

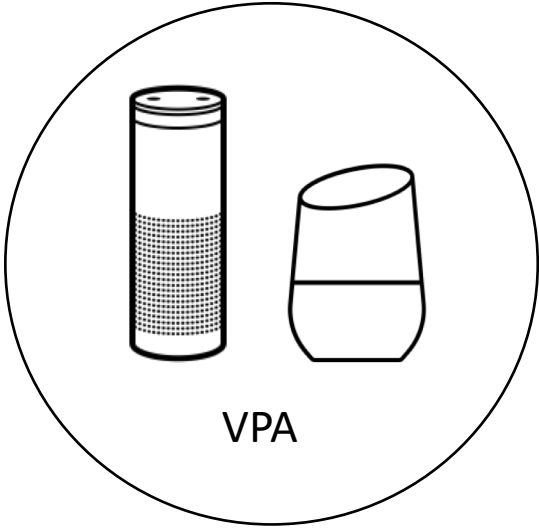
Read Between the Lines:

An Empirical Measurement of Sensitive Applications of Voice Personal Assistant Systems

Faysal Hossain Shezan, Hang Hu, Jiamin Wang, Gang Wang, Yuan Tian



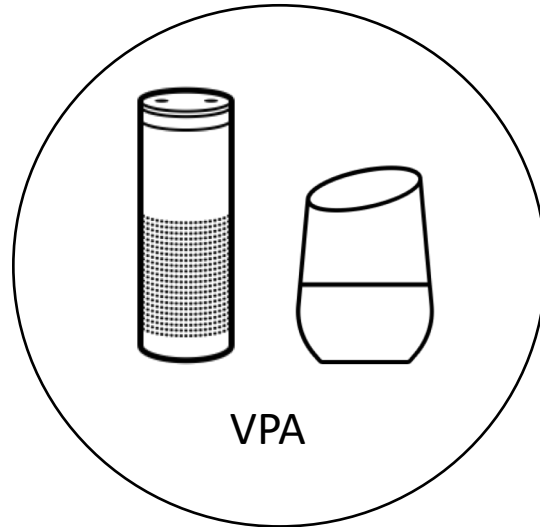
Voice Personal Assistants and Apps (Skills)



Voice Personal Assistants and Apps (Skills)



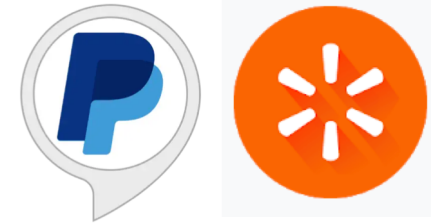
Communication



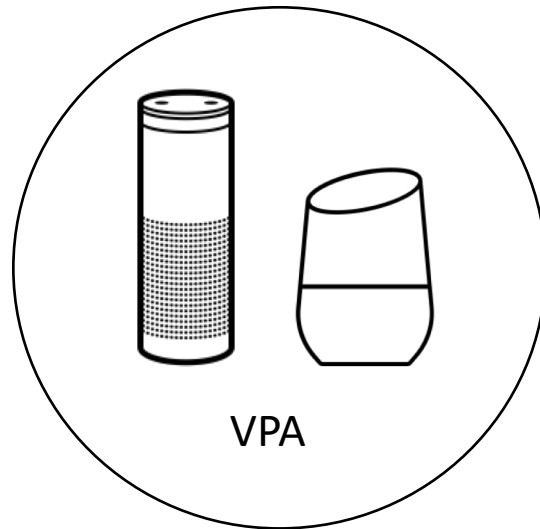
Voice Personal Assistants and Apps (Skills)



Communication



Business & Shopping

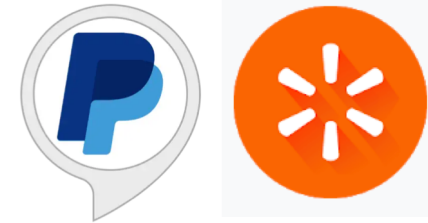


VPA

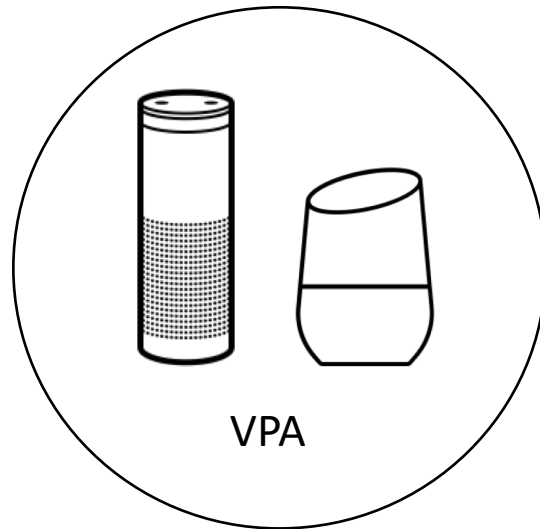
Voice Personal Assistants and Apps (Skills)



Communication



Business & Shopping



VPA

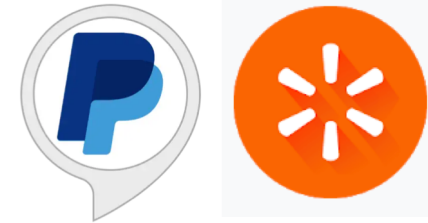


Smart Devices

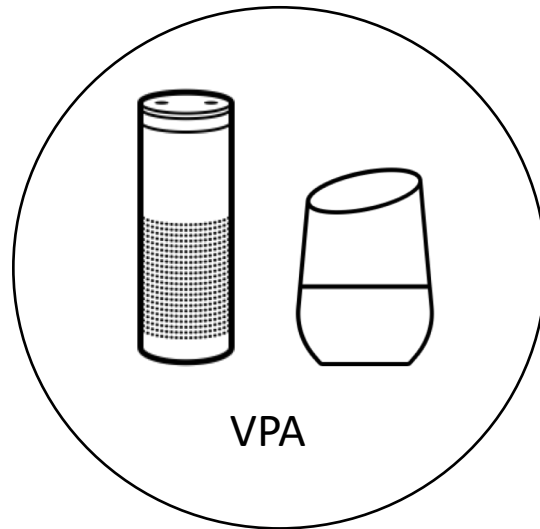
Voice Personal Assistants and Apps (Skills)



