Authentication: Attacks and Defenses

TEJAS SHAH, SUMEDH SAWANT

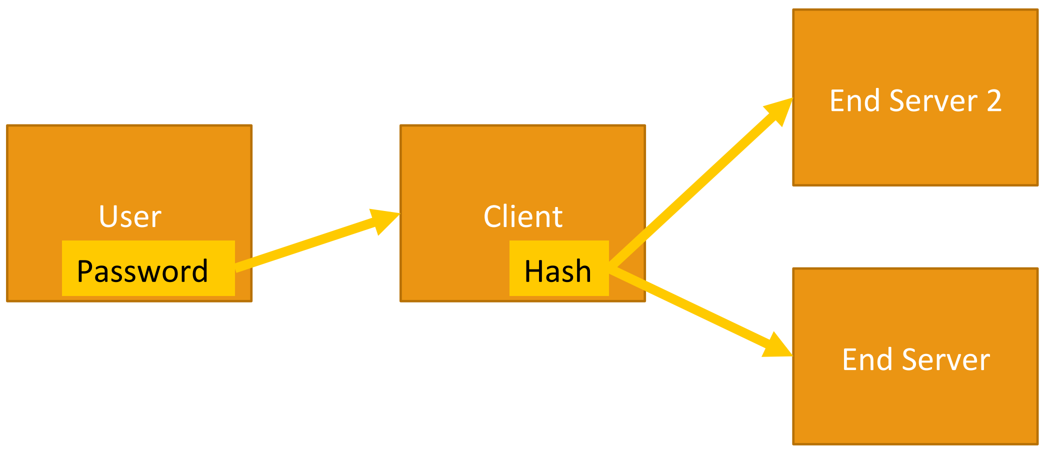
Overview of Authentication

- **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

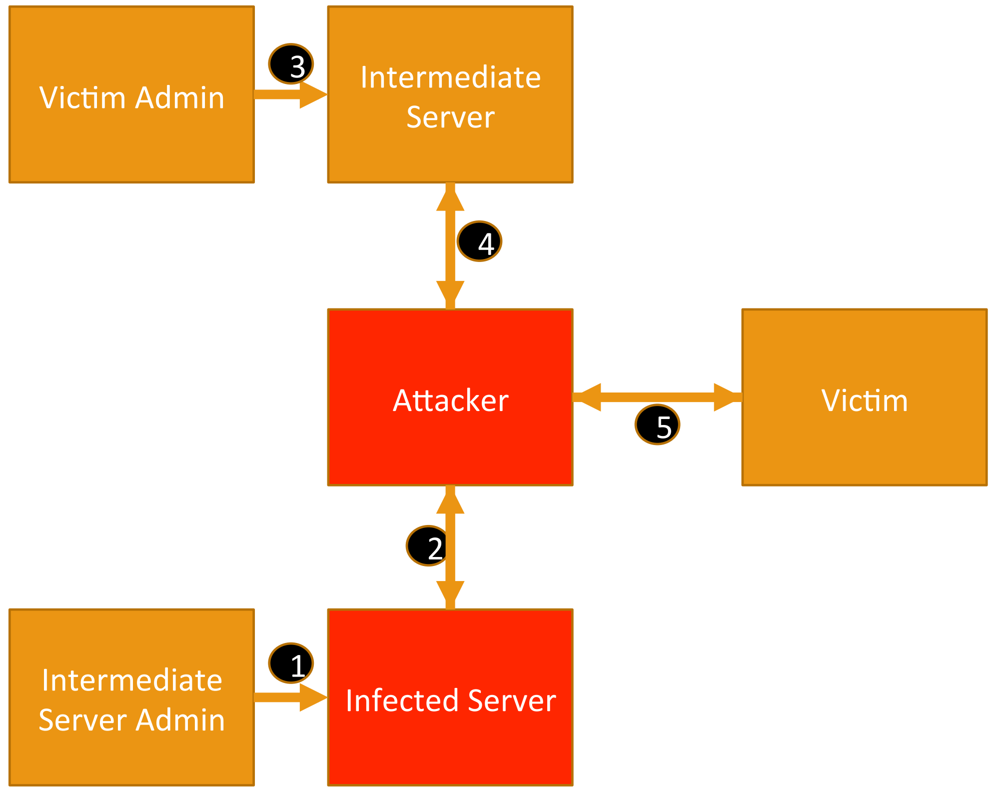
Generic SSO Architecture



Pass the hash

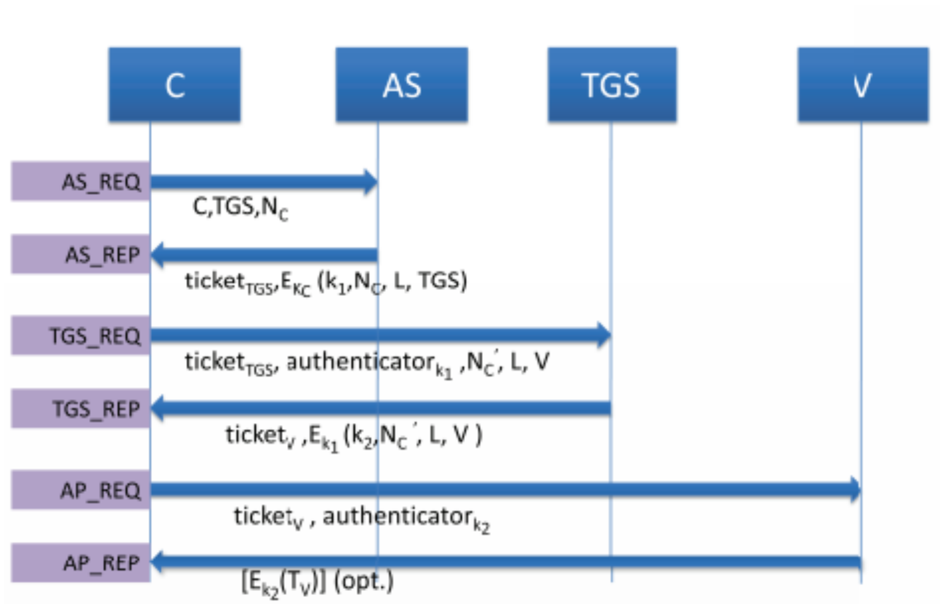
A common attack against single sign on architectures.

