Hey Alexa, is this Skill Safe?

Taking a Closer Look at the Alexa Skill Ecosystem

Christopher Lentzsch[1], Sheel Jayesh Shah[2], Benjamin Andow[3], Martin Degeling[1], **Anupam Das**[2], William Enck[2]

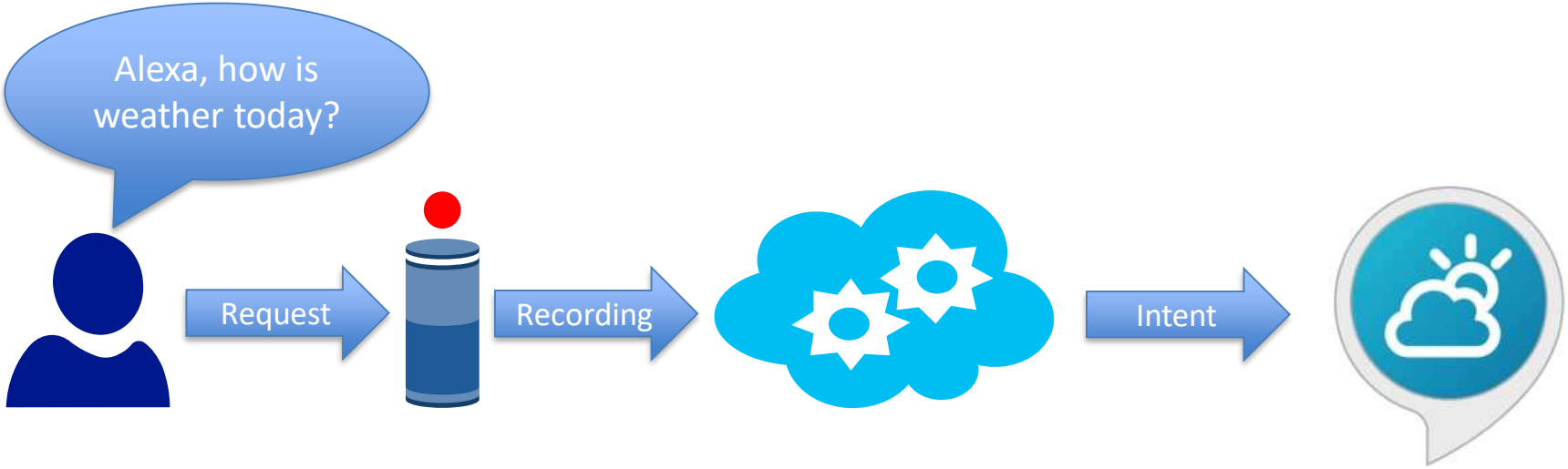
[1] Ruhr-Universität Bochum
[2] North Carolina State University
[3] Google Inc.

LASER Workshop, 2021

Voice Assistants

- Voice-based user interface
- 3+ billion devices
- Placed in personal settings
- New security/privacy risks

How Alexa Works?



How Alexa Works?

Skills: Applications for Voice Assistants

- Enhance basic functionalities
- Integrate with third parties
- Instant activation
  - Alexa, open <*invocation name*>
- Vetted prior to publication

Research Questions

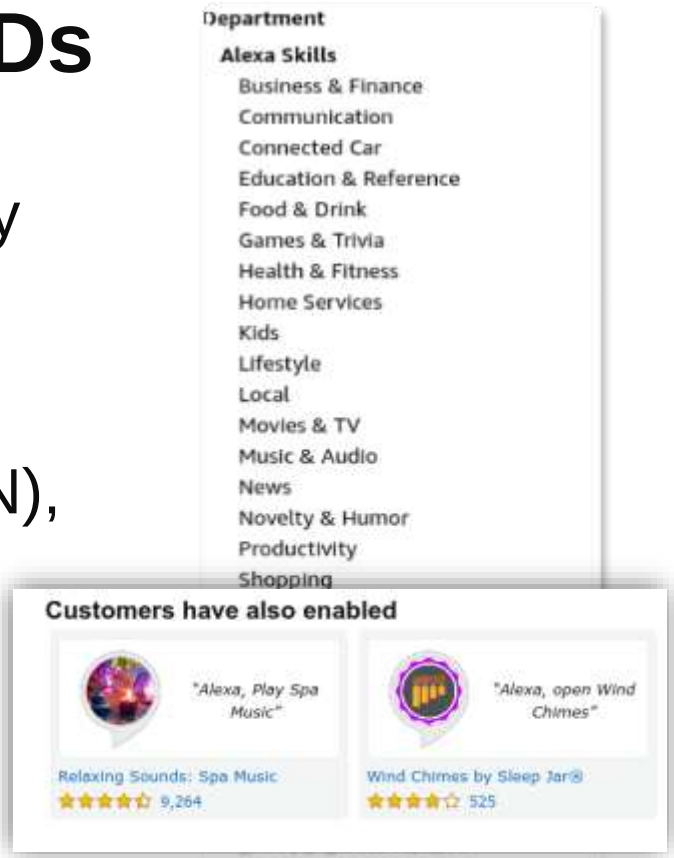
1. What limitations exist in the current skill vetting process?
   - ➢ Duplicate invocation names
   - ➢ Developer names
   - ➢ Registered intents
   - ➢ Permission model
2. How effective are skill squatting attacks?
   - ➢ Which patterns are more effective?
3. Is the requirement of providing a privacy policy effective?
   - ➢ Are data types properly disclosed