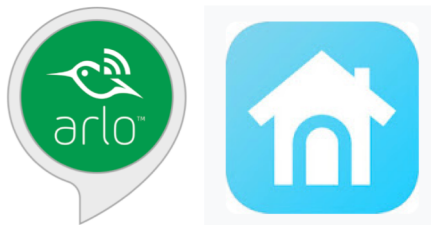
Communication



Business & Shopping



VPA

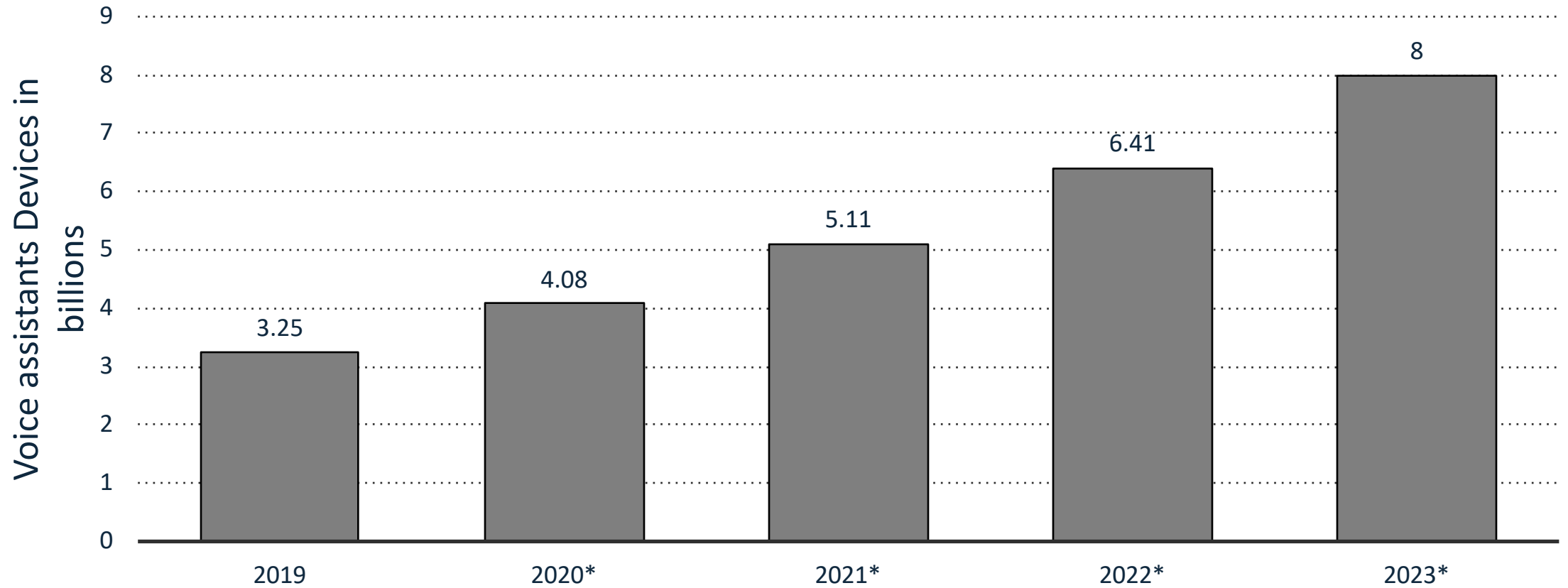


Smart Devices

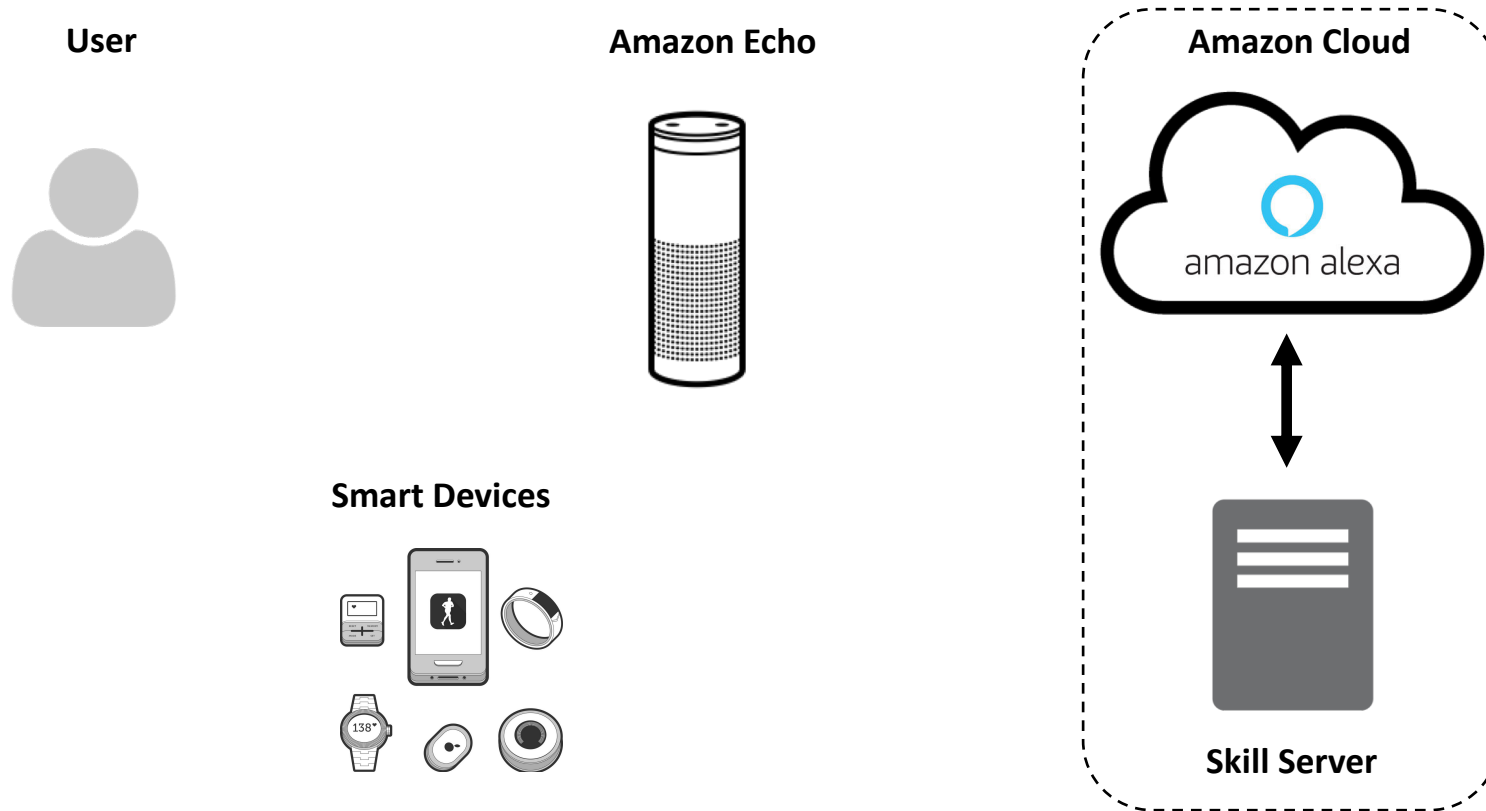


Medical Applications

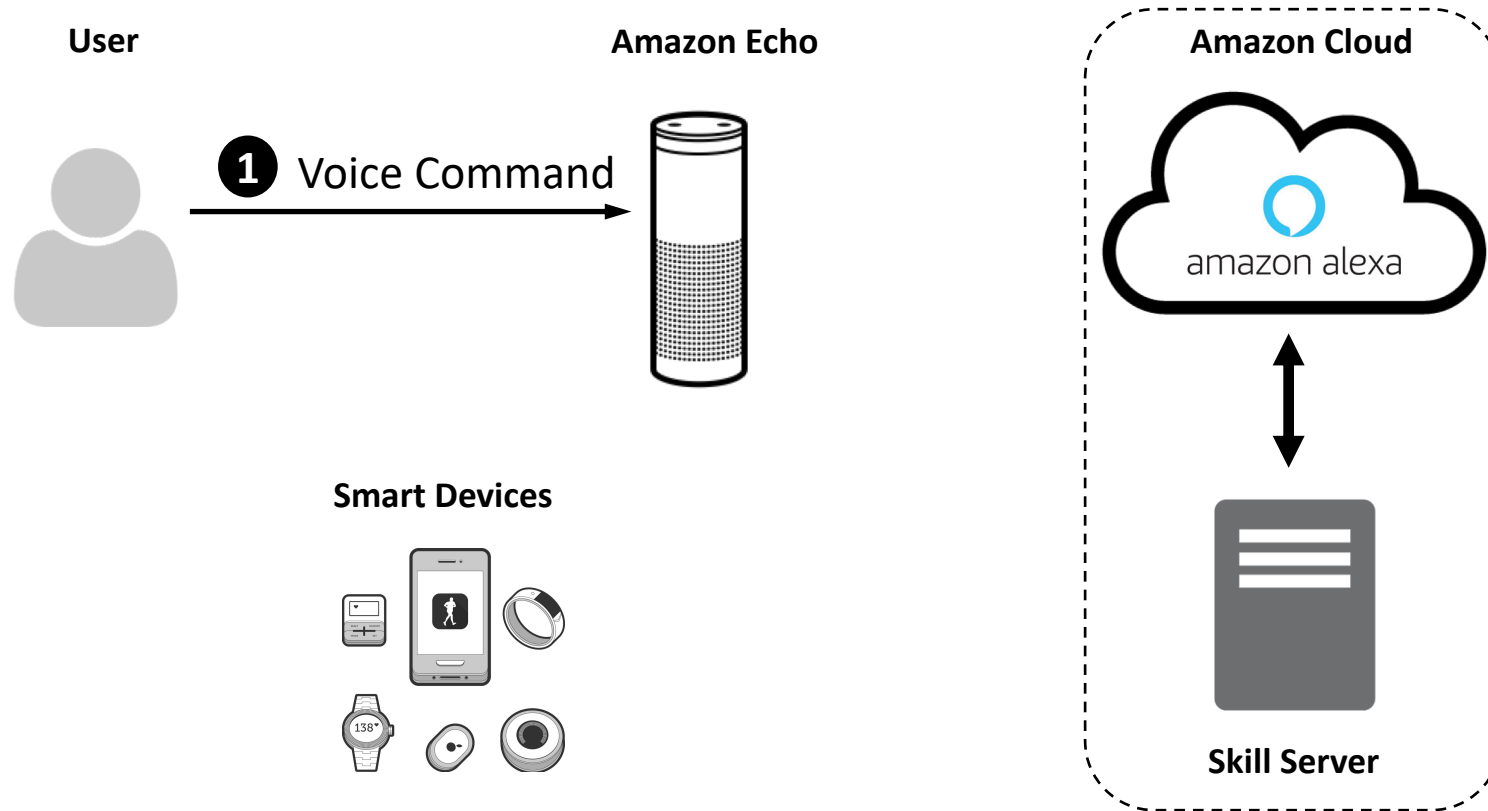
Popularity



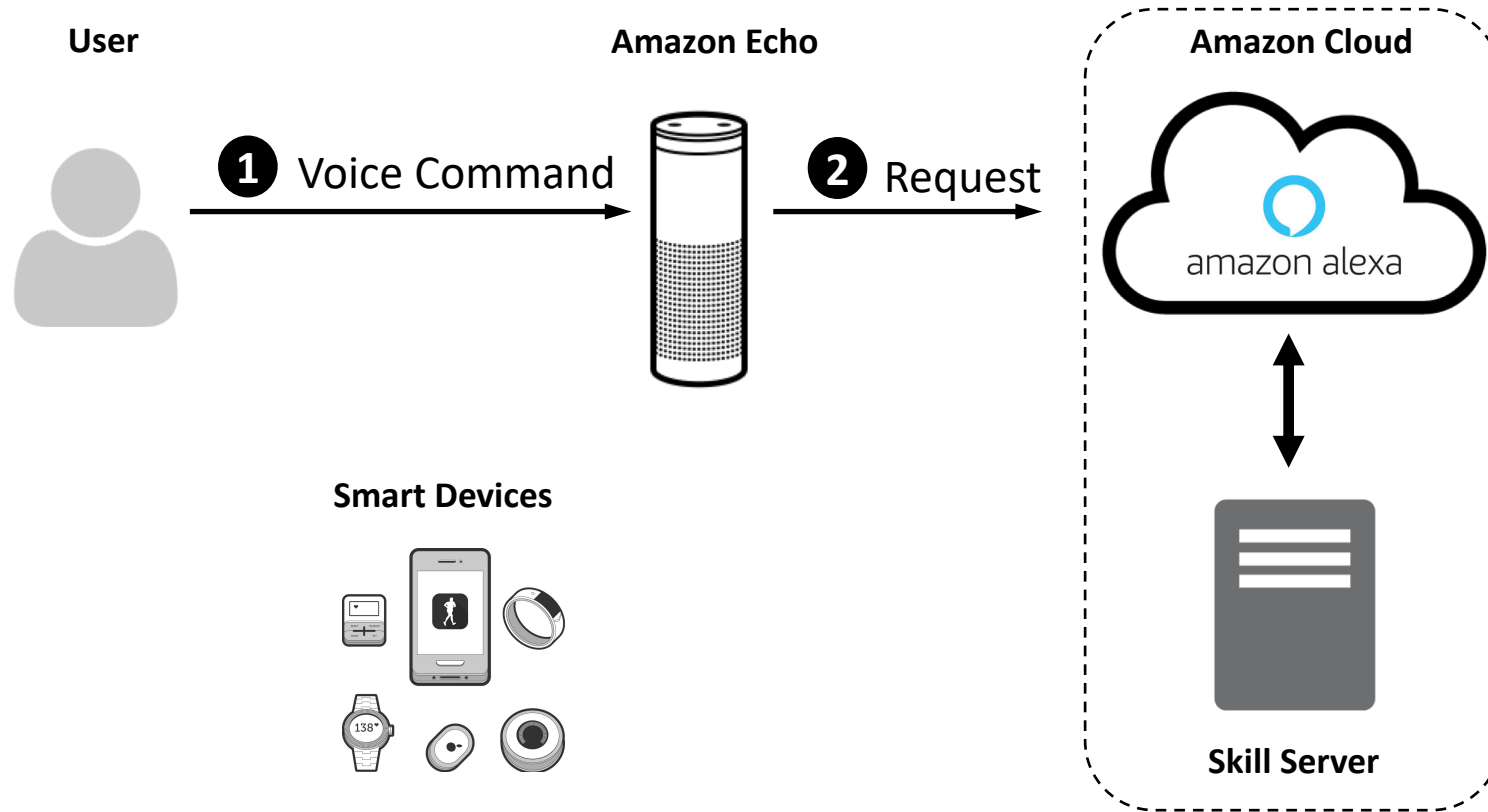
How Does VPA Work?



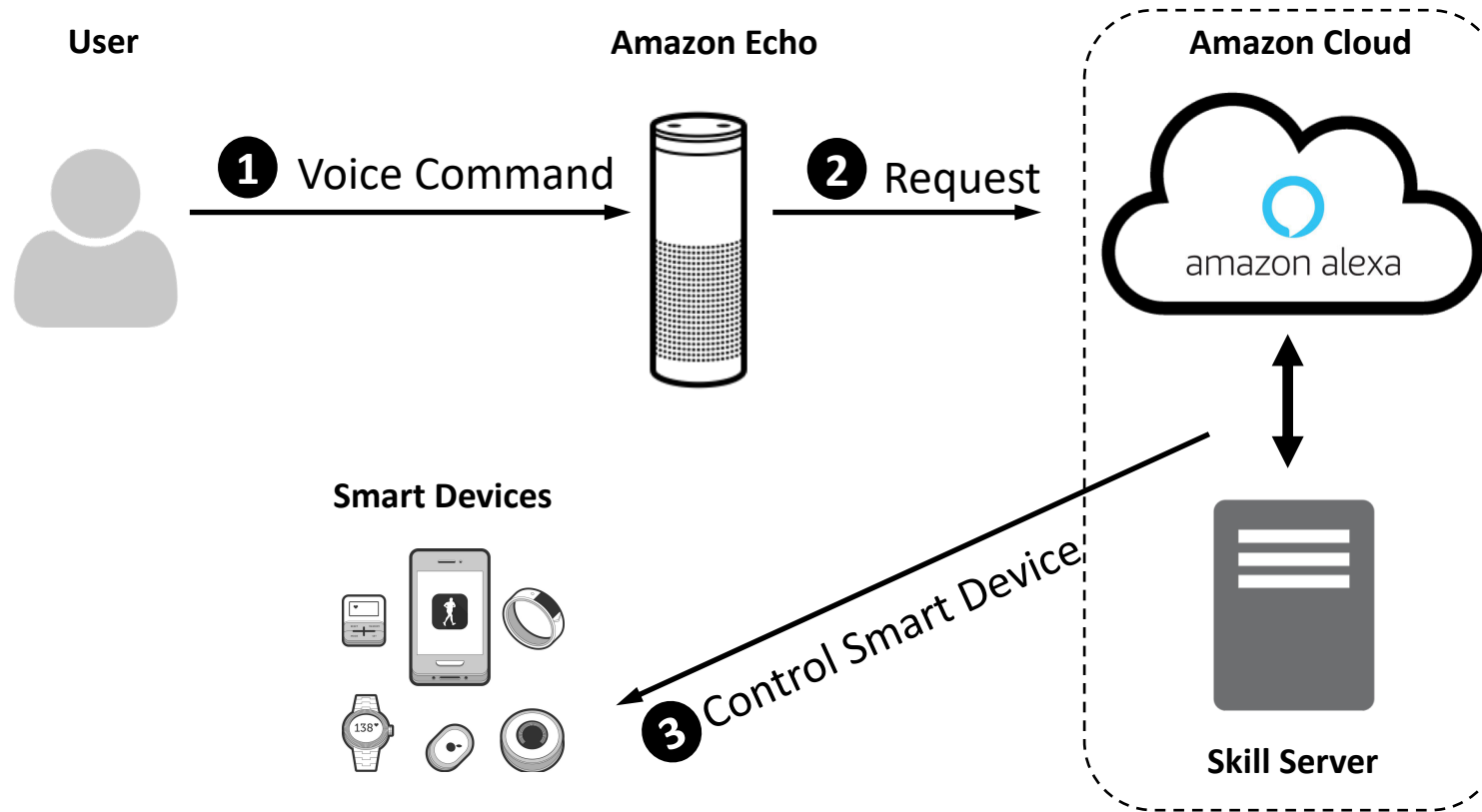
How Does VPA Work?



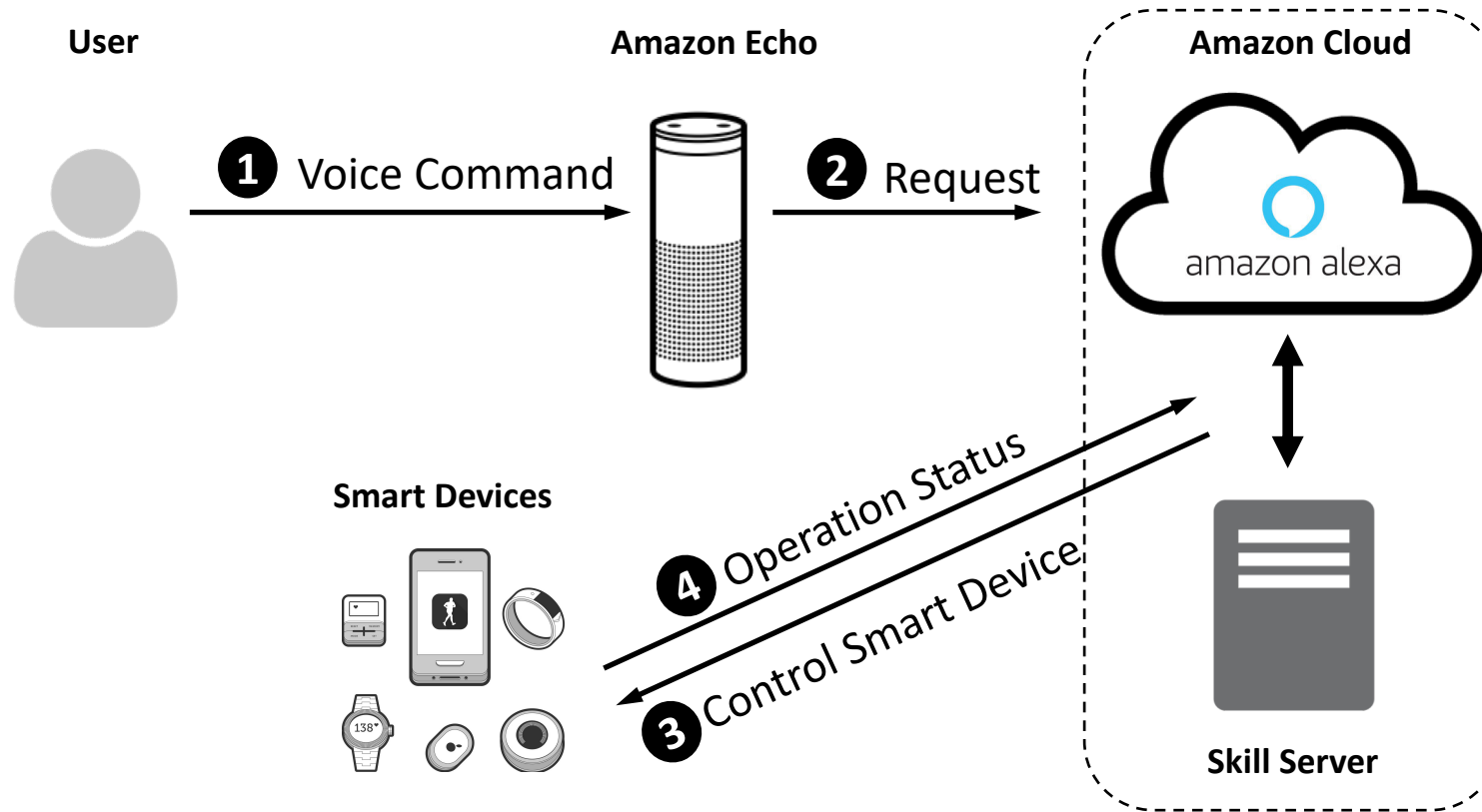
How Does VPA Work?



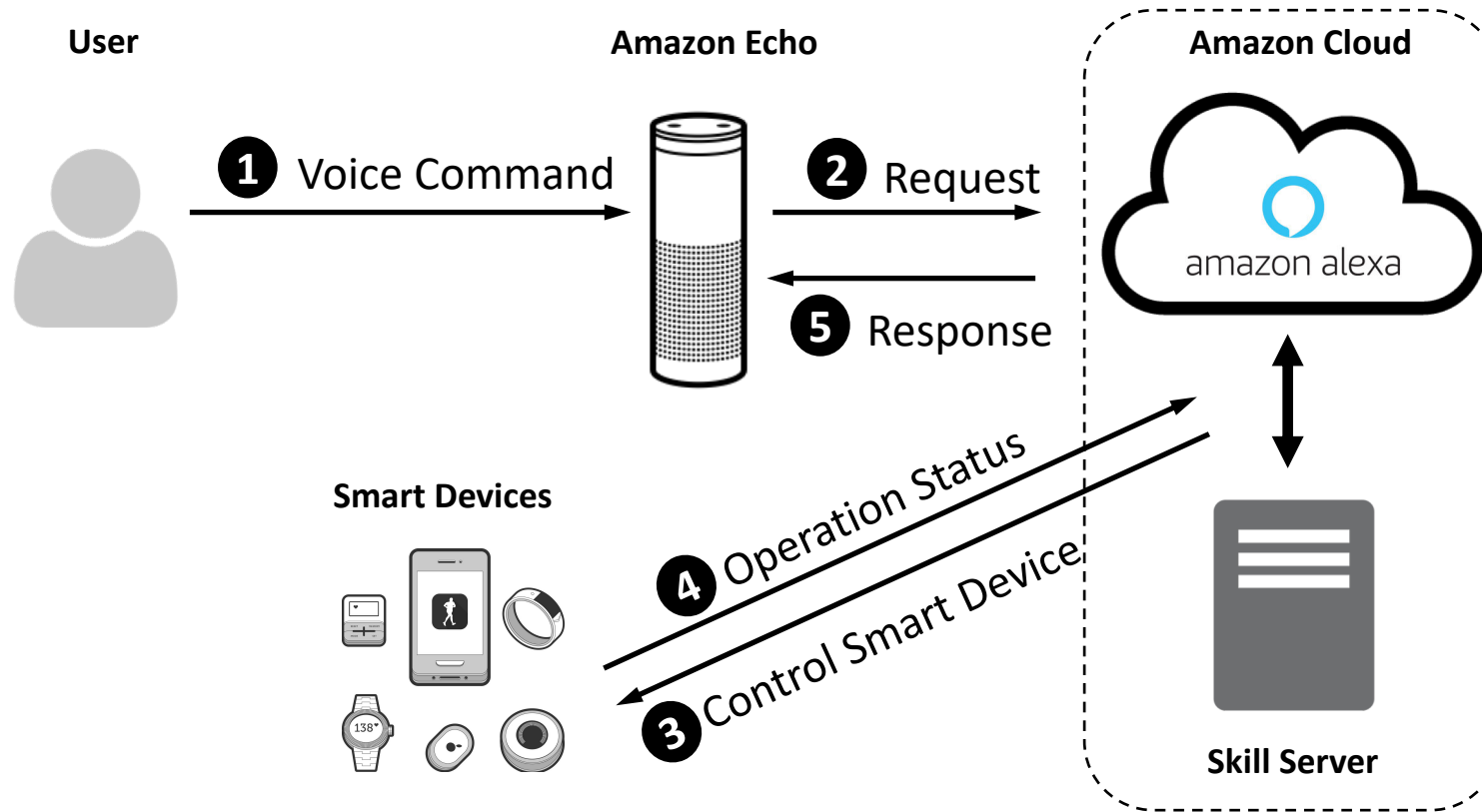
How does VPA work?



