Skill Squatting Attacks on Amazon Alexa

Deepak Kumar
University of Illinois

Riccardo Paccagnella
University of Illinois

Paul Murley
University of Illinois

Eric Hennenfent
University of Illinois

Joshua Mason
University of Illinois

Adam Bates
University of Illinois

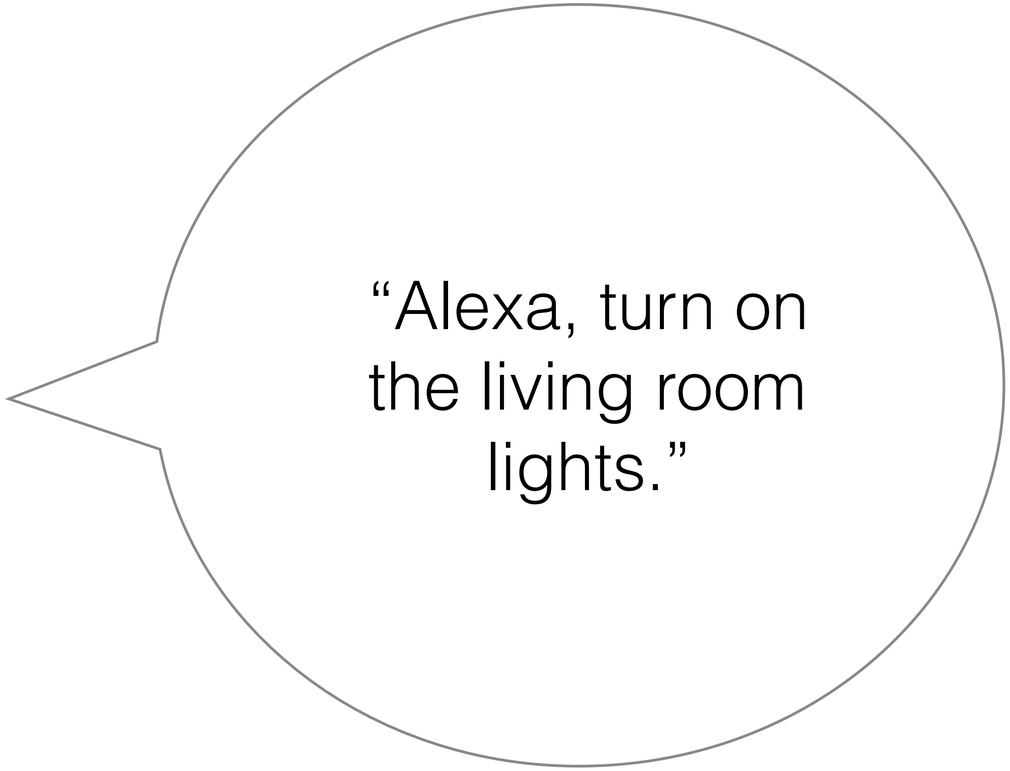
Michael Bailey
University of Illinois

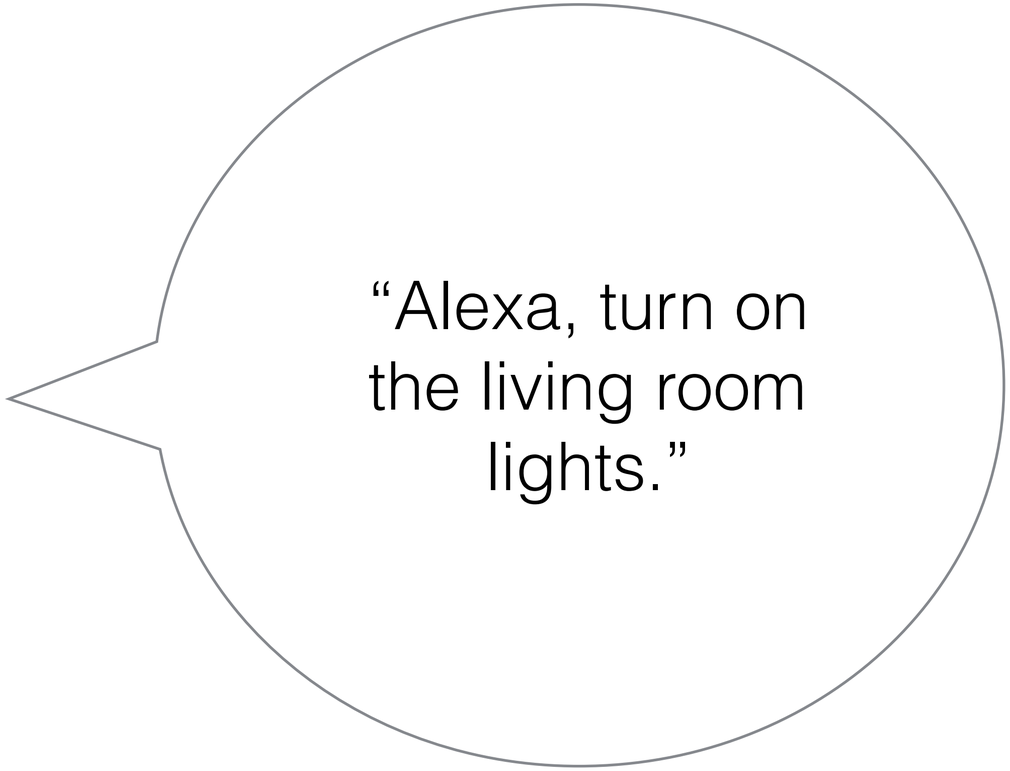
Voice Is the New Touch

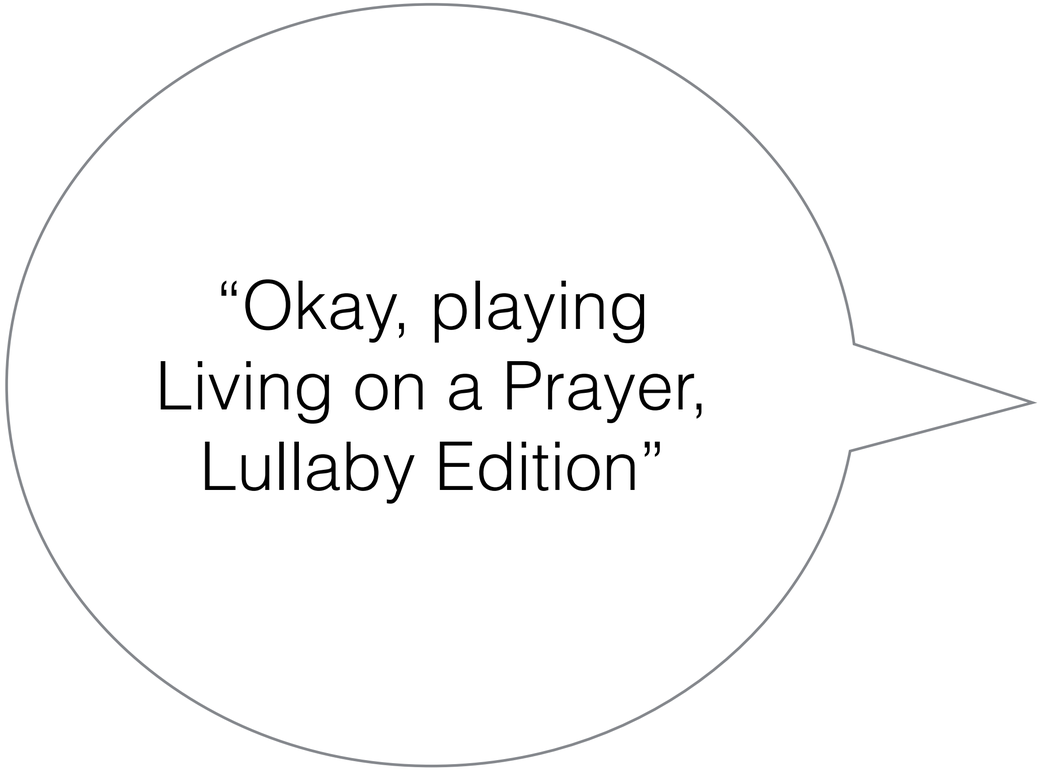
56 Million Smart Speaker Sales in 2018 Says Canalys

