Authentication: Attacks and Defenses

TEJAS SHAH, SUMEDH SAWANT
Overview of Authentication

- **User authentication** is the process of verifying the validity of a claimed user.
  - Knowledge-Based Authenticators (Passwords)
  - Object-Based Authenticators (Tokens/keys)
  - ID-Based Authenticators (Biometric)

- **Machine authentication** is the process of verifying the validity of a machine which is attempting to access or provide a resource.
  - Asymmetric Cryptography (SSL/TLS)
  - Mutual authentication (Kerberos)
Single Sign On

- Allow a user who has been authenticated by one machine to have access to other machines without further user authentication.

- Common use case for machine authentication
  - SSH keys
  - Sign in with Google
  - Kerberos
Generic SSO Architecture
Pass the hash

A common attack against single sign on architectures.

- Assumes attacker has gained admin access to a machine in the network
Kerberos
Microsoft Windows single sign on login service
Pass the ticket

A variant of pass-the-hash for Kerberos
Pass-the-hash traditional defenses

- Grant lowest necessary privileges to users
- Minimize admin log-ins to less secure machines
- Traditional host and network intrusion detection monitoring
  - Monitor for newly created accounts
  - AV process monitoring/automatic restart
  - Anomaly monitoring (e.g., watch for systems making connections to many hosts in a short time)
Pass-the-ticket attack detection algorithm

- Build a graph $G(V,E)$:
  - $V$={IP addresses in network $U$ IP addresses accessing network}
  - $E$={$(u,v)$ | $u$ in $V$, $v$ in $V$, the web log contains an event where source=$u$, dest=$v$}
- Sparsify the graph
  - Random sampling or adjacency matrix sparsification
- Paths in the graph from an external IP to a server containing sensitive information are potential pass-the-hash attacks
Pass-the-ticket attack risk level evaluation

- Build a graph $G(V,E)$:
  - $V=$ {IP addresses in network U IP addresses accessing network}
  - $E=$ {$(u,v) \mid u$ in $V$, $v$ in $V$, there exists a user X who has logged into $u$ and is an admin on $v$}
  - Approximate $E$ by looking for events in the web log that indicate the desired relationship
- Sparsify the graph
  - Random sampling or adjacency matrix sparsification
- Paths in the graph from an external IP to a server containing sensitive information are paths at risk of pass-the-hash attacks
Summary

- Authentication is at the crux of computer security
- Authentication is a huge topic, covering both user and machine authentication
- SSO is a convenient and effective way to save users and admins time, but introduces a single point of failure
- Admins have long tried to increase the security of SSO systems by enforcing best practices, to limited success
- Researchers are now exploring data mining techniques to detect and prevent common attacks against SSO systems
Questions?
Detecting Compromised Accounts

Curran Kaushik, Alex Zamoshchin
Social Network Issues

- Spam, phishing, and malware are real threats on social networking sites
- Large-scale malware campaigns have been carried out over social networks
- 83% of social network users received at least one unwanted message each year
hey
hey
hey

hey, how are you?

haha was that you

5370.meredith8.SkankProfile4.com
Hi. It's me.
Between You and Steve

Steve  
August 24 at 1:18am
Your ass looks not bad in this video.: http://youtube%66iles.%35q%2E%70l/?a=F0F2EFE6E9ECE5AE1EBAEE6E1E3E5E2EFEFEBAE3EFEDEAF6B2B2B7AFB2B9B5AFB2B8AFEED5B9B0B0B3B8B6866BD8FB5B2B0B7AEEAF0E7&b=D3F4E5F6E5A0C8E1F2F4EDE1EE

Joshua Baer  
August 24 at 7:05am
Looks like some kind of virus. Glad I use a Mac. Did you really send this?

Steve  
August 24 at 9:30am
Nope, it's a virus.
Dear Friends, as most of you realize, my fb page has obviously been hacked. I'm sorry you have to see all the stupid, obscene posts that are popping up. Please ignore as we are working with fb to take care of this problem. I appreciate your patience.
#BREAKING: US Air Force One crash feared as air traffic controller loses contact with pilot over Russian air space.
Like · Comment · Share · 👍 19 · 👤 5 · 📢 3

#BREAKING: Vice President Joe Biden to address nation in 15 minutes.
Like · Comment · Share · 👍 89 · 👤 41 · 📢 44

We are aware that our Facebook page was compromised during the last 20 minutes. We have deleted the posts and are looking into it.
Like · Comment · Share · 👍 1,486 · 👤 384 · 📢 252
Breaking: Two Explosions in the White House and Barack Obama is injured
Compromised Accounts

- Accounts of real users that have been compromised
  - Not fake accounts

- For each user, we associate a behavior profile
  - A message that appears to be very different from a user’s typical behavior might indicate a compromise.

- Approach
  1. Check for a set of similar messages
  2. Require that a significant subset of these messages violate the behavioral profiles of their senders
Behavior Profiles

- List of all messages that the user has posted on the social network, in chronological order
  - Need a minimum number $S = 10$ messages

- Features
  - Time (hour of day)
  - Message Source
  - Message Text (Language)
  - *Message Topic
  - *Links in Messages
  - *Direct User Information
  - Proximity

* = optional
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg # conn. of neighbors</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg messages of neighbors</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friends to Followers (F2F)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2F of neighbors</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mutual links</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User distance</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Message Features</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suspicious content</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>URL blacklist</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friends features</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friend name entropy</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of friends</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profile age</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stream Features</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity per day</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Applications used</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Following Rate</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Message length</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Messages sent</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Message similarity</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Message timing</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retweet ratio</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topics</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>URL entropy</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>URL ratio</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>URL repetition</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User interaction</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Features evolving with different sizes of training sets. Each experiment was conducted 25 times on random subsets of 25%, 50%, 70%, 90%, and 99% of the 5,236 labeled training instances. The fraction of positive to negative samples remained constant.
# False Positives

<table>
<thead>
<tr>
<th>Network &amp; Similarity Measure</th>
<th>Twitter Text</th>
<th></th>
<th>Twitter URL</th>
<th></th>
<th>Facebook Text</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Groups</td>
<td>Accounts</td>
<td>Groups</td>
<td>Accounts</td>
<td>Groups</td>
<td>Accounts</td>
</tr>
<tr>
<td>Total Number</td>
<td>374,920</td>
<td>14,548</td>
<td>48,586</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Compromised</td>
<td>9,362</td>
<td>343,229</td>
<td>1,236</td>
<td>54,907</td>
<td>671</td>
<td>11,499</td>
</tr>
<tr>
<td>False Positives</td>
<td>4% (377)</td>
<td>3.6% (12,382)</td>
<td>5.8% (72)</td>
<td>3.8% (2,141)</td>
<td>3.3% (22)</td>
<td>3.6% (412)</td>
</tr>
<tr>
<td># Bulk Applications</td>
<td>12,347</td>
<td></td>
<td>1,569</td>
<td></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td># Compromised Bulk Applications</td>
<td>1,647</td>
<td>178,557</td>
<td>251</td>
<td>8,254</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>False Positives</td>
<td>8.9% (146)</td>
<td>2.7% (4,854)</td>
<td>14.7% (37)</td>
<td>13.3% (1,101)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td># Client Applications</td>
<td>362,573</td>
<td></td>
<td>12,979</td>
<td></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td># Compromised Client Applications</td>
<td>7,715</td>
<td>164,672</td>
<td>985</td>
<td>46,653</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>False Positives</td>
<td>3.0% (231)</td>
<td>4.6% (7,528)</td>
<td>3.5% (35)</td>
<td>2.2% (1,040)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
False Negatives