Our Data Collection Methodology

- Crawled the top 7 skill stores for meta data
- Rented servers in 5 different locations (DigitalOcean)
  - US, Canada, UK, Germany (for DE, FR), Singapore (for AU, JP)
  - To prevent geo-blocking
- Crawler used SELENIUM
  - Firefox Browser, headful, less likely detected as bot

Gathering Skill IDs

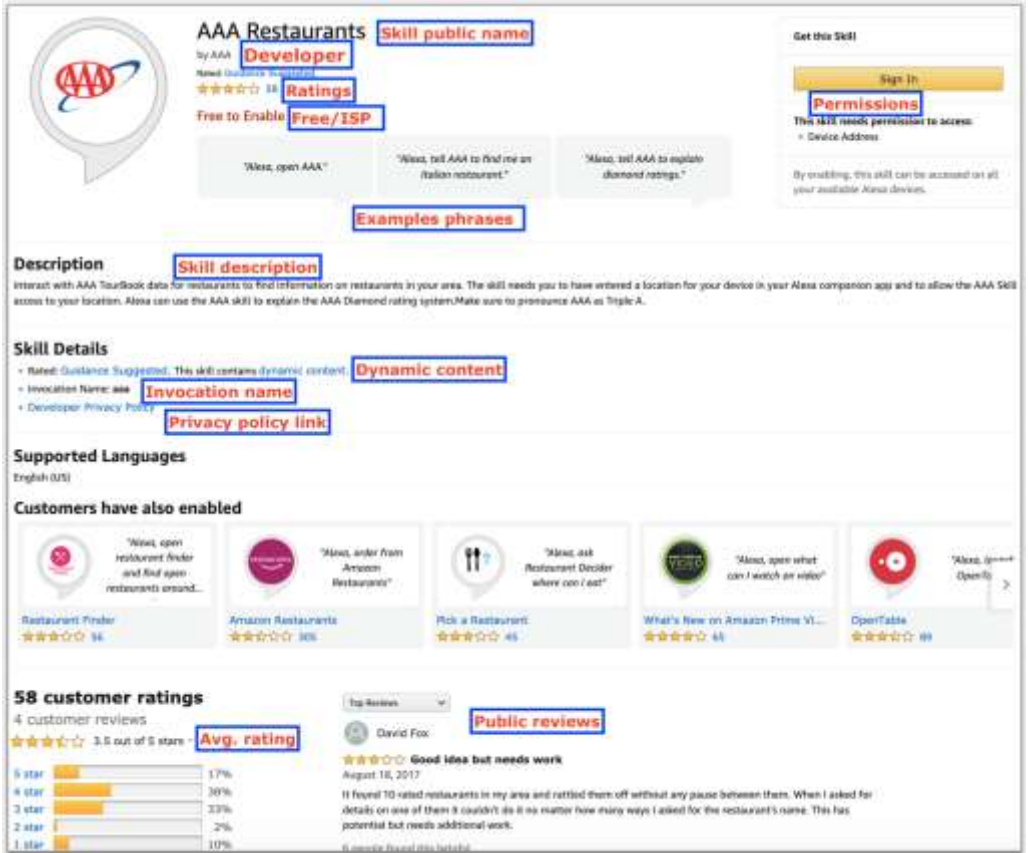
- Traversed each skill category separately
  - 23 categories (in Jan 2020)
  - Max 400 pages per category
  - Track skills through unique IDs (ASIN), e.g. B07KX2CSXM
  - Also downloaded skills appearing in the recommended list