- Assumes attacker has gained admin access to a machine in the network



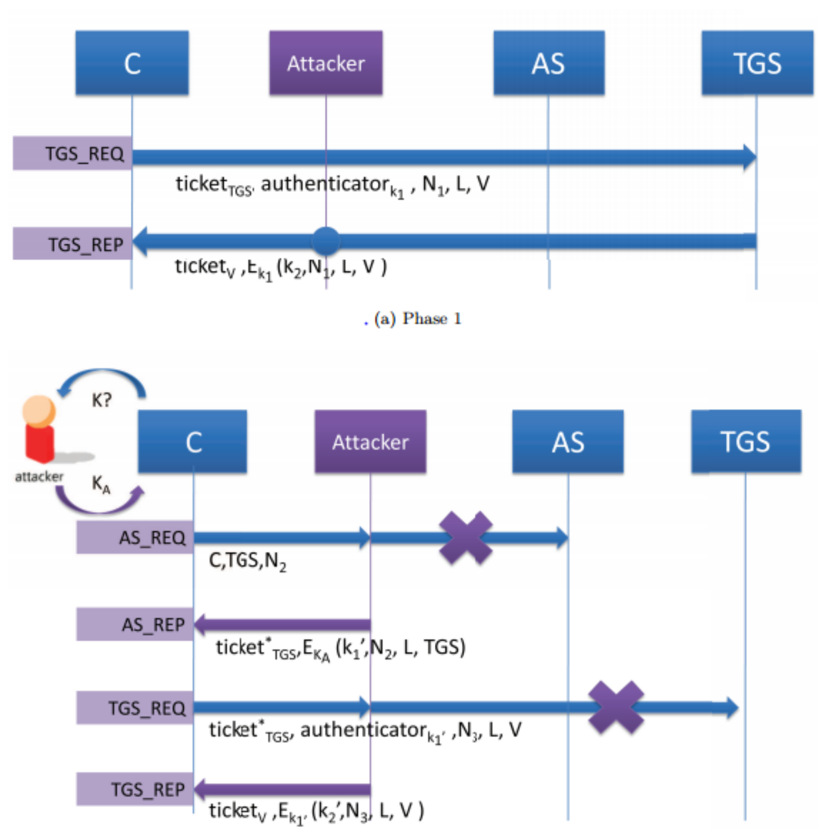
Kerberos

Microsoft Windows single sign on login service



Pass the ticket

A variant of pass-the-hash for Kerberos



(a) Phase 1

(b) Phase 2

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.x. 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)| 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


Jennifer [redacted]
hey
hey
hey
hey, how are you?
haha was that you
5370.meredith8.SkankProfile4co
m


facebook
Joshua Baer
Friends
Applications
Inbox
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=F0F2EFE6E9ECE5AEE1EBAEE6E1E3E5E2EFEFEBAEE
3EFEDAFF6B2B2B7AFB2B9B5AFB
2B8AFEEB5B0B0B0B3B8B6B6B6D
FB5B2B0B7AEEAF0E7&b=D3F4E5
F6E5A0C8E1F2F4EDE1EE
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.


General Colin L. Powell
5 hours ago
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.
Like · Comment · Share
54
1,441 people like this.
View previous comments
2 of 194
Steve Smith It's a shame you have to go through that. Stay as strong as you can be. And we know your a strong General.
18 minutes ago · Like
Twyla Jean It was pretty bad --however entertaining! we knew it was'nt
9 minutes ago via mobile · Like
Write a comment...

The Wall Street Journal
1 min ·

#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

The Wall Street Journal
5 mins ·

#BREAKING: Vice President Joe Biden to address nation in 15 minutes.

Like · Comment · Share · 89 41 44

The Wall Street Journal
Newspaper · 2,831,357 Likes · 10 hrs ·

Profile

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


The Associated Press
@AP
Following
Breaking: Two Explosions in the White House and Barack Obama is injured
Reply
Retweet
Favorite
More
3,146
RETWEETS
149
FAVORITES


Dow Jones Industrial Average (^DJI) - DJI
Add to Portfolio
14,676.57 109.40(0.75%) 2:35PM EDT
Enter name(s) or symbol(s)
GET CHART
COMPARE
EVENTS
TECHNICAL INDICATORS
CHART SETTINGS
RESET
9:30 AM EDT: ^DJI 14645.50
14,700
14,680
14,660
14,640
14,620
14,600
14,580
14,560
10:00 AM
11:00 AM
12:00 PM
1:00 PM
2:00 PM
3:00 PM
4:00 PM

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

	[5]	[3]	[4]	[6]	[7]	[17]	[18]	[19]	**COMPA**
Network Features									
Avg # conn. of neighbors						✔			
Avg messages of neighbors						✔			
Friends to Followers (F2F)	✔	✔			✔				
F2F of neighbors						✔			
Mutual links						✔	✔	✔	
User distance								✔	
Single Message Features									
Suspicious content	✔								
URL blacklist			✔						
Friends features									
Friend name entropy					✔				
Number of friends	✔				✔				
Profile age	✔								
Stream Features									
Activity per day	✔								
Applications used						✔			✔
Following Rate						✔			
Language									✔
Message length	✔								
Messages sent					✔				
Message similarity		✔	✔	✔	✔	✔			
Message timing		✔	✔						✔
Proximity									✔
Retweet ratio	✔								
Topics	✔								✔
URL entropy			✔						
URL ratio	✔	✔		✔	✔	✔			
URL repetition				✔					✔
User interaction	✔	✔		✔					✔

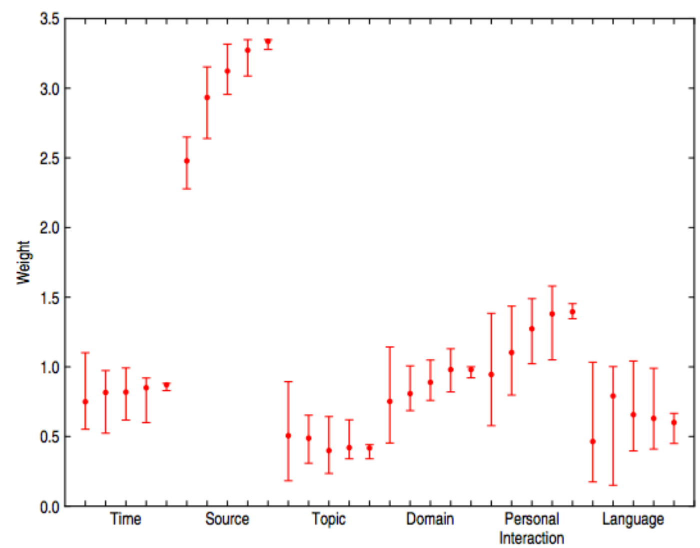


Figure 1. Features evolving with different sizes of training sets. Each experiment was conducted 25 times on random subsets of 25%, 50%, 70%, 90%, and 99% of the 5,236 labeled training instances. The fraction of positive to negative samples remained constant.