Figure 2. Probability of false positives depending on the amount of historical data on Twitter
Limitations

• Impossible if not implemented by the social networks themselves

• Attackers can post messages that align with the sender’s behavior

• Attackers can post messages that evade similarity measures
  o For example, an attacker can use dynamic URLs
    • In response, the system could take into consideration landing page
Sybil Detection

• Simply detecting fake accounts is an equally hard problem

• Due to high false positive rate, manual inspection is needed
  o Can be used to decide when to present the users with CAPTCHAs

• ML performs poorly
  o Able to detect only 20% of fakes deployed, and almost all the detected accounts were flagged by users
Model

• Undirected Graph, $G = (V,E)$

Figure 2: Non-Sybil region, Sybil region, and attack edges in an OSN under a Sybil attack. All Sybils created by malicious users are placed into the Sybil region. The Sybil collective may not be well connected.
Propagating Trust

\[ T^{(0)}(v) = \begin{cases} \frac{T_G}{K} & \text{if node } v \text{ is one of the } K \text{ trust seeds} \\ 0 & \text{else} \end{cases} \]

\[ T^{(i)}(v) = \sum_{(u,v) \in E} \frac{T^{(i-1)}(u)}{\text{deg}(u)} \]

Terminate after log(n) operations
<table>
<thead>
<tr>
<th>Social Network</th>
<th>Nodes</th>
<th>Edges</th>
<th>Clustering Coefficient</th>
<th>Diameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>10,000</td>
<td>40,013</td>
<td>0.2332</td>
<td>17</td>
</tr>
<tr>
<td>ca-AstroPh</td>
<td>18,772</td>
<td>198,080</td>
<td>0.3158</td>
<td>14</td>
</tr>
<tr>
<td>ca-HepTh</td>
<td>9,877</td>
<td>25,985</td>
<td>0.2734</td>
<td>18</td>
</tr>
<tr>
<td>Synthetic</td>
<td>10,000</td>
<td>39,399</td>
<td>0.0018</td>
<td>7</td>
</tr>
<tr>
<td>wiki-Vote</td>
<td>7,115</td>
<td>100,736</td>
<td>0.1250</td>
<td>7</td>
</tr>
<tr>
<td>soc-Epinions</td>
<td>10,000</td>
<td>222,077</td>
<td>0.0946</td>
<td>6</td>
</tr>
<tr>
<td>soc-Slashdot</td>
<td>10,000</td>
<td>153,404</td>
<td>0.0582</td>
<td>4</td>
</tr>
<tr>
<td>email-Enron</td>
<td>10,000</td>
<td>105,343</td>
<td>0.1159</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 1: Social graphs used in our experiments. The last three graphs are 10K-node BFS samples.
Results
Limitations

• Multiple communities

• Sophisticated attackers may obtain knowledge of seeds and established “attack edges” close to the seed nodes
Resources

• Egele, Manuel, et. al, “COMPA: Detecting Compromised Accounts on Social Networks”

• Cao, Qiang, et. al, “Aiding the Detection of Fake Accounts in Large Scale Social Online Services”

data mining

WITH

security
Normalization
Anomaly detection

“Professor Plum did the Study with the 5 machine botnet”
What about the attackers?
DATA MINING!
Think more… analog.
LOW-tech
DATA MINING

Data mining for fun and profit

Derrick Liu + Daniel Chiu
War dialing
War driving
Search engines
War dialing
What is war dialing?
<table>
<thead>
<tr>
<th></th>
<th>650-555-2363</th>
</tr>
</thead>
<tbody>
<tr>
<td>04</td>
<td>650-555-2364</td>
</tr>
<tr>
<td>05</td>
<td>650-555-2365</td>
</tr>
<tr>
<td>06</td>
<td>650-555-2366</td>
</tr>
<tr>
<td>07</td>
<td>650-555-2367</td>
</tr>
<tr>
<td>08</td>
<td>650-555-2368</td>
</tr>
<tr>
<td>09</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>

Sounds like a modem!

Flag it for later human reference.
War dialing can reveal significant lapses in security.
Many sensitive systems are still connected to the internet via dial-up.
ATM demo video
Dillinger – Remote ATM Attack/Admin Tool

- Exploits an authentication bypass vuln in remote monitoring
- Remote monitoring enabled by *default*
- Attack is 100% reliable
- Supports dial-up and TCP/IP
- ~95% of retail ATMs using dial-up
- Voip based dialing tools (ex: WarVox) make exchange scanning practical
LOW-tech DATA MINING

War dialing
War driving
Search engines
War driving
In a nutshell
WiFi security is getting better.
There’s another wireless target out there.

With

“secret” and “secure” networking and

“patented, proprietary encryption”
Infrastructure!
Credit where it’s due.

The tech is (sometimes) really old.

Wireless is really cheap compared to wires / people.

The use cases didn’t have a good standard to follow.

Product of evolution.

Infrastructure is really expensive.
Lots of problems, though

Everyone has physical access to these things
Everyone uses different tech and different “standards”
Lack of security mindset
Needs to survive for a very long time, slow updates
An example

$40 + $0
LOW-tech
DATA MINING

- War dialing
- War driving
- Search engines
Search engines are really good at finding things.
Unsecured surveillance cameras

2130R PTZ Network Camera - Site Web officiel
www.sony.ca/view/view.shtml
For this result is not available because of this site's robots.txt – learn

Straight Bourbon-Cam (LIVE) / Sonstige-Cams
www.trace.com/view/view.shtml
For this result is not available because of this site's robots.txt – learn

World, Wide, Collection, Internet
www.onde.de/view/view.shtml
Double with online offline status, add new to our collection.

view.shtml" - LIVE webcam directory
vue.com/s.asp?s=368&search=inurl:view/view.shtml
Om Fullscreen webcams: Live webcams.

PTZ Network Camera
dcam.com/webcam-index.php?var.../view/view.shtml
• Close • Smooth Iris • Iris Open, Open. AutoIris. FOCUS Near, Focus

axis-cgi/jpg - AMOS - Washington University in St. Louis
amos.cse.wustl.edu/browse?query=/axis-cgi/jpg
Browse Cameras: /axis-cgi/jpg. Next Pages. [+] Ignore dead cameras. Order by size:

axis-cgi/jpg: AXIS 206 Network Camera ,AXIS 207 Network ...
axis-cgi-jpg.blogspot.com/2011/.../axis-206-network-camera-axis-207.ht...
AXIS 206 Network Camera ,AXIS 207 Network Camera ,Axis 2100 Network Cameras ,AXIS 211 Network Camera ,AXIS 215 PTZ Network Cameras ,AXIS 221 ...

inurl:axis-cgi/jpg - LIVE webcam directory - webcamVue.com
www.webcamvue.com/s.asp?s=186&search=inurl:axis-cgi/jpg
webcamVue.com Fullscreen webcams: Live webcams.