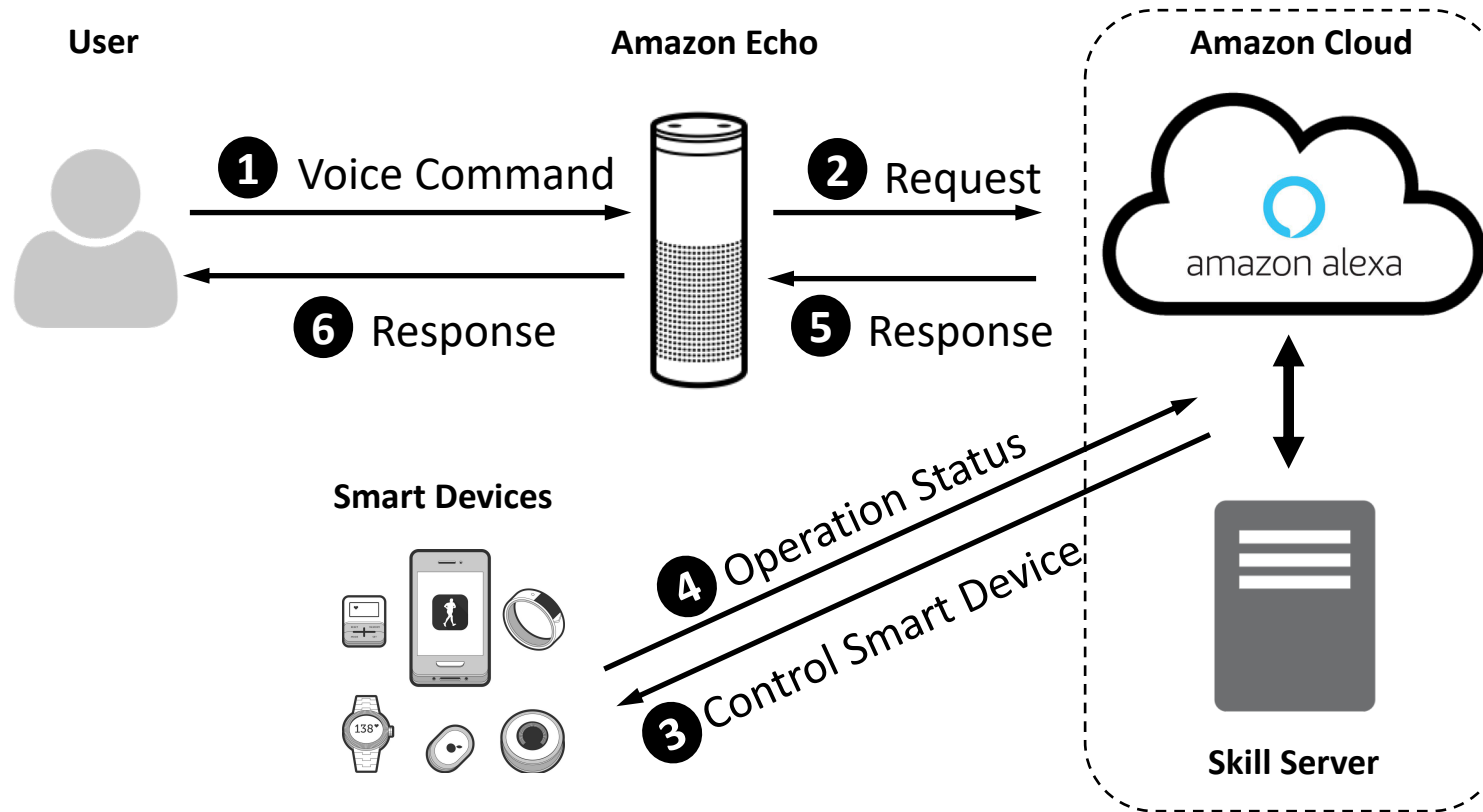
How does VPA work?



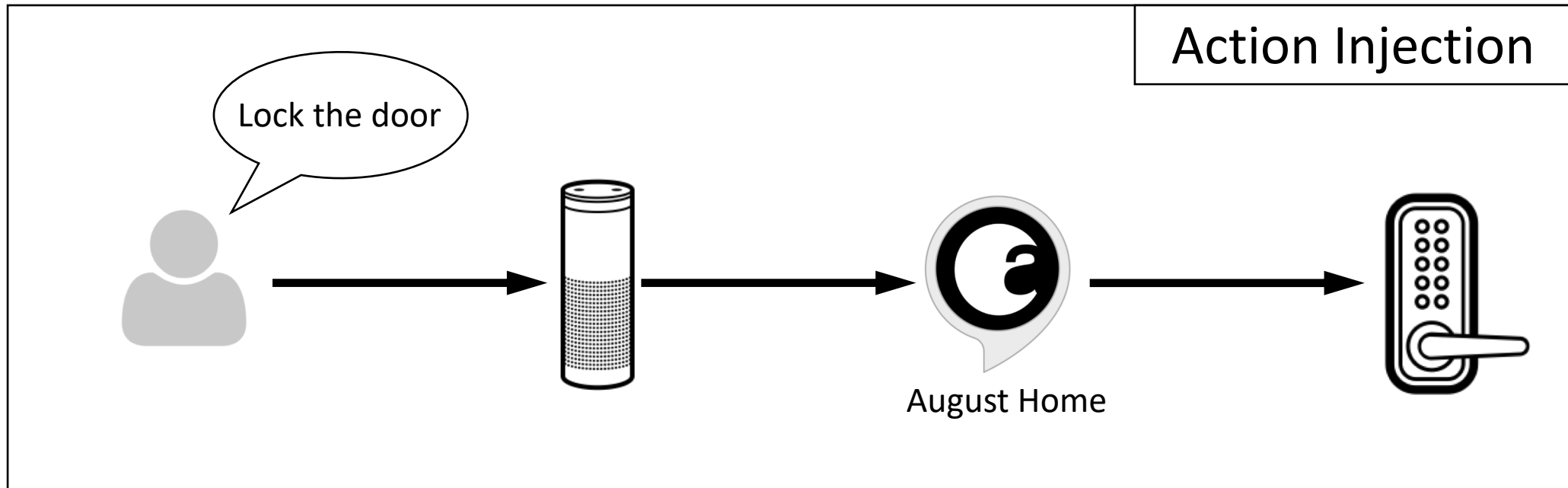
How does VPA work?



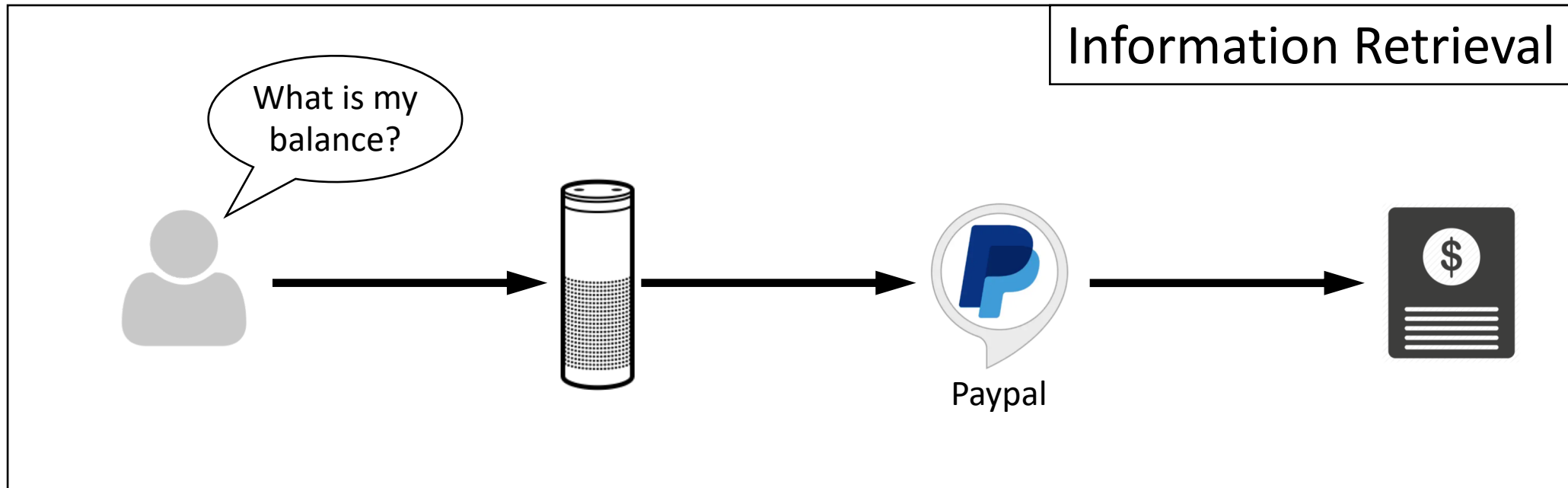
How does VPA work?



Skills With Sensitive Activities



Skills With Sensitive Activities

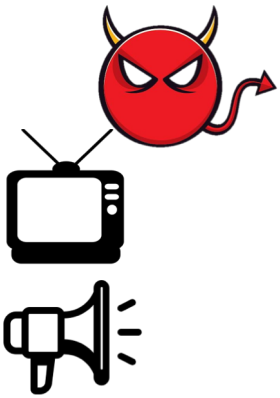


Threat Models

Threat Model #1. Hidden Voice Command Attack

Threat Models

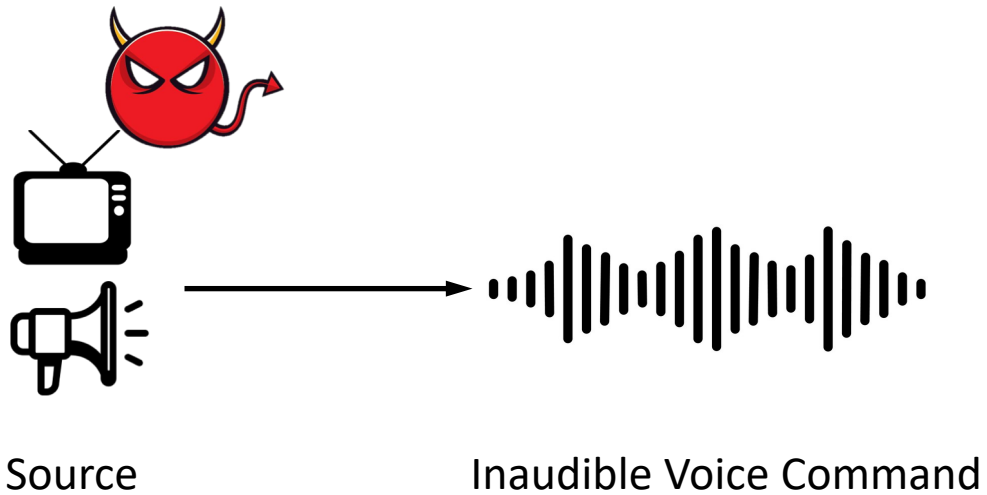
Threat Model #1. Hidden Voice Command Attack



Source

Threat Models

Threat Model #1. Hidden Voice Command Attack



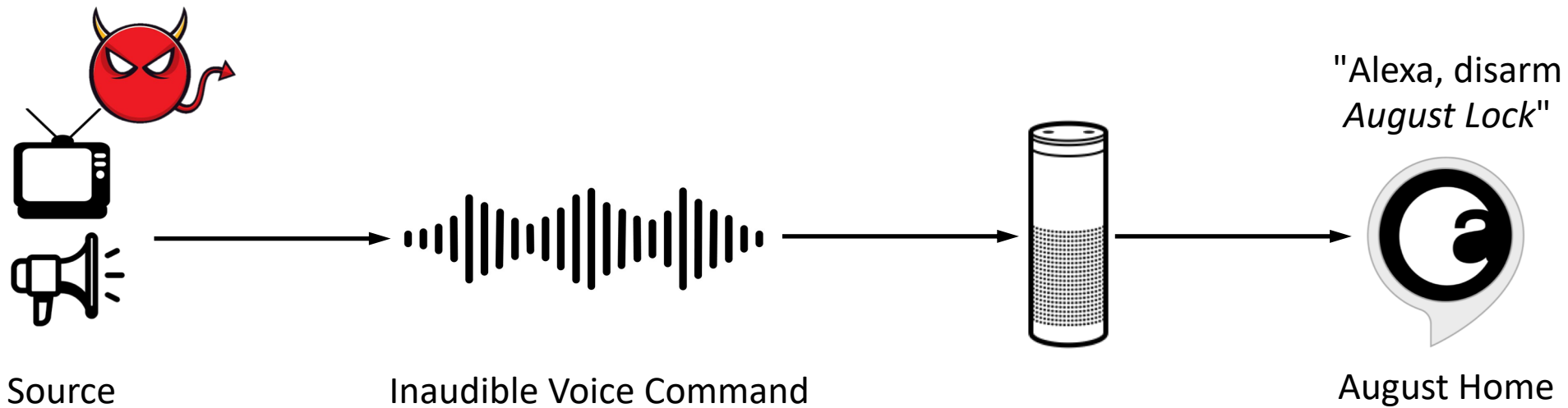
Threat Models

Threat Model #1. Hidden Voice Command Attack



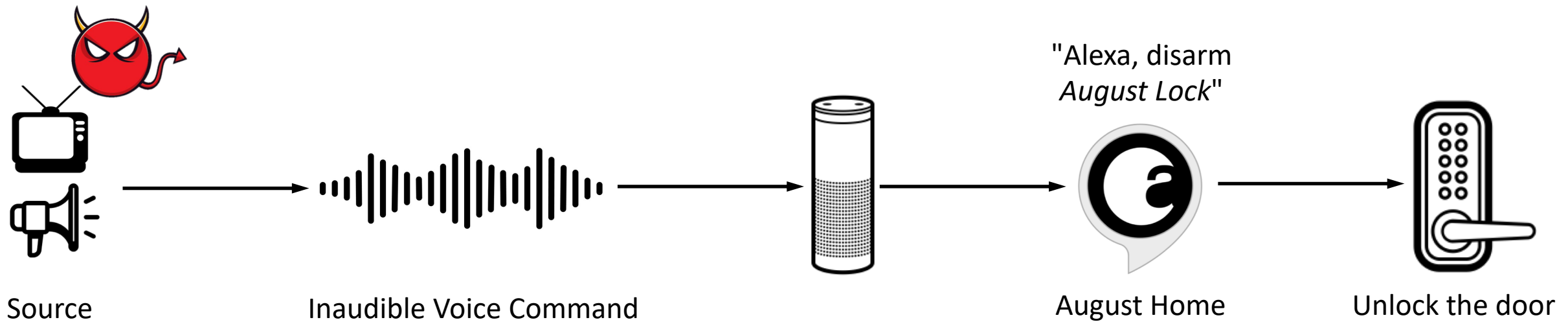
Threat Models

Threat Model #1. Hidden Voice Command Attack



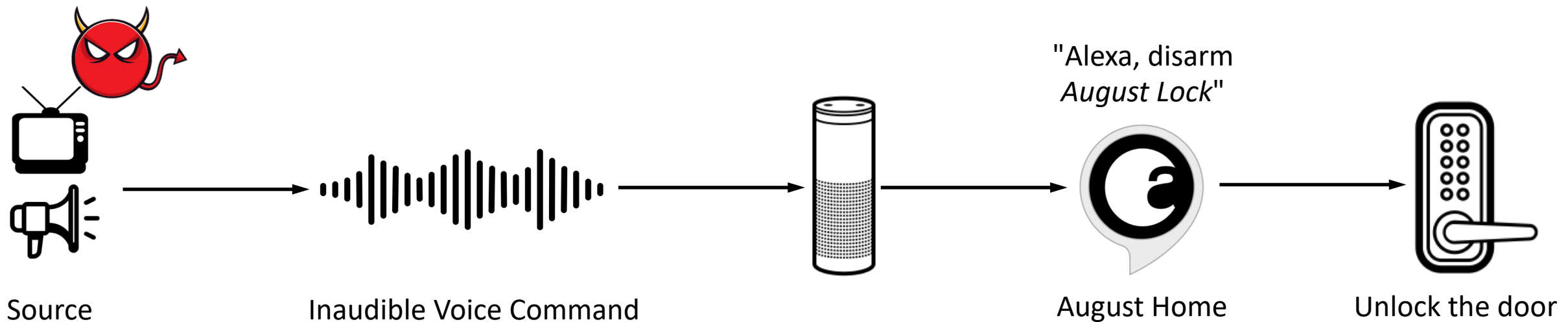
Threat Models

Threat Model #1. Hidden Voice Command Attack



Threat Models

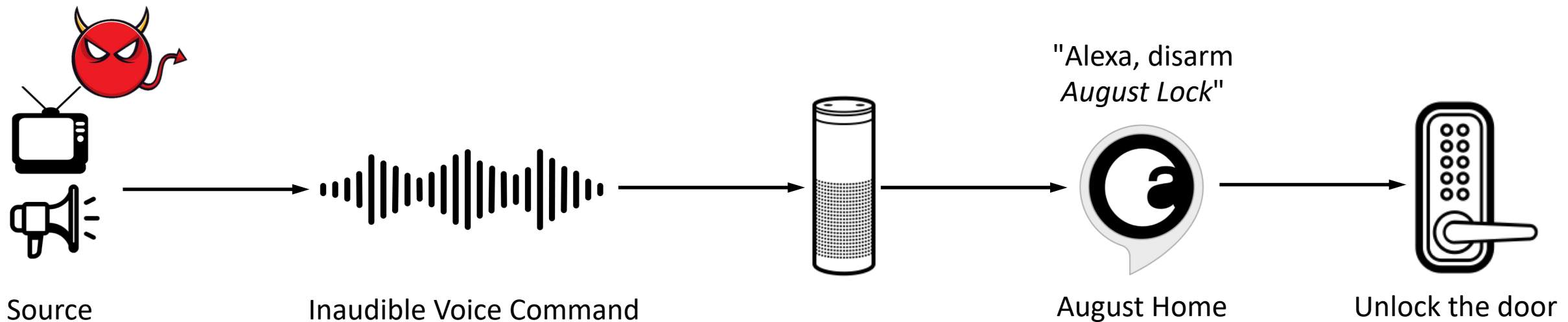
Threat Model #1. Hidden Voice Command Attack