Extracting Skill Metadata

- Downloaded metadata from skill information page
  - Parsed using beautiful soup
  - Downloaded privacy policies (if available)
- Took around 9 days for the US Skill store
  - At time our crawler was blocked and had to wait

Example of Metadata



Brief Summary of Data

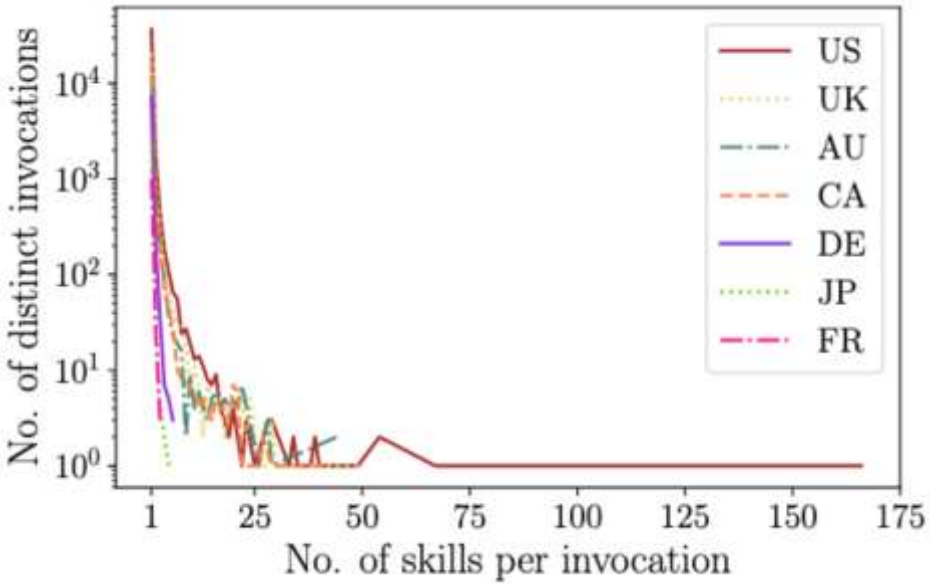
- Total 90,194 unique skills

Store		# of common developers						
		AU	CA	DE	FR	JP	UK	US
# of common skills	AU	**(3023 / 948)**	7634	506	182	113	8164	8175
	CA	15151	**(2243 / 229)**	636	423	180	7838	8091
	DE	904	911	**(8558 / 2278)**	357	146	887	937
	FR	475	722	563	**(1189 / 499)**	120	440	455
	JP	234	246	226	196	**(3022 / 1019)**	191	247
	UK	16556	16815	1322	655	262	**(8557 / 2465)**	9796
	US	14916	16294	1295	601	299	19688	**(35698 / 13090)**

In the US store there are over 35k unique skills with over 13k unique developers

Question: How does Alexa select skills?

- Many duplicate skill invocation names
  - For *space facts* there are 81 skills
  - Skills are auto enabled



Methodology to Test Skill Selection Process

- Find skills that have same invocation name
  - Differ in other observable attributes
- Test invocation names
  - Alexa, open <invocation name>
  - Rerun test three times
- Analyze results

Example: Skills with Same Invocations

Invocation	UID	Name	Developer	# Ratings	Rating
grandfather clock	**B071DWVZQW**	**Sleep Sounds: Grandfather Clock**	**Voice Apps, LLC.**	**115.0**	**4.5 out of 5**
grandfather clock	B076452X2P	Grandfather Clock	ut666	57.0	3.8 out of 5

- Example: <grandfather clock>
- Different number of ratings
- Different avg. rating

What Factors to Test?

- Available metadata
  - Number of ratings
  - Avg. rating
  - Age of skill
  - Category
  - Dynamic (true/false)
  - Guidance (true/false)
  - Permissons
  - Privacy policy (present/none)
  - Terms of service (present/none)

Note: not all properties provided enough statistically signifcant samples

Automate Testing using Amazon Polly

- TTS-Service Amazon Polly
  - 'Salli', female, en-US voice
  - "Alexa, start unicorn facts"

Create User Accounts for Testing

- Tested across three new independent accounts
- Created Accounts with US-bound IPs (VPN-Service)
  - To avoid localization, e.g., german language or store
  - To avoid rate limiting