Brig Saltinabrücke
81.201.204.198/axis-cgi/jpg/image.cgi?resolution=640x480
A description for this result is not available because of this site's robots.txt – learn more.

Penn State Hazelton Campus - Site Web officiel
hncam1.hn.psu.edu/axis-cgi/jpg/image.cgi?resolution=320x240
Exposed server info

About 12,900 results (0.23 seconds)

**PHP: phinfo - Manual**
Outputs a large amount of information about the current state of PHP. This includes information about PHP compilation options and extensions, the PHP version, ...

**How can I create a phinfo.php page? - KnowledgeBase**
kb.mediatemple.net/.../764/How+can+i+create+a+phinfo.php+page%3F
Overview. You can use a phinfo() page to view the current PHP information for your server. This file outputs a large amount of information, such as: Information ...

**4WebHelp - Scripts: PHP: phinfo**
www.4webhelp.net/scripts/php/phinfo.php
Comments. Name: Mike T, Email mike dot tilder at drop dot mail dot netropic dot co dot uk. Thanks. I've been searching for a while to find out how to do this!

**ubuntu - localhost/phinfo.php - Stack Overflow**
stackoverflow.com/questions/11682662/localhost-phinfo-ph
Jul 27, 2012 - I had a similar problem but the reason was because I had just restored my files into www from a Windows NTFS backup drive. Naturally, with NTFS ...

**Secure your phinfo.php files with .htaccess**
Exposed server info

**PHP: phpinfo - Manual**
php.net/manual/en/function.phpinfo.php ➜ PHP ➜
Outputs a large amount of information about the current state of PHP. This includes information about PHP compilation options and extensions, the PHP version, ...

**How can I create a phpinfo.php page? - KnowledgeBase**
kb.mediatemple.net/.../764/How+can+I+create+a+phpinfo.php+page%3F ➜
Overview. You can use a phpinfo() page to view the current PHP information for your server. This file outputs a large amount of information, such as: Information ...

**4WebHelp - Scripts: PHP: phpinfo**
www.4webhelp.net/scripts/php/phpinfo.php ➜
Comments. Name: Mike T, Email mike dot tilder at drop dot mail dot netropic dot co dot uk. Thanks. I've been searching for a while to find out how to do this!

**ubuntu - localhost/phpinfo.php - Stack Overflow**
stackoverflow.com/questions/11682662/localhost-phpinfo-php ➜
Jul 27, 2012 - I had a similar problem but the reason was because I had just restored my files into www from a Windows NTFS backup drive. Naturally, with NTFS ...

Secure your phpinfo.php files with .htaccess | Parsible
Some search engines are tailor made for finding vulnerabilities.
SHODAN

EXPOSE ONLINE DEVICES.
WEBCAMs. ROUTERS. POWER PLANTS. iPHONES. WIND TURBINES. REFRIGERATORS. VOIP PHONES.

TAKE A TOUR FREE SIGN UP

Popular Search Queries: IPads - IPads. Think different. Think no security.

DEVELOPER API
Find out how to access the Shodan database with Python, Perl or Ruby.

LEARN MORE
Get more out of your searches and find the information you need.

FOLLOW ME
Contact me and stay up to date with the latest features of Shodan.

IN THE PRESS
Shodan pinpoints shoddy industrial controls.
It greatly lowers the technical bar needed to canvas the Internet...
'Shodan for Penetration Testers' presented at DEF CON 18
It's a reminder to many to know what's on your network...

The Register threatpost DEFCON18 darkREADING
How do we fix this stuff?
We’re getting better at it

Adopt better disclosure practices for new vulnerabilities

Regular audits and pen-testing, especially for non-traditional attack vectors

Companies need to prioritize secure designs and be more transparent

Need to update faster in order to keep up with attackers

Physical security is no longer enough!

Use robots.txt
THANKS
Abstract

We show that the MEMS gyroscopes found on modern smartphones are sufficiently sensitive to measure acoustic signals in the vicinity of the phone. The resulting signals contain only very low-frequency information (<200Hz). Nevertheless, we show, using signal processing and machine learning, that this information is sufficient to identify speaker information and even parse speech. Since iOS and Android require no special permissions to access the gyroscope, our results show that apps and active web content that cannot access the microphone can nevertheless eavesdrop on speech in the vicinity of the phone.

1 Introduction

Modern smartphones and mobile devices have many sensors that can be used to make the user experience better. Being generally put to good use, they can sometimes unintentionally expose information that the user does not want to share. While the privacy risks associated with some sensors like a microphone (eavesdropping), camera or GPS (tracking) are obvious and well understood, some of the risks remained under the radar for users and application developers. In particular, access to motion sensors such as gyroscopes and accelerometers is unmitigated by mobile operating systems. Namely, every application installed on a phone and every web page browsed over it can measure and record these sensors without the user being aware of it.

Recently, a few research works pointed out unintended information leaks using motion sensors. In Ref. [34] the authors suggest a method for user identification from gait patterns obtained from a mobile device's accelerometers. The feasibility of utilizing the gyroscopic sensors has been shown in Ref. [35]. In Ref. [21], the authors demonstrate the potential of using gyroscope measurements to improve the accuracy of key patterns on mobile devices. In Ref. [20], the authors utilize the gyroscope information to construct a model for speech recognition.

In this paper, we reveal a new way to extract information from gyroscope measurements. We show that gyroscopes are sufficiently sensitive to measure acoustic vibrations and that they can be used to record voice signals. This leads to the possibility of recovering speech from gyroscope readings, namely using the gyroscope as a crude microphone. We show that the sampling rate of the gyroscope is up to 200 Hz which covers some of the audible range. This raises the possibility of eavesdropping on speech in the vicinity of a phone without access to the real microphone.