Voice-First Devices Are The Next Big Thing -- Here's Why

Speech recognition systems frequently make errors, even in normal use


“Alexa, turn on the living room lights.”


“Alexa, turn on the living room lights.”


“Okay, playing Living on a Prayer, Lullaby Edition”


hue
PHILIPS
P
BBC
NEW
WIKIPE
The Free Encyc
a
pr
ews
The
Economist
IF
wink
we
mo
amazon

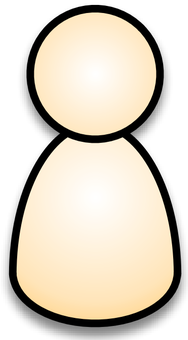
Alexa Skills

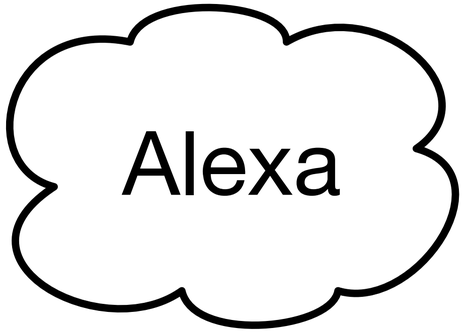
Alexa Skills

Alexa Skills

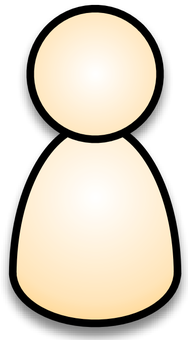
Skills: Apps, but for Alexa

“Alexa, tell me some cat facts!”





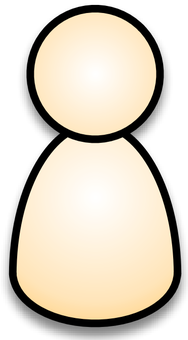
“Alexa, tell me some cat facts!”



Alexa

Skills
…
…
cat forks
cat fast
cat facts
…
…
…
…

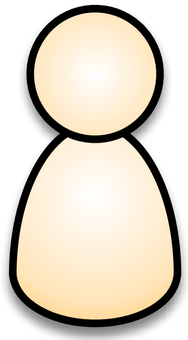
“Alexa, tell me some cat facts!”



Alexa

Skills
…
…
cat forks
cat fast
cat facts
…
…
…
…

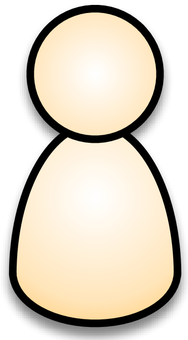
“Alexa, tell me some cat facts!”



Alexa

Skills
…
…
cat forks
cat fast
cat facts
…
…
…
…

“Alexa, tell me some cat facts!”



Alexa

Skills

…

…

cat forks

cat fast

cat facts

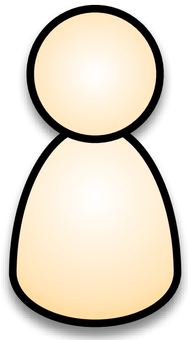
…

…

…

…

“Alexa, tell me some cat facts!”

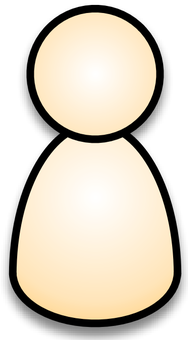


Alexa

cat fact pls

Skills
…
…
cat forks
cat fast
cat facts
…
…
…
…

“Alexa, tell me some cat facts!”



Alexa

cat fact pls

Skills
…
…
cat forks
cat fast
cat facts
…
…
…
…

“A group of cats is called a clowder!”

1. Alexa makes mistakes

1. Alexa makes mistakes
2. Skills are the new apps

1. Alexa makes mistakes
2. Skills are the new apps

What could go wrong?

Fish Geek

Matt Mitchell

"Alexa ask Fish Geek to tell me a fact"

"Alexa ask Fish Geek to tell me trivia"

Phish Geek

EP

"Alexa, open Phish Geek"

"Alexa, launch Phish Geek and tell me a fact"

Rachel

★☆☆☆☆ **It just gives "fish" facts, not "Phish" facts**

February 8, 2017

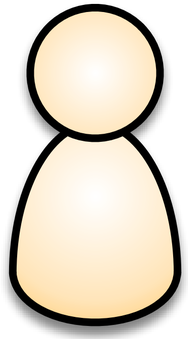
I would love it if this actually gave facts about the band. But instead it tells you things like, "some fish have fangs"

Can Alexa errors be leveraged to cause harm to end users?

Skill Squatting Attacks

- An attacker can leverage predictable errors in Alexa to route users to skills that they didn't intend to go to

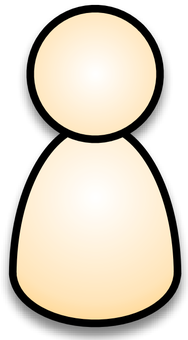
“Alexa, tell me some cat facts!”



Alexa

Skills
…
…
cat forks
cat fast
cat facts
cat fax
…
…
…

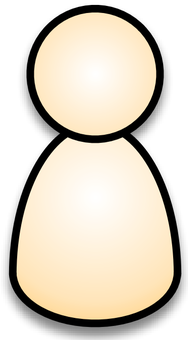
“Alexa, tell me some cat facts!”



Alexa

Skills
…
…
cat forks
cat fast
cat facts
cat fax
…
…
…

“Alexa, tell me some cat facts!”