Action Injection voice command to control system

Threat Models

Threat Model #1. Hidden Voice Command Attack



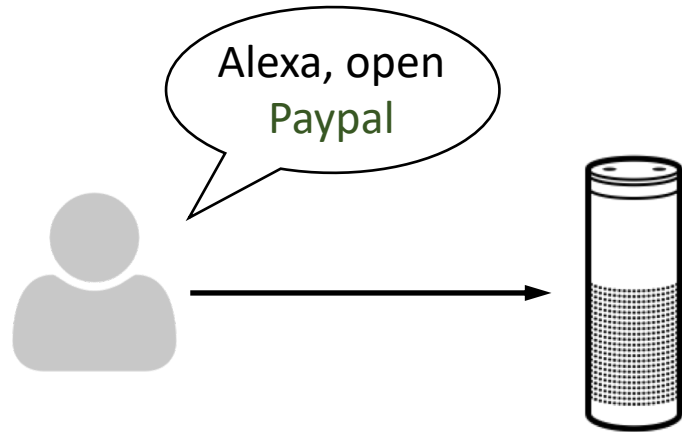
Action Injection voice command to control system

Threat Models

Threat Model #2. Skill Squatting Attack

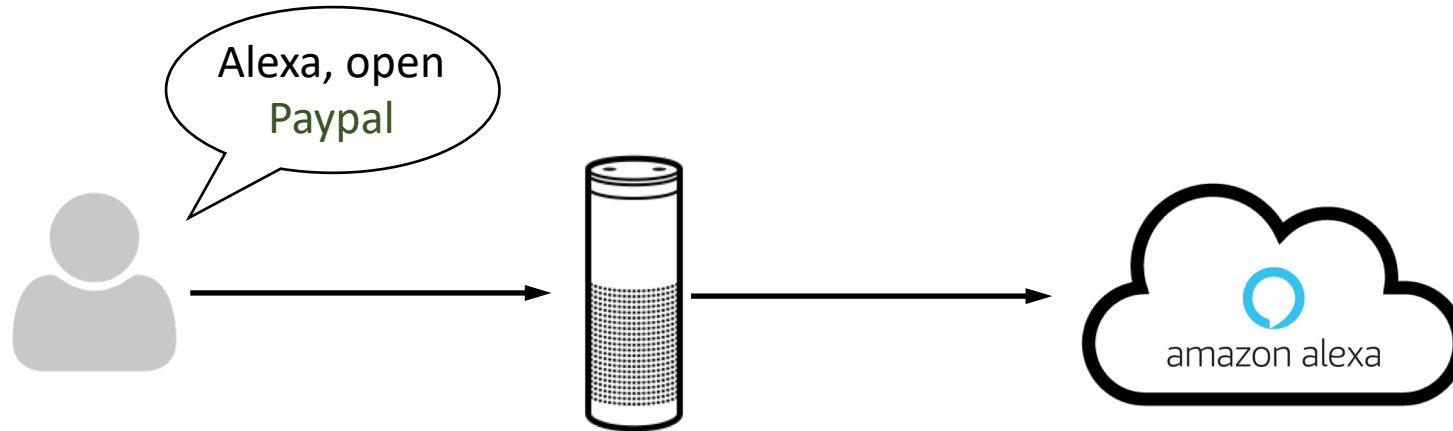
Threat Models

Threat Model #2. Skill Squatting Attack



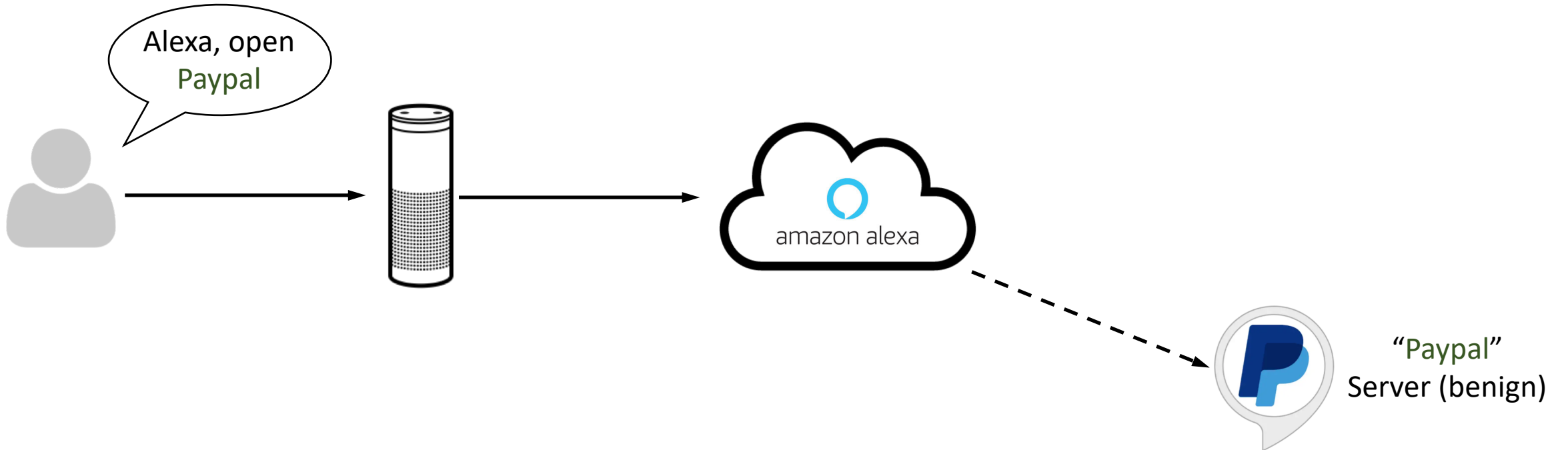
Threat Models

Threat Model #2. Skill Squatting Attack



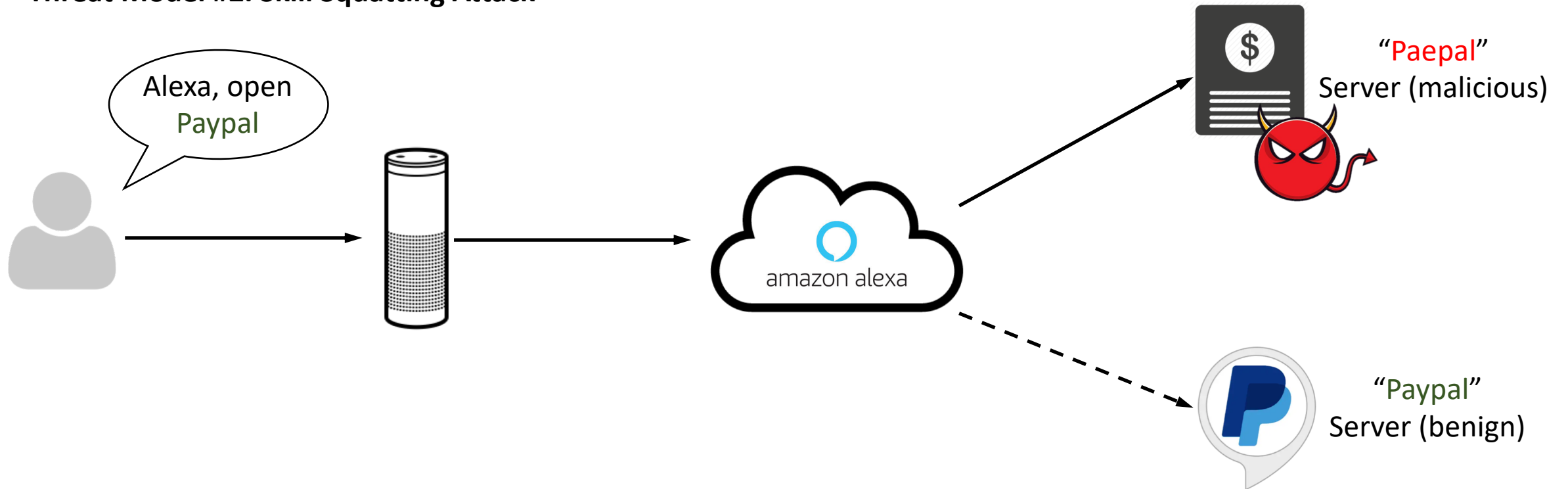
Threat Models

Threat Model #2. Skill Squatting Attack



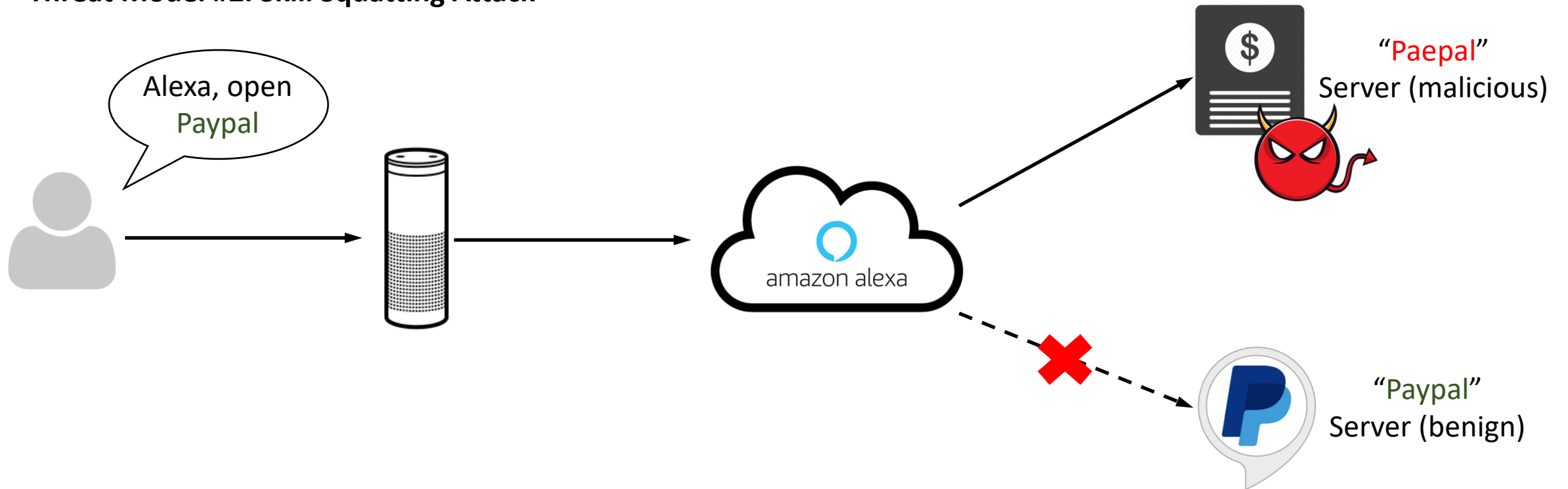
Threat Models

Threat Model #2. Skill Squatting Attack



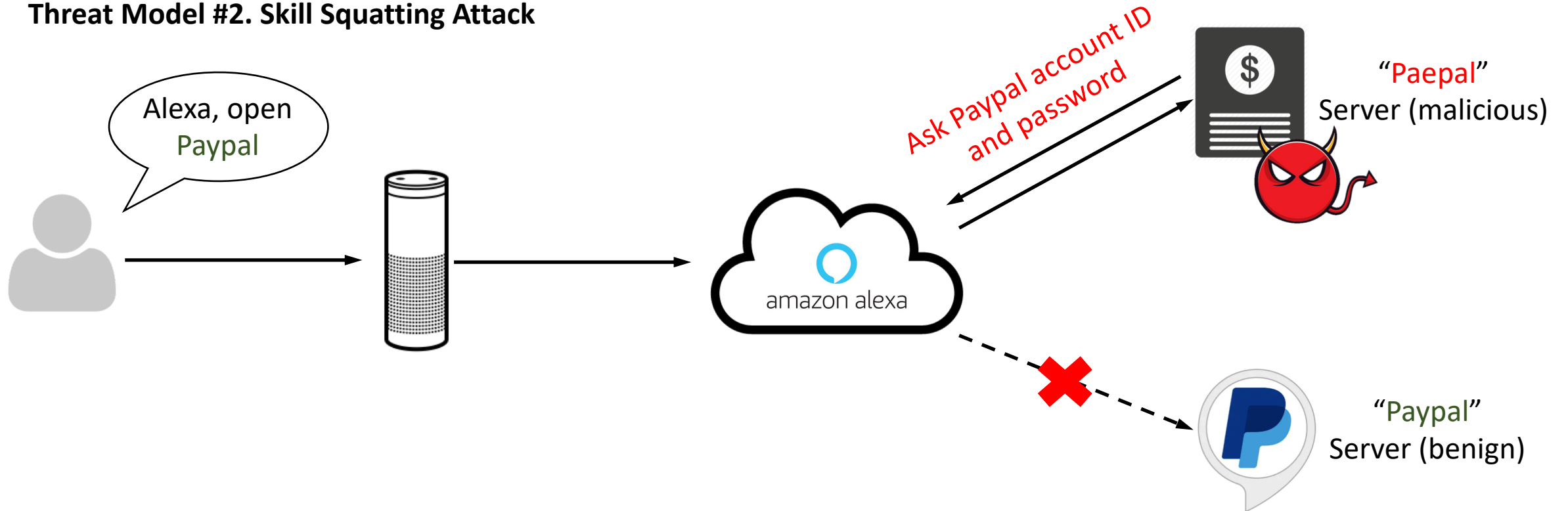
Threat Models

Threat Model #2. Skill Squatting Attack



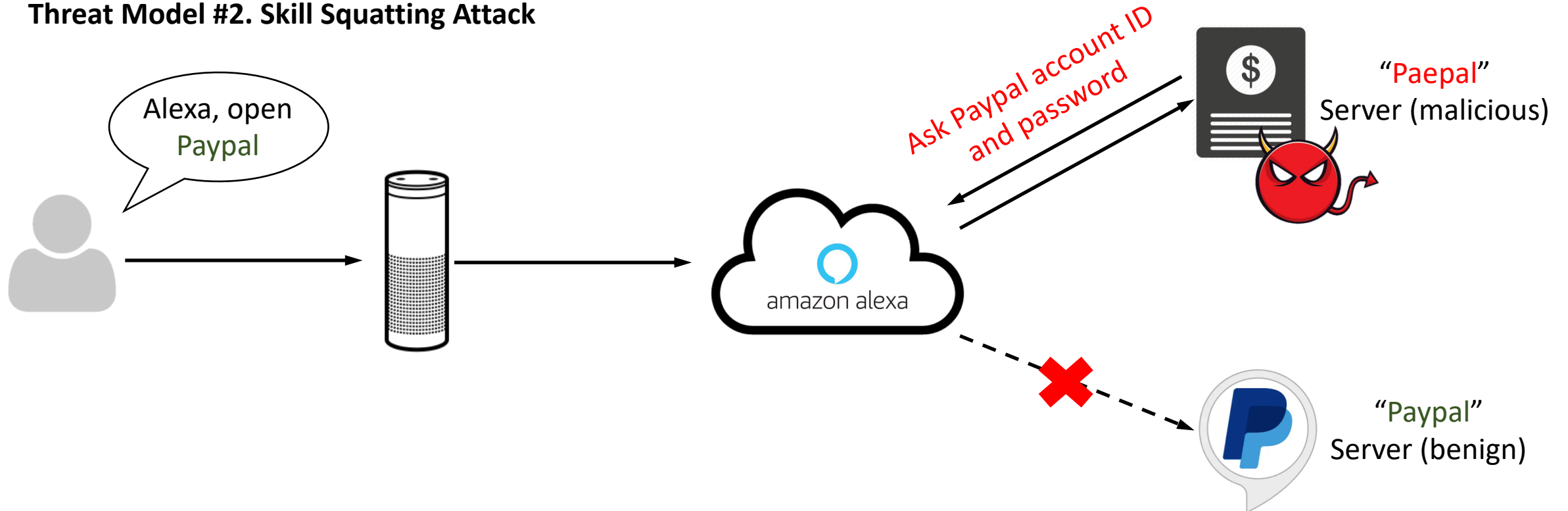
Threat Models

Threat Model #2. Skill Squatting Attack



Threat Models

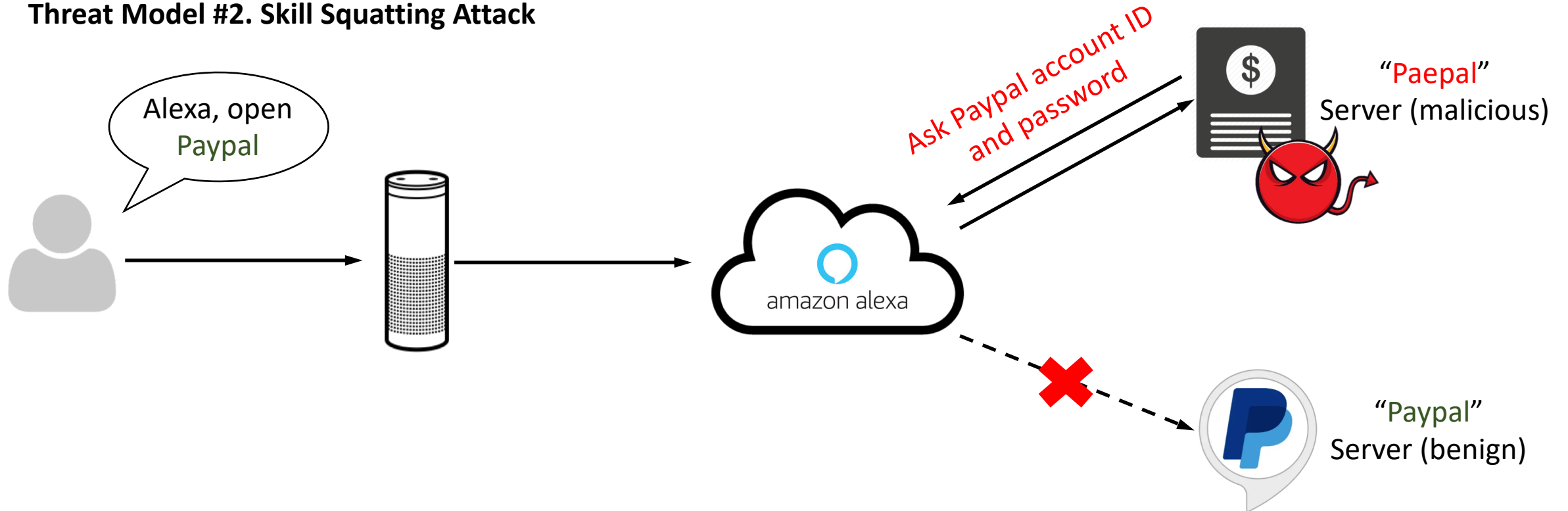
Threat Model #2. Skill Squatting Attack



Information Retrieval voice command to steal sensitive information

Threat Models

Threat Model #2. Skill Squatting Attack



Information Retrieval voice command to steal sensitive information

Security Consequences

Attack #1. Hidden Voice Command Attack

Attack #2. Skill Squatting Attack

Security Consequences

Attack #1. Hidden Voice Command Attack

Attack #2. Skill Squatting Attack

No systematic measurement

Security Consequences

Attack #1. Hidden Voice Command Attack

Attack #2. Skill Squatting Attack

Action Injection command
Information Retrieval command

Examples

Command	Type
"Stop the camera"	Action Injection
"Unlock the door"	Action Injection
"Start my car"	Action Injection

Examples

Command	Type
"Stop the camera"	Action Injection
"Unlock the door"	Action Injection
"Start my car"	Action Injection



Maliciously
control the
system

Examples

Command	Type
“Stop the camera”	Action Injection
“Unlock the door”	Action Injection
“Start my car”	Action Injection
“What is my account balance”	Information Retrieval
“What are the details of the upcoming events”	Information Retrieval
“Where is my order”	Information Retrieval

Examples

Command	Type
"Stop the camera"	Action Injection