False Positives

Network & Similarity Measure	Twitter Text		Twitter URL		Facebook Text	
	Groups	Accounts	Groups	Accounts	Groups	Accounts
Total Number	374,920		14,548		48,586	
# Compromised	9,362	343,229	1,236	54,907	671	11,499
False Positives	4% (377)	3.6% (12,382)	5.8% (72)	3.8% (2,141)	3.3% (22)	3.6% (412)
# Bulk Applications	12,347		1,569		N/A	N/A
# Compromised Bulk Applications	1,647	178,557	251	8,254	N/A	N/A
False Positives	8.9% (146)	2.7% (4,854)	14.7% (37)	13.3% (1,101)	N/A	N/A
# Client Applications	362,573		12,979		N/A	N/A
# Compromised Client Applications	7,715	164.672	985	46,653	N/A	N/A
False Positives	3.0% (231)	4.6% (7,528)	3.5% (35)	2.2% (1,040)	N/A	N/A

False Negatives

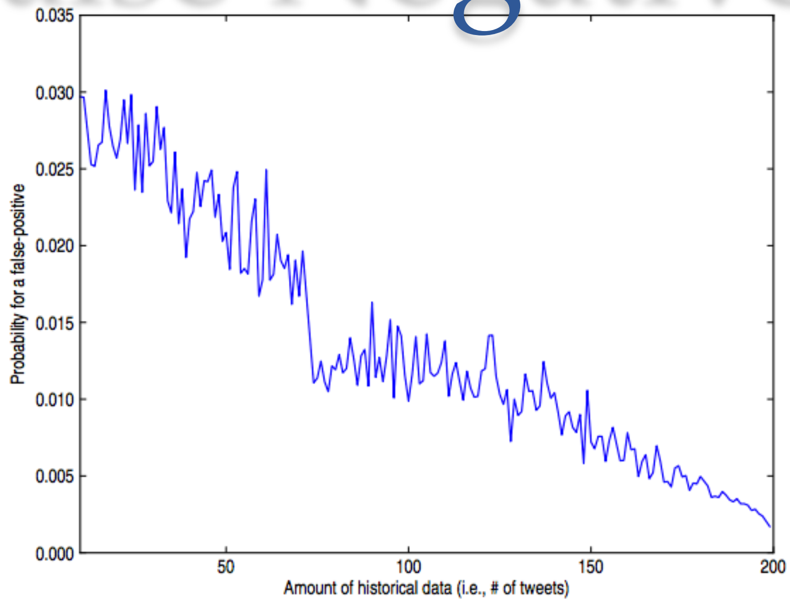


Figure 2. Probability of false positives depending on the amount of historical data on Twitter

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
  - 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
  - Can be used to decide when to present the users with CAPTCHAs

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

Model

- Undirected Graph, $G = (V,E)$

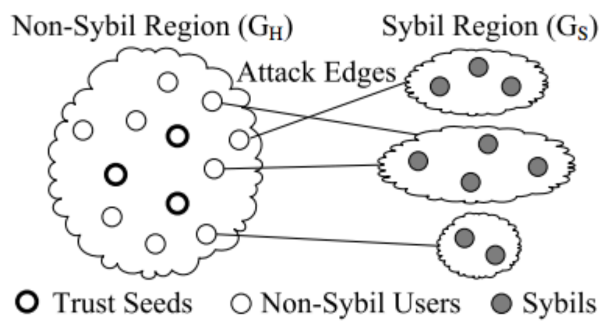


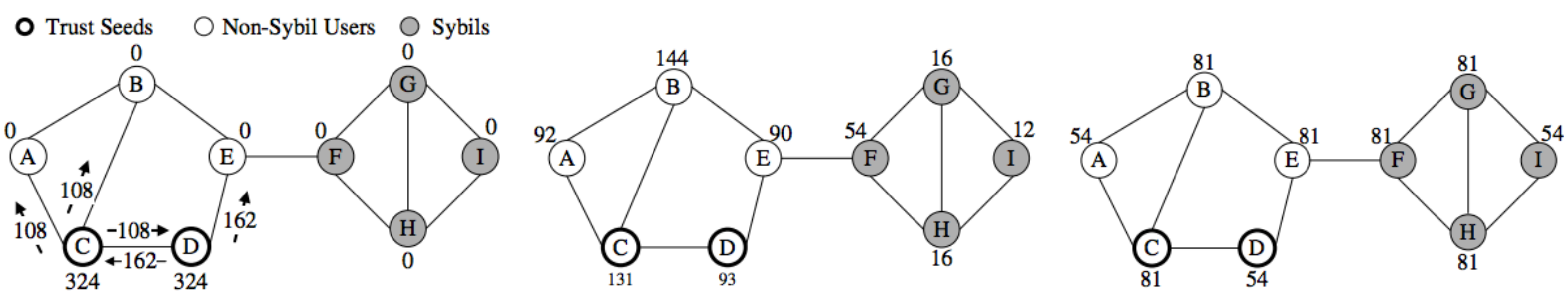
Figure 2: Non-Sybil region, Sybil region, and attack edges in an OSN under a Sybil attack. All Sybils created by malicious users are placed into the Sybil region. The Sybil collective may not be well connected.

Propagating Trust

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

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

Terminate after log(n) operations

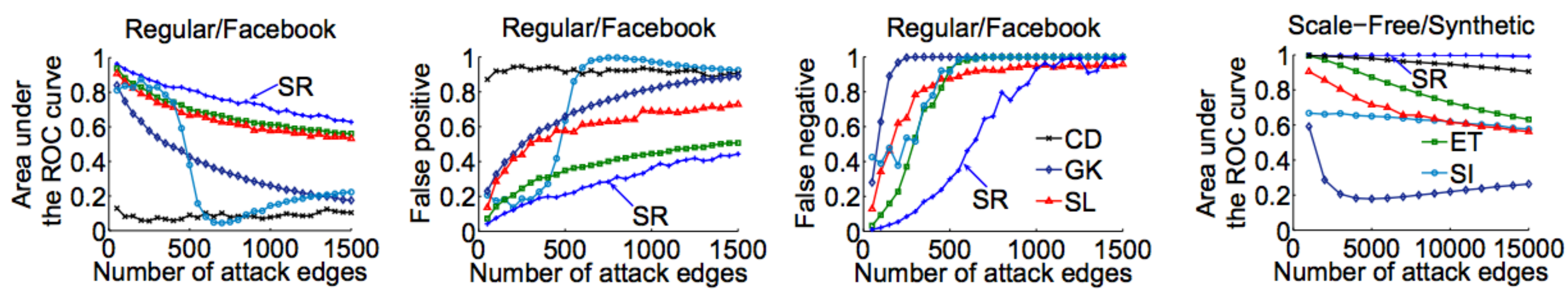


(a) Initialization (b) After 4 power iterations (c) Stationary distribution

Social Network	Nodes	Edges	Clustering Coefficient	Diameter
Facebook	10,000	40,013	0.2332	17
ca-AstroPh	18,772	198,080	0.3158	14
ca-HepTh	9,877	25,985	0.2734	18
Synthetic	10,000	39,399	0.0018	7
wiki-Vote	7,115	100,736	0.1250	7
soc-Epinions	10,000	222,077	0.0946	6
soc-Slashdot	10,000	153,404	0.0582	4
email-Enron	10,000	105,343	0.1159	6