Example: Testing

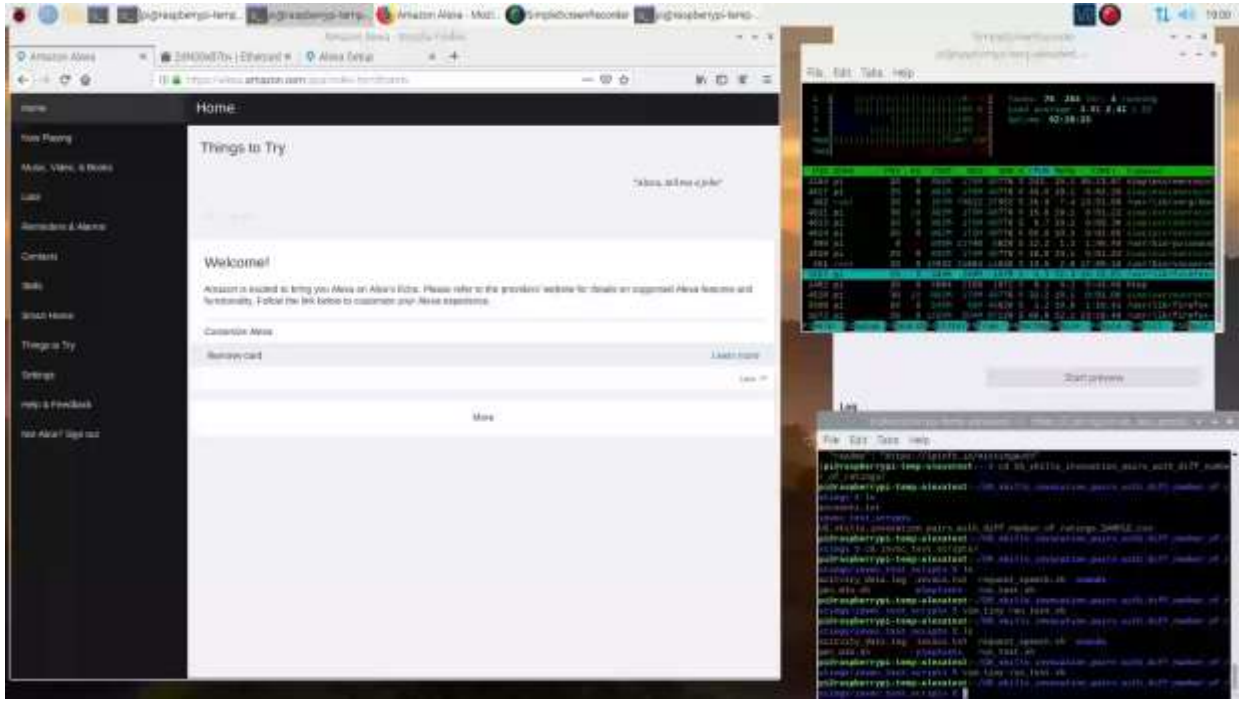
Play invocation
- "Alexa, start unicorn facts"

Wait 5 seconds
- "Alexa, stop"

Wait 3 seconds
- "Alexa, exit"

Wait 3 seconds

Get Activity-Log
- Curl-Call to REST-API



Activity Log


```
{
  "cards": [
    {
      "cardType": "A2SEnableSkillCard",
},
      "developerName": "Envy Eden",
      "examplePhrase": "Alexa, open Unicorn Facts",
"hint": null,
"originIntentType": "LaunchNativeAppIntent",
      "playbackAudioAction": {
        "actionType": "PlayAudioAction",
        "mainText": "Alexa heard: \"alexa start unicorn facts\"",
},
      "primaryActions": [
        {
          "actionType": "NavigateAction",
          "mainText": "View Skill Details",
          "route": "skills/dp/B07DL8X97K",
          "routeAddOnComponent": null,
          "serviceName": null,
          "subText": "Description, additional phrases, reviews, developer terms of use, privacy policy, and other details",
          "subTextRoute": null
        }
      ],
      "skillIconUrl": "https://images-na.ssl-images-amazon.com/images/I/618xoR-pTFL.png",
      "skillName": "Unicorn Facts",
      },
}
```

Fisher's Exact Test

for # of ratings

	more ratings	less ratings
activated	40	10
not activated	10	40

- Odds ratio = 16.0
- p-value <0.0001

Other Factors

Attribute 1	Attribute 2	Favored Attribute	p-value
Different number of ratings		**more ratings**	< 0.0001 ****
Different avg. rating		**higher avg. rating**	0.00012 ***
Age of skill			0.84162
Content advisory			0.54874
Same number of ratings	Different avg. rating	**higher avg. rating**	0.03476 *
Same number of ratings	Age of skill		0.31734
Same number of ratings	Content advisory		0.84161

Manipulating Attributes

- We launched our own skill pair
  - Identical invocation name
  - Increased num. of ratings and interaction
    - Failed to change skill selection

Takeaway: Positive correlation with rating, but it does not necessarily imply causation.

