Spying through your Voice Assistants: Realistic Voice Command Fingerprinting

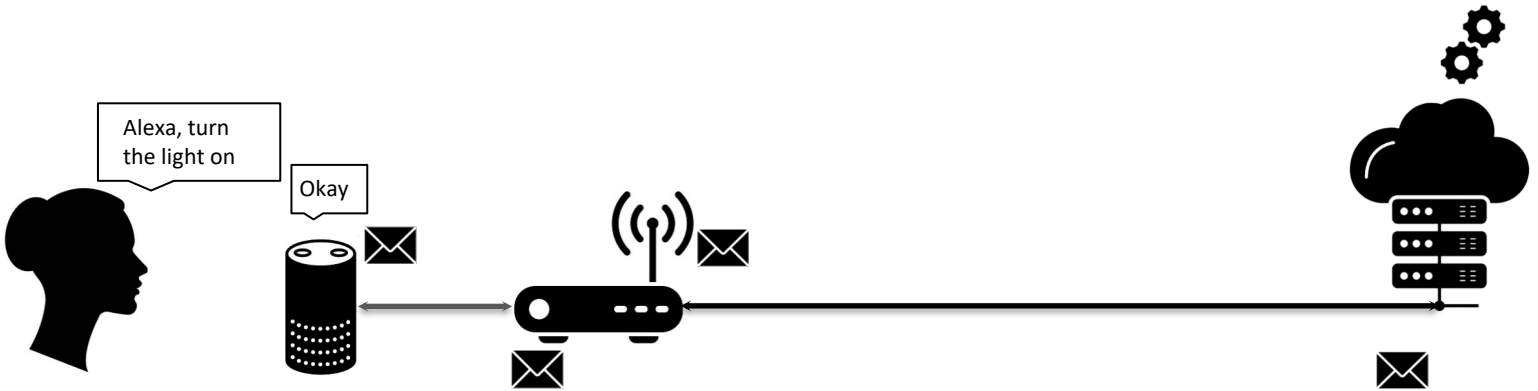
USENIX Security ‘23

Dilawer Ahmed, Aafaq Sabir, Anupam Das

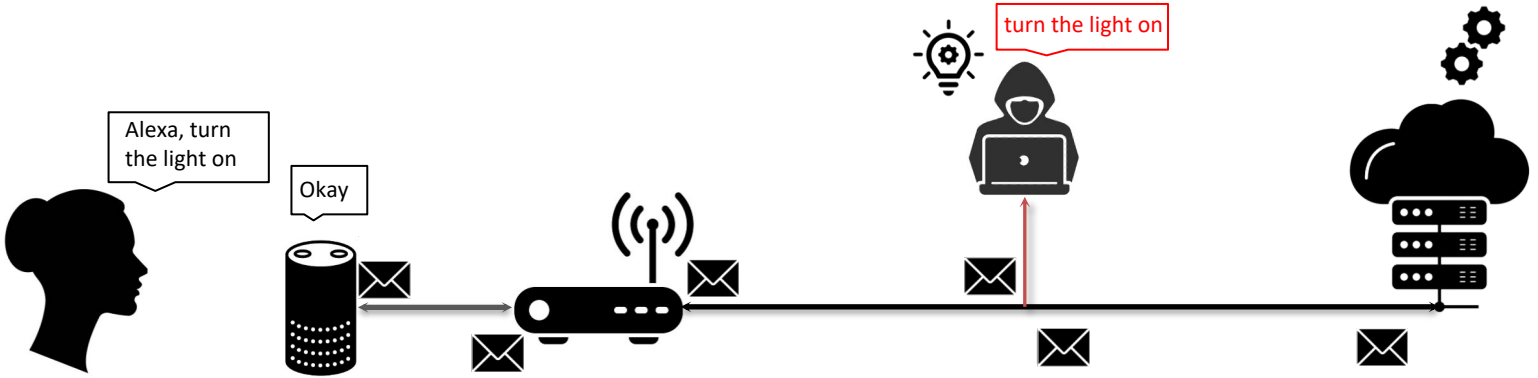
What is voice command fingerprinting?