Alexa

cat fact pls

Skills
…
…
cat forks
cat fast
cat facts
cat fax
…
…
…

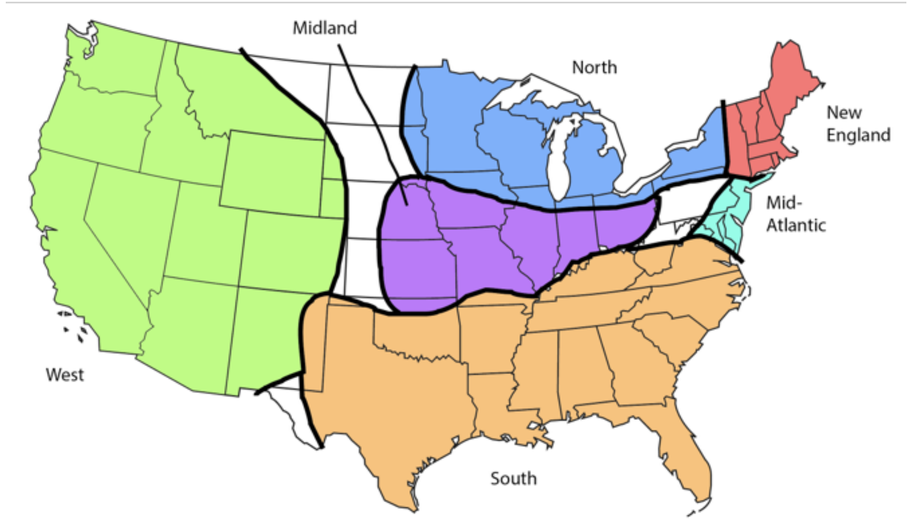
“A group of cats is called a chowder!”

How can you tell which errors are predictable?

Send speech samples to Alexa, figure out where it goes wrong

Speech Corpus

- Leveraged the **NSP Dataset**
  - 60 speakers, 188 unique words each (11,460 audio samples)
  - Speakers were representative of 6 US dialect regions



Measuring Interpretation Errors

- We sent each speech sample to Alexa 50 times, providing us 573,000 transcriptions across the 60 speakers

Predictable Errors

Word	Prediction
Sail	Sale
Rip	Rap
Outshine	Outshyne
Lung	Lang
Accelerate	Xcelerate
Mill	No
Preferably	Preferrably
Earthy	Fi
Calm	Com
Coal	Call
Outdoors	Out Doors
Loud	Louder

Word	Prediction
Superhighway	Super Highway
Wet	What
Main	Maine
Boil	Boyle
Sell	Cell
Full	Four
Dime	Time
Bean	Been
Dull	Doll
Sweeten	Sweden
Luck	Lock
Con	Khan

Can we use our predictable errors to route users to unintended skills?

Validating the Skill Squatting Attack

- Split speakers into two sets: “training” set and the “testing” set
- For each word with predictable error, we built two skills: the word, and the predictable error
  - Skill A: Boil
  - Skill B: Boyle
- Sent the testing set through to Alexa, observed how many times skill B was triggered instead of skill A

A Brief, Ethical Note….

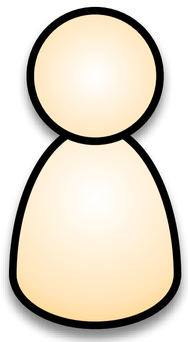
- We validated this attack strictly in a developer environment, no real skills were targeted or tested in the wild
- This is a fundamental limitation, but it’s what we thought was the right thing to do

Successfully squatted 25 of 27 (93%) predictable errors at least once

I would never want a cat fact.
Why does this matter?