"Unlock the door"	Action Injection
"Start my car"	Action Injection
"What is my account balance"	Information Retrieval
"What are the details of the upcoming events"	Information Retrieval
"Where is my order"	Information Retrieval

Leak sensitive information

Research Question

- Understanding the attack surface of VPA

Research Question

- Understanding the attack surface of VPA
- Identify the category with the most sensitive voice commands

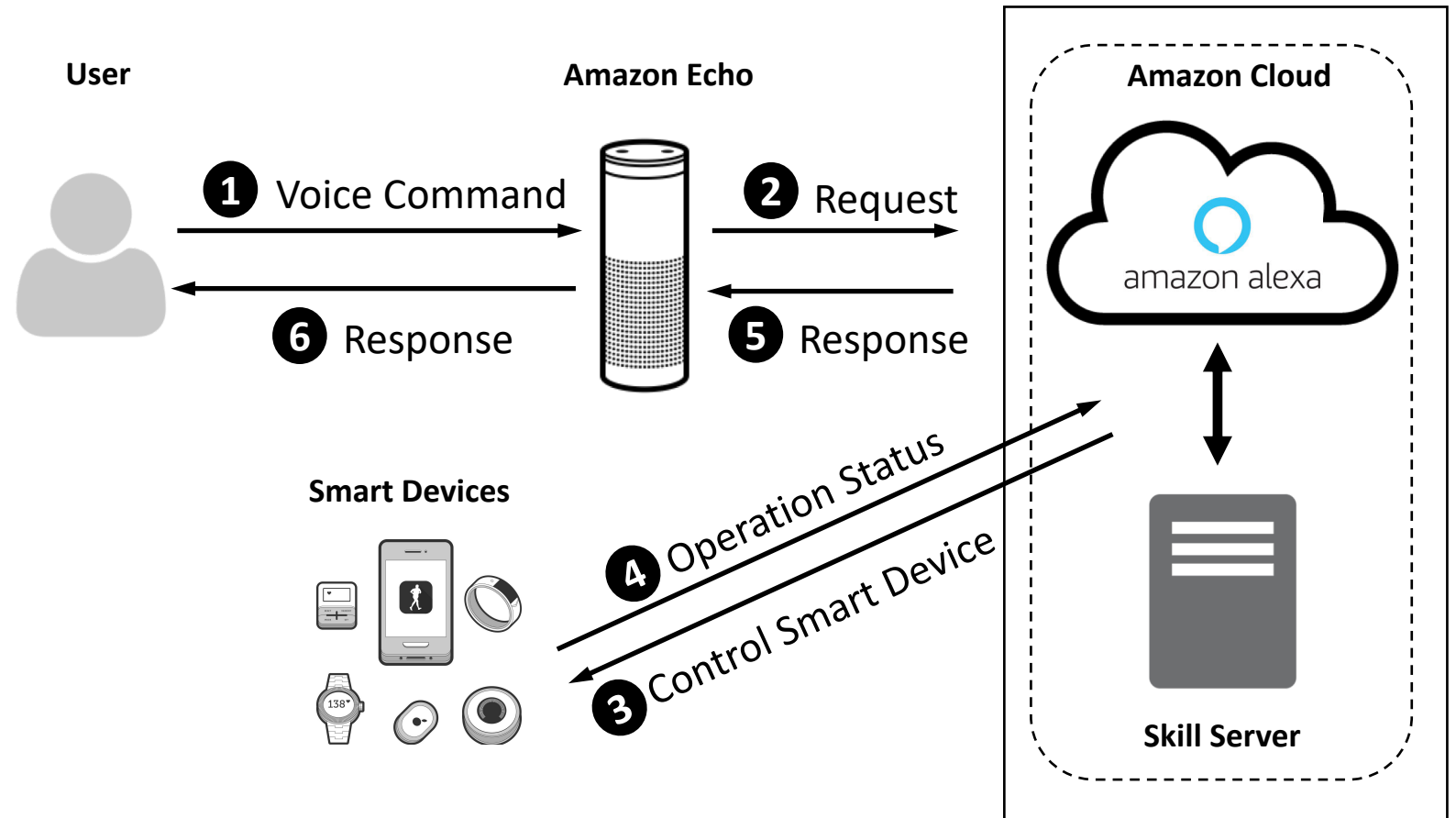
Research Question

- Understanding the attack surface of VPA
- Identify the category with the most sensitive voice commands
- Large scale measurement of the sensitive skills

Outline

1. Introduction & Problem definition
2. Methodology
3. Measurement analysis

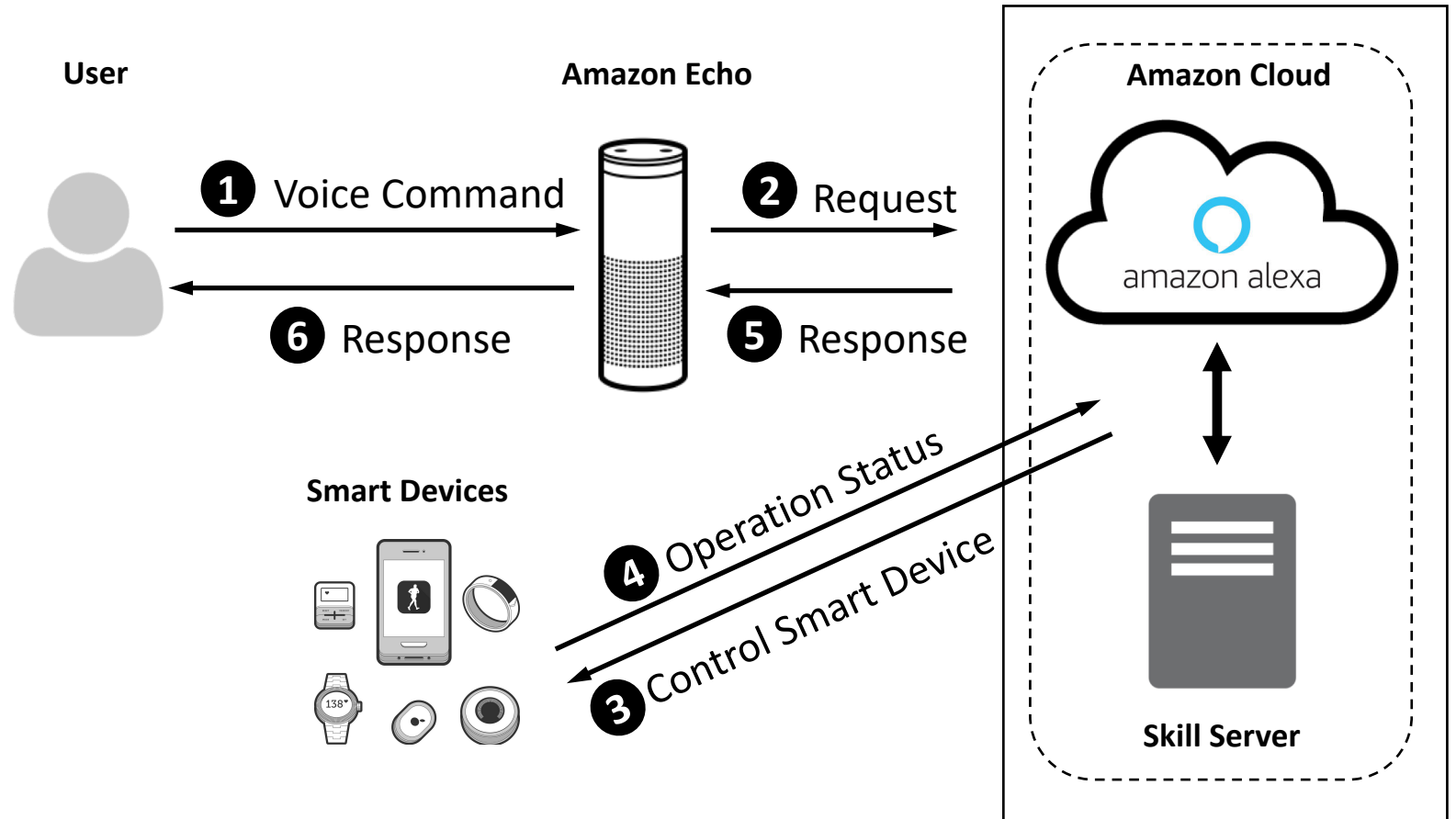
Challenges



Challenges

Four main key factors-

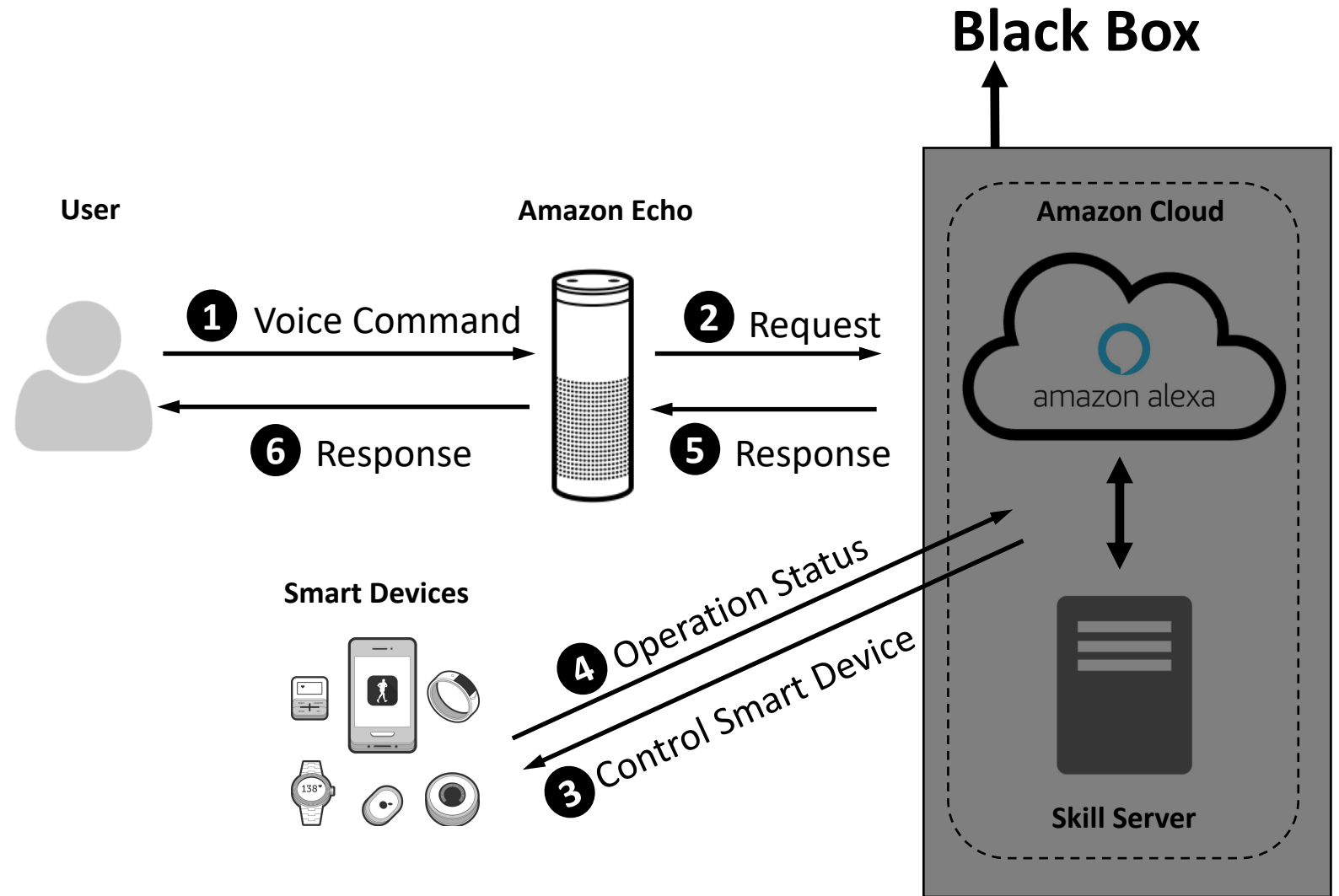
1. No executable files



Challenges

Four main key factors-

1. No executable files



Challenges

Four main key factors-

1. No executable files
2. **Short length of voice commands**

Challenges

Four main key factors-

1. No executable files
- 2. Short length of voice commands**

Skill	Command
Blink SmartHome	“Stop the camera”
Schlage Sense	“Lock the door”
FordPass	“Start my car”

Challenges

Four main key factors-

1. No executable files
2. Short length of voice commands
3. **Lacking of labeled data**

Challenges

Four main key factors-

1. No executable files
2. Short length of voice commands
- 3. Lacking of labeled data**

	Alexa US 2019	Google 2019
Skill	31,413	3,148
Command	80,129	9,096

Challenges

Four main key factors-

1. No executable files
2. Short length of voice commands
3. Lacking of labeled data
4. **Sensitivity being subjective**

Challenges

Four main key factors-

1. No executable files
2. Short length of voice commands
3. Lacking of labeled data
4. **Sensitivity being subjective**

“Set the room temperature to 25 degree”

	Sensitivity	Reason
User1	Nonsensitive	Temperature is within the comfort zone.
User2	Sensitive	Wrong temperature setting will make it worse.
User3	Sensitive	It will trigger other devices dependent on the temperature.

