Video demo
Continuous Authentication for Voice Assistants

Huan Feng, Kassem Fawaz*, Kang G. Shin

The University of Michigan, Real Time Computing Laboratory
An Emerging Threat

Attacking Voice Assistants

Design Overview

System & Threat Models

Implementation

VAuth

Preliminary Results

Precision, Effectiveness, Security
Authentication Designed for Touch Screen

- Assumes *direct* physical contact
- Assumes *exclusive* control of the device
  - No need to distinguish the operator and the owner
Why They Fail

• Assumes *direct* physical contact (Fails on usability)

• Assumes *exclusive* control of the device (Fails on security)
Existing Attempts

• Google Now & Siri try to provide voice authentication
  • Very insecure, and subject to replay attacks
  • Only for usability purposes

• Banks use voice fingerprint for continuous authentication
  • Requires extensive learning
  • Also subject to replay attacks

• We call for: a contactless (for usability) and continuous (for security) authentication for voice assistants
Design Overview

An Emerging Threat
Attacking Voice Assistants

System & Threat Models

Implementation
VAuth

Evaluation
Effectiveness & Security
Threat Model

• Attackers that steal private information or perform unauthorized operations by exploiting voice assistants

• Scenario A – Stealthy Attack
  • Inaudible or incomprehensible

• Scenario B – Biometric-override Attack
  • Impersonation or replay

• Scenario C – Acoustic Injection Attack
  • High energy vibrations
System Model

• Utilize *on-body vibrations* as physical assurance

• A wearable secure token
  • Mounted on chest, neck, facial area
  • Continuously uploading data from accelerometer
  • Do not share with others

• Extended voice assistant (Siri, Google Now)
  • Performs real-time matching between *acc* and *mic*
  • Only issues commands when match
Design Overview

An Emerging Threat
Attacking Voice Assistants

Design Overview
System & Threat Models

Implementation
VAuth

Evaluation
Effectiveness & Security
Architecture

[Diagram: Architecture flowchart showing components such as Bluetooth Chip, Controller, AD Converter, Accelerometer, Hotword Detector, Audio Session, Acc Session, Matching Engine, Voice To Text, and Query Engine, connecting to External Wearable (e.g., earbuds, necklace) and Voice Assistant.]
Wearable Component

(a) Wireless
• Smaller than a coin
• Sensitive (11k Hz)

(b) Eyeglasses
• Bluetooth Transmitter
  • Analog-to-digital conversion
  • Bluetooth transmission
Voice Assistant Extension

• How to integrate VAuth with Google Now
  • Start collecting acc signal after “OK Google”
  • Real-time matching between acc and mic signal
  • Block the execution of voice command

• Google Now is closed source!
  • VAuth as a standalone user-level service
  • Two RPC methods: start() and end()
  • start() -> #onHotWordDetection event (logcat)
  • end() -> intercept Intent action
Usability

• VAuth requires wearing a security-assisting device
  • Wear an additional device for security
  • Embed VAuth in existing wearable products

(a) Earbuds  (b) Eyeglasses  (c) Necklace
Usability Survey

• Study the acceptance of VAuth on **952** individuals
  • Amazon Mechanical Turk (US, 40% female)
  • Prior experience with voice assistants
  • How much users agree with a statement (scale of 1-7)

• Key findings:
  • 71% find this type of attack *dangerous*
  • 75% are willing to *take steps* to mitigate
  • 70% are willing to *wear* VAuth (1 of the 3 wearables)
Usability Survey
Matching Algorithm

- Pre-processing:
  - Highpass filter -> Re-sample -> Normalization
Matching Algorithm

• Per-segment Analysis
  • Identify glottal cycles (periodical patterns)
  • Keep segments that match human speech
Matching Algorithm

• Matching Decision
  • Input: the *surviving* segments of *acc* and *mic* signals
  • Output: boolean decision (matching/non-matching)
Matching Algorithm

• SVM-based decision making
  • Feature: the vector of XCorr
  • Training Set: One user on all 44 English phonemes
Phonetic-level Analysis

• The basic unit of human speech
  • 44 phonemes (20 vowels and 24 consonants)
  • Two speakers (one male and one female)
  • A lower-bound of whole-command matching

<table>
<thead>
<tr>
<th>microphone</th>
<th>accelerometer</th>
<th>TP (%)</th>
<th>FP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>consonants</td>
<td>consonants</td>
<td>90</td>
<td>0.2</td>
</tr>
<tr>
<td>consonants</td>
<td>vowels</td>
<td>-</td>
<td>1.0</td>
</tr>
<tr>
<td>vowels</td>
<td>consonants</td>
<td>-</td>
<td>0.2</td>
</tr>
<tr>
<td>vowels</td>
<td>vowels</td>
<td>100</td>
<td>1.7</td>
</tr>
<tr>
<td>all</td>
<td>all</td>
<td>94</td>
<td>0.7</td>
</tr>
</tbody>
</table>
Phonetic-level Analysis

• Idle Detection
  • No single match of a false matching

(a) Low energy and white noise

(b) High energy and periodic noise
Design Overview

An Emerging Threat
Attacking Voice Assistants

Implementation
VAuth

Design Overview
System & Threat Models

Evaluation
Effectiveness & Security
User Study

• Matching Accuracy (True Positives) & False Positives

• 18 users speaking 30 commands under 6 scenarios
  • Three positions (eyeglasses, earbuds, necklace)
  • Two mobility patterns (still and jogging)
Still

- Matching Accuracy (>97%) & False Positives (0.09%)
Jogging

• Comparable performance, in some cases better
Different Languages

• 4 languages (Arabic, Chinese, Korean, Persian)
Takeaway

• Voice is an open channel

• Properties that work on touch interface does not naturally transcend to voice interface.

• Couple the open channel with physical assurances