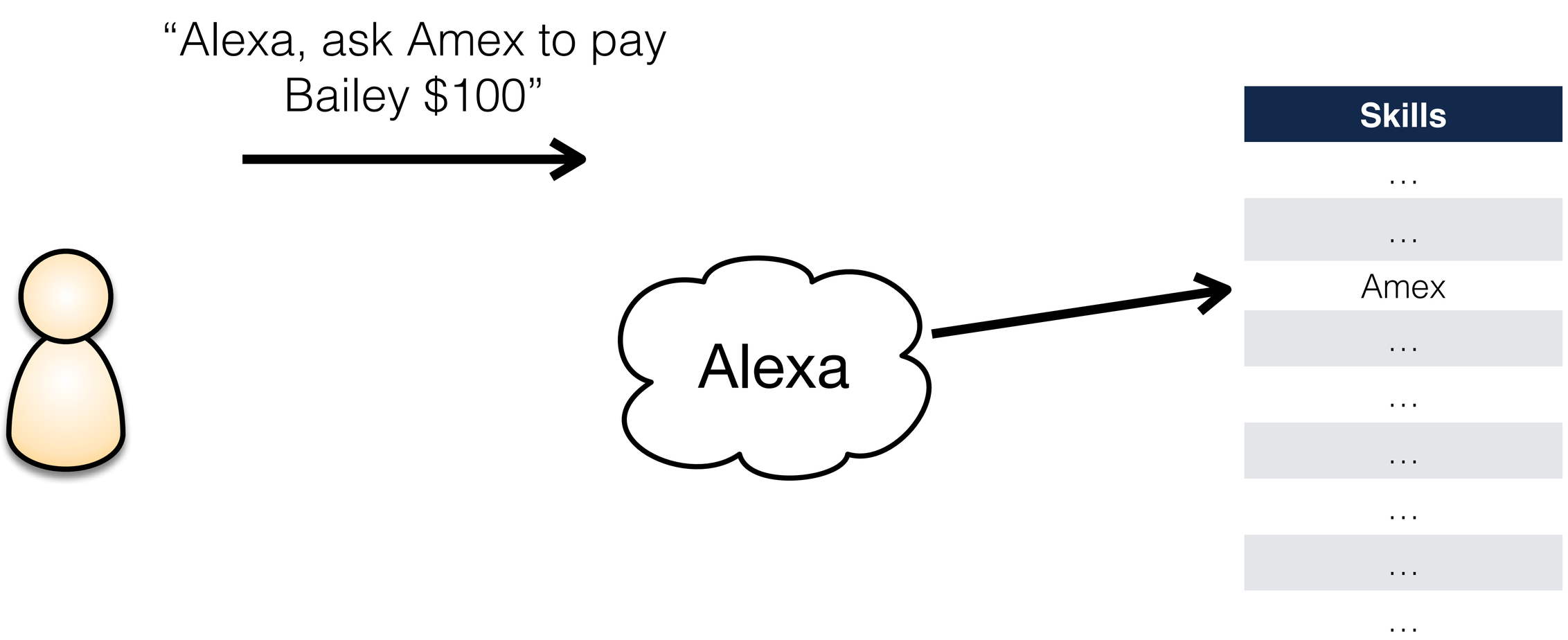

AMEX

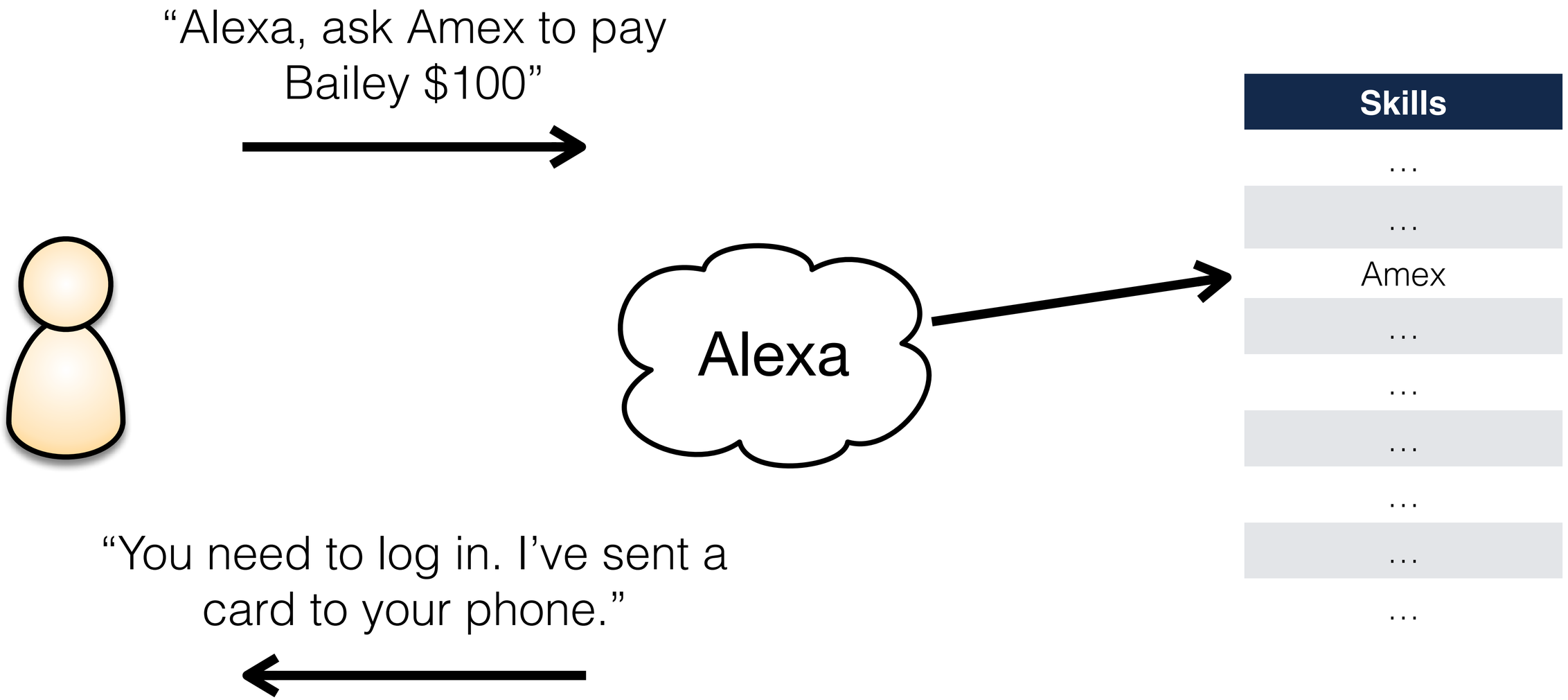
“Alexa, ask Amex to pay Bailey $100”

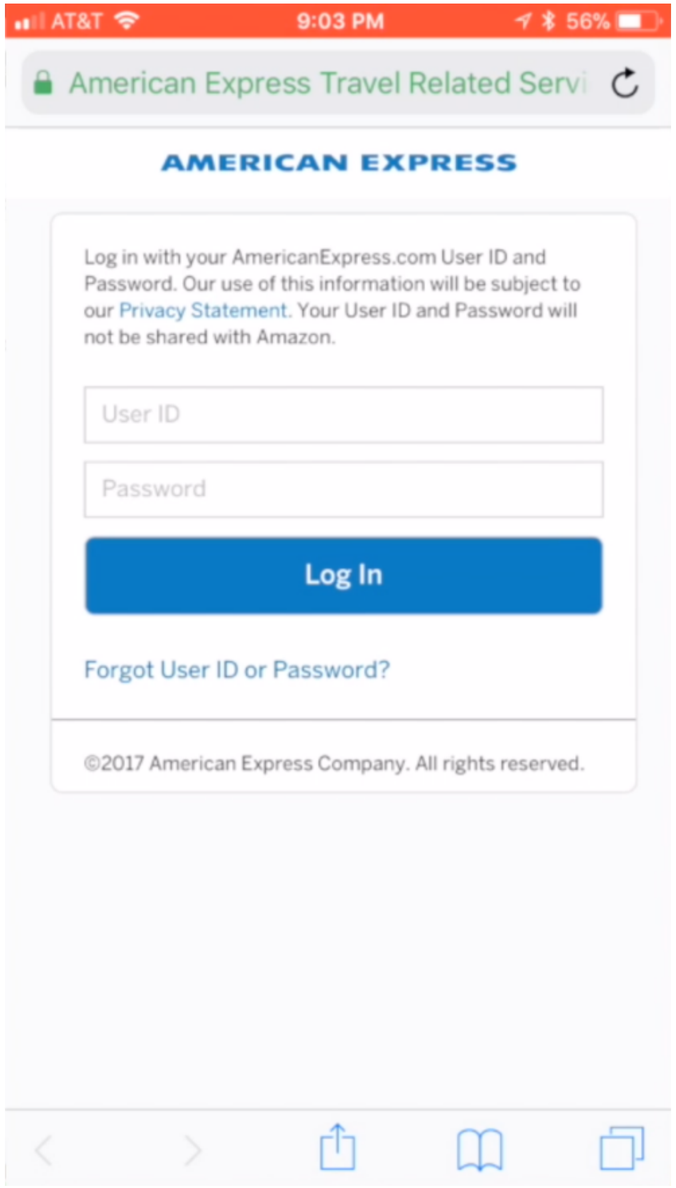


Alexa

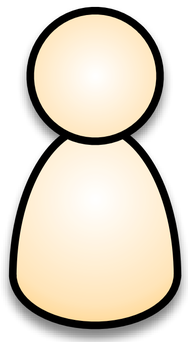
Skills
…
…
Amex
…
…
…
…
…
…


“Alexa, ask Amex to pay Bailey $100”
Alexa
Skills
…
…
Amex
…
…
…
…
…
…


“Alexa, ask Amex to pay Bailey $100”
Alexa
Skills
…
…
Amex
…
…
…
…
…
“You need to log in. I’ve sent a card to your phone.”