Dataset

	Alexa US 2019	Alexa UK 2019	Alexa US 2018	Google 2019	Google 2018
Skill	31,413	20,213	26,331	3,148	1,665
Command	80,129	51,922	66,148	9,096	4,548

Dataset

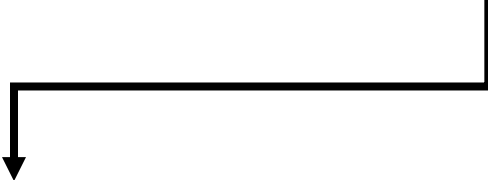
	Alexa US 2019	Alexa UK 2019	Alexa US 2018	Google 2019	Google 2018
Skill	31,413	20,213	26,331	3,148	1,665
Command	80,129	51,922	66,148	9,096	4,548

↓

Ground-Truth Voice Commands			
Capability	Action Injection	647	2,285
	Information Retrieval	1,638	
Sensitivity	Sensitive	515	1,927
	Nonsensitive	1,412	

Dataset

	Alexa US 2019	Alexa UK 2019	Alexa US 2018	Google 2019	Google 2018
Skill	31,413	20,213	26,331	3,148	1,665
Command	80,129	51,922	66,148	9,096	4,548



Ground-Truth Voice Commands			
Capability	Action Injection	49	200
	Information Retrieval	151	
Sensitivity	Sensitive	57	200
	Nonsensitive	143	

Data Process

- Removing commands for enabling skills

Data Process

- Removing commands for enabling skills

**Enabling Voice
Commands**

Open Amex

Start Song Quiz

Data Process

- Removing commands for enabling skills
- Extracting action

Data Process

- Removing commands for enabling skills
- Extracting action

Ask mastermind to text Kelly Miller

Data Process

- Removing commands for enabling skills
- Extracting action

Ask mastermind to text Kelly Miller



Action

Data Process

- Removing commands for enabling skills
- Extracting action
- Data format

Data Process

- Removing commands for enabling skills
- Extracting action
- Data format
 - Remove punctuation
 - Convert Lowercase
 - Numeric value to alphabetic value (1 to “one”)

Data Process

- Removing commands for enabling skills
- Extracting action
- Data format
- Category context

Data Process

- Removing commands for enabling skills
- Extracting action
- Data format
- Category context

Unlock the door

Data Process

- Removing commands for enabling skills
- Extracting action
- Data format
- Category context

Unlock the door **smarthome**



Category of the
corresponding skill

Data Process

- Removing commands for enabling skills
- Extracting action
- Data format
- Category context
- Remove redundancy

Data Process

- Removing commands for enabling skills
- Extracting action
- Data format
- Category context
- Remove redundancy

Ask Doctor Who Facts for a fact

Ask Unofficial Stargate Facts for a fact

Data Process

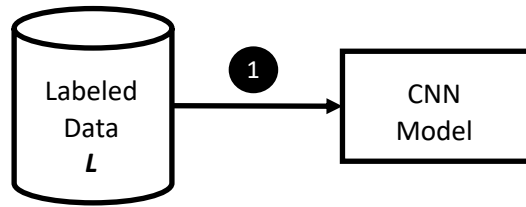
- Removing commands for enabling skills
- Extracting action
- Data format
- Category context
- Remove redundancy

Ask Doctor Who Facts for a fact

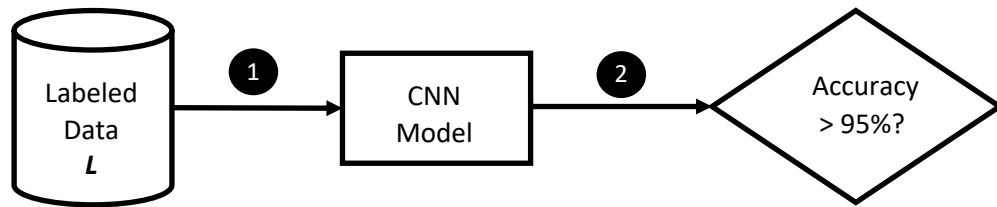
Ask Unofficial Stargate Facts for a fact

