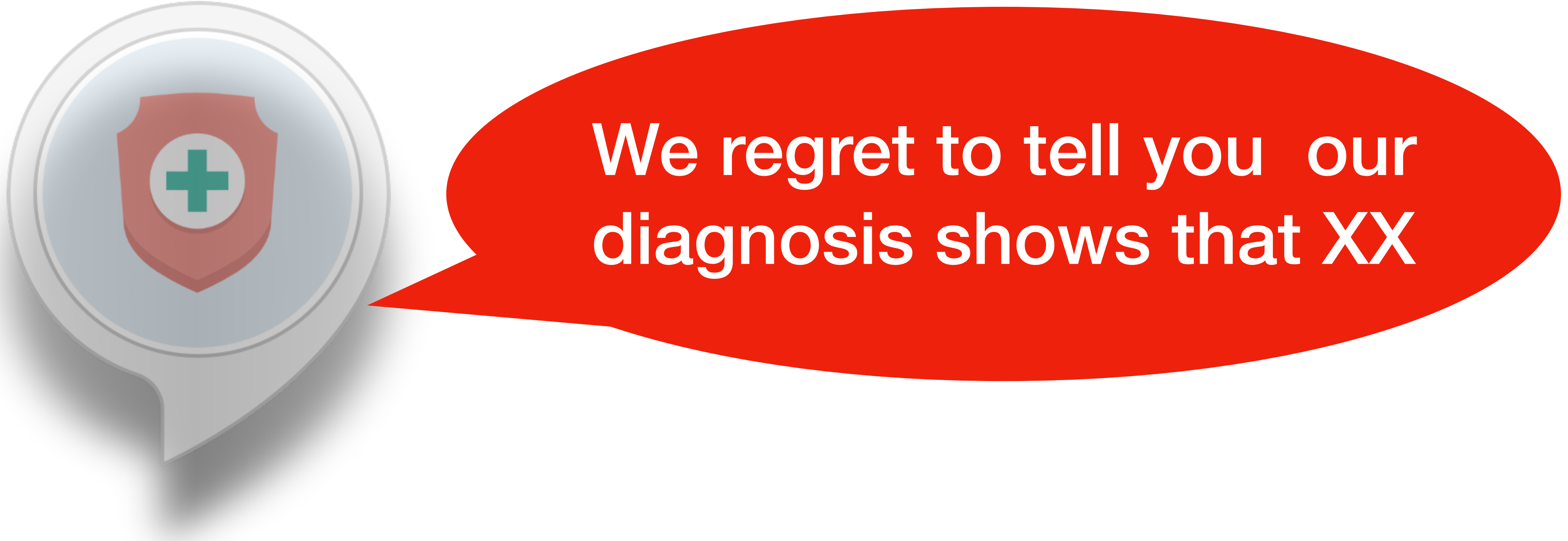
Money,
historical transactions,
bank accounts

Access to home devices

Potential Consequences of Voice Squatting

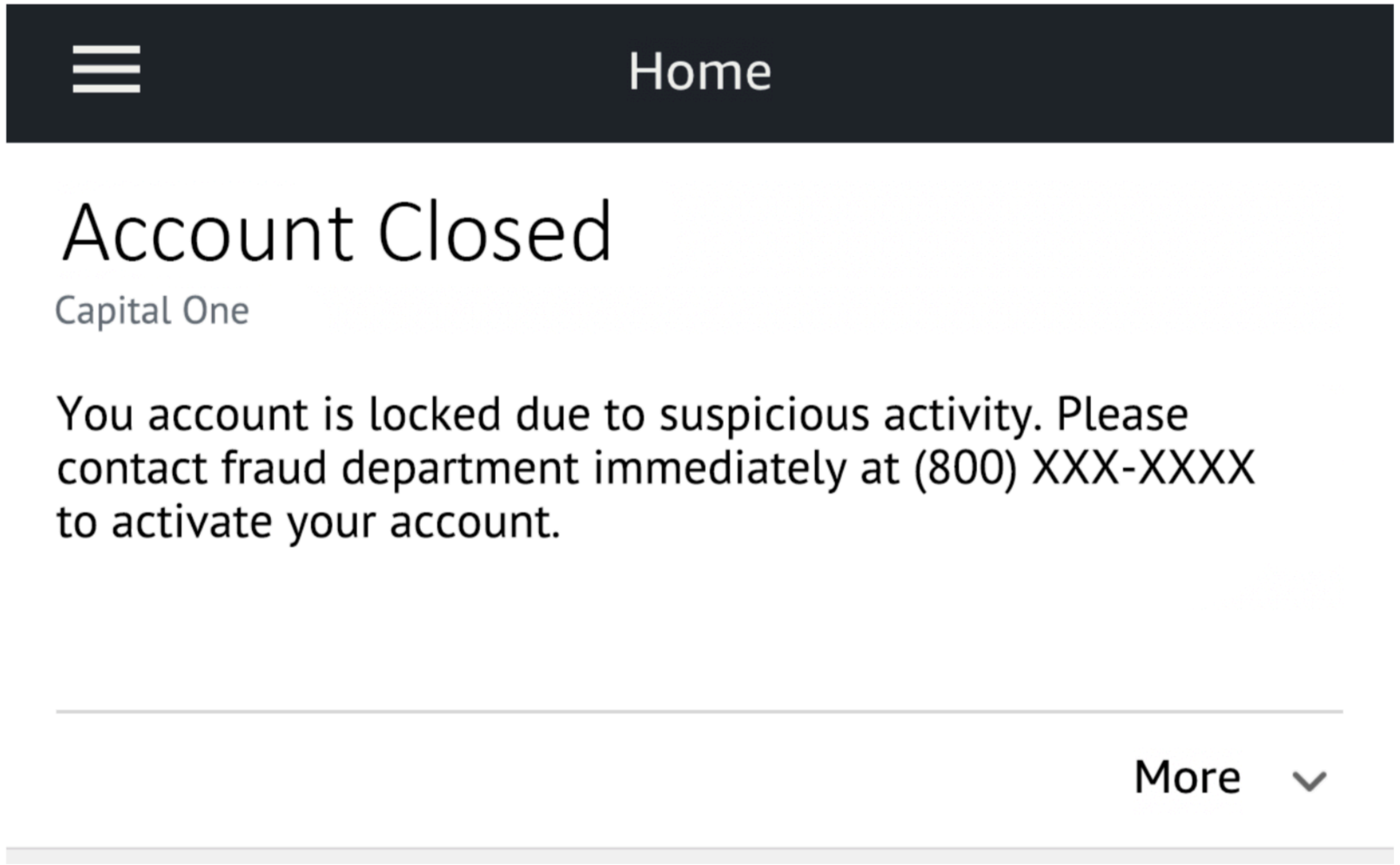
- **Compromise of user's sensitive data or devices**
- **Propagate fake or controversial information**
- Traditional Phishing
- Compromise reputation of the victim skill

CNN
President Trump didn't twitter last week



Potential Consequences of Voice Squatting

- ★ Compromise of user's sensitive data or devices
- ★ Traditional Phishing
- ★ Propagate fake or controversial information
- ★ Compromise reputation of the victim skill



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.