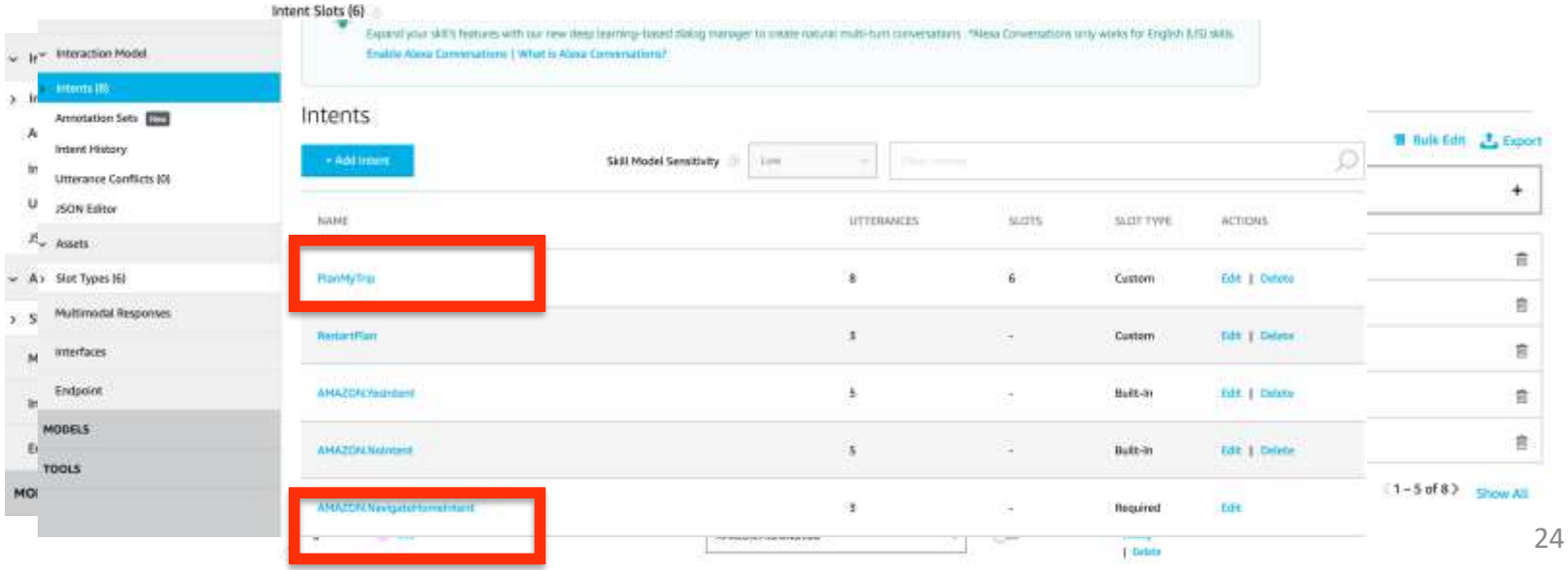
Question: Can I use any Developer Name?

- Skill page shows a developer name
- Published our skills as:
  - *Microsoft*, *Ring, Samsung* and *Withings*
- *Philips* got rejected

Takeaway: An attacker can getaways with publishing skills using well-known company names.

Question: Can I Register Dormant Intents?

- Skills have intents
  - Each intent can many slots (i.e., data type)



Code Change after Approval

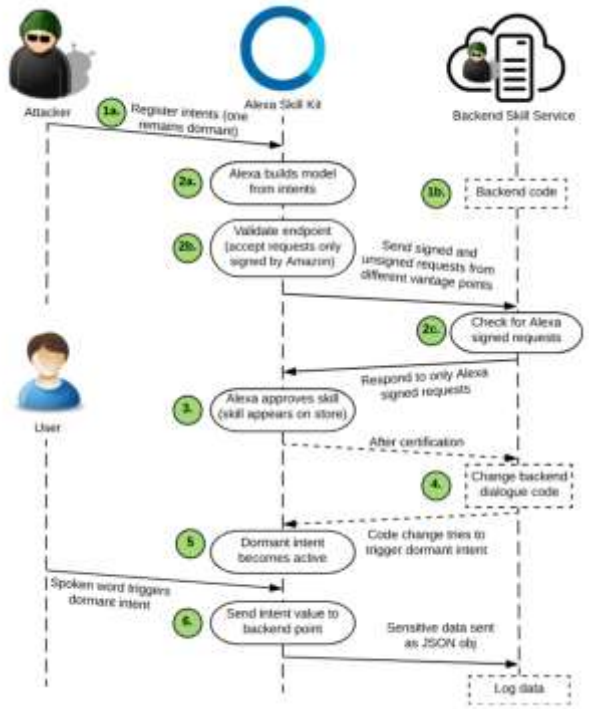
Generate intent model (dormant intent present)