As the sampling rate of the gyroscope is limited, one cannot fully reconstruct a comprehensible speech signal from gyroscope readings. However, we show that by combining the signals from two or more gyros, we can increase the effective sampling rate of the acoustic signal and achieve better speech recognition rates. In our experiments, we achieved 77% successful recognition rate for speaker-dependent cases and 65% success rate for speaker-independent cases. This capability allows an attacker to substantially leak information about numbers spoken over a phone (e.g., credit card numbers or social security numbers). In Section 4, we suggest a method for audio signal recovery using samples from multiple devices. In Section 5, we discuss more directions for exploitation of gyroscopes' acoustic sensitivity. Finally, in Section 6, we discuss mitigation measures of this unexpected threat.
2.2 Acoustic Effects

MEMS gyros are susceptible to acoustic noise which degrades their accuracy [22, 24, 25]. An acoustic signal affects the gyroscope measurement by making the driving masses vibrate in the sensing axis (the axis which senses the angular rate of different axes, while others use multiple orthogonal to the primary vibration movement. Some MEMS gyros are susceptible to acoustic noise which degrades their accuracy [22, 24, 25]. An acoustic signal affects the gyroscope measurement by making the driving masses vibrate in the sensing axis (the axis which senses the angular rate of different axes, while others use multiple orthogonal to the primary vibration movement. Some MEMS gyros are susceptible to acoustic noise which degrades their accuracy [22, 24, 25]. An acoustic signal affects the gyroscope measurement by making the driving masses vibrate in the sensing axis (the axis which senses the angular rate of different axes, while others use multiple orthogonal to the primary vibration movement. Some MEMS gyros are susceptible to acoustic noise which degrades their accuracy [22, 24, 25]. An acoustic signal affects the gyroscope measurement by making the driving masses vibrate in the sensing axis (the axis which senses the angular rate of different axes, while others use multiple orthogonal to the primary vibration movement. Some MEMS gyros are susceptible to acoustic noise which degrades their accuracy [22, 24, 25]. An acoustic signal affects the gyroscope measurement by making the driving masses vibrate in the sensing axis (the axis which senses the angular rate of different axes, while others use multiple orthogonal to the primary vibration movement. Some MEMS gyros are susceptible to acoustic noise which degrades their accuracy [22, 24, 25]. An acoustic signal affects the gyroscope measurement by making the driving masses vibrate in the sensing axis (the axis which senses the angular rate of different axes, while others use multiple orthogonal to the primary vibration movement. Some MEMS gyros are susceptible to acoustic noise which degrades their accuracy [22, 24, 25]. An acoustic signal affects the gyroscope measurement by making the driving masses vibrate in the sensing axis (the axis which senses the angular rate of different axes, while others use multiple orthogonal to the primary vibration movement. Some MEMS gyros are susceptible to acoustic noise which degrades their accuracy [22, 24, 25]. An acoustic signal affects the gyroscope measurement by making the driving masses vibrate in the sensing axis (the axis which senses the angular rate of different axes, while others use multiple orthogonal to the primary vibration movement. Some MEMS gyros are susceptible to acoustic noise which degrades their accuracy [22, 24, 25]. An acoustic signal affects the gyroscope measurement by making the driving masses vibrate in the sensing axis (the axis which senses the angular rate of different axes, while others use multiple orthogonal to the primary vibration movement. Some MEMS gyros are susceptible to acoustic noise which degrades their accuracy [22, 24, 25]. An acoustic signal affects the gyroscope measurement by making the driving masses vibrate in the sensing axis (the axis which senses the angular rate of different axes, while others use multiple orthogonal to the primary vibration movement. Some MEMS gyros are susceptible to acoustic noise which degrades their accuracy [22, 24, 25]. An acoustic signal affects the gyroscope measurement by making the driving masses vibrate in the sensing axis (the axis which senses the angular rate of different axes, while others use multiple orthogonal to the primary vibration movement. Some MEMS gyros are susceptible to acoustic noise which degrades their accuracy [22, 24, 25]. An acoustic signal affects the gyroscope measurement by making the driving masses vibrate in the sensing axis (the axis which senses the angular rate of different axes, while others use multiple orthogonal to the primary vibration movement. Some MEMS gyros are susceptible to acoustic noise which degrades their accuracy [22, 24, 25]. An acoustic signal affects the gyroscope measurement by making the driving masses vibrate in the sensing axis (the axis which senses the angular rate of different axes, while others use multiple orthogonal to the primary vibration movement.
2. Characteristics of a gyro as a microphone

2.3.1 Sampling

Sampling resolution is measured by the number of bits per sample. More bits allow us to sample the signal more accurately at any given time. All the latest generations of gyroscopes have a sample resolution of 16 bits [9, 12]. This is comparable to a microphone's sampling resolution used in most mobile devices and operating systems. In our experiments, we sample the angular velocity at a rate of up to 1 Hz. A gyroscope's sampling frequency is the rate at which a signal is captured. According to the Nyquist sampling theorem, sampling resolution is 0 Hz.

Due to the gyro's acoustic susceptibility, one can treat gyroscope readings as if they were sound pressure level signals.

Figure 2: InvenSense 3-axis gyro design (Taken from [43]. Figure copyright of InvenSense. Used with permission.)

Table 1 summarizes the results of our experiments.
This is the design of the gyro built into Galaxy S III.

We do not see the aliases corresponding to 180 - 200 Hz, which perhaps correspond to their harmonics. Figure 3(c) depicts a recording of a chirp in the range of 420 - 480 Hz. The aliased chirp is detectable in the range of 20 - 80 Hz, whereas harmonics corresponding to 100 Hz are not. The aliased frequencies can be seen at the base frequencies of the recorded tones, and per-
spective axis, i.e. the signal travels in parallel to the phone's axes (see Figure 4). The gyroscope senses in three axes, allowing the signal to hit it parallel to one of its three axes, the resulting samples are indis-
sible with frequency $f_s$. The values corresponding to the base frequencies of the recorded tones, and per-
spective axis, i.e. the signal travels in parallel to the phone's axes (see Figure 4). The gyroscope senses in three axes, allowing the signal to hit it parallel to one of its three axes, the resulting samples are indis-

Using the gyro, we recorded a single 280 Hz tone. An ultrasound was played at 80 Hz. The FFT magnitude at 80 Hz. The tone was recorded at each time by the Galaxy S III gyro while the phone rested at a different orientation from the gyroscope recordings gives a noticeable peak at the corresponding frequency and time values. It can be clearly seen that there is a strong signal sensed at frequency 80 Hz and 200 Hz. Again, a strong signal can be seen at the strong signals which are lower than 100 Hz.

The tone was played at a volume of 75 dB, and the sound level of a normal conversation is 57 dB which is below the sound level of a normal conversational speaking frequency beyond the limit imposed by the operating system, while WebKit and Blink based browsers does not impose stricter limits on the recording of a chirp in the range of 420 - 480 Hz. The aliased chirp is detectable in the range of 20 - 80 Hz, whereas harmonics corresponding to 100 Hz are not. The aliased frequencies can be seen at the base frequencies of the recorded tones, and per-
spective axis, i.e. the signal travels in parallel to the phone's axes (see Figure 4). The gyroscope senses in three axes, allowing the signal to hit it parallel to one of its three axes, the resulting samples are indis-

Table 1: Maximum sampling frequencies on different platforms

<table>
<thead>
<tr>
<th>Platform</th>
<th>Chrome</th>
<th>Safari</th>
<th>Opera</th>
<th>Firefox</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampled Readings</td>
<td>2.0 Hz</td>
<td>2.0 Hz</td>
<td>2.0 Hz</td>
<td>2.0 Hz</td>
</tr>
<tr>
<td>Application</td>
<td>100 Hz</td>
<td>100 Hz</td>
<td>100 Hz</td>
<td>100 Hz</td>
</tr>
</tbody>
</table>

The self noise characteristic of a microphone indicates what is the most quiet sound, in decibels, a microphone can pick up, i.e. the sound that is just over its self noise. The self noise of the gyro for aliased tones we played 150 Hz and 250 Hz tones. The lowest level of sound the gyro picked up was 67 dB and 77 dB, respectively. These are high values considering the precision of the gyro. The value of $\theta$ plays an important role in determining the upper limit of the frequency $f$ which can be correctly sensed by the gyro. For this experiment we used a Gecko based browser does not limit the sampling frequency beyond the limit imposed by the operating system, while WebKit and Blink based browsers does impose stricter limits on it.