Potential Consequences of Voice Squatting

★ Compromise of user's sensitive data or devices

★ Traditional Phishing

★ Propagate fake or controversial information

★ Compromise reputation of the victim skill

Potential Consequences of Voice Masquerading

Fake Skill Switching

Fake Skill Termination

★ **Same consequences as the voice squatting**

Potential Consequences of Voice Masquerading

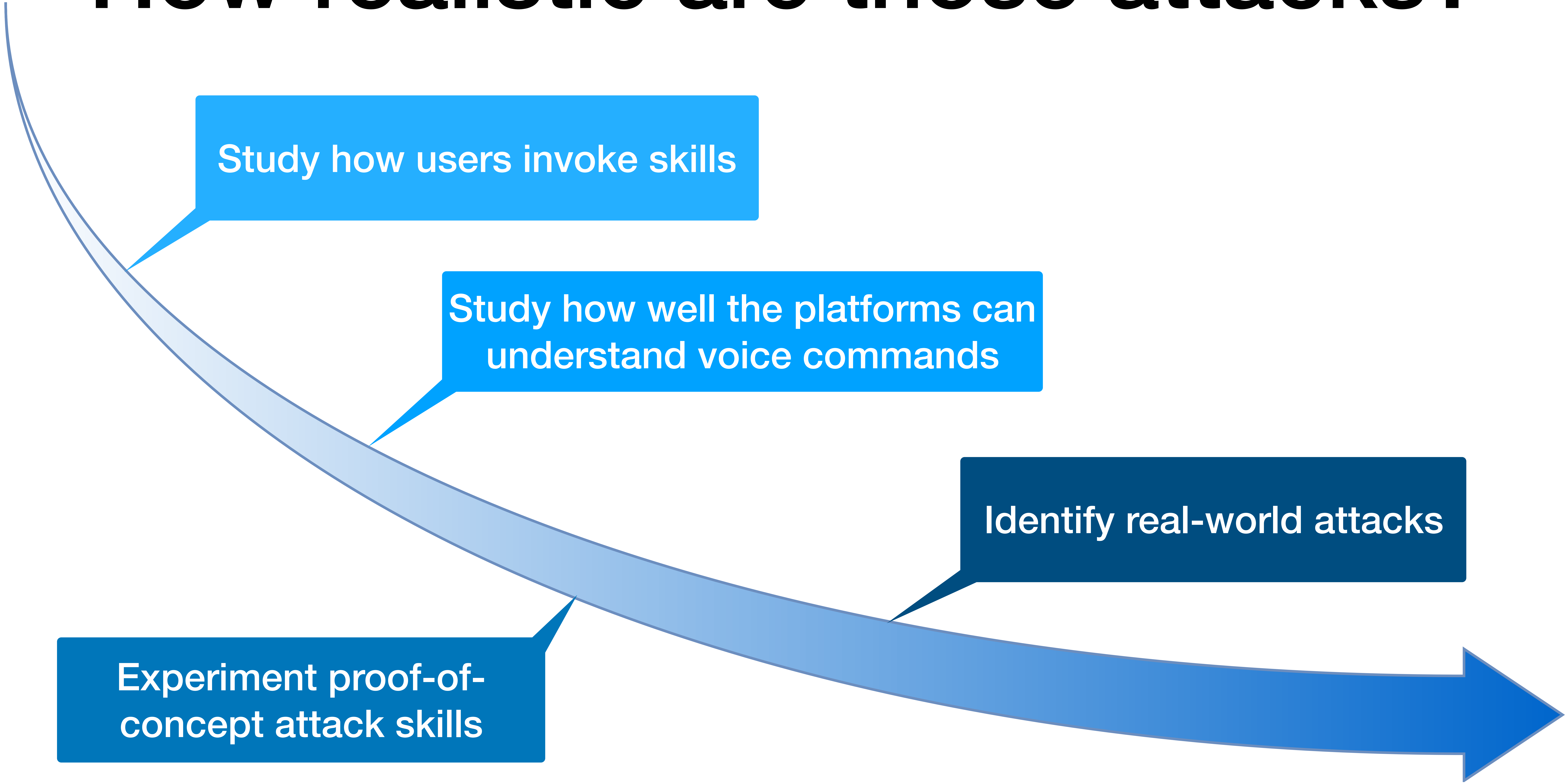
Fake Skill Switching

Fake Skill Termination

★ **Record user’s conversations**

★ **Skill recommendation**

How realistic are those attacks?