Skill is vetted and approved

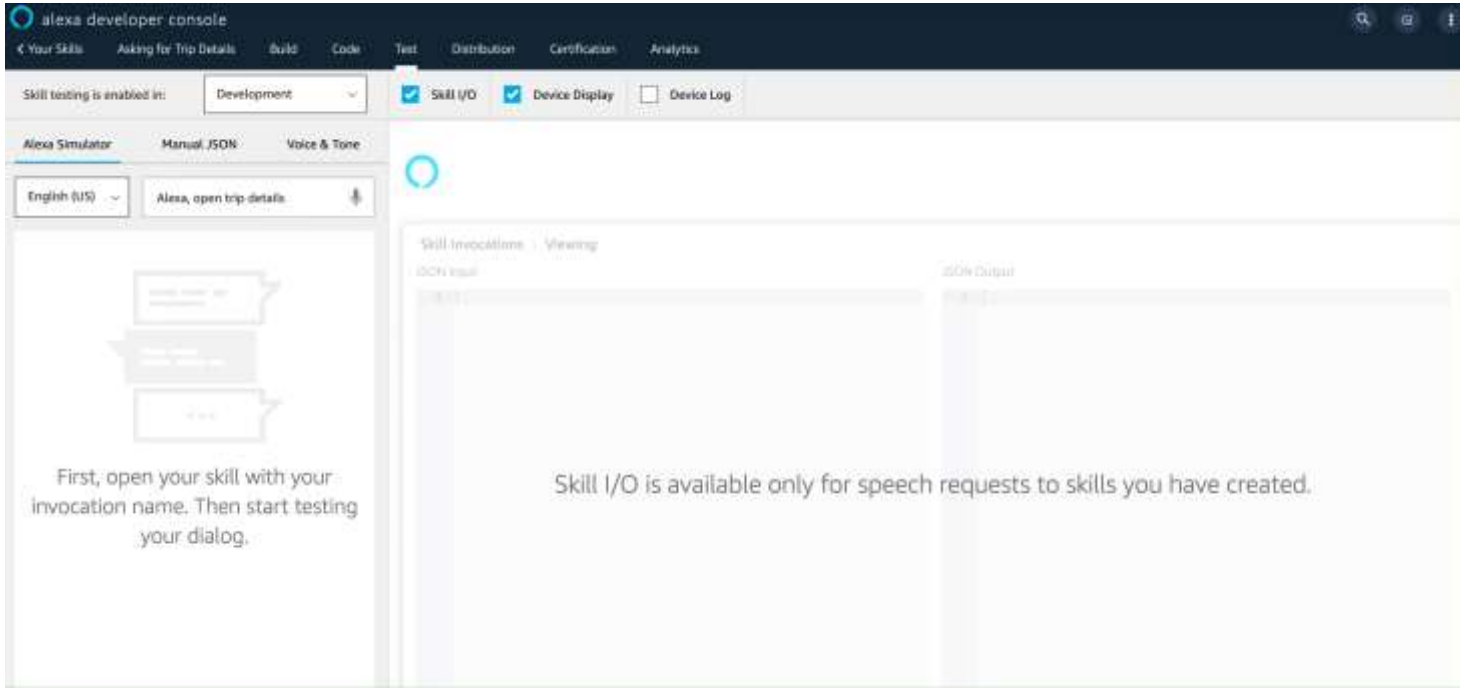
Backend code is changed

User is directed to trigger dormant intent

Adversary gets user‘s phone number



Demo Video



Takeaway: An adversary can change the backend code after approval to coax users into revealing sensitive information

Question: Do Skills Bypass Permission?