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

Results



(a) (b) (c) (d)

Limitations

- Multiple communities

- Sophisticated attackers may obtain knowledge of seeds and established "attack edges" close to the seed nodes

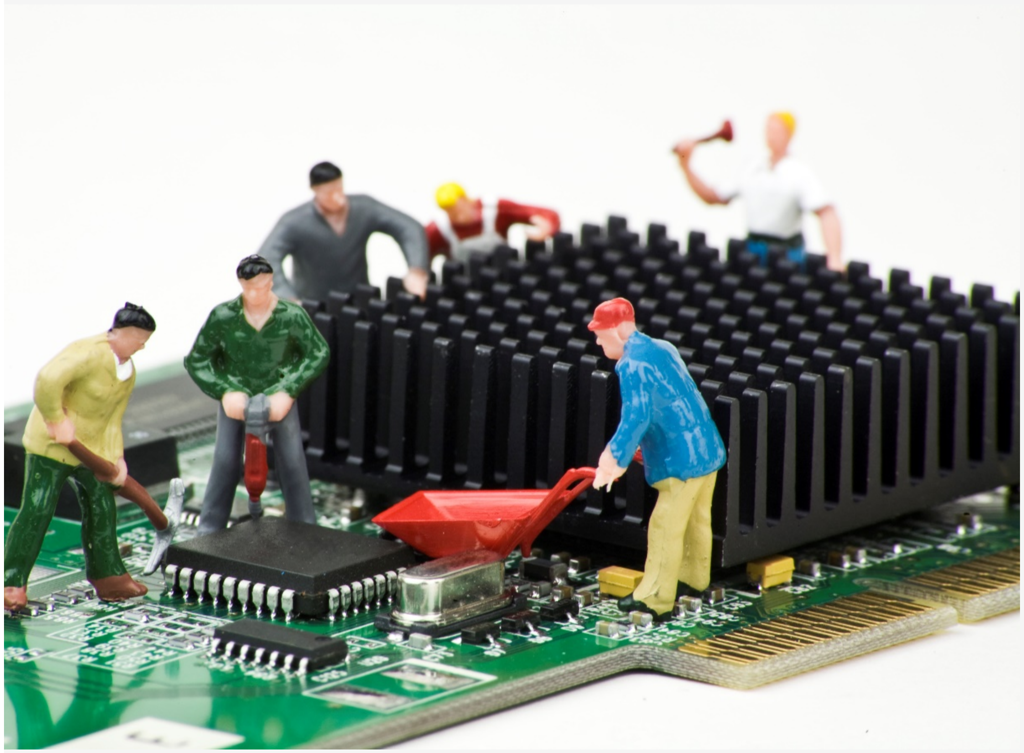
Resources

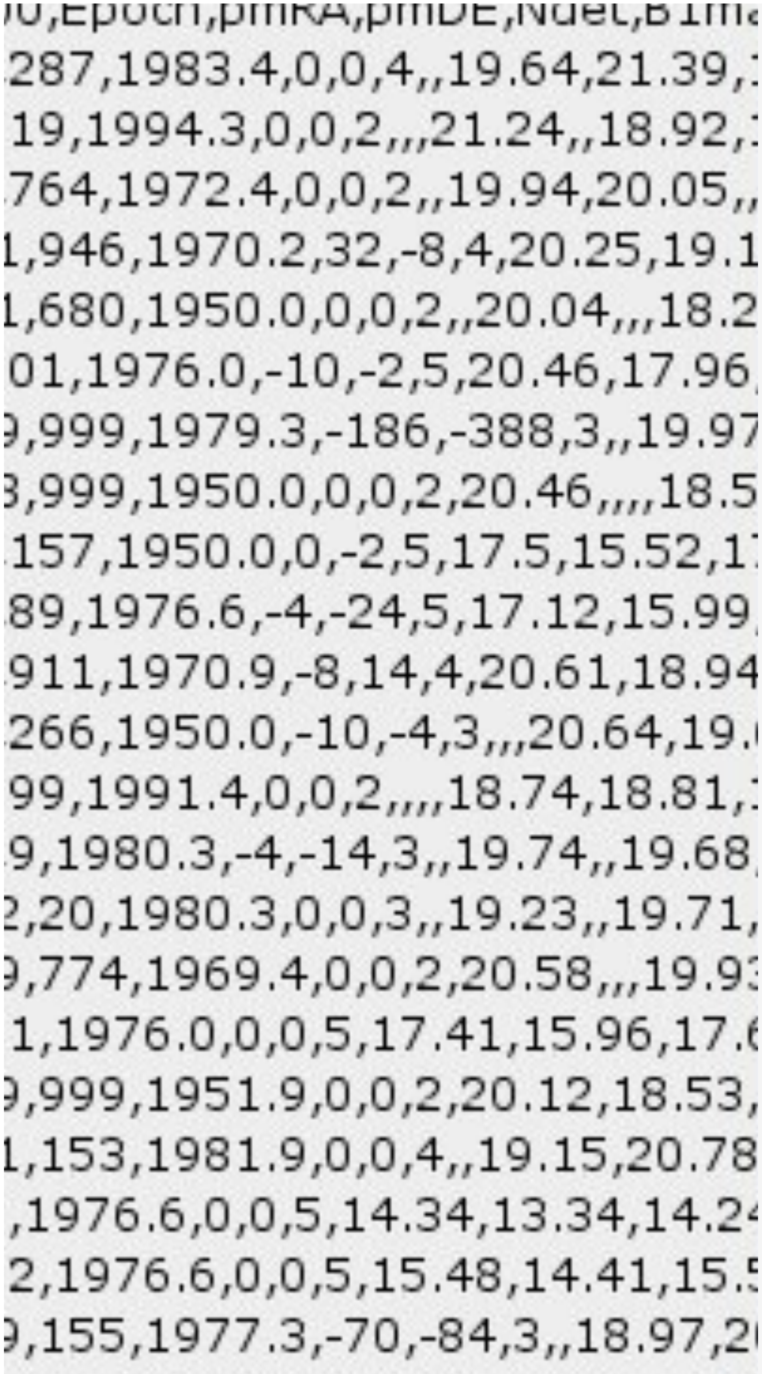
- Egele, Manuel, et. al, "COMPA: Detecting Compromised Accounts on Social Networks"
- Cao, Qiang, et. al, "Aiding the Detection of Fake Accounts in Large Scale Social Online Services"
- Harris Interactive Public Relations Research, "A Study of Social Networks Scams," 2008.