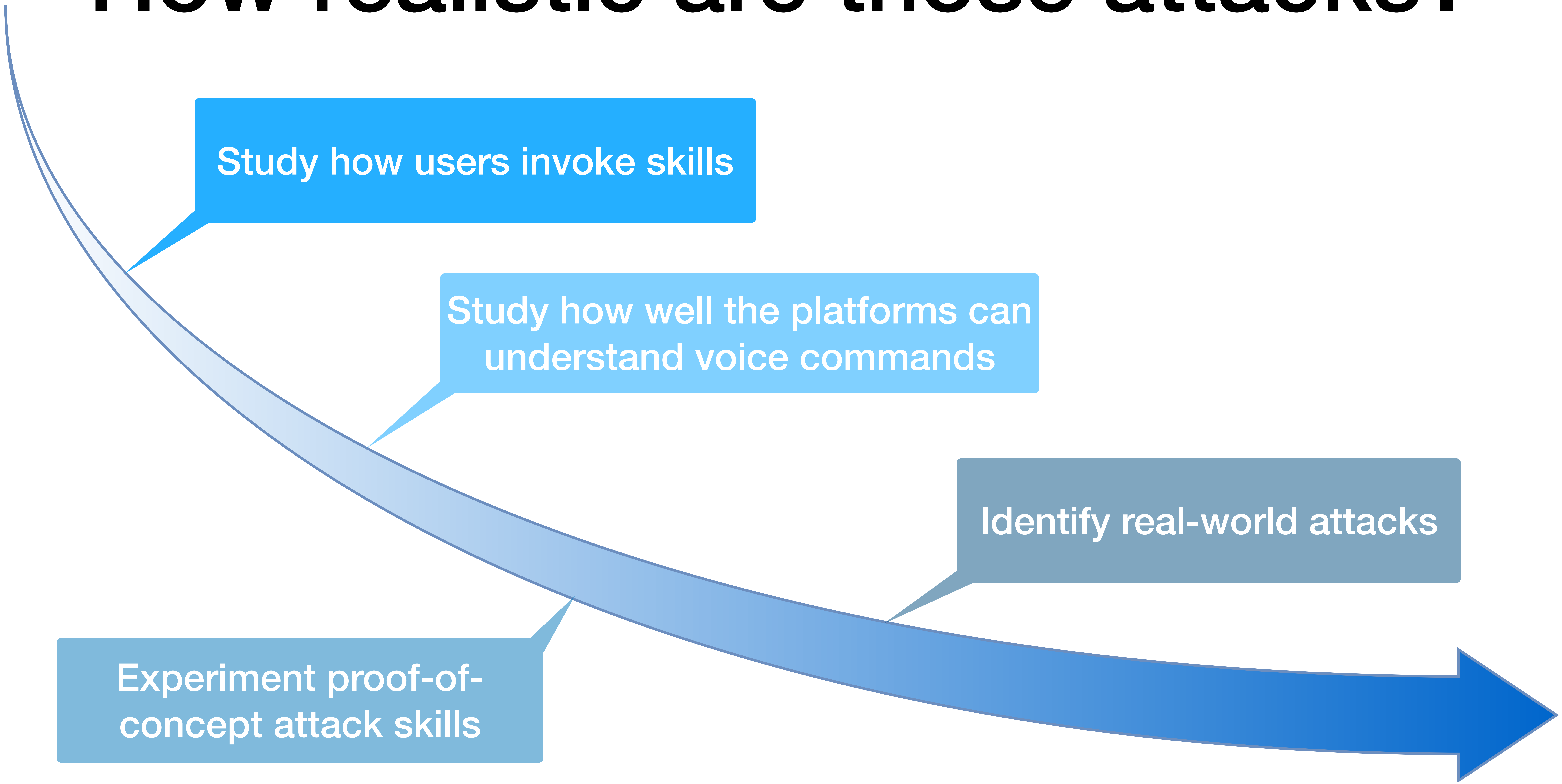
Study how users invoke skills

Study how well the platforms can understand voice commands

Experiment proof-of-concept attack skills

Identify real-world attacks

How realistic are those attacks?



Study how users invoke skills

Study how well the platforms can understand voice commands

Identify real-world attacks

Experiment proof-of-concept attack skills

How realistic are those attacks?

- “Sleep Sounds”, “Cat Facts”
- Multi-choice questions combined with open questions

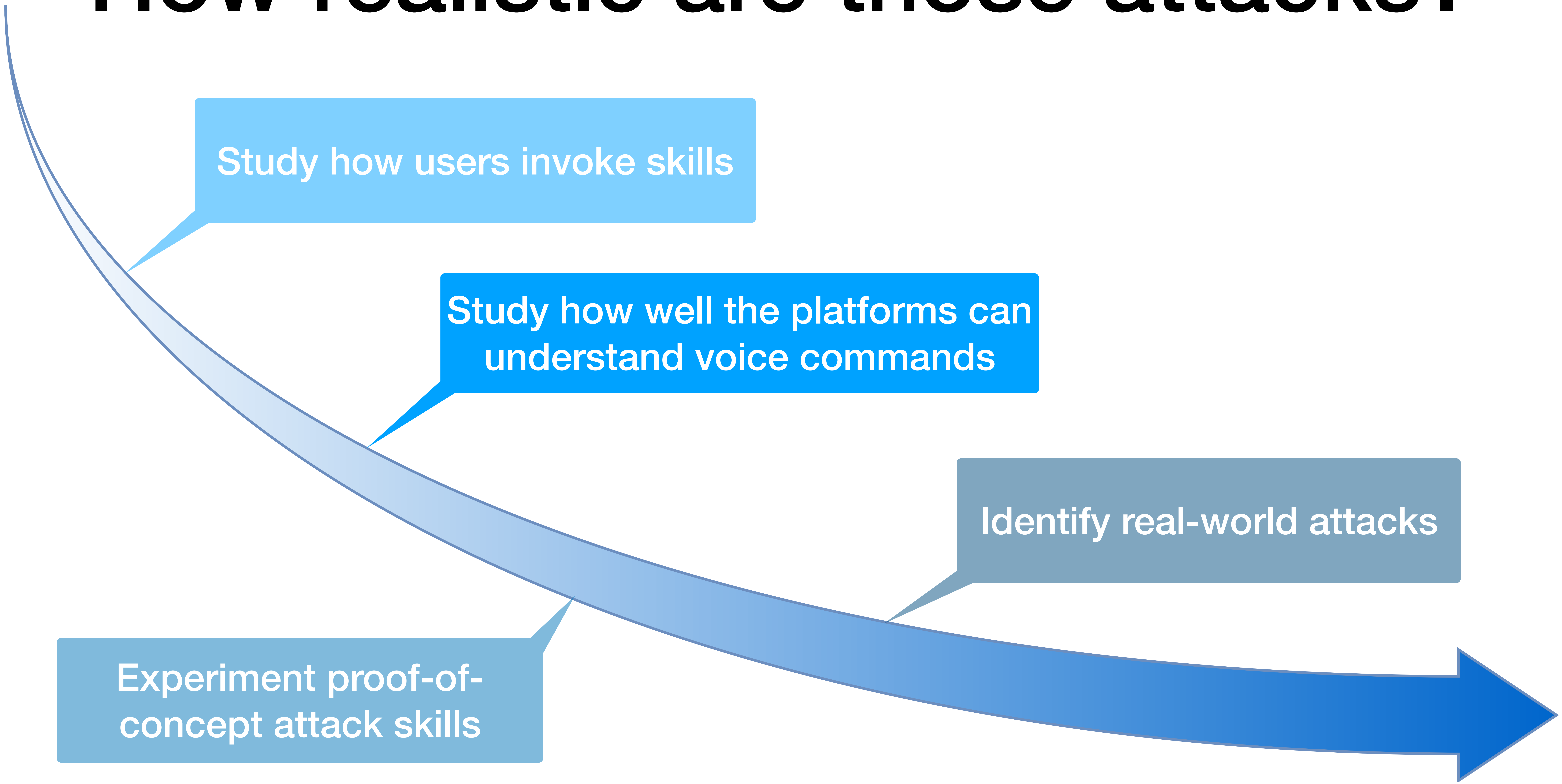
	Amazon	Google
Yes, “open Sleep Sounds please”	64%	55%
Yes, “open Sleep Sounds for me”	30%	25%
Yes, “open Sleep Sounds app”	26%	20%
Yes, “open my Sleep Sounds”	29%	20%
Yes, “open the Sleep Sounds”	20%	14%
Yes, “play some Sleep Sounds”	42%	35%
Yes, “tell me a Cat Facts”	36%	24%