The self noise of the gyro for aliased tones we played 150 Hz and 250 Hz tones. The lowest level of sound the gyro picked up was 67 dB and 77 dB, respectively. These are high values considering the precision of the gyro. The value of $\theta$ plays an important role in determining the upper limit of the frequency $f$ which can be correctly sensed by the gyro. For this experiment we used a Gecko based browser does not limit the sampling frequency beyond the limit imposed by the operating system, while WebKit and Blink based browsers does impose stricter limits on it.

An ultrasound was played at 80 Hz. The FFT magnitude at 80 Hz. The tone was recorded at each time by the Galaxy S III gyro while the phone rested at a different orientation from the gyroscope recordings gives a noticeable peak at the corresponding frequency and time values. It can be clearly seen that there is a strong signal sensed at frequency 80 Hz and 200 Hz. Again, a strong signal can be seen at the

The table below shows the maximum sampling frequencies on different platforms:

<table>
<thead>
<tr>
<th>Platform</th>
<th>Chrome</th>
<th>Safari</th>
<th>Opera</th>
<th>Firefox</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampled Readings</td>
<td>2.0 Hz</td>
<td>2.0 Hz</td>
<td>2.0 Hz</td>
<td>2.0 Hz</td>
</tr>
<tr>
<td>Application</td>
<td>100 Hz</td>
<td>100 Hz</td>
<td>100 Hz</td>
<td>100 Hz</td>
</tr>
</tbody>
</table>

As noted above, the sampling frequency of a given platform impacts the ability of the gyro to correctly sense audio signals of up to 100 Hz. Aliasing is a phenomenon where for a sinusoid of frequency $f$ and sampling frequency $N$ of the sample rate, the resulting samples are indis-

In the next section of this paper, we will look at how the maximum sampling frequencies on different platforms.
In this section, we show that the acoustic signal measured by a single gyroscope is sufficient to extract information about the speech signal, such as speaker characteristics and identity. Speech recognition tasks can be classified into several types according to the setup. Speech recognition can handle fluent speech or isolated words (or phrases); operate on a closed set of words (finite dictionary) or an open set; and can be speaker dependent (in which case the recognizer is trained per speaker) or speaker independent. Speech recognition is an essential component for many applications, such as virtual assistants, smart home devices, and security systems. However, speech recognition from a single device is challenging due to various factors, including noise, reverberation, and the limited bandwidth of the microphone.

Table 2: Sensed amplitude for every direction of a tone played at different orientations relative to the phone. For each orientation, the dominant sensed directions are emphasized.

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>0</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>600</th>
<th>700</th>
<th>800</th>
<th>900</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone direction</td>
<td>x</td>
<td>z</td>
<td>x</td>
<td>z</td>
<td>x</td>
<td>z</td>
<td>x</td>
<td>z</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: Coordinate system of Android and iOS.

3 Speech analysis based on a single gyroscope

Speech recognition and speaker identification by a single gyroscope is a promising approach to speech recognition from mobile devices. When the signal is recorded in parallel to the phone's axis, the gyroscope primarily senses a rotation at the z-axis. When the signal is recorded in perpendicular to the phone's z-axis, the gyroscope primarily senses a rotation at the y-axis. These findings indicate that a single gyroscope is an omni-directional audio sensor allowing it to pick up audio signals from every direction.

Speech recognition is a complex task that involves several steps, including feature extraction, modeling, and decoding. Feature extraction is the process of transforming the raw signal into a set of features that can be processed by the subsequent stages. Modeling involves estimating the probability of a speech signal given the model parameters. Decoding is the process of finding the most likely sequence of words given the acoustic model and the observed features.

Table 2: Sensed amplitude for every direction of a tone played at different orientations relative to the phone. For each orientation, the dominant sensed directions are emphasized.

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>0</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>600</th>
<th>700</th>
<th>800</th>
<th>900</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone direction</td>
<td>x</td>
<td>z</td>
<td>x</td>
<td>z</td>
<td>x</td>
<td>z</td>
<td>x</td>
<td>z</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: Coordinate system of Android and iOS.

3 Speech analysis based on a single gyroscope

Speech recognition and speaker identification by a single gyroscope is a promising approach to speech recognition from mobile devices. When the signal is recorded in parallel to the phone's axis, the gyroscope primarily senses a rotation at the z-axis. When the signal is recorded in perpendicular to the phone's z-axis, the gyroscope primarily senses a rotation at the y-axis. These findings indicate that a single gyroscope is an omni-directional audio sensor allowing it to pick up audio signals from every direction.

Speech recognition is a complex task that involves several steps, including feature extraction, modeling, and decoding. Feature extraction is the process of transforming the raw signal into a set of features that can be processed by the subsequent stages. Modeling involves estimating the probability of a speech signal given the model parameters. Decoding is the process of finding the most likely sequence of words given the acoustic model and the observed features.

Table 2: Sensed amplitude for every direction of a tone played at different orientations relative to the phone. For each orientation, the dominant sensed directions are emphasized.

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>0</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>600</th>
<th>700</th>
<th>800</th>
<th>900</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone direction</td>
<td>x</td>
<td>z</td>
<td>x</td>
<td>z</td>
<td>x</td>
<td>z</td>
<td>x</td>
<td>z</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: Coordinate system of Android and iOS.

3 Speech analysis based on a single gyroscope

Speech recognition and speaker identification by a single gyroscope is a promising approach to speech recognition from mobile devices. When the signal is recorded in parallel to the phone's axis, the gyroscope primarily senses a rotation at the z-axis. When the signal is recorded in perpendicular to the phone's z-axis, the gyroscope primarily senses a rotation at the y-axis. These findings indicate that a single gyroscope is an omni-directional audio sensor allowing it to pick up audio signals from every direction.

Speech recognition is a complex task that involves several steps, including feature extraction, modeling, and decoding. Feature extraction is the process of transforming the raw signal into a set of features that can be processed by the subsequent stages. Modeling involves estimating the probability of a speech signal given the model parameters. Decoding is the process of finding the most likely sequence of words given the acoustic model and the observed features.
It is common for various feature extraction methods to view speech as a process that is stationary for short time windows. Therefore speech processing usually involves segmentation of the signal into short (10 – 30 ms) overlapping or non-overlapping windows and operation on them. The results in a time-series of features (such as MFCCs) that capture the time-dependent behavior of the signal. If we are interested in time-independent properties we shall use spectral features or the statistics of those time-series (such as mean, variance, skewness and kurtosis).