} **for a fact**

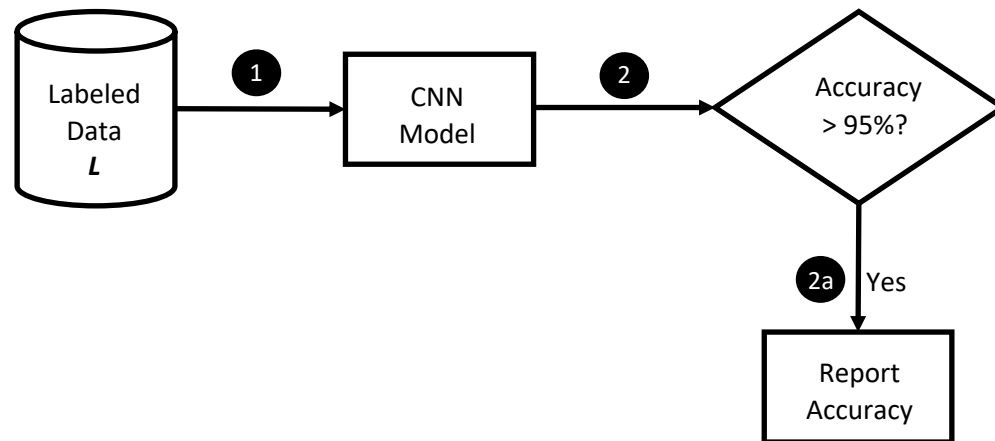
Capability Analysis Model



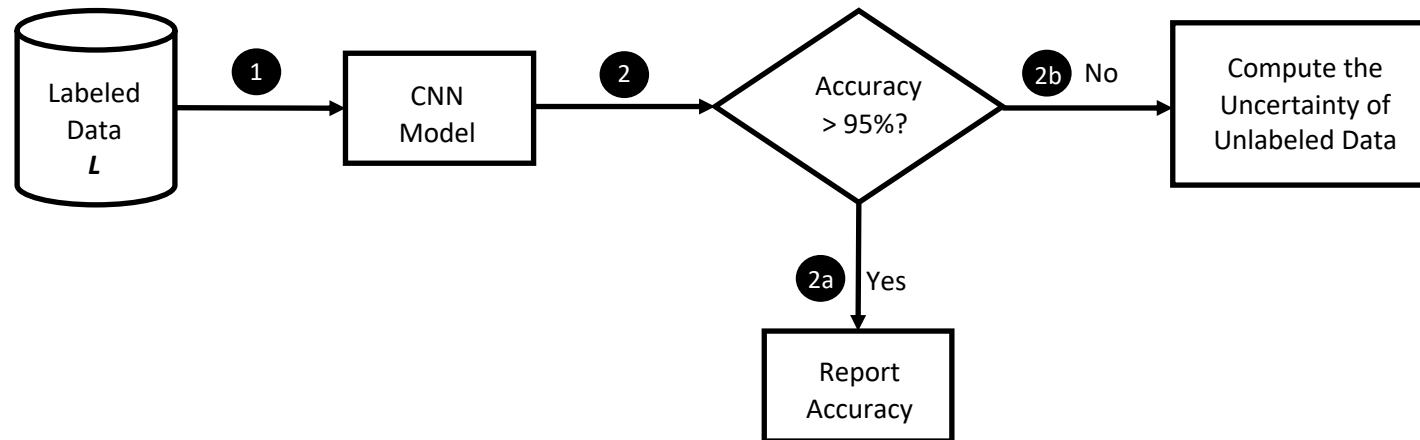
Capability Analysis Model



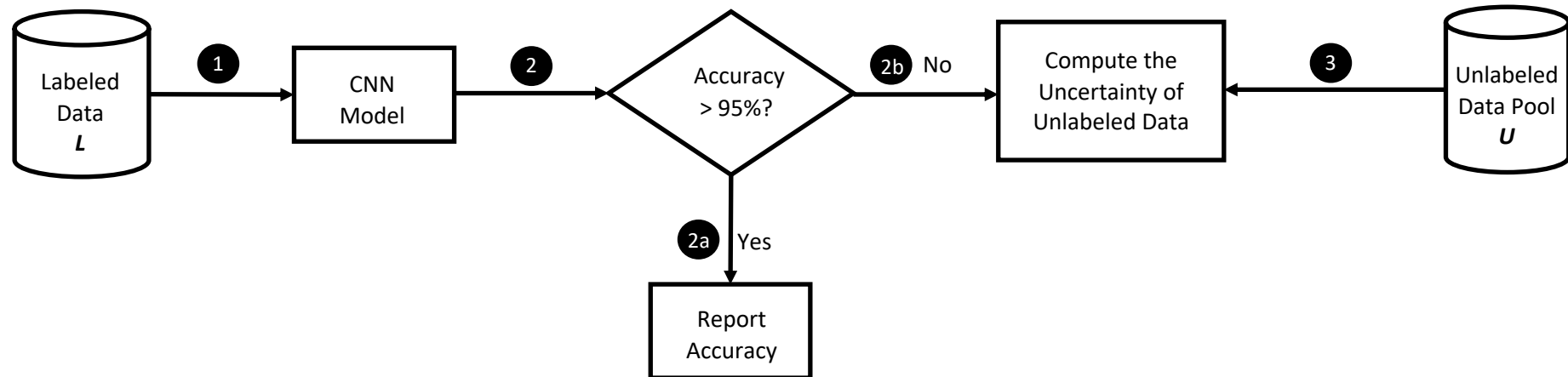
Capability Analysis Model



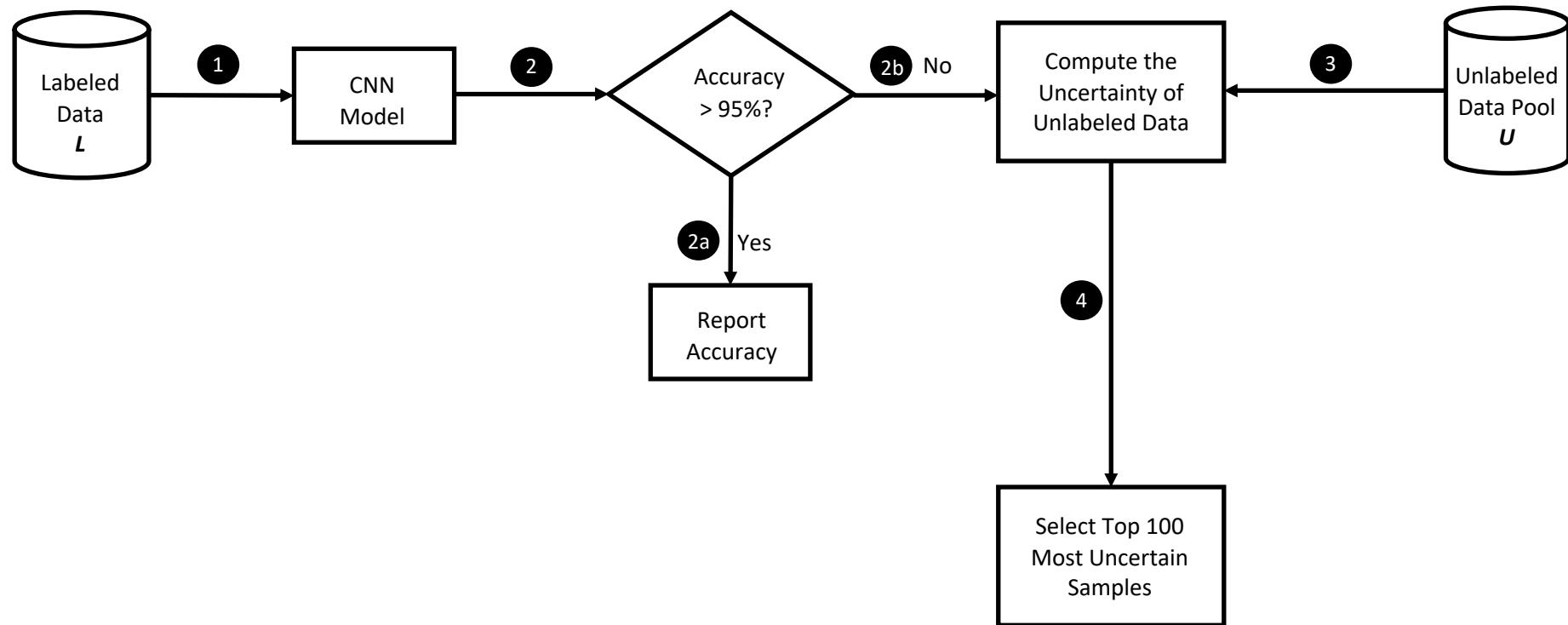
Capability Analysis Model



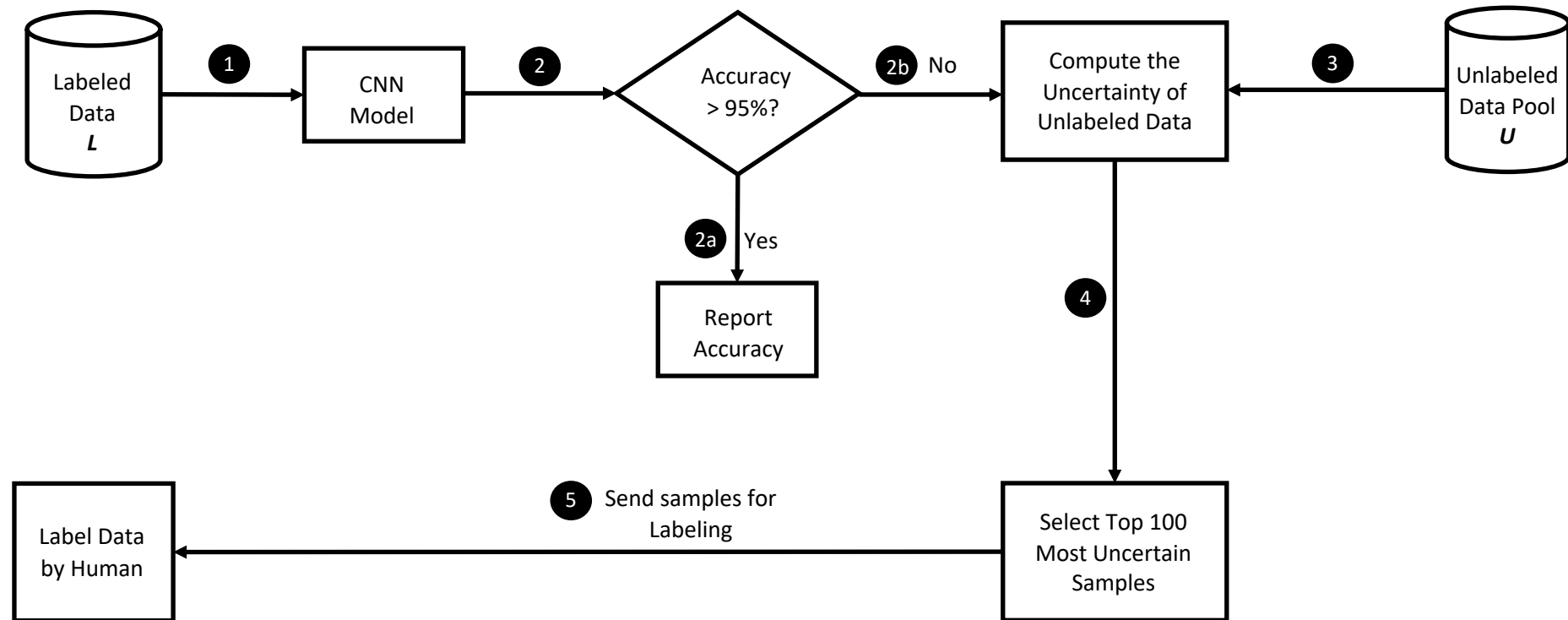
Capability Analysis Model



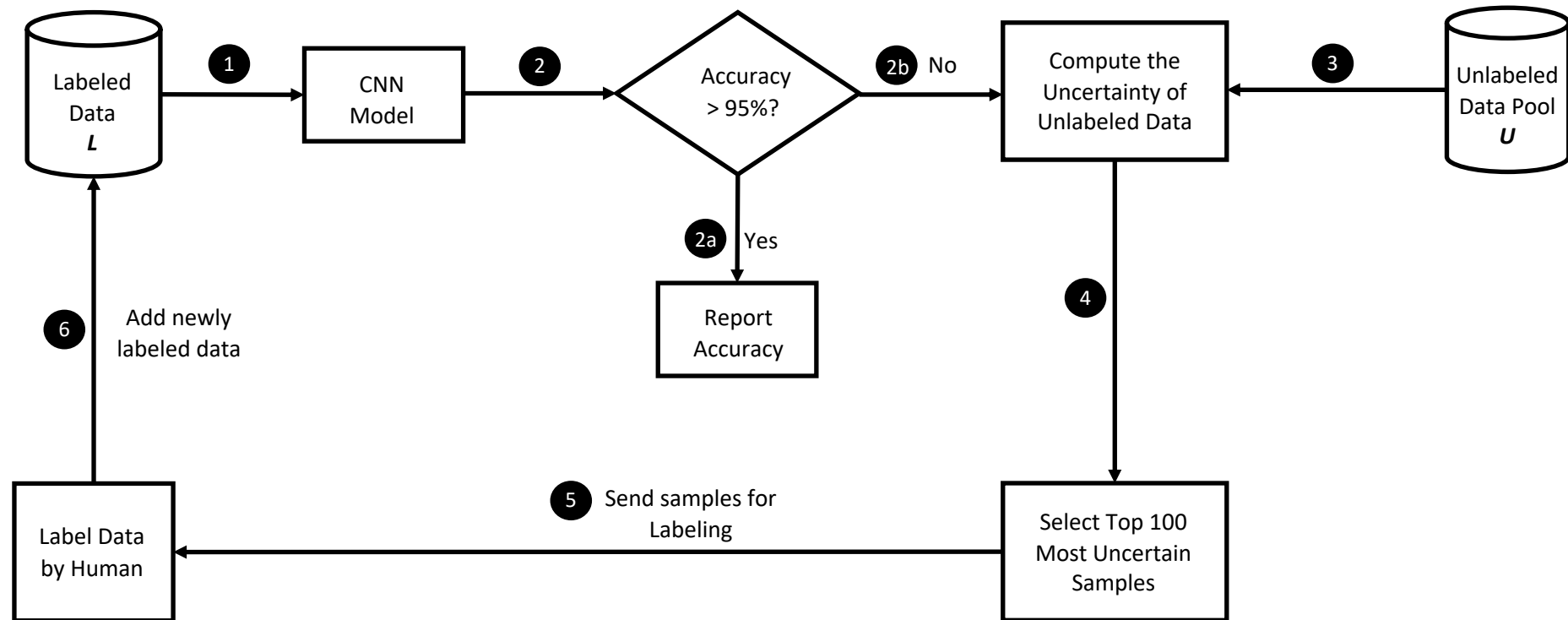
Capability Analysis Model



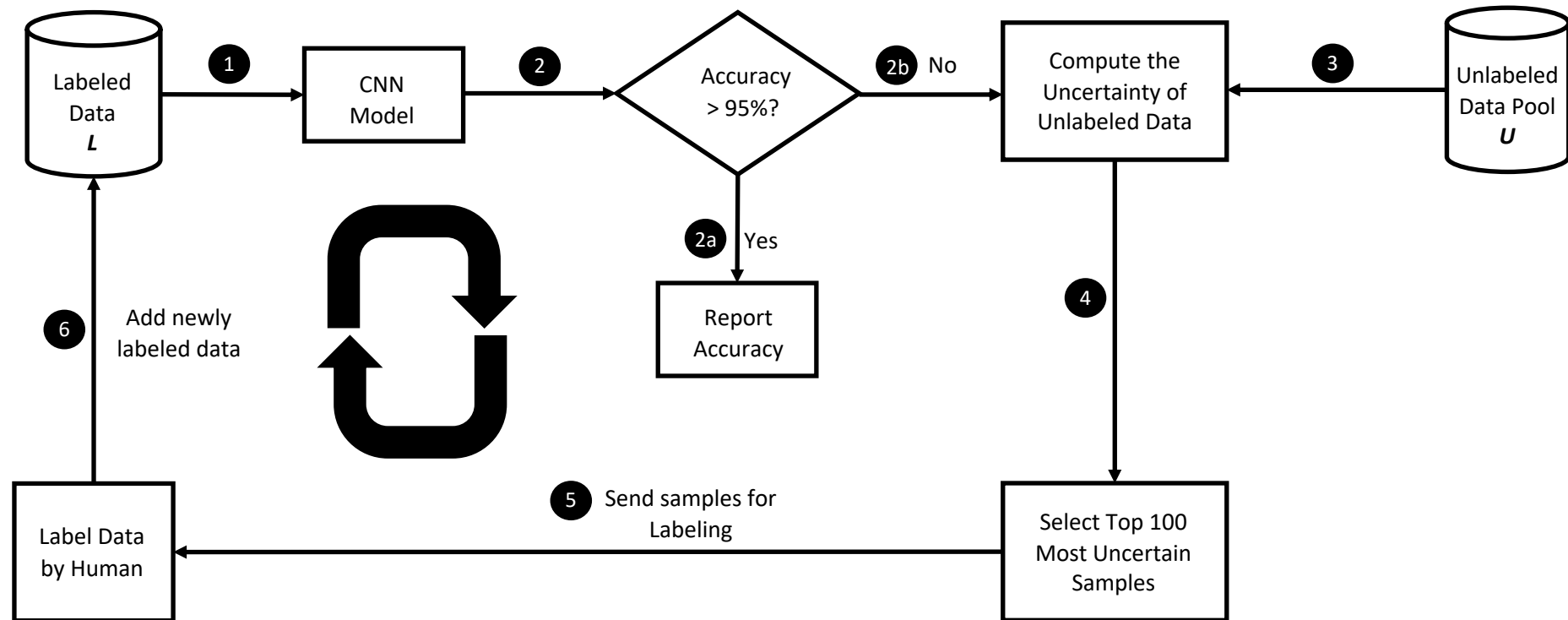
Capability Analysis Model



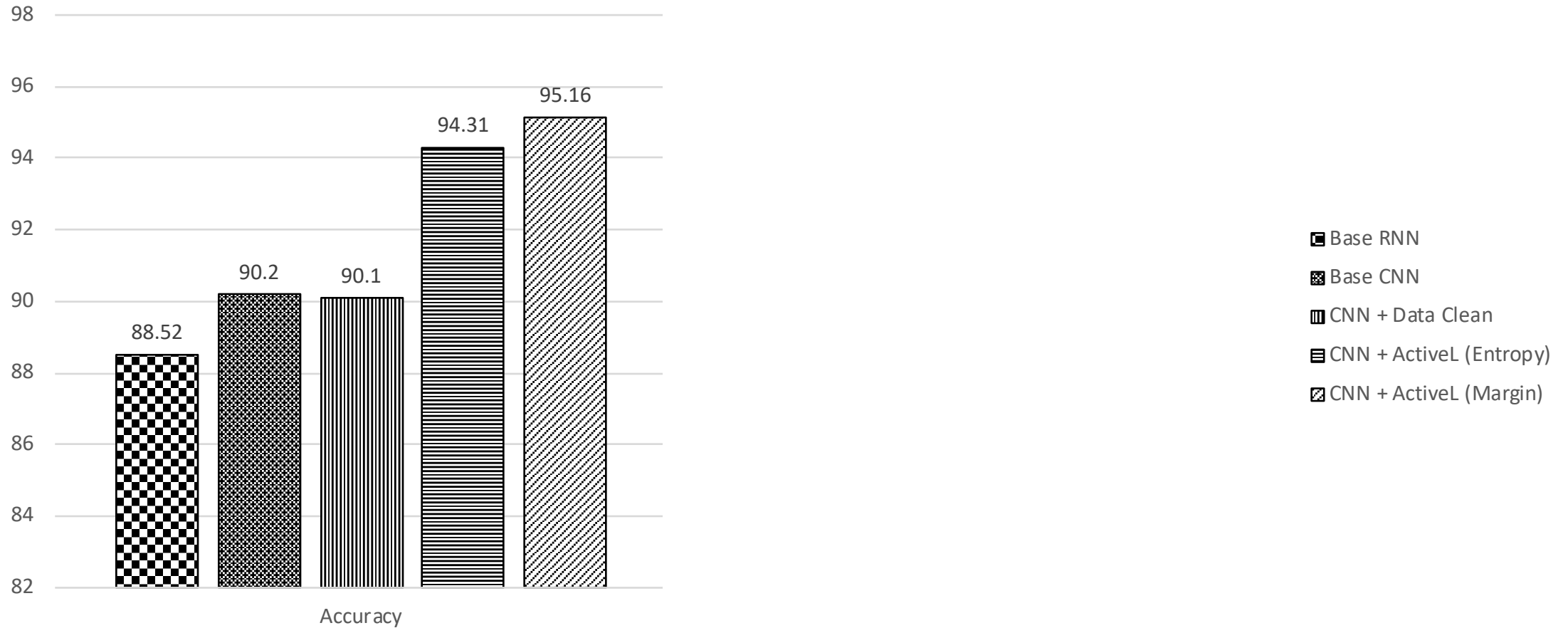
Capability Analysis Model



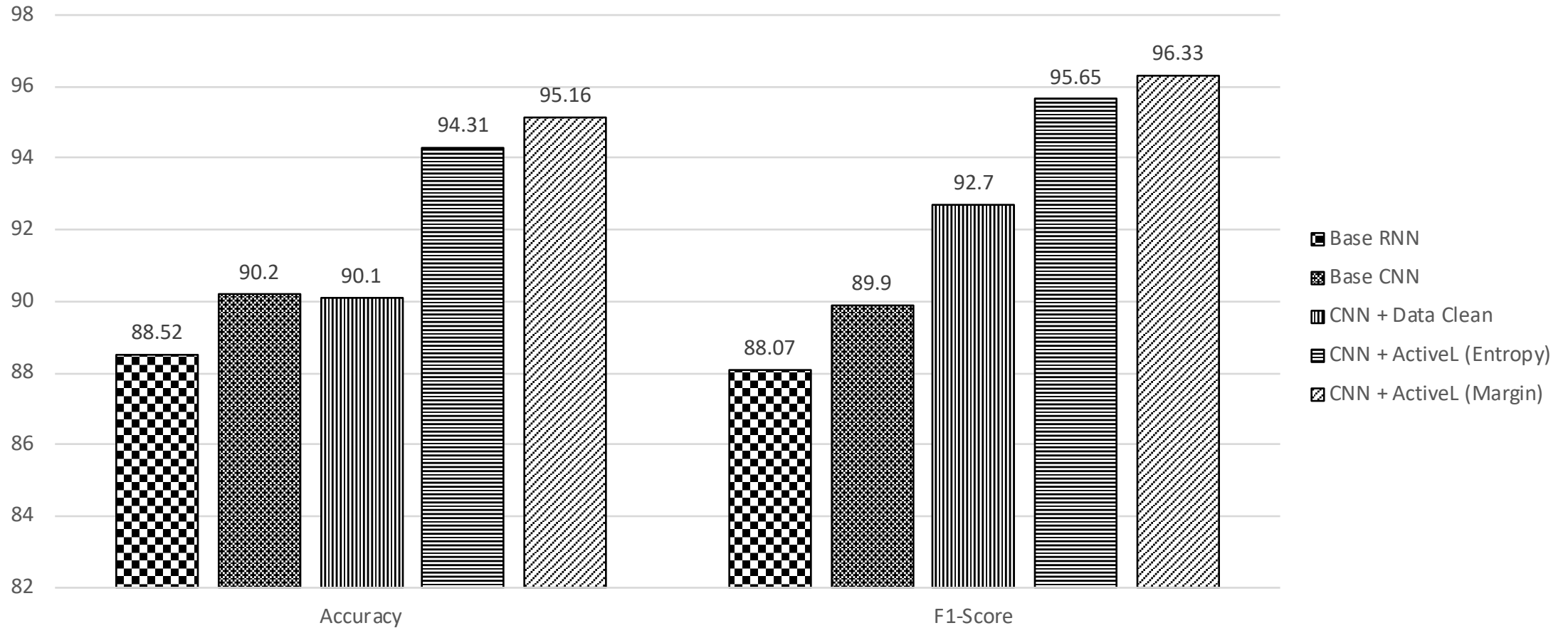
Capability Analysis Model



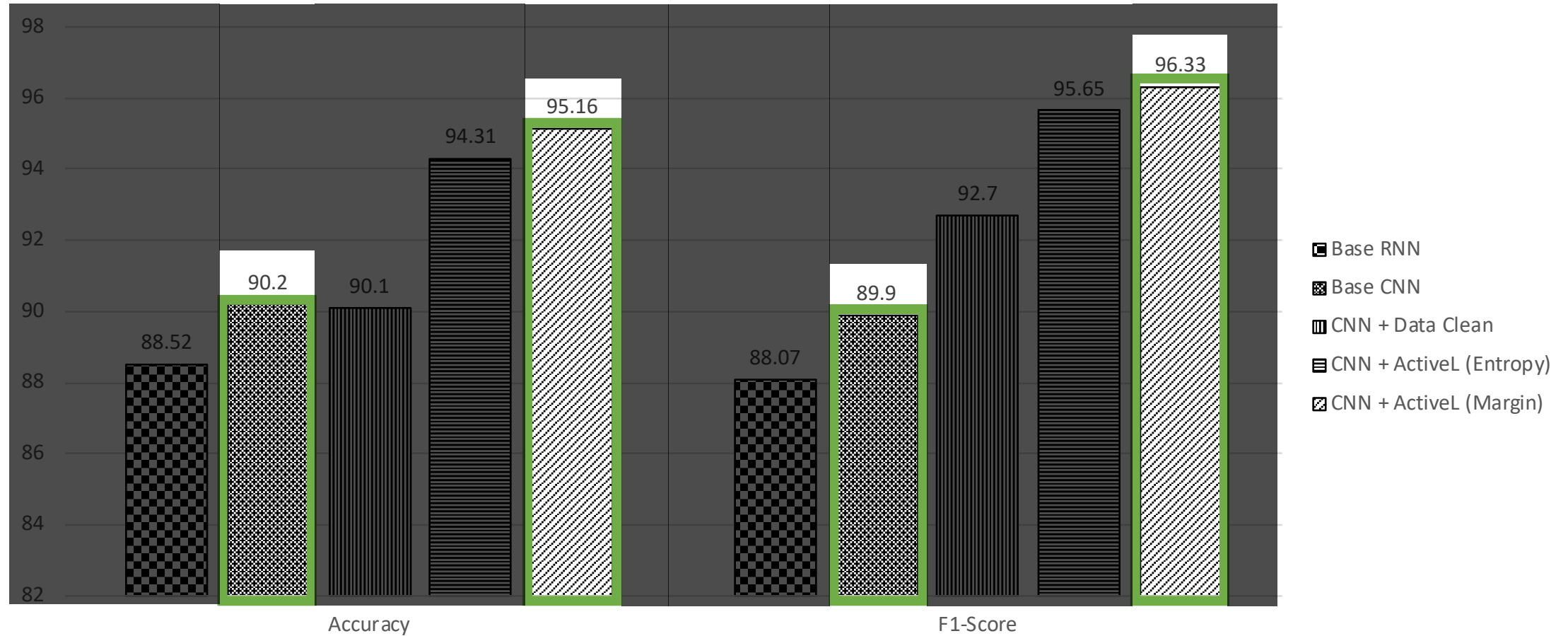
Evaluation of Capability Analysis Model



Evaluation of Capability Analysis Model



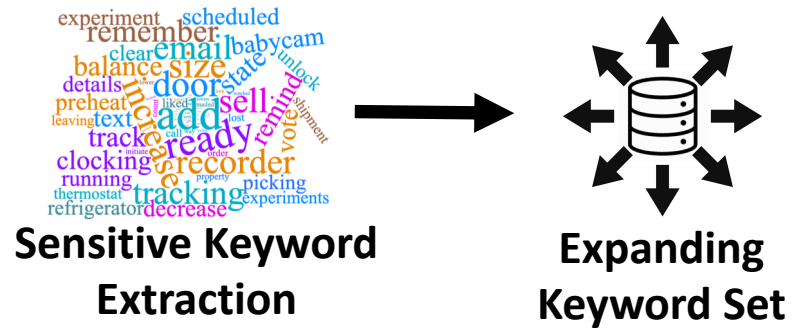
Evaluation of Capability Analysis Model



Sensitivity Analysis Model



Sensitivity Analysis Model



Survey Design

Surveying with 106 keywords

Survey Design

Surveying with 106 keywords

Recruiting using Mechanical Turk

Survey Design

Surveying with 106 keywords

Recruiting using Mechanical Turk

Highlighting functional keywords

Survey Design

Surveying with 106 keywords

Recruiting using Mechanical Turk

Highlighting functional keywords

Alexa, tell Virtual Keypad to arm my system away

Survey Design

Surveying with 106 keywords

Recruiting using Mechanical Turk

Highlighting functional keywords

Alexa, tell Virtual Keypad to “**arm**” my system away

Survey Design

Surveying with 106 keywords

Recruiting using Mechanical Turk

Highlighting functional keywords

Voting scale

Survey Design

Surveying with 106 keywords

Recruiting using Mechanical Turk

Highlighting functional keywords

Voting scale

Not sensitive

Less sensitive

Neutral

Sensitive

Most sensitive

Survey Design

Surveying with 106 keywords

Recruiting using Mechanical Turk

Highlighting functional keywords

Voting scale

Ranking

Survey Design

Surveying with 106 keywords

Recruiting using Mechanical Turk

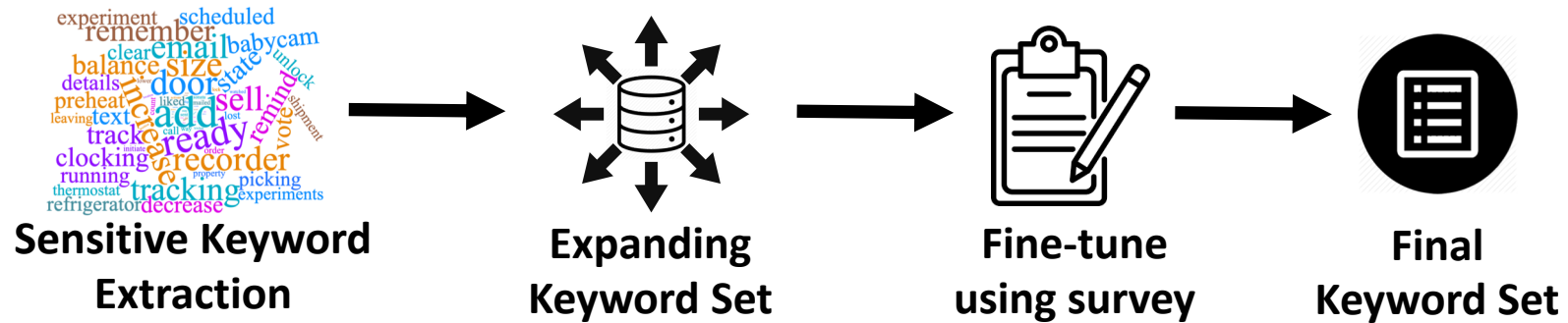
Highlighting functional keywords

Voting scale

Ranking

- 1 If $(\# \text{most sensitive} + \# \text{sensitive}) > (\# \text{not sensitive} + \# \text{less sensitive} + \# \text{neutral})$
- 0 Otherwise

Sensitivity Analysis Model



Outline

1. Introduction & Problem definition
2. Methodology
3. **Measurement analysis**

Top Sensitive Skills Category

Alexa US 2019

Category	Sensitive Voice Command		Nonsensitive Voice Command	
	Action Injection	Information Retrieval	Action Injection	Information Retrieval
Smart Home	8,095 (75.4%)	2,218 (20.66%)	157 (1.46%)	265 (2.47%)
Connected Car	118 (29.5%)	207 (51.75%)	13 (3.25%)	62 (15.5%)
Productivity	84 (2.42%)	64 (1.84%)	990 (28.51%)	2,335 (67.23 %)
Lifestyle	45 (0.96%)	60 (1.28%)	1,648 (35.03%)	2,952 (62.74%)
Business & Finance	35(0.64%)	146 (2.68%)	1,422 (26.12%)	3,842 (70.56%)
Health & Fitness	6 (0.44%)	19 (1.4%)	432 (31.83%)	900 (66.32%)
Kids	-	-	592 (35.45%)	1,078 (64.55%)

Top Sensitive Skills Category

Alexa US 2019

Category	Sensitive Voice Command		Nonsensitive Voice Command	
	Action Injection	Information Retrieval	Action Injection	Information Retrieval
Smart Home	8,095 (75.4%)	2,218 (20.66%)	157 (1.46%)	265 (2.47%)
Connected Car	118 (29.5%)	207 (51.75%)	13 (3.25%)	62 (15.5%)
Productivity	84 (2.42%)	64 (1.84%)	990 (28.51%)	2,335 (67.23 %)
Lifestyle	45 (0.36%)	50 (1.28%)	1,648 (35.03%)	2,552 (52.74%)
Business & Finance	35 (0.54%)	146 (1.68%)	1,432 (25.42%)	3,842 (70.56%)
Health & Fitness	6 (0.44%)	19 (1.4%)	432 (31.83%)	900 (66.32%)
Kids	-	-	592 (35.45%)	1,078 (64.55%)

Smart home and Connected Car contain the most sensitive voice commands

Top Sensitive Skills Category

Alexa US 2019

Category	Sensitive Voice Command		Nonsensitive Voice Command	
	Action Injection	Information Retrieval	Action Injection	Information Retrieval
Smart Home	8,095 (75.4%)	2,218 (20.66%)	157 (1.46%)	265 (2.47%)
Smart TV	1,000 (1.5%)	20 (0.05%)	13 (0.25%)	57 (1.1%)
Productivity	84 (2.42%)	64 (1.84%)	990 (28.51%)	2,335 (67.23%)
Lifestyle	45 (0.96%)	60 (1.28%)	1,648 (35.03%)	2,952 (62.74%)
Business & Finance	35(0.64%)	146 (2.68%)	1,422 (26.12%)	3,842 (70.56%)
Health & Fitness	6 (0.44%)	19 (1.4%)	432 (31.83%)	900 (66.32%)
Kids	-	-	592 (35.45%)	1,078 (64.55%)

Intuitively Health and Kids category should contain more sensitive voice commands, but they don't