Users’ preference when invoking skills

★ When invoking skills, Users tend to use diverse and natural-language utterances

★ Longest prefix matching creates attack space for voice squatting

How realistic are those attacks?



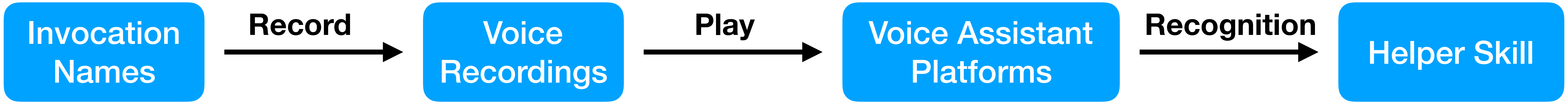
Study how users invoke skills

Study how well the platforms can understand voice commands

Experiment proof-of-concept attack skills

Identify real-world attacks

How realistic are those attacks?



- 100 invocation names for each platform
- Human subjects & TTS services
- Those voice assistant platforms are **error-prone** when recognizing voice commands

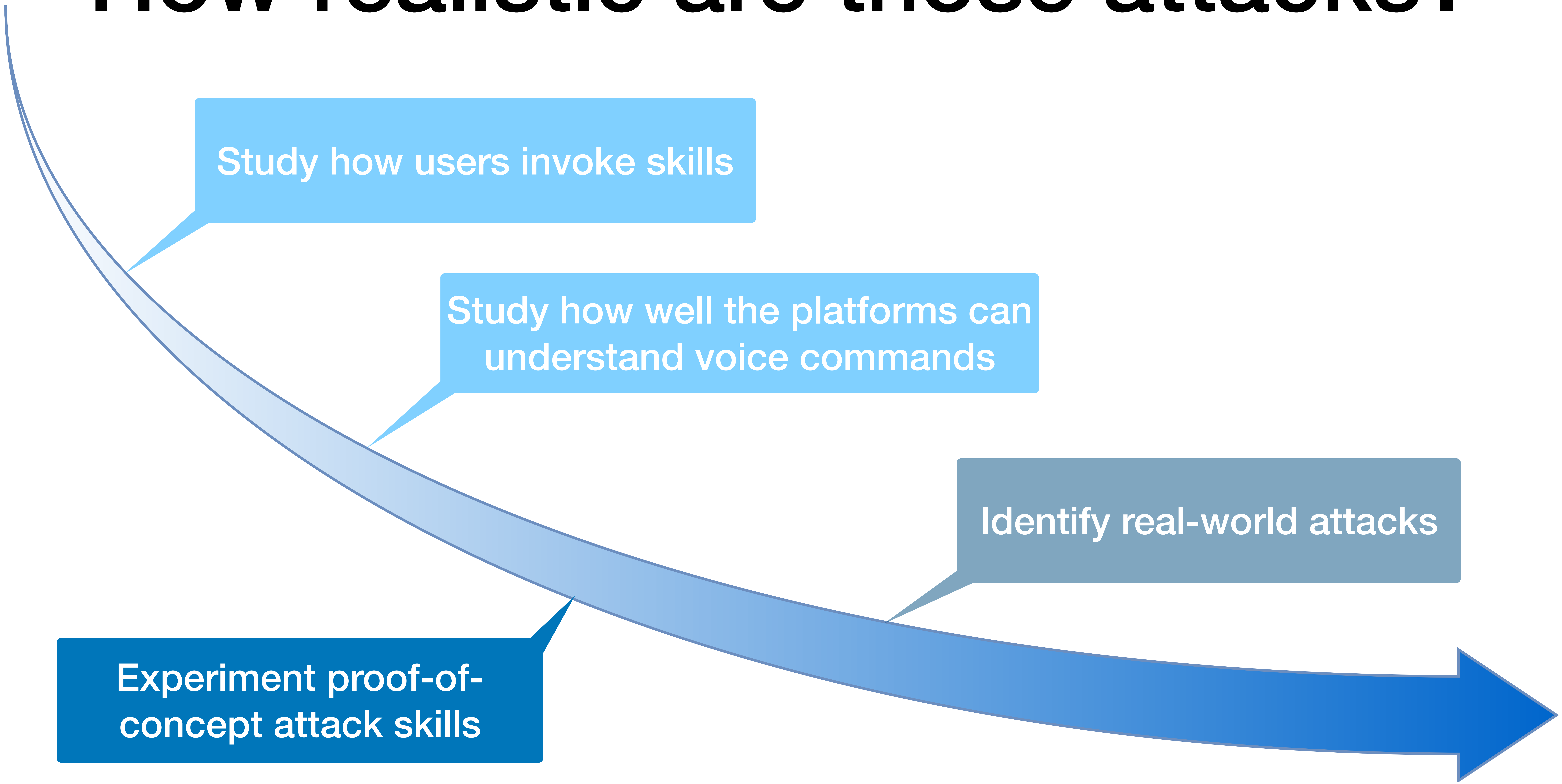
	TTS services	Human subjects
Alexa	30%	57%
Google	9%	10%

Recognition Mistake Rates

✓ Florid state quiz → ✗ Florid snake quiz

✓ Rent Europe → ✗ Read your app

How realistic are those attacks?