- Identify the **activity** being performed on the voice assistant
- Privacy attack which can result in sensitive information leakage
- Using **passively** sniffed **encrypted** network traffic

How a typical voice assistant works



Voice Command Fingerprinting

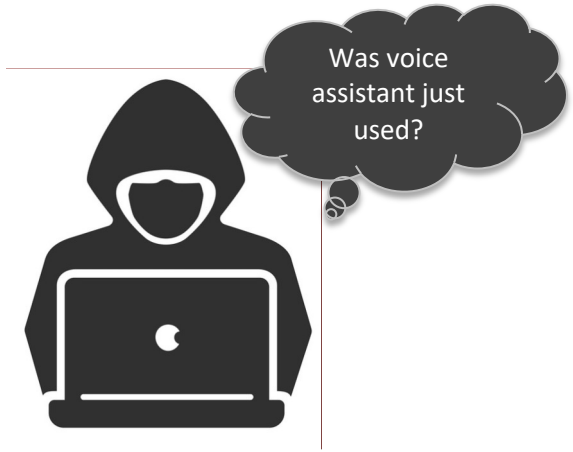


Non-local adversary challenges

- No effortless way to tell when voice assistant was used
- Due to NAT, traffic from devices isn't easily distinguishable
- We split our attack into Invocation Detection and Activity Detection
- We introduce Traffic Flow Filtering to filter noisy traffic

Invocation Detection

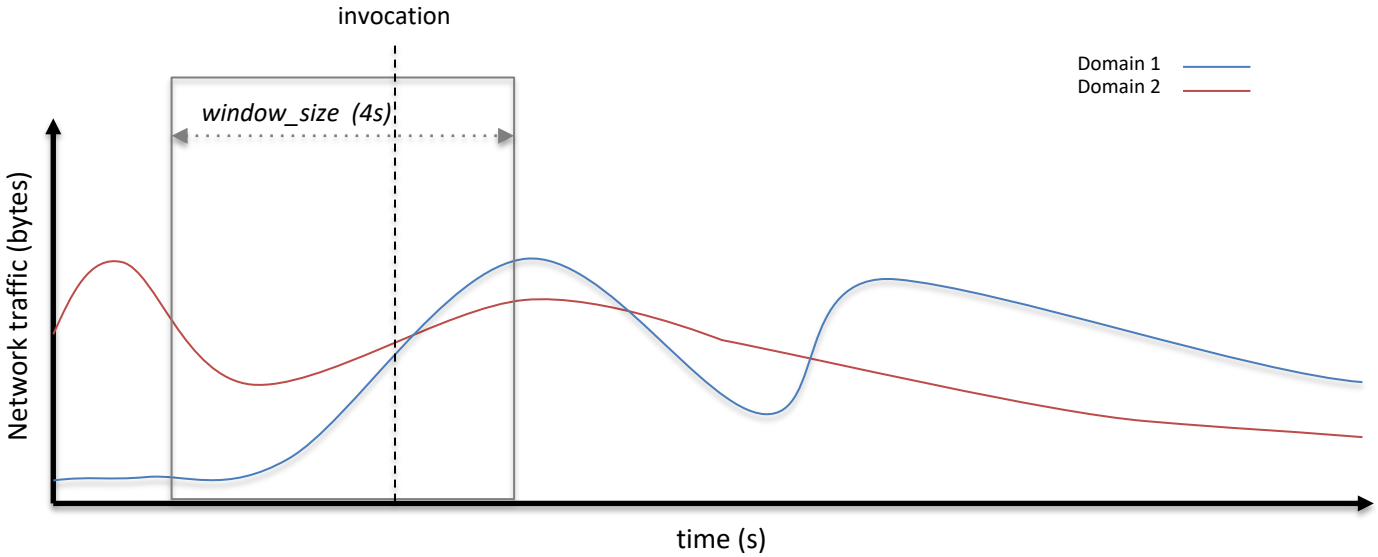
- To detect ‘activation’ (invocation) of voice assistants
- Desired properties:
  - Continuous real-time detection
  - Low or no false positives
  - Lightweight



Activity Detection

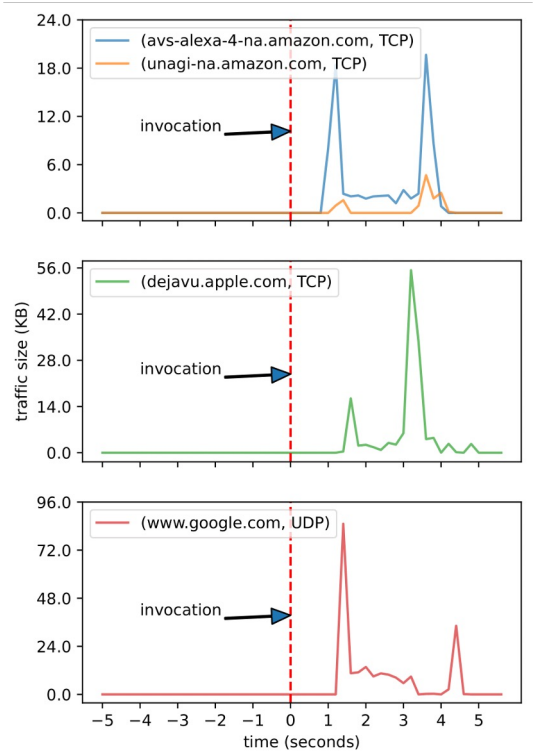
- To detect the actual ‘activity’ performed on device after ‘invocation’
- Desired Properties:
  - Performance (Accuracy)
  - Captures varied commands
  - Resistant to noise