Top Sensitive Skills Category

Google 2019

Category	Sensitive Voice Command		Nonsensitive Voice Command	
	Action Injection	Information Retrieval	Action Injection	Information Retrieval
Home Control	642 (24.51%)	281 (10.73%)	1,092 (41.7%)	604 (23.06%)
Productivity	10 (3.45%)	14 (4.83%)	119 (41.03%)	147 (50.69%)
Shopping	6 (1.22%)	20 (4.07)	135 (27.44%)	331 (67.28%)
Travel & Transportation	1 (0.23%)	6 (1.4%)	96 (22.33%)	327 (76.05%)
Business & Finance	-	11 (3.01%)	29 (7.95%)	325 (89.04%)
Health & Fitness	4 (0.94%)	-	163 (38.44%)	257 (60.61%)
Kids	-	-	14 (19.18%)	59 (80.82%)

Top Sensitive Skills Category

Google 2019

Category	Sensitive Voice Command		Nonsensitive Voice Command	
	Action Injection	Information Retrieval	Action Injection	Information Retrieval
Home Control	642 (24.51%)	281 (10.73%)	1,092 (41.7%)	604 (23.06%)
Productivity	10 (3.45%)	14 (4.83%)	119 (41.03%)	147 (50.69%)
Shopping	6 (1.22%)	20 (4.07)	135 (27.44%)	331 (67.28%)
Transportation	1 (0.23%)	6 (1.46%)	90 (22.52%)	327 (79.51%)
Business & Finance	-	11 (5.55%)	11 (7.52%)	325 (89.04%)
Health & Fitness	4 (0.94%)	-	163 (38.44%)	257 (60.61%)
Kids	-	-	14 (19.18%)	59 (80.82%)

Home Control and Productivity contain the most sensitive voice commands

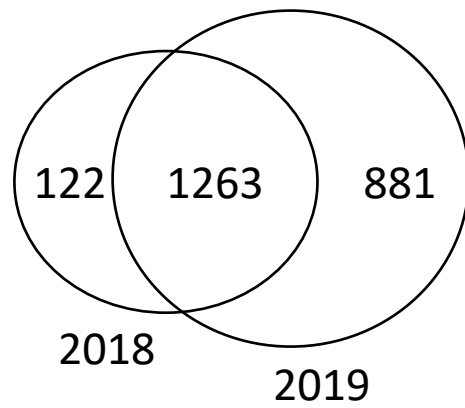
Top Sensitive Skills Category

Google 2019

Category	Sensitive Voice Command		Nonsensitive Voice Command	
	Action Injection	Information Retrieval	Action Injection	Information Retrieval
Home Control	642 (24.51%)	281 (10.73%)	1,092 (41.7%)	604 (23.06%)
Productivity	10 (3.45%)	14 (4.83%)	119 (41.03%)	147 (50.69%)
Shopping	1 (0.22%)	20 (4.07%)	35 (27.45%)	331 (67.28%)
Travel & Transportation	1 (0.23%)	6 (1.4%)	96 (22.33%)	327 (76.05%)
Business & Finance	-	11 (3.01%)	29 (7.95%)	325 (89.04%)
Health & Fitness	4 (0.94%)	-	163 (38.44%)	257 (60.61%)
Kids	-	-	14 (19.18%)	59 (80.82%)

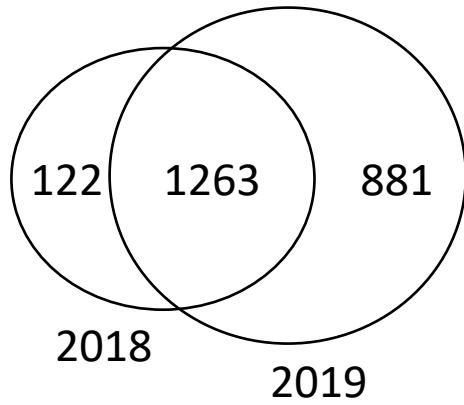
Similar behavior like Alexa US 2019

Evolution of Sensitive Skills

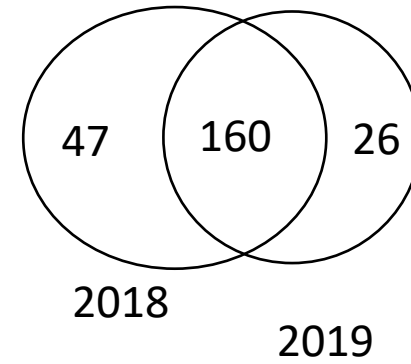


Sensitive Skills of
Alexa US

Evolution of Sensitive Skills



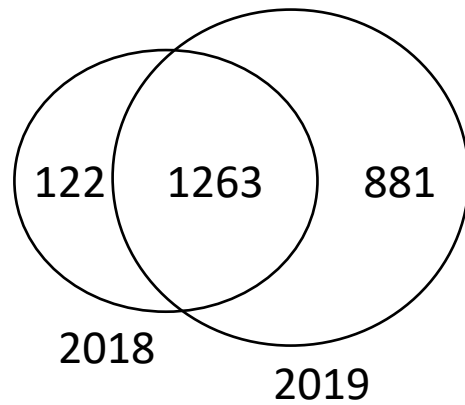
Sensitive Skills of
Alexa US



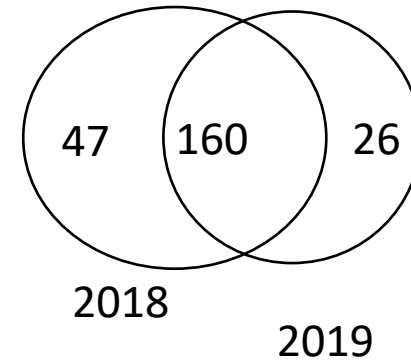
Sensitive Skills of
Google

Evolution of Sensitive Skills

Google's growth of sensitive skills is low compared to Alexa



Sensitive Skills of
Alexa US



Sensitive Skills of
Google

Region Based Analysis

Platform		Sensitive Voice Commands		Nonsensitive Voice Commands	
		Action Injection	Information Retrieval	Action Injection	Information Retrieval
Alexa	US 2019	8,503 (10.61%)	2,939 (3.67%)	20,559 (25.66%)	48,128 (60.06%)
	UK 2019	3,397 (6.54%)	757 (1.46%)	13,538 (26.07%)	34,230 (65.93%)

Region Based Analysis

Alexa UK store has less sensitive voice commands compared to Alexa US store

Platform		Sensitive Voice Commands		Nonsensitive Voice Commands	
		Action Injection	Information Retrieval	Action Injection	Information Retrieval
Alexa	US 2019	8,503 (10.61%)	2,939 (3.67%)	20,559 (25.66%)	48,128 (60.06%)
	UK 2019	3,397 (6.54%)	757 (1.46%)	13,538 (26.07%)	34,230 (65.93%)

Conclusion

- Large-scale empirical analysis of 82,770 skills and 211,843 voice commands

Conclusion

- Large-scale empirical analysis of 82,770 skills and 211,843 voice commands
- Identified 5.55% sensitive skills

Conclusion

- Large-scale empirical analysis of 82,770 skills and 211,843 voice commands
- Identified 5.55% sensitive skills
- Common characteristics of the skills between cross platforms

Conclusion

- Large-scale empirical analysis of 82,770 skills and 211,843 voice commands
- Identified 5.55% sensitive skills
- Common characteristics of the skills between cross platforms
- Platform should enforce authentication for sensitive skills

Conclusion

- Large-scale empirical analysis of 82,770 skills and 211,843 voice commands
- Identified 5.55% sensitive skills
- Common characteristics of the skills between cross platforms
- Platform should enforce authentication for sensitive skills
- Public dataset



Thank You!

Contact: Faysal Hossain Shezan (Email-fs5ve@virginia.edu)

Conclusion

- Large-scale empirical analysis of voice commands
- Identified 5.55% sensitive commands
- Common characteristics across platforms
- Platform should enforce sensitive skills



Dataset

s and 211,843 voice

en cross platforms

sensitive skills

Email-fs5ve@virginia.edu