data mining

WITH

security



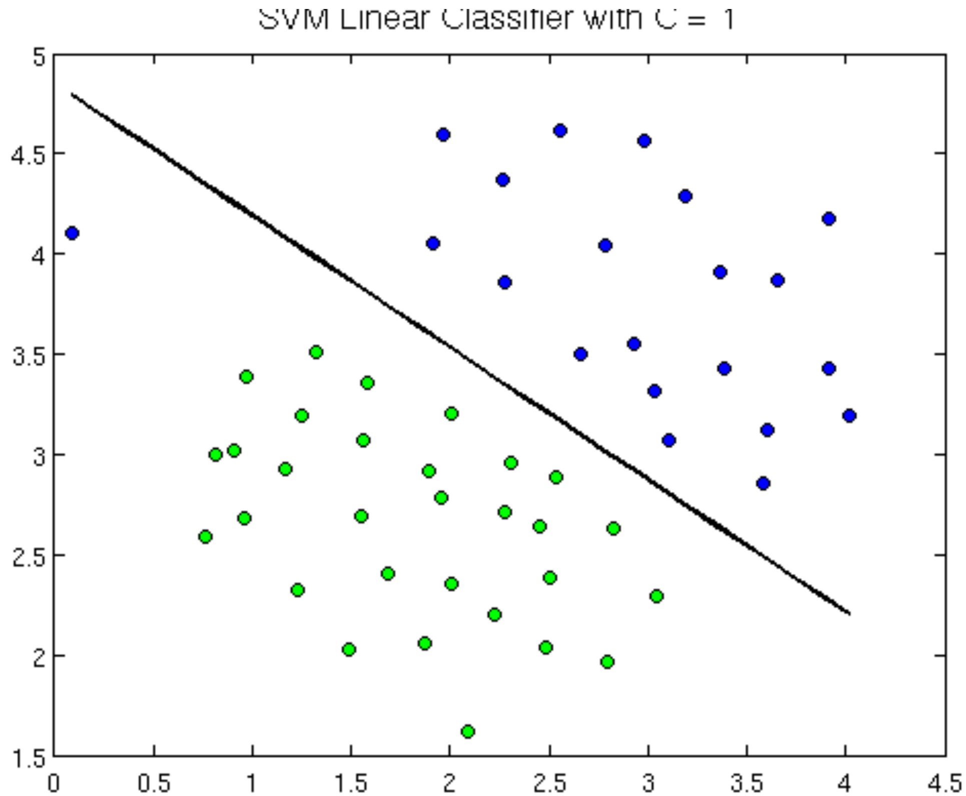



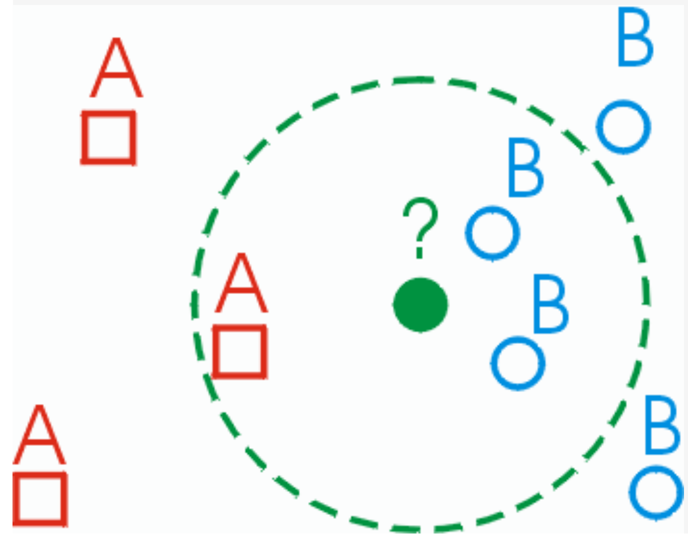
Normalization
Anomaly detection

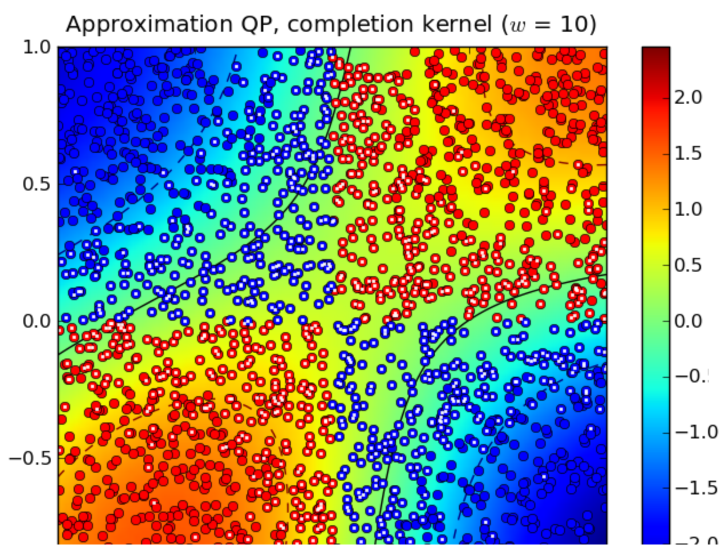
“Professor Plum did the Study with the 5 machine botnet”

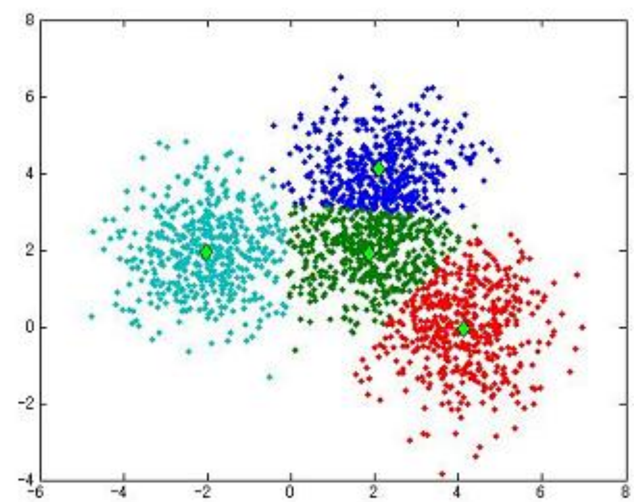
What about the attackers?