Invocation Detection: Windows



Spikes due to Invocation

- Alexa
  - avs-alexa-4-na.amazon.com (443, TCP)
  - unagi-na.amazon.com (443, TCP)
- Siri
  - dejavu.apple.com (443, TCP)
- Google Assistant
  - www.google.com(443, UDP)

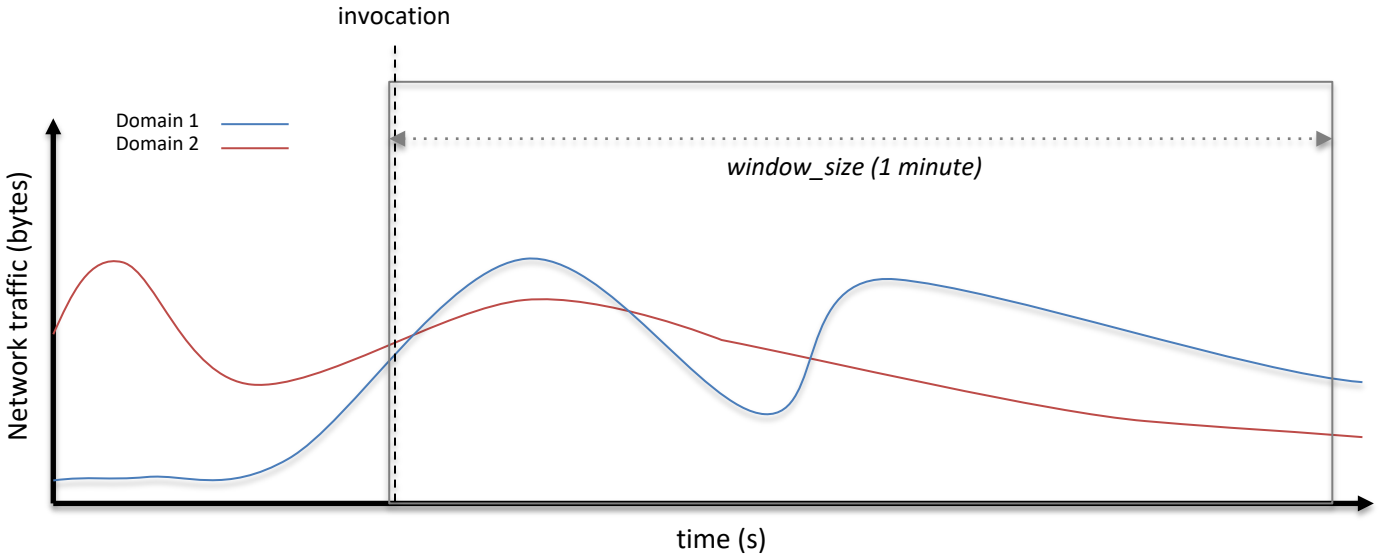


Invocation Detection: All 3 Platforms

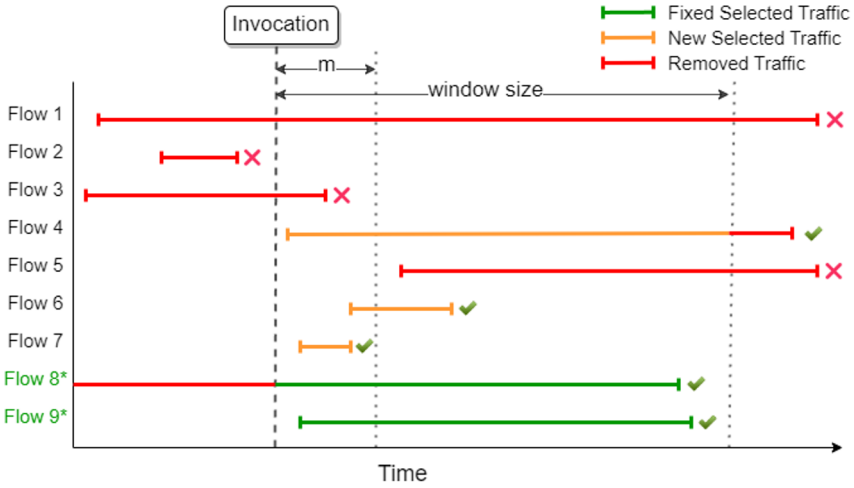
Voice Assistant	Accuracy	Precision	Recall
Alexa	99.81	99.63	100.0
Google Assistant	99.70	99.71	99.71
Siri	99.50	99.71	99.32