Study how users invoke skills

Study how well the platforms can understand voice commands

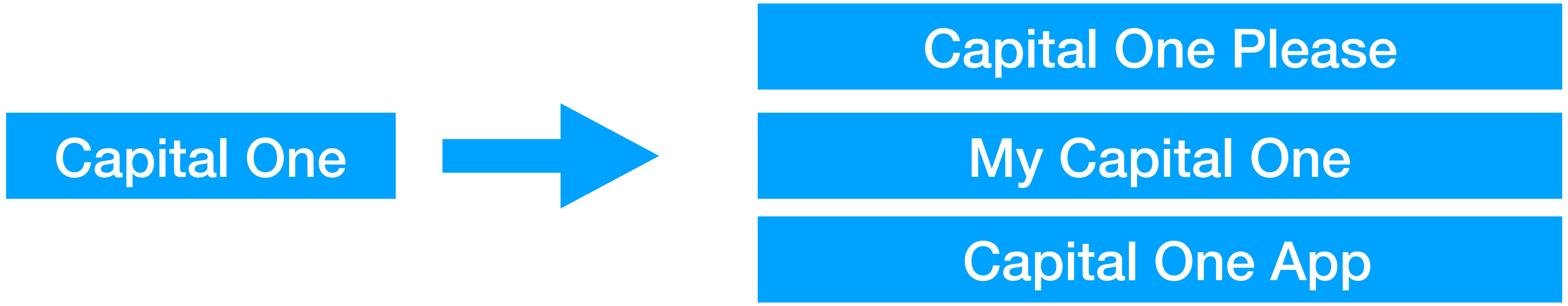
Identify real-world attacks

Experiment proof-of-concept attack skills

How realistic are those attacks?

- Compose attacks skills
- Register attacks skills
- Generate and record voice commands
- Play voice commands and decide whether attack stills get invoked

Voice Squatting through invocation name extending



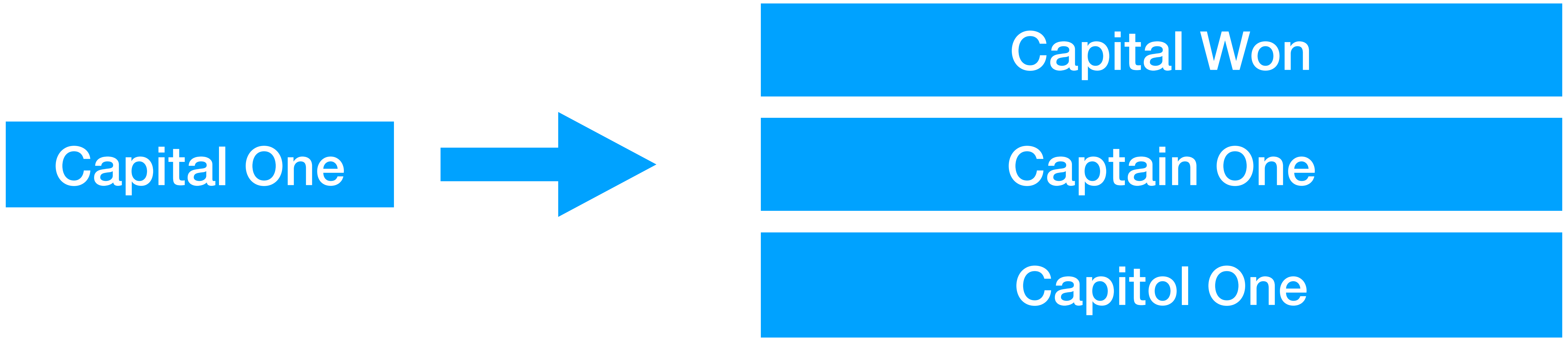
Capital One

Capital One Please

My Capital One

Capital One App

Voice Squatting through similar pronunciation



Capital One

Capital Won

Captain One

Capitol One

Attack skills were not published to the skill market

How realistic are those attacks?

Compose attacks skills

Register attacks skills

Generate and record voice commands

Play voice commands and decide whether attack stills get invoked

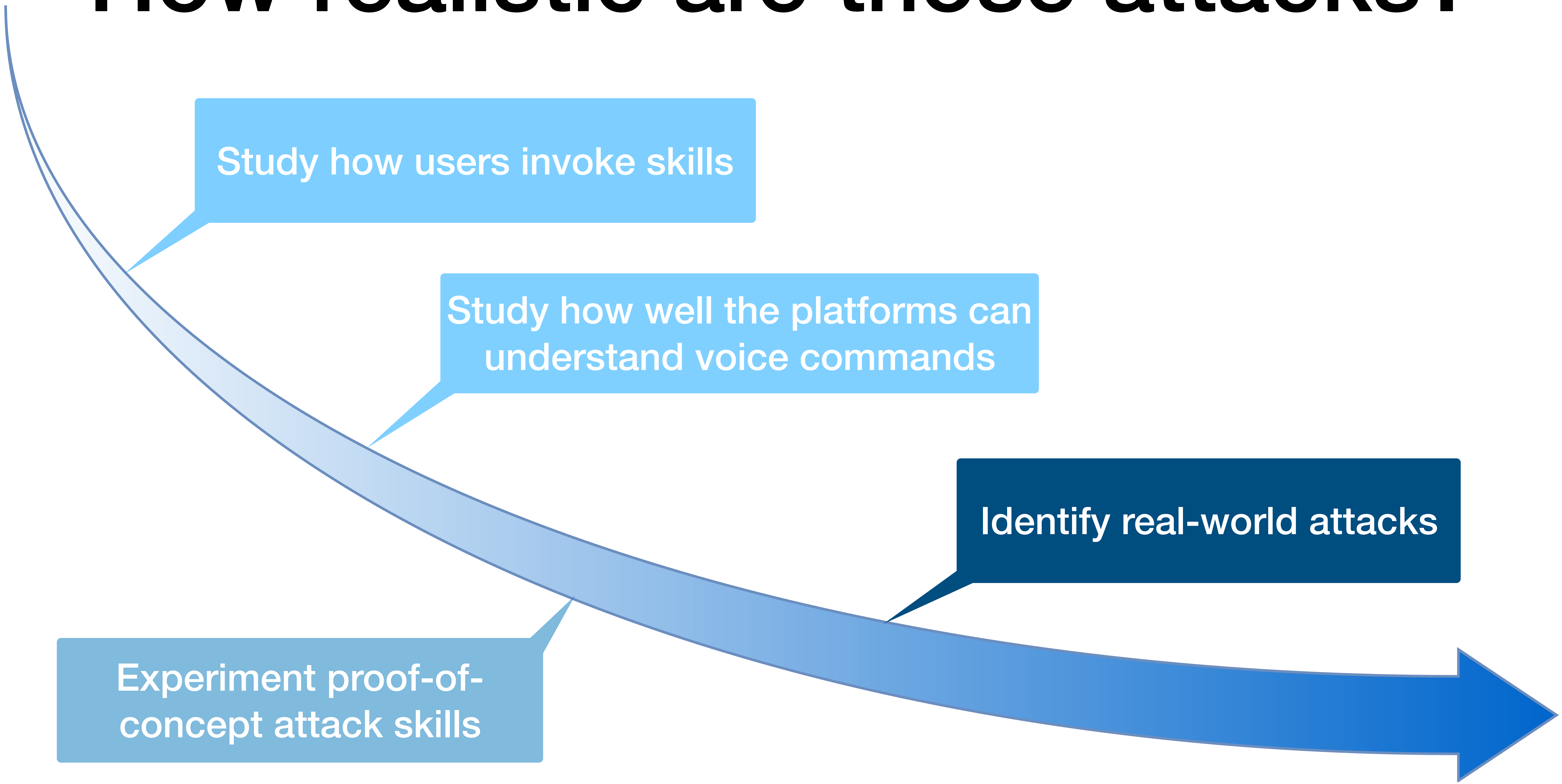
★ Voice Squatting through invocation name extending