Mel-Frequency Cepstral Coefficients (MFCC) are widely used features in audio and speech processing applications. The Mel-scale, originally used for the auditory perception, is also common in the analysis of the vocal signal. It is also common to take the first and second derivatives of the MFCC as additional features, indicative of temporal changes. Short Time Fourier Transform (STFT) is essentially a spectrogram of the signal. Windowing is applied to short overlapping segments of the signal and FFT is computed. The result captures both spectral and time-dependent features of the signal.

3.1.2 Classifiers

Support Vector Machine (SVM) is a general binary classifier, trained to distinguish two groups. We use SVM to distinguish male and female speakers. Multi-class SVMs can be constructed using multiple binary SVMs to distinguish male and female speakers, and speaker-dependent SVMs to distinguish between multiple groups. We used multi-class SVM to recognize words from a limited dictionary.

Dynamic Time Warping (DTW) is a time-series matching and alignment technique. It can be used to match time-dependent features in presence of misalignment or when the series are of different lengths. One of the challenges in word recognition is that the samples may differ in length, resulting in different number of segments used to extract features.

3.2 Speaker identification algorithm

Prior to processing we converted the gyroscope recordings to audio files in WAV format while upsampling them to 8 KHz. We applied silence removal to include only relevant information and minimize noise. Although upsampling the signal from 200 Hz to 8 KHz does not increase the accuracy of audio signal, it is more convenient to handle the WA V file at higher sampling rate with standard speech processing tools.

We used statistical features based on the first 13 MFCC computed on 40 sub-bands. For each MFCC we computed the mean and standard deviation. Those features reflect the spectral properties which are independent of the pronounced word. We also use delta-MFCC (the derivatives of the MFCC). RMS Energy and MFCC (the derivatives of the MFCC) are used to capture the time-dependent features of the signal. Short Time Energy (STEnergy) and Spectral Centroid features computed on short segments of the speech signal. Note that the gyroscope's zero-offset yields particularly noisy recordings even during unvoiced segments.

Prior to processing we converted the gyroscope recordings to audio files in WAV format while upsampling them to 8 KHz. We applied silence removal to include only relevant information and minimize noise. Although upsampling the signal from 200 Hz to 8 KHz does not increase the accuracy of audio signal, it is more convenient to handle the WA V file at higher sampling rate with standard speech processing tools.

We used statistical features based on the first 13 MFCC computed on 40 sub-bands. For each MFCC we computed the mean and standard deviation. Those features reflect the spectral properties which are independent of the pronounced word. We also use delta-MFCC (the derivatives of the MFCC). RMS Energy and MFCC (the derivatives of the MFCC) are used to capture the time-dependent features of the signal. Short Time Energy (STEnergy) and Spectral Centroid features computed on short segments of the speech signal. Note that the gyroscope's zero-offset yields particularly noisy recordings even during unvoiced segments.
Spectral Centroid statistical features. We used MIRToolbox [32] for the feature computation. It is important to note that while MFCC have a physical meaning for real speech signal, in our case of a narrow-band aliased signal, MFCC don't necessarily have an advantage, and were used partially because of availability in MIRToolbox. We attempted to identify the gender of the speaker, distinguish between different speakers of the same gender and distinguish between different speakers in a mixed set of male and female speakers. For gender identification we used a binary SVM, and for speaker identification we used multi-class SVM and GMM. We also attempted gender and speaker recognition using DTW with STFT features. All STFT features were computed with a window of 512 samples which, for sampling rate of 8 kHz, corresponds to 64 ms.

## 3.3 Speech recognition algorithm

The preprocessing stage for speech recognition is the same as for speaker identification. Silence removal is particularly important here, as the noisy unvoiced segments can confuse the algorithm, by increasing similarity with irrelevant samples. For word recognition, we are less interested in the spectral statistical features, but rather in the development of the features in time, and therefore suitable features could be obtained by taking the full spectrogram. In the classification stage we extract the same features for a sample $y$. For each possible label $l$ we obtain a similarity score of the $y$ with each sample $X_l$ corresponding to that guess in the training set. Let us denote this similarity function by $D(y, X_l)$. Since different samples of the same word can differ in length, we use DTW. We sum the similarities to obtain a total score for that guess $S_l = \sum_i D(y, X_{l,i})$. After obtaining a total score for all possible words, the sample is classified according to the maximum total score $C(y) = \text{argmax}_l S_l$.

## 3.4 Experiment setup

Our setup consisted of a set of loudspeakers including a sub-woofer and two tweeters (depicted in Figure 5). The sub-woofer was particularly important for experimenting with low-frequency tones below 200 Hz. The playback was done at volume of approximately 75 dB, which is a reasonable volume for a mobile setting. We used a microphone to record the speech signals. The microphone was positioned at the front of the room to capture the speech signal directly.

### 3.4.1 Data

Due to the low sampling frequency of the gyro, a recognition of speaker-independent general speech would be an ambitious long-term task. Therefore, in this work we focus on the recognition of isolated digits, i.e., 11 words per speaker where each speaker recorded each word twice. There are 10 speakers (5 female and 5 male). In total, there are $10 \times 11 \times 2 = 220$ recordings. Each recording was digitized at 20 kHz.

### 3.4.2 Mobile devices

We primarily concentrated on experiments using the following mobile devices:

- **DROID**: This is a sample of a corpus published in [33]. It includes speech of isolated digits, i.e., 11 words per speaker where each speaker recorded each word twice. There are 10 speakers (5 female and 5 male). In total, there are $10 \times 11 \times 2 = 220$ recordings. Each recording was digitized at 20 kHz.

- **STACI**: This setup was particularly important for our experiments. The STACI is a smartphone (decked in Figure 5) with a loudspeaker and two microphones (depicted in Figure 5). We used MIROToolbox as the recognition engine.

### 3.5 Speech recognition algorithm

The preprocessing stage for speech recognition is the same as for speaker identification. Silence removal is particularly important here, as the noisy unvoiced segments can confuse the algorithm, by increasing similarity with irrelevant samples. For word recognition, we are less interested in the spectral statistical features, but rather in the development of the features in time, and therefore suitable features could be obtained by taking the full spectrogram. In the classification stage we extract the same features for a sample $y$. For each possible label $l$ we obtain a similarity score of the $y$ with each sample $X_l$ corresponding to that guess in the training set. Let us denote this similarity function by $D(y, X_l)$. Since different samples of the same word can differ in length, we use DTW. We sum the similarities to obtain a total score for that guess $S_l = \sum_i D(y, X_{l,i})$. After obtaining a total score for all possible words, the sample is classified according to the maximum total score $C(y) = \text{argmax}_l S_l$.

## 3.6 Conclusion

In this work, we have demonstrated the feasibility of using a mobile device to recognize speaker-independent general speech. Our results show that it is possible to achieve high recognition accuracy using a simple recognition engine. However, there is still much work to be done to improve the performance of the system, particularly in terms of robustness to noise and variations in pronunciation. Future work will focus on developing more sophisticated recognition algorithms, as well as improving the hardware and software used in the experiments. Our goal is to develop a practical and easy-to-use speech recognition system that can be integrated into mobile devices.