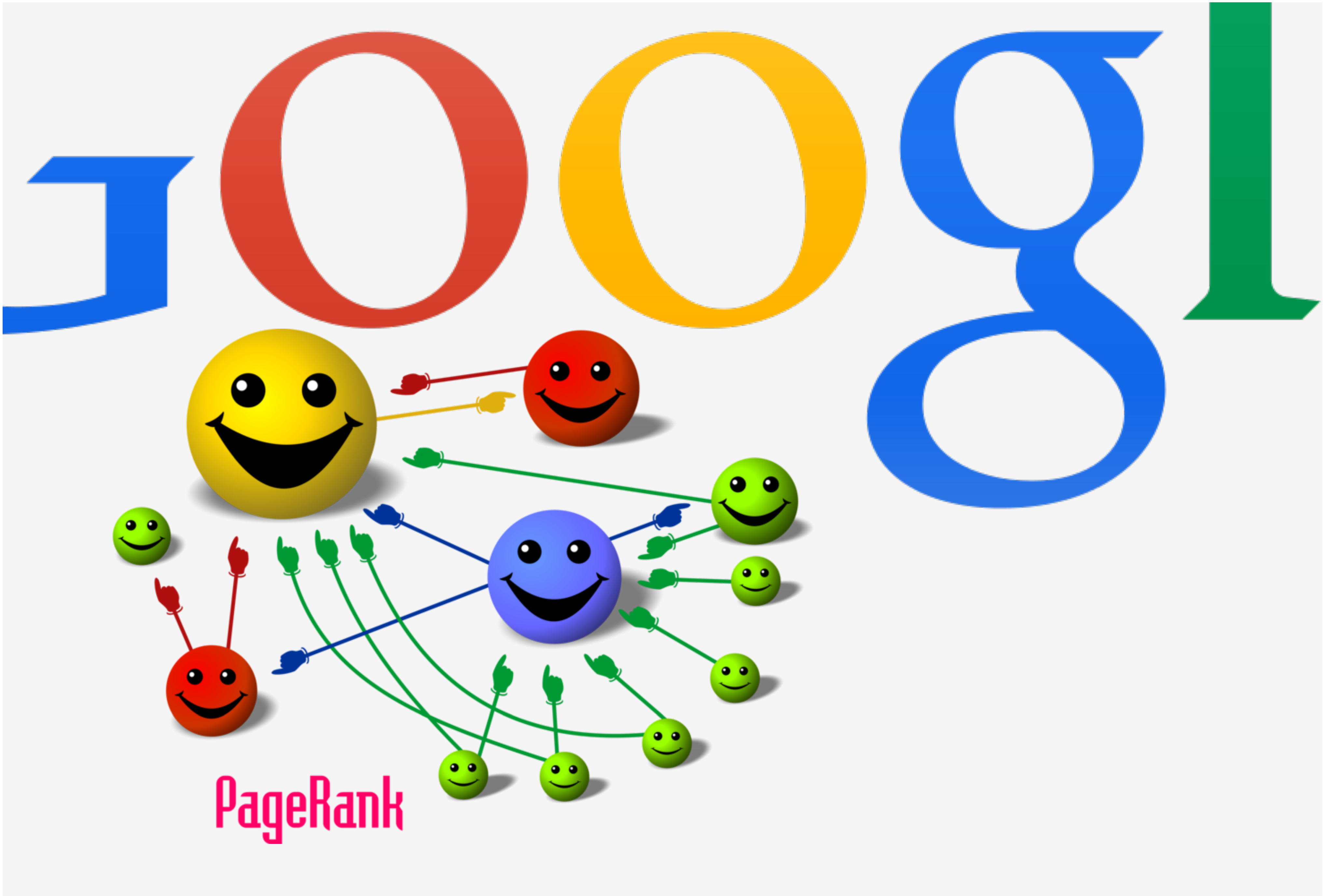

DATA MINING!


SVM Linear Classifier with C = 1
5
4.5
4
3.5
3
2.5
2
1.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5


A
A
A
B
B
B
B
?


Approximation QP, completion kernel (w = 10)
1.0
0.5
0.0
−0.5
2.0
1.5
1.0
0.5
0.0
−1.0
−1.5


8
6
4
2
0
-2
-4
-6
-4
-2
0
2
4
6
8


PageRank

Think more… analog.