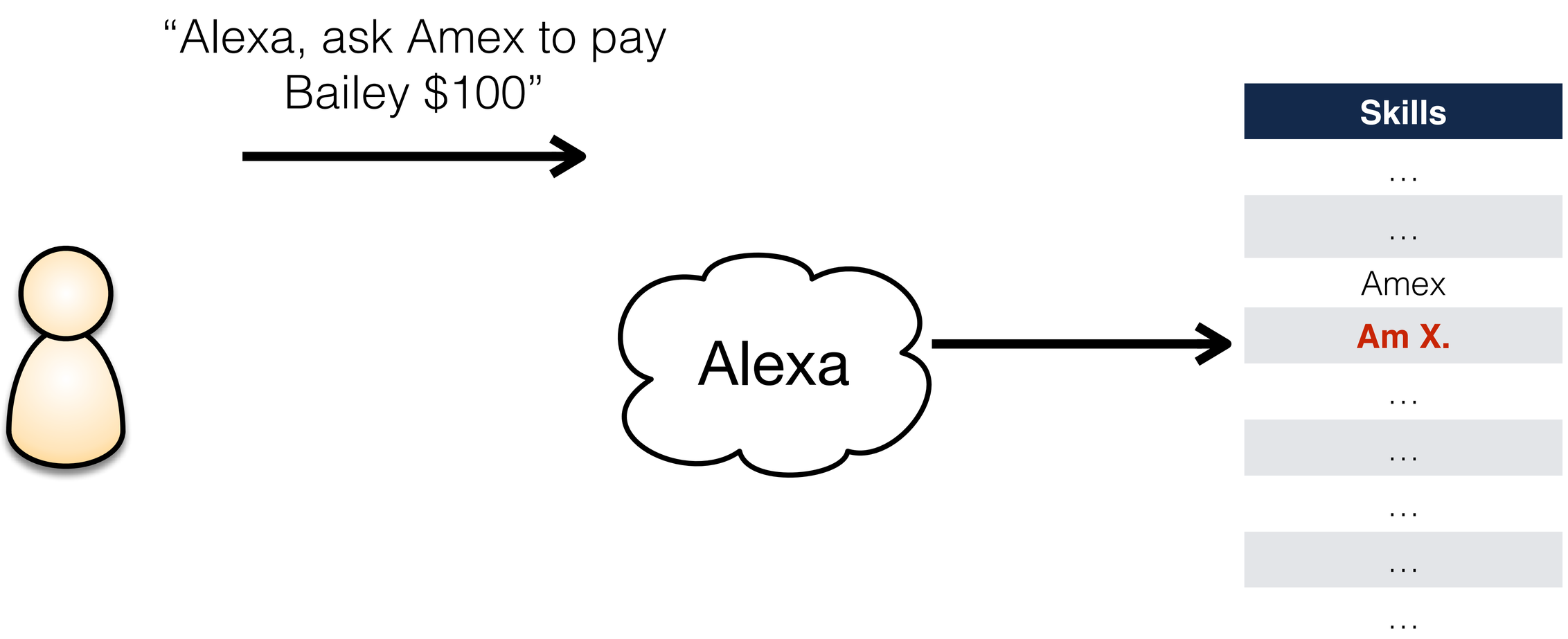

AT&T
9:03 PM
56%
American Express Travel Related Servi
AMERICAN EXPRESS
Log in with your AmericanExpress.com User ID and Password. Our use of this information will be subject to our Privacy Statement. Your User ID and Password will not be shared with Amazon.
User ID
Password
Log In
Forgot User ID or Password?
©2017 American Express Company. All rights reserved.

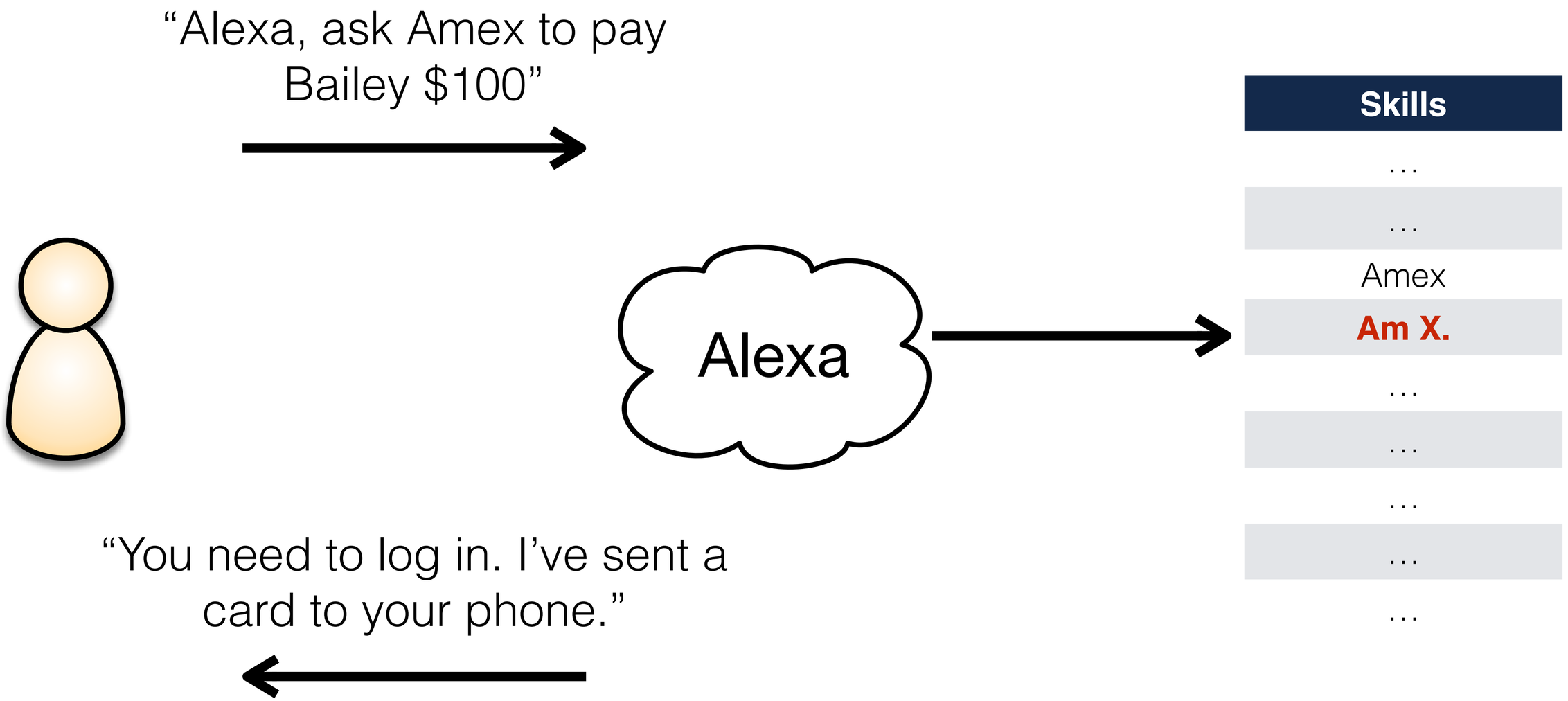
“Alexa, ask Amex to pay Bailey $100”