Extract skill description	**Skill description define what the skill does**
Use regular expressions to search for permission protected data types	**Phone number, location, e-mail, name**
Manually vet and activate skill for verification	**Ignore account linking skills and skills requesting permissions**

Regex Used

Data Type	Regular Expression
Name	\b(your)\s+(((whole|entire|first/last|full|given|first|last|legal|first\sand\slast)\s+)? |(sur)?name)\b
Location	\b(your)\s+((home|work|personal|physical|billing|mailing|business|device('s)?)\s+)? (city|state|province| area|(postal\s+)?address|(zip|postal)\scode|((gps|device|geographic(al)?) \s+)?location| latitude|longitude|lat(itude)?/lon(gitude)?|lat(itude)?\sand\slon(gitude)? |region|country)\b
Phone Number	\b(your)\s+((home|work|personal|billing|business|device ('s)?)\s+)?(phone|telephone|mobile|cellular|cell (\s*phone)?)\s+number\b
Email	\b(your)\s+((home|work|personal|billing|business|valid| school|device('s)?)\s+)?((e|g)(\-)?mail(\saddress)?)\b

Some Skills ask Sensitive Data

Filtering mechanism		Data Type				Unique skills *	w/o PP
		Name	Email	Phone	Location		
Skills detected through regular expression		432	417	242	416	1,482	668
After manually inspecting skill description		109	26	108	133	358	169
Activation	Verbally request data	65	4	33	76	166	99
	Non-verbally request data	1	1	1	0	3	2
	Does not request data	20	7	4	22	52	34
	Skill invocable but non-functional	19	12	62	24	113	25
	Skill not available in store	4	2	8	11	24	9

* Some skills access multiple data types, hence the summation across different data types will be slightly higher than the number of unique skills.

An example of True Positive

Developer Quotes

by Tony

Rated: Guidance Suggested

☆☆☆☆☆ 0

Free to Enable

"Alexa open geek quotes"

"tell me a quo

Shown in: English (US) See all supported languages

Get this Skill

Enable

By enabling, this skill can be accessed on all your available Alexa devices.

Description

A fun app that tells geeky computer jokes. When you start up it tries to get to know you by asking your name and favorite computer language, then it tells you geeky computer jokes.

An example of False Positive

Presidential Quest

by Will Mundy

★★★★★ 1

Free to Enable

"Alexa, launch Presidential Quest"

"Alexa, ask Presidenti... start quest"

Shown in: English (US) See all supported languages

Get this Skill

To use kid skills, a parent or guardian needs to give permission. Learn more.

Enable

By enabling, this skill can be accessed on all your available Alexa devices.

Description

Think you know our presidents? Think again.

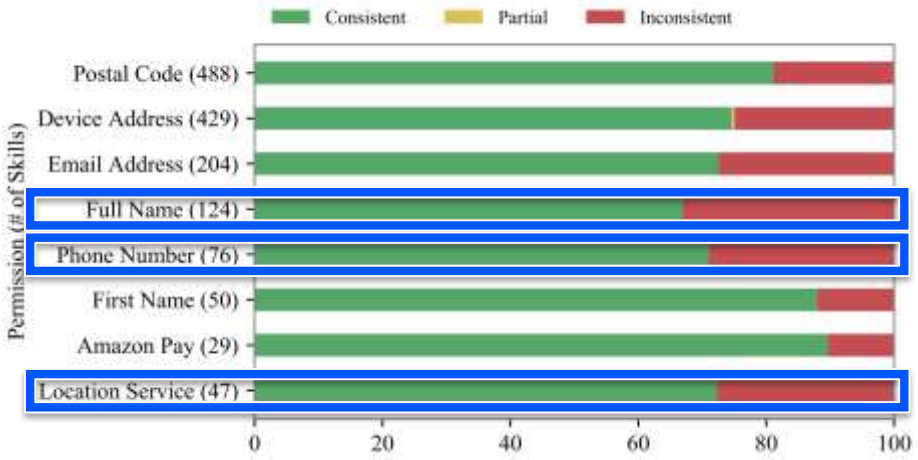
Imagine this: your name is Thomas Jefferson, and you've just received word that Aaron Burr is planning to create a new country within your borders. What do you do? Do you sit back and watch as your former Vice-President takes your land? Or do you fight for your country by answering his Presidential Trivia?

Question: Do Skills Disclose Data Practices?