555-5922
HOLD


GMC

LOW-tech DATA MINING

Data mining for fun and profit

Derrick Liu + Daniel Chiu

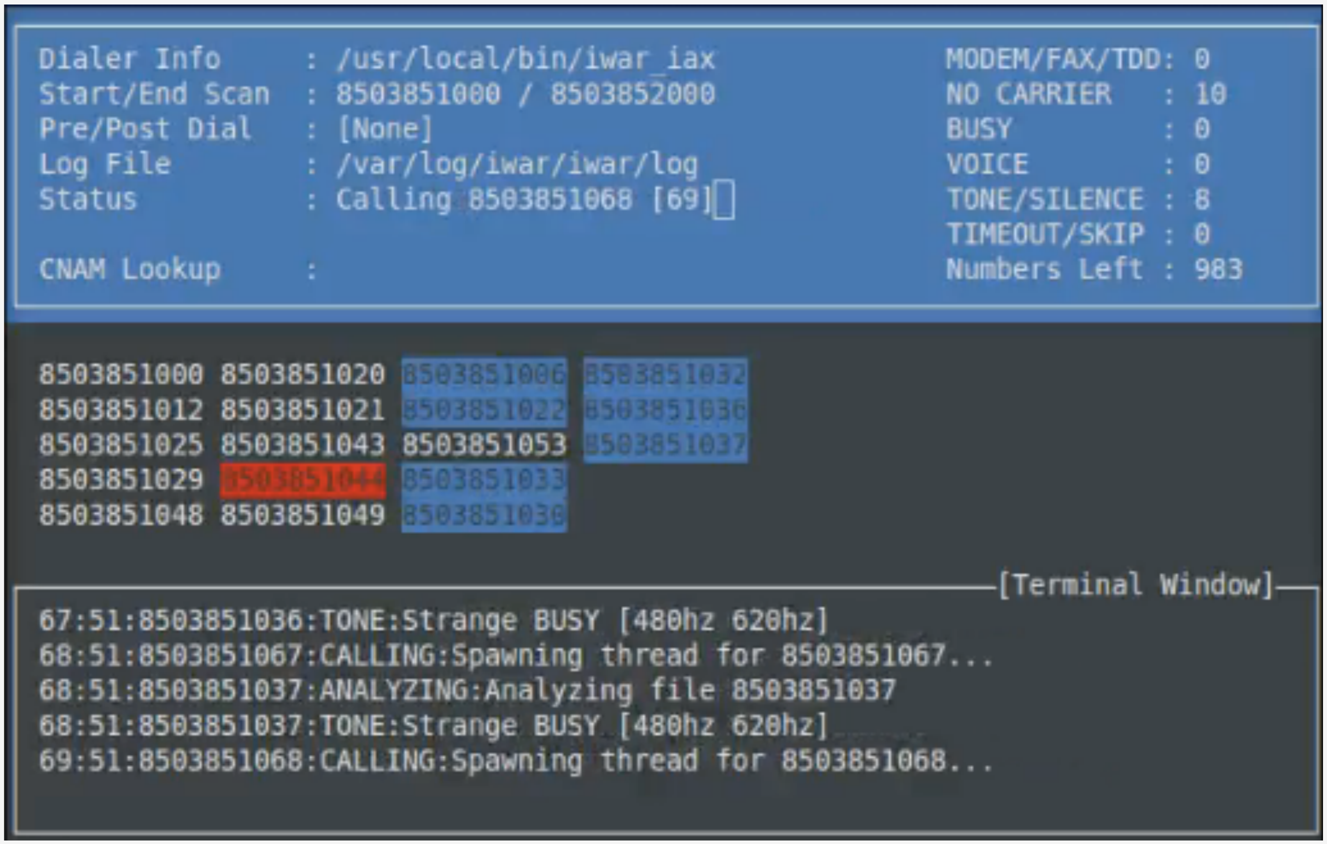
LOW-tech
DATA MINING

War dialing

War driving

Search engines

War dialing


Dialer Info : /usr/local/bin/iwar_iax
Start/End Scan : 8503851000 / 8503852000
Pre/Post Dial : [None]
Log File : /var/log/iwar/iwar/log
Status : Calling 8503851068 [69]
CNAM Lookup :
MODEM/FAX/TDD: 0
NO CARRIER : 10
BUSY : 0
VOICE : 0
TONE/SILENCE : 8
TIMEOUT/SKIP : 0
Numbers Left : 983
8503851000 8503851020
8503851012 8503851021
8503851025 8503851043 8503851053
8503851029
8503851048 8503851049
[Terminal Window]
67:51:8503851036:TONE:Strange BUSY [480hz 620hz]
68:51:8503851067:CALLING:Spawning thread for 8503851067...
68:51:8503851037:ANALYZING:Analyzing file 8503851037
68:51:8503851037:TONE:Strange BUSY [480hz 620hz]
69:51:8503851068:CALLING:Spawning thread for 8503851068...

What is war dialing?

03 650-555-2363

04 650-555-2364

05 650-555-2365

06 650-555-2366

07 650-555-2367

08 650-555-2368 Sounds like a modem!

09

10

11

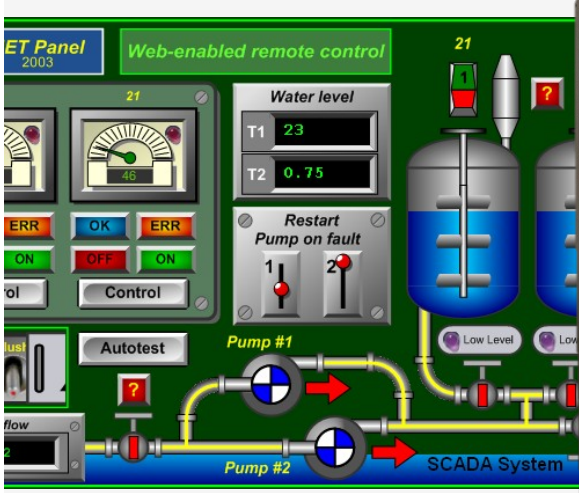
12

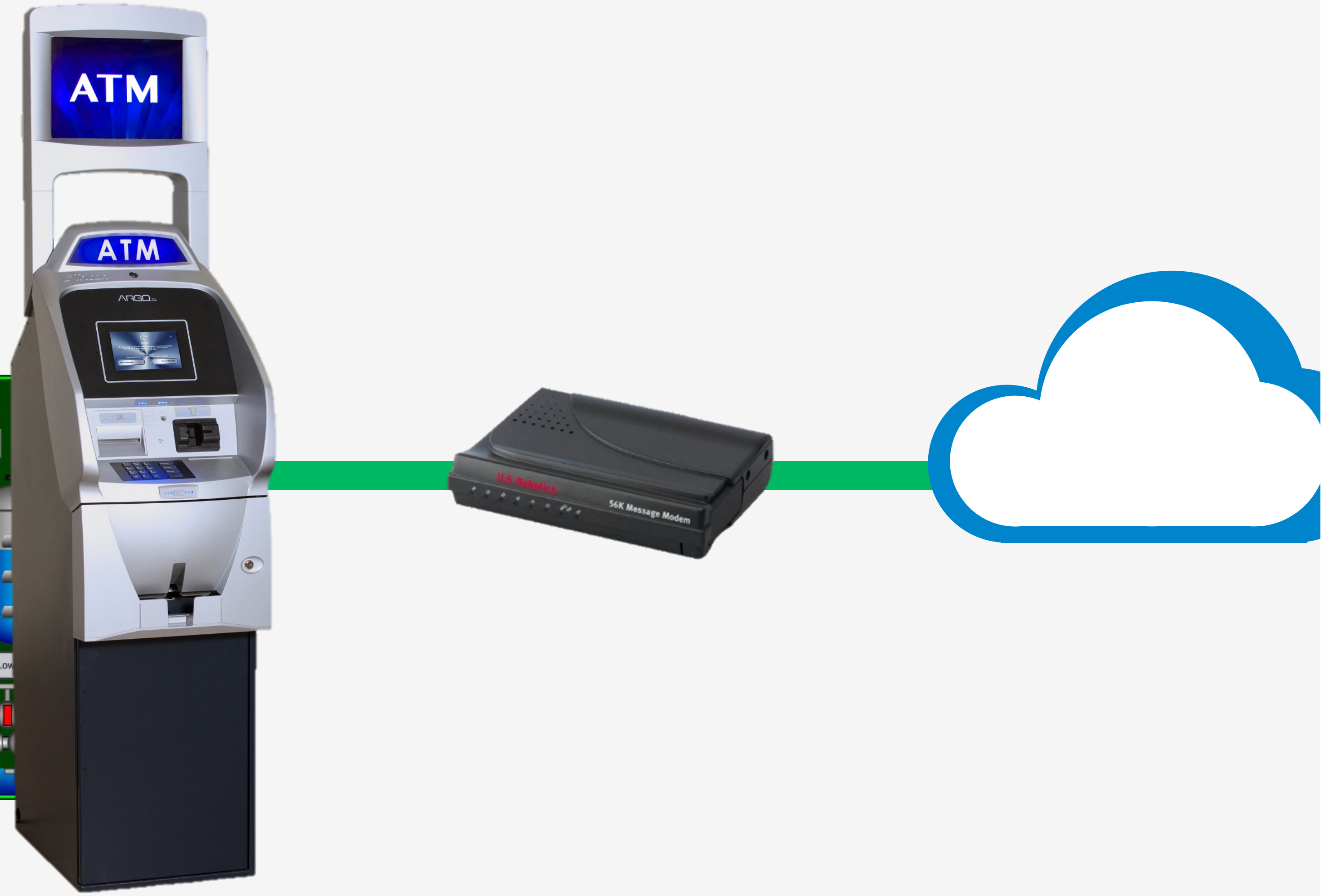
13

Flag it for later human referenc

War dialing can reveal significant lapses in security.