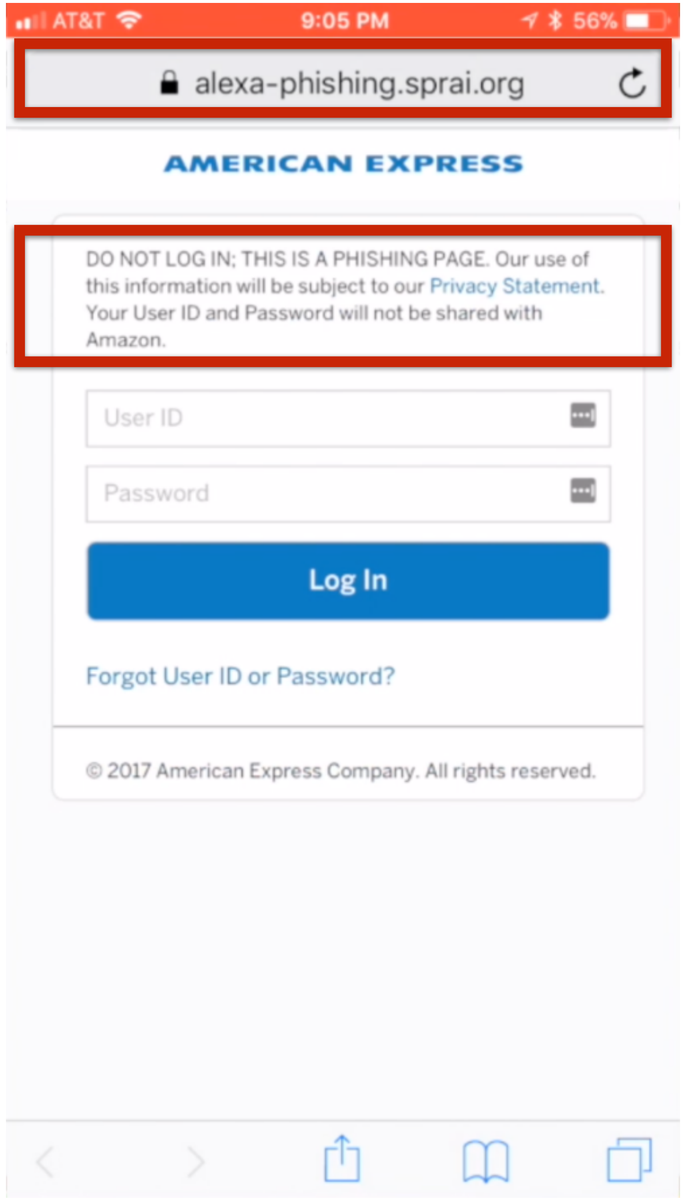


Alexa

Skills
…
…
Amex
Am X.
…
…
…
…
…


“Alexa, ask Amex to pay Bailey $100”
Alexa
Skills
…
…
Amex
Am X.
…
…
…
…
…

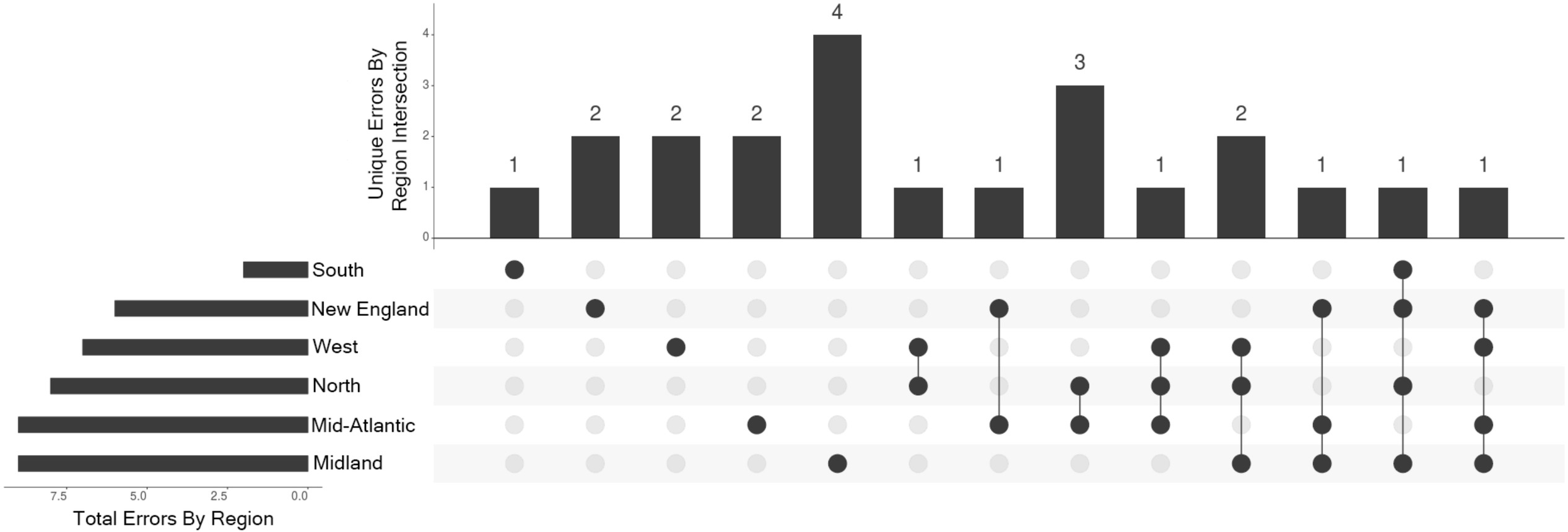

“Alexa, ask Amex to pay Bailey $100”
Alexa
Skills
…
…
Amex
Am X.
…
…
…
…
…
“You need to log in. I’ve sent a card to your phone.”


AT&T
9:05 PM
56%
alexa-phishing.sprai.org
AMERICAN EXPRESS
DO NOT LOG IN; THIS IS A PHISHING PAGE. Our use of this information will be subject to our Privacy Statement. Your User ID and Password will not be shared with Amazon.
User ID
Password
Log In
Forgot User ID or Password?
© 2017 American Express Company. All rights reserved.