2. Nexus 7 tablet which according to a teardown analysis [14] is equipped with an InverSense MPU-6050 gyroscope and accelerometer.


3.5 Sphinx

We first try to recognize digit pronunciations using general-purpose speech recognition software. We used Sphinx-4 [47] – a well-known open-source speech recognizer and trainer developed in Carnegie Mellon University. Our aim for Sphinx is to recognize gyro-recordings of the TIDIGITS corpus. As a first step, in order to test the waters, instead of using actual gyro recordings we downsampled the recordings of the TIDIGITS corpus to 200 Hz; then we trained Sphinx based on the modified recordings. The aim of this experiment is to understand whether Sphinx detects any useful information from the sub-100 Hz band of human speech. Sphinx had a reasonable success rate, recognizing about 40% of pronunciations.

Encouraged by the above experiment we then recorded the TIDIGITS corpus using a gyro – both for Galaxy S III and Nexus 4. Since Sphinx accepts recording in WAV format we had to convert the raw gyro recordings. Note that at this point for each gyro recording we had 3 WAV files, one for each gyro axis. The final stage is silence removal. Then we trained Sphinx to create a model based on a training subset of the TIDIGITS, and tested it using the complement of this subset.

The recognition rates for either axes and either Nexus 4 or Galaxy S III were rather poor: 14% on average. This presents only marginal improvement over the expected success of a random guess which would be 9%.

This poor result can be explained by the fact that Sphinx's recognition algorithms are geared towards standard speech recognition tasks where most of the voice-band is present and is less suited to speech with very low sampling frequency.

3.6 Custom recognition algorithms

In this section we present the results obtained using our custom algorithm. Based on the TIDIGITS corpus we randomly performed a 10-fold cross-validation. We refer mainly to the results obtained using Nexus 4 gyroscope.

<table>
<thead>
<tr>
<th>Table 3: Speaker's gender identification results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4: Speaker identification results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Mixed</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Male</td>
</tr>
</tbody>
</table>

These results for speaker-independent isolated word recognition are summarized in Table 5. We had correct classification rate of about 10% using multi-class SVM and GMM trained with MFCC statistical features, which is almost equivalent to a random guess. Using DTW with STFT features we got 23% correct classification for male speakers, 26% for female speakers and 17% for a mixed set of both female and male speakers.

The confusion matrix in Figure 6, corresponding to the mixed speaker-set recorded on a Nexus 4 device, explains the not so high recognition rate, exhibiting many false positives for the words "6" and "9".

The poor performance of Sphinx and our custom algorithm may be explained by the fact that high but quite recent gyro accelerometers have different performance characteristics. The same gender in the same condition. The same gender in the same condition. The results for speaker-independent isolated word recognition include the confusion matrix for the mixed speaker-set recorded on a Nexus 4 device.
In this experiment, speaker-independent word recognition is used to determine the number of correct words identified. The confusion matrix shown in Figure 6 displays the results of the speaker-independent word recognition system. The confusion matrix is a heat map, where the number of samples from each speaker that were classified as a particular speaker is represented. This allows for a clear visualization of the performance of the system.

The results for speaker-dependent word recognition are presented in Figure 7. The confusion matrix for this case also shows the number of samples from each speaker that were classified as a particular speaker. These results indicate a higher accuracy compared to the speaker-independent case, as expected.

Table 6 presents the results for speaker-dependent word recognition using the Nexus 4 device. The table shows the accuracy of different classifiers for recognizing isolated words pronounced by a single speaker. SVM and GMM perform similarly, with accuracies of 65% and 5%, respectively. DTW, on the other hand, achieves an accuracy of 15%, which is significantly higher than the other methods.

Further improvement

For speaker-dependent word recognition, we suggest the following improvements:

- **Phoneme recognition instead of whole words**: This would be more suitable since different pronunciations have different durations, and an HMM could better model these changes.
- **Pre-filtering of the signal**: This could reduce irrelevant noise and improve the performance of the system.
- **A larger training set**: This is likely to yield better results, as we used relatively small training and evaluation sets in our tests.

The experimental results show that DTW outperforms SVM and GMM in most cases. While this is true for the Nexus 4 device, it did not hold for measurements taken with the Galaxy III device. The low-pass filtering on the Galaxy III device rendered all methods quite ineffective, resulting in a success rate equivalent to a random guess.

For gender and speaker identification, statistical spectral features based methods (SVM and GMM) perform at least as good as DTW. However, for the Galaxy S III mixed speaker set and gender identification cases, DTW performed better than SVM. This is somewhat surprising and requires further experimentation to confirm whether this phenomenon is consistent and whether it is related to capturing the high frequencies.
4.1 Reconstruction algorithm

We have explored methods of recognizing a single speaker using multiple microphones. To achieve this, we first apply gain mismatch correction, then we use the Whittaker-Shannon interpolation formula to reconstruct the signal.

4.1.1 Signal offset correction

We correct for signal offset by taking the mean of the gyro samples and comparing it to 0 to get the constant offset.

4.1.2 Gain mismatch correction

Gain mismatch correction is crucial for a successful reconstruction. We correct for gain mismatch by normalizing the signal to have a standard deviation equal to 1. In case there is a constant offset, we can take the mean of the samples to compensate for it.

4.2 Reconstruction using multiple devices

In this section, we suggest that isolated word recognition can be performed using all of the recordings, and we can afford the computational overhead of processing the data from multiple microphones. Unlike audio recordings, gyroscope recordings are highly correlated while the noise of gyroscope readings is not. Hence, we can use a larger feature vector, or have the classification algorithm take into account the score obtained for all readings from multiple devices.

Moreover, we hinted at the more ambitious task of recognizing other signals using multiple microphones. For instance, we can use existing methods for enhancing signals in signal processing and machine learning literature. In a case where we gain access to the signal to have a standard deviation equal to 1, we can use existing techniques to reconstruct the signal.

Another possible improvement is to leverage the 3-axis recordings. It is obvious that the three recordings are correlated while the noise of gyro readings is not. Hence, we can use a larger feature vector, or have the classification algorithm take into account the score obtained for all readings from multiple devices.

4.2.1 Speaker independence

While many papers on the subject exist, one may take advantage of this to get a composed signal that is more speaker-independent. Even for small microphones, the composite signal may be useful. In case of using multiple microphones, one may take advantage of this to get a composed signal that is more speaker-independent.

4.2.2 Speaker adaptation

For speaker-dependent speech recognition, it would be easier to use existing audio corpora for training a speech recognition engine than to produce gyroscope recordings for a large set of words. For that purpose, it would be interesting to test how well the recognition can perform using low-cost and energy-efficient low-frequency A/D converters. While many papers on the subject exist, one may take advantage of this to get a composed signal that is more speaker-independent.

For speaker-dependent speech recognition, we used samples recorded by the Whittaker-Shannon interpolation formula to reconstruct the signal.