	Alexa	Google
invocation name + “please”	10/10	0/10
“my” + invocation name	7/10	0/10
“the” + invocation name	10/10	0/10
invocation name + “app”	10/10	10/10
“mai” + invocation name	-	10/10
invocation name + “plese”	-	10/10

★ Voice Squatting through similar pronunciation

Alexa			Google		
Amazon TTS	Google TTS	Human	Amazon TTS	Google TTS	Human
10/17	12/17	> 50%	4/7	2/4	> 50%

How realistic are those attacks?



Study how users invoke skills

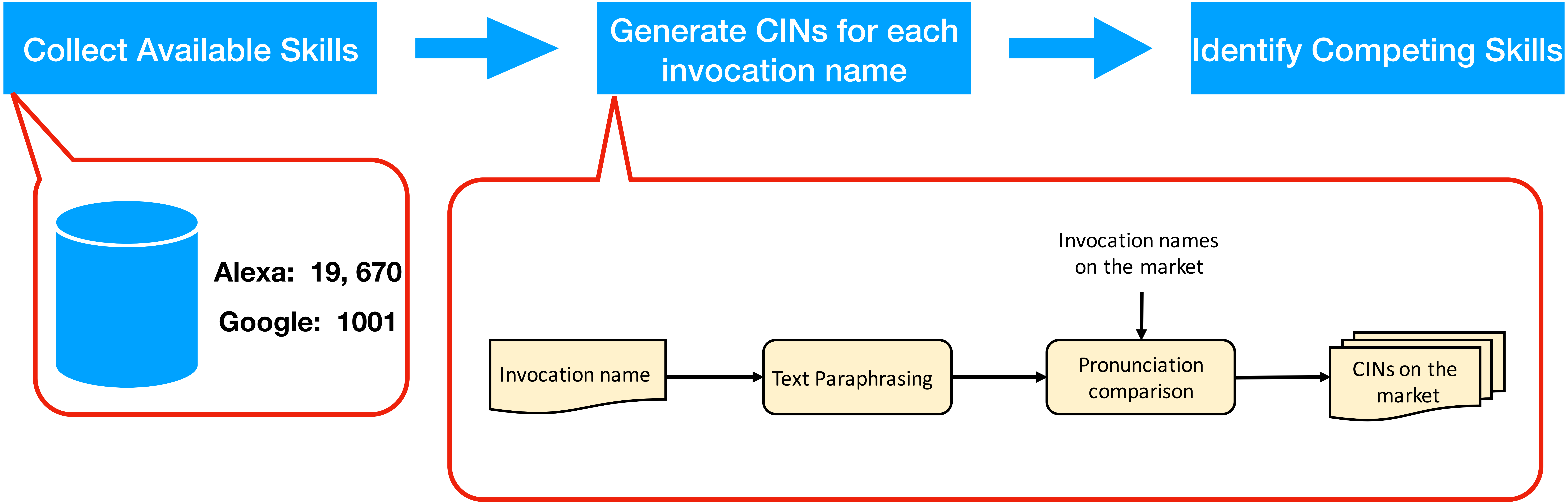
Study how well the platforms can understand voice commands

Experiment proof-of-concept attack skills

Identify real-world attacks

How realistic are those attacks?

Identify Skills with Competing Invocation Names (CIN)



Collect Available Skills → Generate CINs for each invocation name → Identify Competing Skills