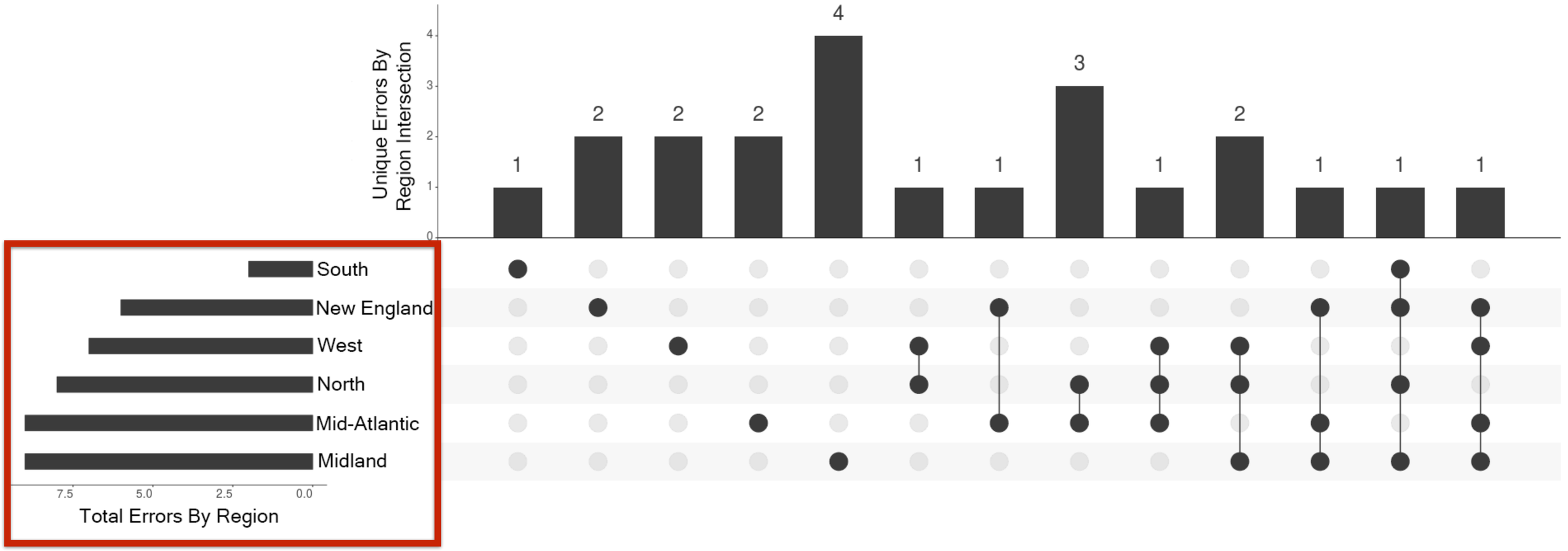
VOICE IS THE NEXT BIG PLATFORM, UNLESS YOU HAVE AN ACCENT

Do different regions exhibit unique predictable interpretation errors?