4.2.3 Speaker adaptation

For speaker-independent speech recognition, we used samples recorded by the Whittaker-Shannon interpolation formula to reconstruct the signal.
4.1.4 Signal reconstruction from non-uniform samples

...
Table 7: Evaluation of the method of reconstruction from multiple devices. Results obtained via "leave-one-out" cross-validation on 44 recorded words pronounced by a single speaker. Recorded using a Nexus 4 device.

<table>
<thead>
<tr>
<th>Method</th>
<th>SVM</th>
<th>GMM</th>
<th>DTW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>18%</td>
<td>14%</td>
<td>77%</td>
</tr>
</tbody>
</table>
overall consideration of the whole system and a clear definition of the power of the attacker against whom we defend. To defend against an attacker that has only user-level access to the device (an application or a web-site), it might be enough to apply low-pass filtering to the raw samples provided by the gyroscope. Judging by the sampling rate available for Blink and WebKit based browsers, it is enough to pass frequencies in the range 0 – 20 Hz. If this rate is enough for most of the applications, the filtering can be done by the driver or the OS, subverting any attempt to eavesdrop on higher frequencies that reveal information about surrounding sounds. In case a certain application requires an unusually high sampling rate, it should appear in the list of permissions requested by that application, or require an explicit authorization by the user. To defend against attackers who gain root access, this kind of filtering should be performed at the hardware level, not being subject to configuration. Of course, it imposes a restriction on the sample rate available to applications.

Another possible solution is some kind of acoustic masking. It can be applied around the sensor only, or possibly on the case of the mobile device.

7 Conclusion

We show that the acoustic signal measured by the gyroscope can reveal private information about the phone’s environment such as who is speaking in the room and, to some extent, what is being said. We use signal processing and machine learning to analyze speech from very low frequency samples. With further work on low-frequency signal processing of this type it should be possible to further increase the quality of the information extracted from the gyro.

This work demonstrates an unexpected threat resulting from the unmitigated access to the gyro: applications and active web content running on the phone can eavesdrop sound signals, including speech, in the vicinity of the phone. We described several mitigation strategies. Some are backwards compatible for all but a very small number of applications and can be adopted by mobile device manufacturers. Other approaches, such as acoustic masking, should be used for applications that need access to high frequencies, such as security or audio applications, to prevent eavesdropping.

A general conclusion we suggest following this work is that access to all sensors should be controlled by the permissions framework, possibly differentiating between low and high sampling rates.

Acknowledgements

We would like to thank Nimrod Peleg, from the Signal and Image Processing Lab (SIPL) at the Technion, for providing assistance with the TIDIGITS corpus. We also gratefully acknowledge Sanjay Kumar Sindhi, from IIT Madras, for providing implementation and testing of several signal reconstruction algorithms. We would also like to thank Prof. Jared Tanner, from UC Davis, and Prof. Yonina Eldar, from the Technion, for advising on reconstruction of non-uniformly sampled signals. We thank Hriso Bojinov for taking part in the initial brainstorming related to this research and finally, Katharina Roesler, for proofreading and advising on writing and formulation.

This work was supported by NSF and the DARPA SAFER program. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF or DARPA.

Figure 8: Filterbank reconstruction scheme

This diagram illustrates the filterbank reconstruction scheme used in our implementation. The expression in the figure is used in the reconstruction process as shown in Figure 8.

Here we present the derivation of the discrete-time interpolation filters used in our implementation. The notation in the expressions corresponds to the notation in [28]. The continuous time expression for the interpolation filters according to Eq. 18 in [28] is given by:

\[ h_p(t) = \sum_{q=-\infty}^{\infty} \frac{1}{N} \sum_{b=-\infty}^{\infty} \frac{1}{2\pi} \sin \left( \frac{\pi}{N} \left( \frac{q}{b} \right) \right) \delta \left( \frac{t}{T} - \frac{q}{b} \right) \]

The number of samples used in the reconstruction process is \( N \), and \( b \) is the number of ADCs. The expression above is used to reconstruct the signal from its non-uniformly sampled versions. We then sample this expression at times \( t = nT \) and calculate the filter coefficients for 48 taps. Given these filters, the reconstruction process consists of up-sampling the input signals by factor \( N \), filtering and summation of the outputs of all filters (as shown in Figure 8).

B Code for sampling a gyroscope via a HTML web-page

For a web page to sample a gyro the DeviceMotion class needs to be utilized. In the following we included a JavaScript snippet that illustrates this:

```javascript
13
```

For Android

C Code Scope Rate Limitation on Android

Figure 9 depicts measurements of the above code run on Firefox (Android) while sampling an audio chirp.

References


Fraud Detection

Han Wang, CS 259D, Fall 2014

Stanford

November 20, 2014
Introduction

- Financial fraud accounts for over $50 billion dollars
- Fraud is a highly heterogeneous field with multiple subfields
  - Bank fraud
  - Insurance fraud
  - Securities and commodities fraud
  - Corporate fraud
- Research methods into fraud detection overlaps well with data mining for cyber security
  - Classification
  - Clustering
  - Prediction
  - Outlier detection
  - Regression
  - Visualization
Focus of published articles by field [1]
- Credit card fraud 14.3%
- Money laundering 2.0%
- Healthcare insurance fraud 10.2%
- Automobile insurance fraud 34.7%
- Corporate fraud 34.7%

Focus of published articles by method [1]
- Classification 61.2%
- Clustering 6.1%
- Prediction 6.1%
- Outlier detection 2.0%
- Regression 16.3%
- Visualization 2.0%
Case study 1: Detection of 10-K fraud

- Model based off of senior management’s knowledge of fraud (i.e. Enron) [2]
- Analysis focused on Management’s Discussion and Analysis (MDA) of the 10-k filing.
- NLP feature processing
- Feature reduction via SVD
- Clustering
Case study 1

Feature processing

- Convert all MDAs’ to raw text
- Stem all words but differentiate parts of speech
- Bin words + parts of speech into synonyms via SAS Enterprise Miner
- Treat bins as features and perform SVD on term data (essentially PCA without mean subtraction)
Clustering algorithm

- Tried expectation maximization to no avail
- Hierarchical clustering performed much better
- Documents stabilized into two clusters of fraud and no fraud irrespective of allowed maximum clusters (5, 10, and 40 tried)
Results

- 95.6% training accuracy
- Validation Fraud MDA: 10 Correct — 1 Wrong
- Validation Non-Fraud MDA: 16 Correct — 4 Wrong
- Impressive results given simplicity of model
Case study 2: Credit card fraud detection with Hidden Markov model [3]

- Online method for fraud detection
- Low penalty of false negative, assumed patrons would be asked a credit question
Model

- Assume HMM have observables $O_l, O_m, O_h$ which are the price bins of the purchases
- 1) Cluster the users into low, medium, high spending patterns
- 2) Train a HMM for each spending category with EM
- 3) Form running window on customers’ purchase histories
- 4) Calculate $\Delta$’s in probabilities of emission between running window updates
- 5) Threshold on large $\Delta$’s for rejection/acceptance
Results

- Paper did not provide a results table
Conclusions

- Fraud detection spans multiple subfields which pose unique problems
- Fraud detection can be used in realtime detection of financial fraud