Extract data flows from privacy policy using **PoliCheck (NLP tool)**	Extract statements regarding the collection and share of data
Re-map data flows to permission requests	Manually adapt the data type ontology of PoliCheck to match Alexa permissions
Classify each data flow	Three groups: consistent, partial, inconsistent

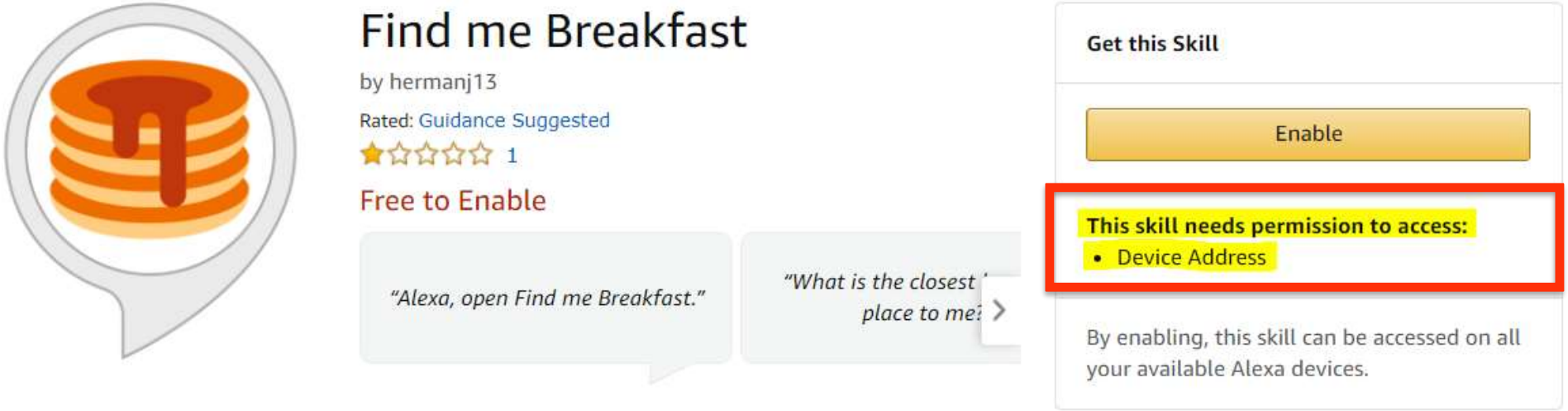
Data Practice Disclosure in Privacy Policy

- Are requested permissions covered?
  - Analyzed 1,124 skills with 1,447 permission requests



Takeaway: Sensitive data types are not fully disclosed by ~23% of the skills‘ privacy policies.

Skill: “Find me Breakfast” (B07K9NQX5B)



Privacy Policy of “Find me Breakfast” (B07K9NQX5B)

General

When you use our skills you have to talk to Alexa. This voice input is sent to Amazon and us where we use it to understand what our skill should do for you. This is absolutely necessary for our service to give you an appropriate answer.

Data

We never collect or share personal data with our skills.

To improve our services we analyze automatically how often utterances are spoken and other analytics. This is done automatically by Amazon in the Amazon Developer Portal.

Skill: “Wear Assistant” (B072KL1S3G)

Wear Assistant

by FluiBex

Rated: Guidance Suggested

4.0 out of 5 stars 249

Free to Enable

“Alexa, open Wear Assistant”

“Alexa, ask Wear Assistan… should I dress today… York”

Shown in: English (US) See all supported languages

Get this Skill

Enable

This skill needs permission to access:

- Device Address

By enabling, this skill can be accessed on all your available Alexa devices.

Privacy Policy of “Wear Assistant” (B072KL1S3G)

General

When you use our skills you have to talk to Alexa. This voice input is sent to Amazon and us where we use it to understand what our skill should do for you. This is absolutely necessary for our service to give you an appropriate answer.

Data

We never collect or share personal data with our skills.

For some of our skills, after your consent, we may use your address information in order to automatically detect your location and speed up the interaction. Your location data is never collected on our server, but it is used on the fly during the skill response built. In any moment you can revoke the use of your location information to our skills from the Amazon Alexa app.

What did we learned from our methodology?

- Automated skill activation
  - Tried building a chatbot to interact with skills, but failed to process and respond before timeout
- Detecting squatting skills wasn't as easy we had thought
  - Had to manually go through similar skill pairs after identifying phonetically similar invocations
- Existing NLP techniques were readily applicable
  - PoliCheck

Going Forward

- Develop systems to automatically interact with skills to determine if they access sensitive data
- Conduct user studies to understand people's perception of how skills work
- Better security indicators for the voice interface

Contact

Responsible Disclosure

Reported our findings to Amazon and have talked with them.

Our data set is public:

https://alexa-skill-analysis.org

Christopher Lentzsch <christopher.lentzsch@rub.de>
Sheel Jayesh Shah <sshah28@ncsu.edu>
Benjamin Andow <andow@google.com>
Martin Degeling <martin.degeling@rub.de>
Anupam Das <anupam.das@ncsu.edu>
William Enck <whenck@ncsu.edu>