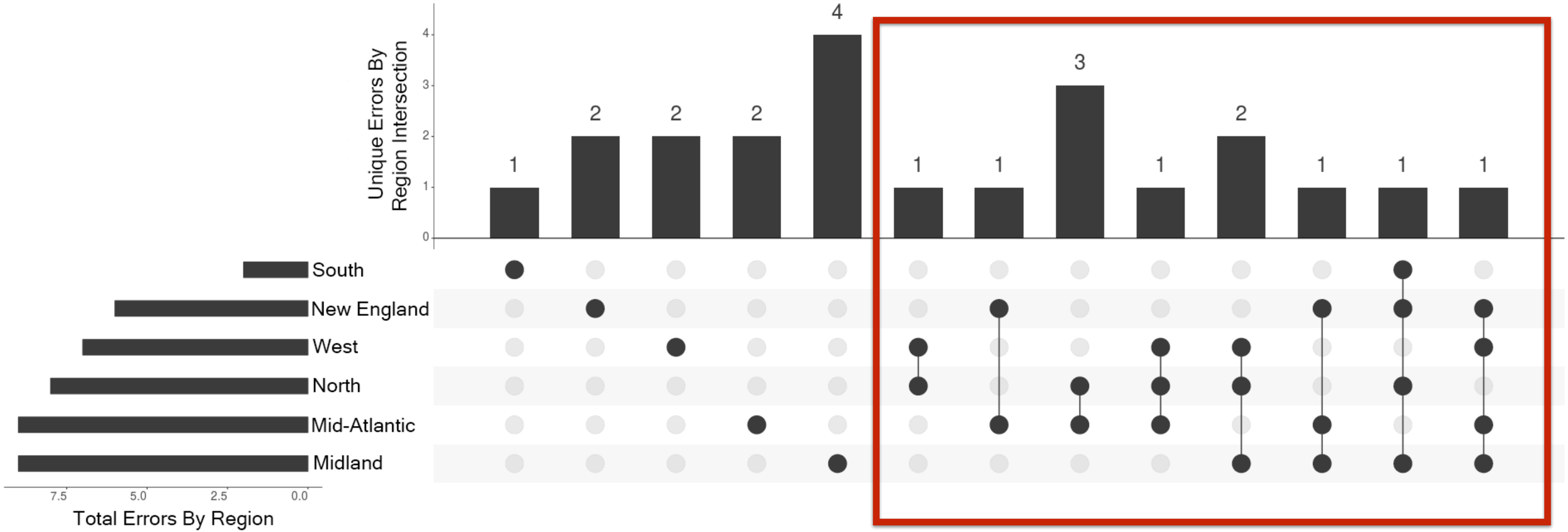
Predictable Errors by Region



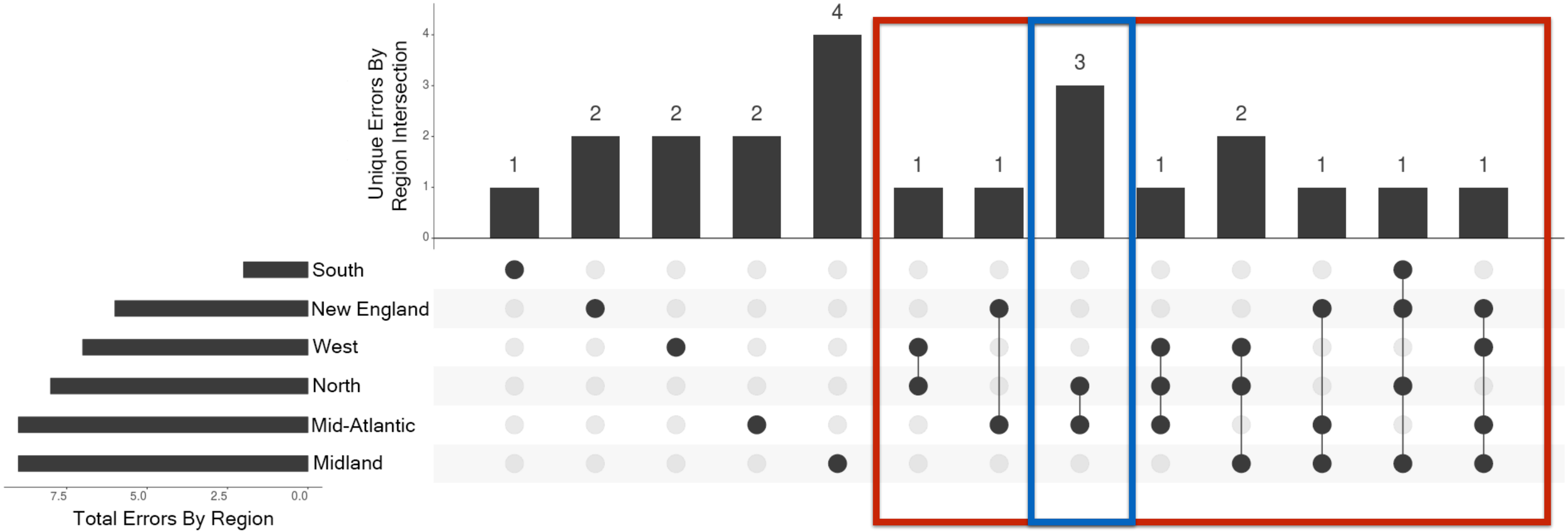
Predictable Errors by Region



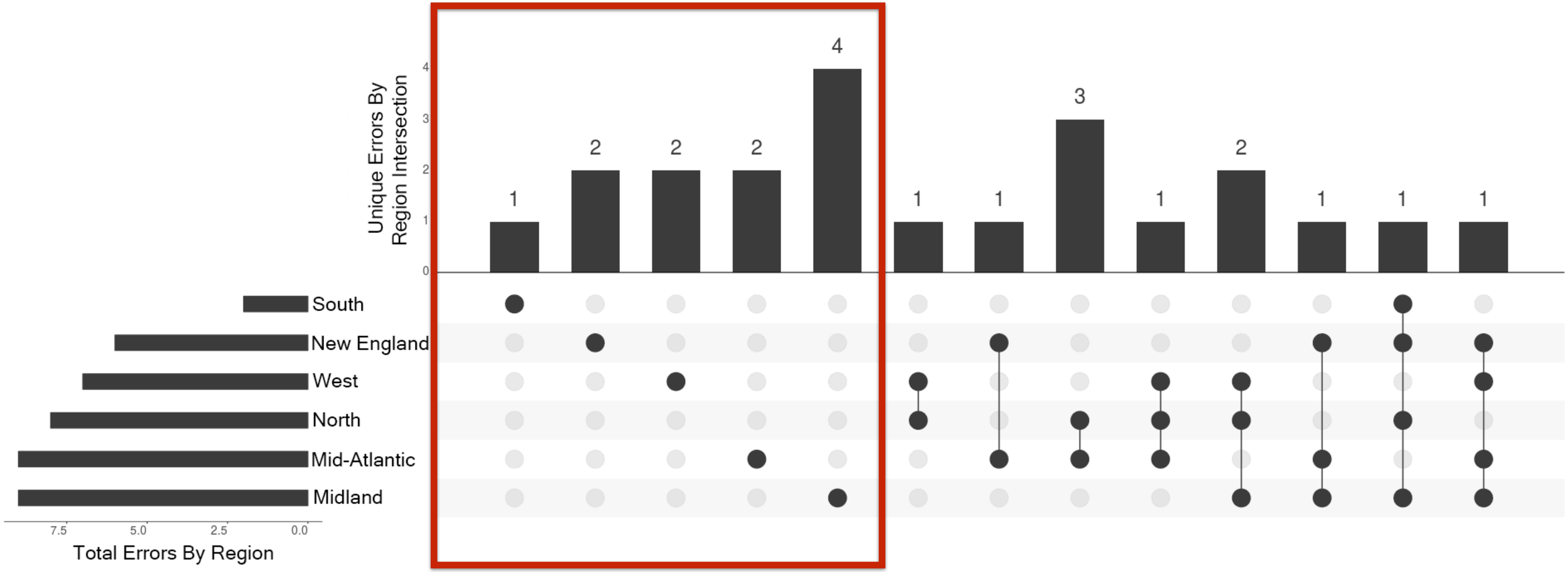
Predictable Errors by Region



Predictable Errors by Region



Predictable Errors by Region



Spear Skill Squatting Attacks

- An attacker can leverage accent-specific predictable errors in Alexa to route *specific* users to skills that they didn't intend to go to

Validating the Spear Skill Squatting Attack

Squatted Pair	Region	Target %	Overall %	Significant?
Tool/Two	South	34.0%	14.1%	Yes
Dock/Doc	West	97.4%	81.6%	No
Mighty/My T.	West	20.0%	4.1%	Yes
Exterior/Xterior	New England	42.9%	22.5%	Yes
Meal/Meow	New England	55.6%	34.3%	Yes
Wool/Well	Midland	50%	32.4%	No
Pal/Pow	Midland	65.9%	37.7%	Yes
Accuser/Who's There	Midland	26.0%	4.9%	Yes
Pin/Pen	Midland	26.3%	10.0%	Yes
Malfunction/No Function	Mid-Atlantic	36.0%	27.5%	No
Fade/Feed	Mid-Atlantic	59.0%	14.7%	Yes

Validating the Spear Skill Squatting Attack

Squatted Pair	Region	Target %	Overall %	Significant?
Tool/Two	South	34.0%	14.1%	Yes
Dock/Doc	West	97.4%	81.6%	No
Mighty/My T.	West	20.0%	4.1%	Yes
Exterior/Xterior	New England	42.9%	22.5%	Yes
Meal/Meow	New England	55.6%	34.3%	Yes
Wool/Well	Midland	50%	32.4%	No
Pal/Pow	Midland	65.9%	37.7%	Yes
Accuser/Who's There	Midland	26.0%	4.9%	Yes
Pin/Pen	Midland	26.3%	10.0%	Yes
Malfunction/No Function	Mid-Atlantic	36.0%	27.5%	No
Fade/Feed	Mid-Atlantic	59.0%	14.7%	Yes

Validating the Spear Skill Squatting Attack

Squatted Pair	Region	Target %	Overall %	Significant?
Tool/Two	South	34.0%	14.1%	Yes
Dock/Doc	West	97.4%	81.6%	No
Mighty/My T.	West	20.0%	4.1%	Yes
Edging/Ketchup	New England	42.9%	12.2%	Yes
Meal/Meow	New England	55.6%	34.5%	Yes
Wool/Well	Midland	50%	32.4%	No
Pen/Pin	Midland	26.3%	10.0%	Yes
Accuser/Who's There	Midland	26.0%	4.9%	Yes
Pin/Pen	Midland	26.3%	10.0%	Yes
Malfunction/No Function	Mid-Atlantic	36.0%	27.5%	No
Fade/Feed	Mid-Atlantic	59.0%	14.7%	Yes

Successfully squatted 8 out of 11 spear squattable pairs

Limitations

- Scale + Representativeness of the dataset

Limitations

- Scale + Representativeness of the dataset
- Skill behavior *outside* of a development environment

Takeaways

- New medium, same problems
  - “Typosquatting” in the land of IoT

Takeaways

- New medium, same problems
  - “Typosquatting” in the land of IoT
- Opaque ML for *decision making* is still nascent
  - Interface quirks can and will be exploited to cause abuse

Moving Forward

- Working with Amazon to fix these issues in their platform

Moving Forward

- Working with Amazon to fix these issues in their platform
- Measuring the widespread harms of skill squatting

Moving Forward

- Working with Amazon to fix these issues in their platform
- Measuring the widespread harms of skill squatting
- Investigating IoT trust relationships
  - Do users intrinsically trust voice-based devices more than online?

Questions?
dkumar11@illinois.edu
@_kumarde

backup