- Compared Multiple lightweight ML models
- Random Forest was selected based on overall performance

Activity Detection: Window



Flow Filtering Method



- Predefined fixed flows are always included
- ‘m’ second window for inclusion of flows

Activity Detection: Across platforms

Type	Accuracy	Precision	Recall	# Labels
Alexa	87.70	87.46	88.20	50
Google Assistant	92.67	92.66	92.96	50
Siri	92.80	92.91	93.18	50

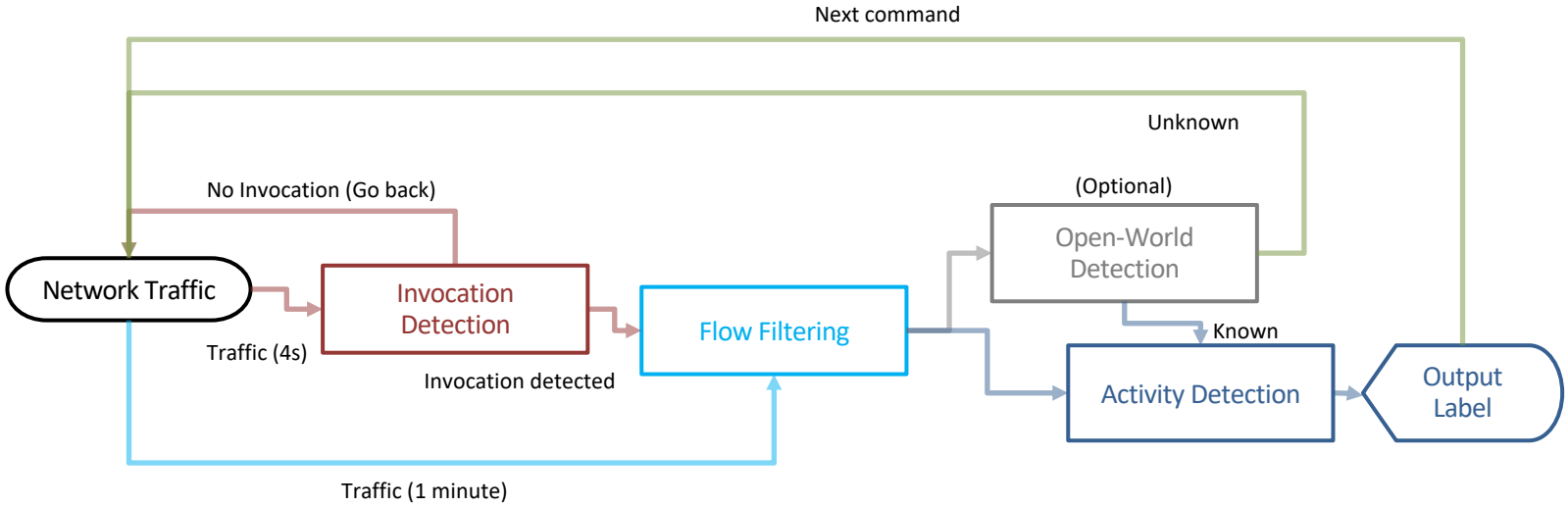
- Slightly better results for Google Assistant and Siri
- Overall good performance for all platforms

Activity Detection: Command types

Type	Accuracy	Precision	Recall	# Labels
Simple	80.30	80.58	81.44	100
Skills	82.76	85.33	82.42	100
Stream	99.39	99.30	99.34	15

- Very high accuracy for streaming commands
- Performance suffers slightly for simple commands and skills
  - Similar commands
  - Same vendor/functionality skills

Design: End-to-End Classification



End-to-End: Real-world test

- Conducted an IRB-approved study across 5 days with 15 participants using Alexa
- Used pre-trained Invocation Detection and Activity Detection models
- Realistic background noise added by participants using laptop, TV and smart phone.

End-to-End: Evaluation

- All invocations were correctly detected without any false positives (100% accuracy)
- 91% precision and 92% recall in distinguishing novel unknown voice commands
- 77% End-to-End accuracy in detecting voice commands

Limitations

- Unable to fingerprint two same-platform voice assistants if active within ‘1-minute’ of each other
- Only focus on ~100 command set at one time
- Domains for Invocation Detection are region-specific and need manual work for selection in other regions (e.g., *unagi-**na**.amazon.com)*

Contributions

- Focused on top 3 most popular platforms
- Introduced Flow Filtering and Invocation Detection
- Improved state-of-the-art in voice command fingerprinting
- Used multiple types of commands e.g., skills, streaming
- Designed an End-to-End fingerprinting method
- Code and data is open-sourced

Dilawer Ahmed (dahmed2@ncsu.edu)

https://github.com/dilawer11/va-fingerprinting