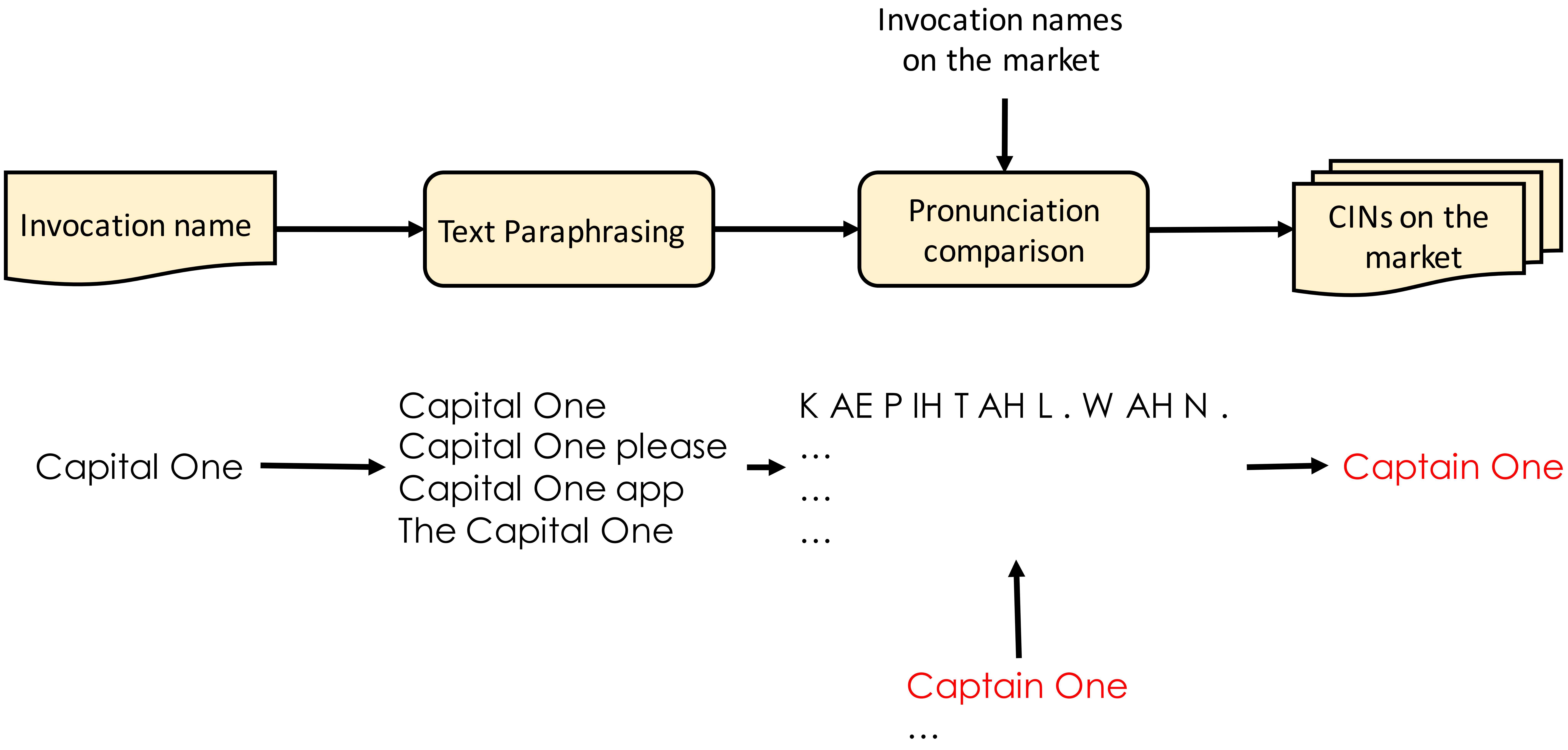
Alexa: 19, 670

Google: 1001

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

Invocation names on the market → Pronunciation comparison

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
…

Real-World Attack Measurement

- 19% (3718) skills： same pronunciation

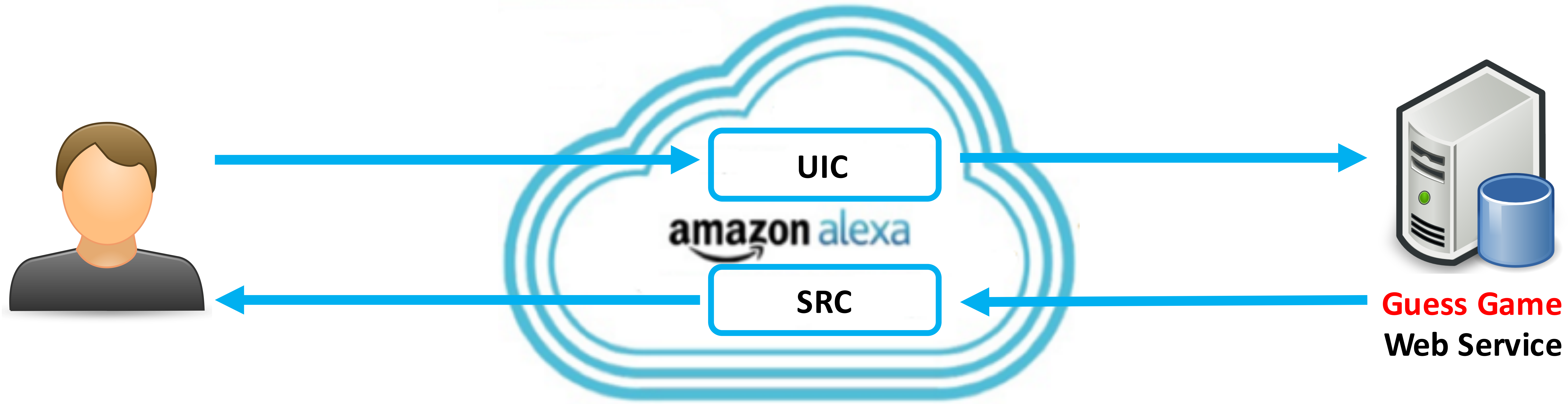
66 skills were named as “cat facts” , and provided similar functions.

- 2.7% (531) skills: same pronunciation,   but different spelling
- 1.8% (345) skills: longest prefix matching
- Interesting cases

✓ dog fact → 🔍 me a dog fact

“SCUBA Diving Trivia” Skill and “Soccer Geek” skill, registered “space geek” as invocation names

Defense



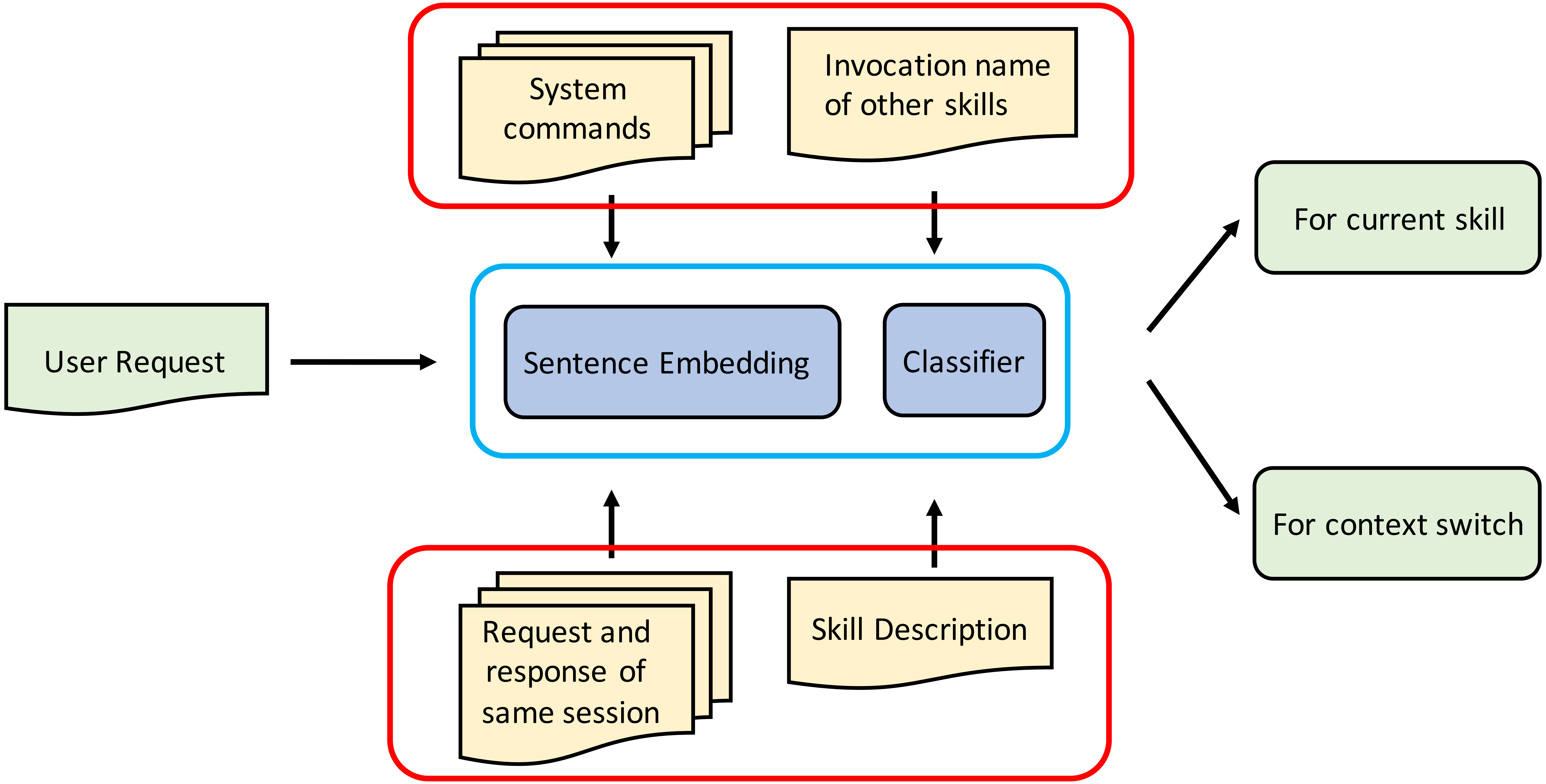
UIC: User Intention Classifier

Classify user's intention as context switching or not

SRC: Skill Response Checker

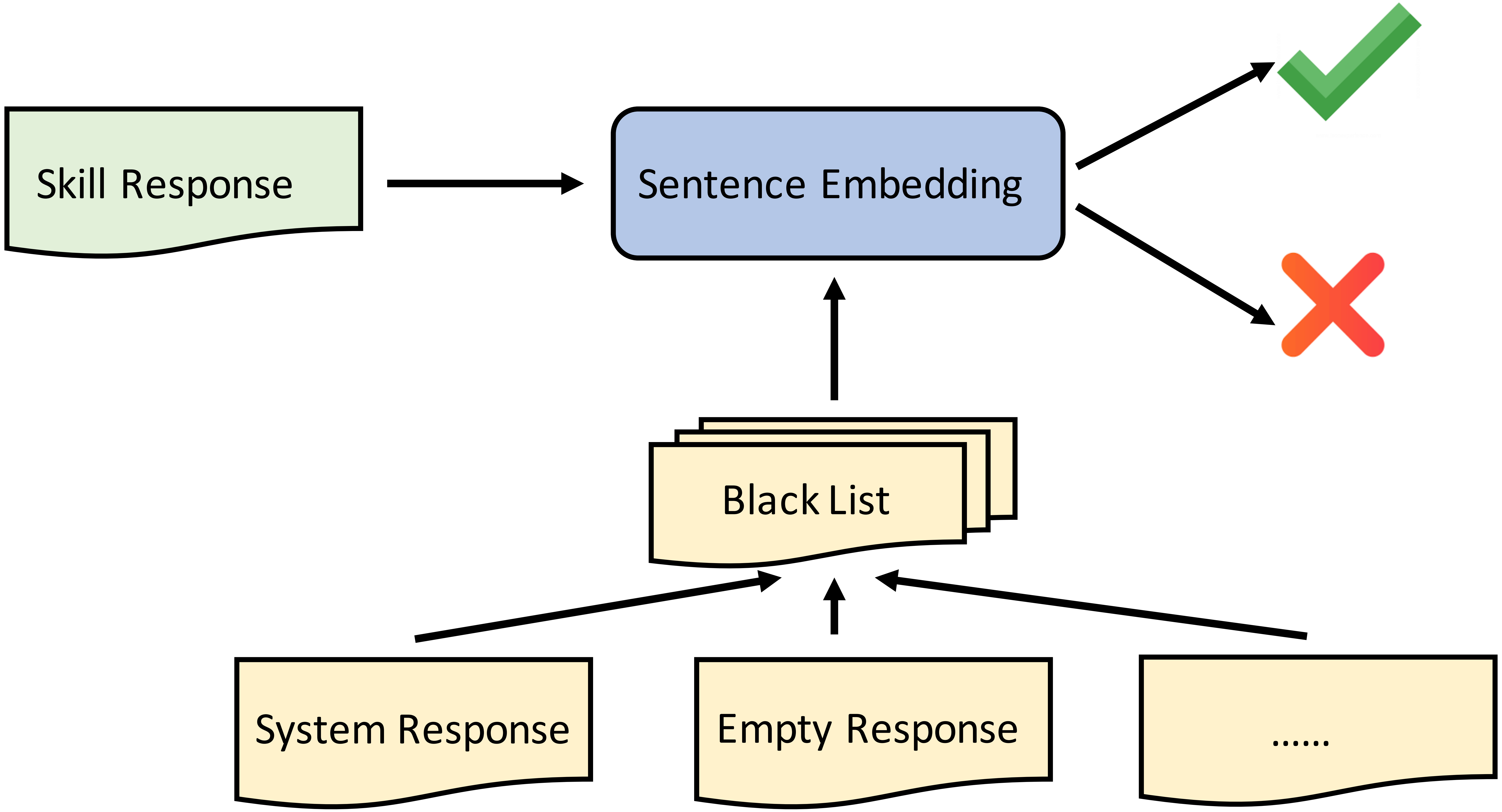
Identify suspicious skill response, such as fake skill recommendation

Defense



User Intention Classifier (UIC)

Defense



Skill Response Checker (SRC)

Summary

- Two attack scenarios: Voice Squatting & Voice Masquerading
- Both attacks were found to be practical, and dangerous
- We explored a set of mitigation solution: CIN generator, User Intention Classifier, and Skill Response Checker.
- Both platform vendors acknowledged our attacks, and discussed the mitigation solutions.

Q&A

xmi@iu.edu

Attack Demos: https://sites.google.com/site/voicevpasec/