Industrial control systems



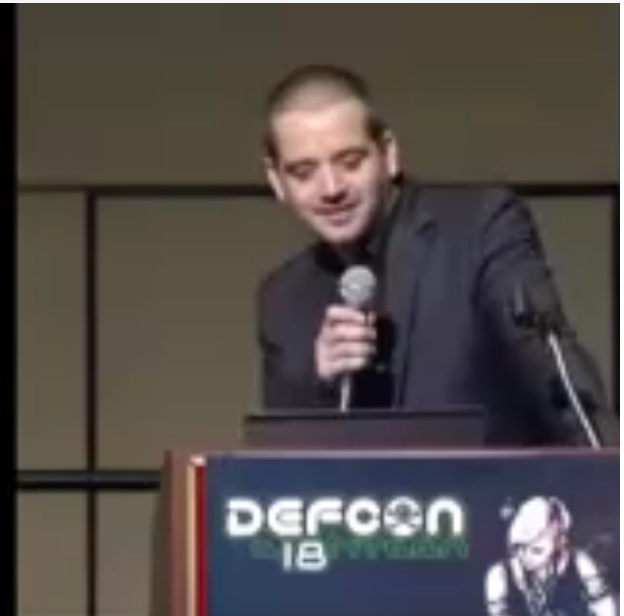


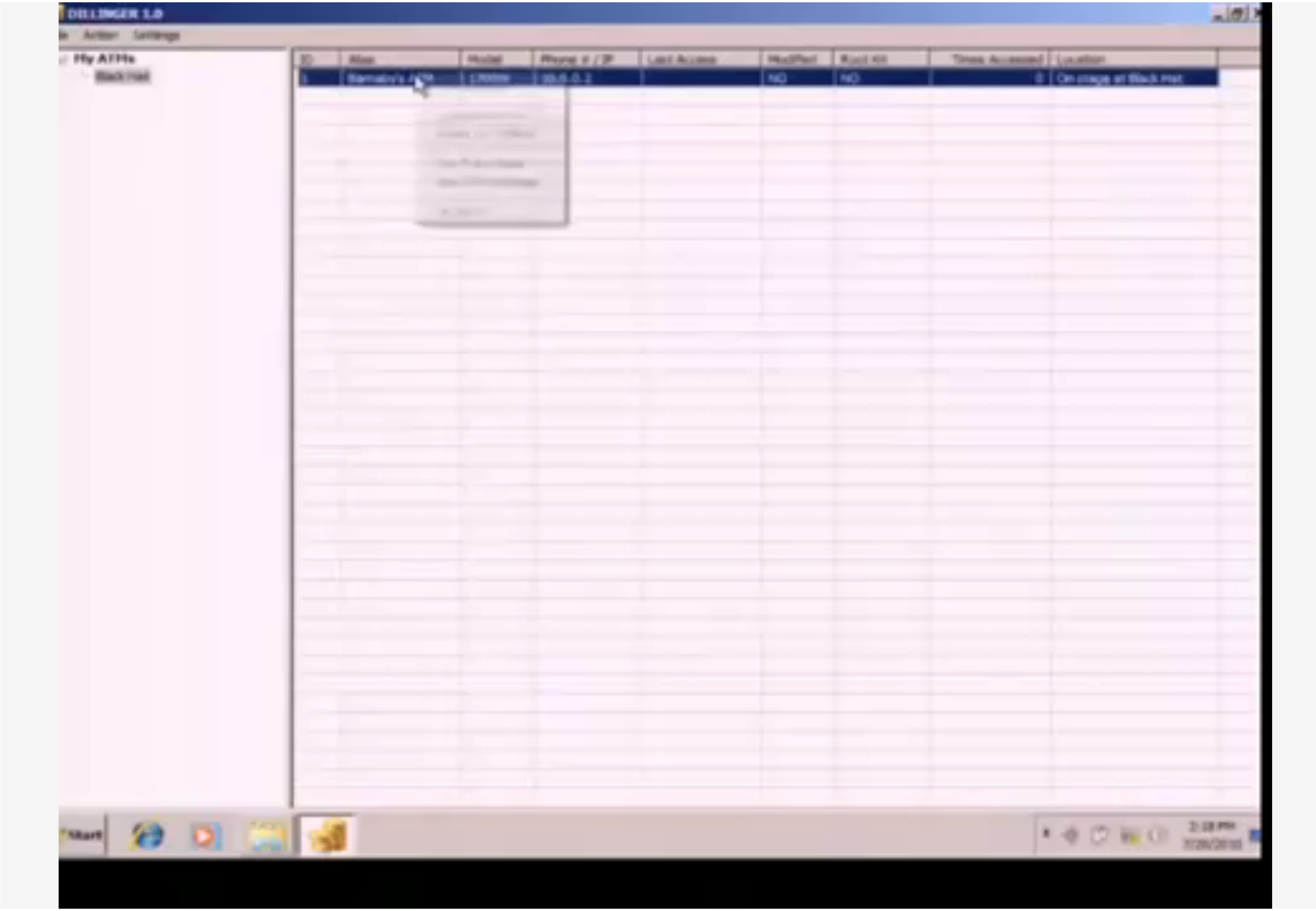
any sensitive systems are still connected to the internet via dial-u

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




DILLINGER 1.0
My ATMs
Times Accessed
Location
NO
On stage at Black Hat

LOW-tech

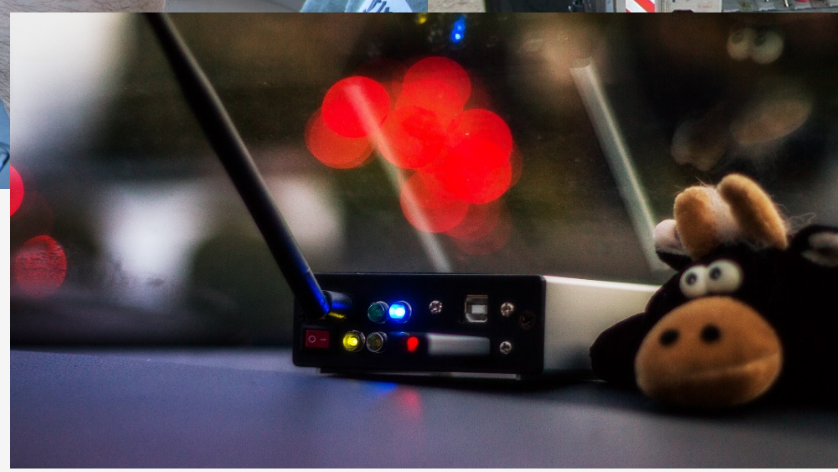
DATA MINING

War dialing

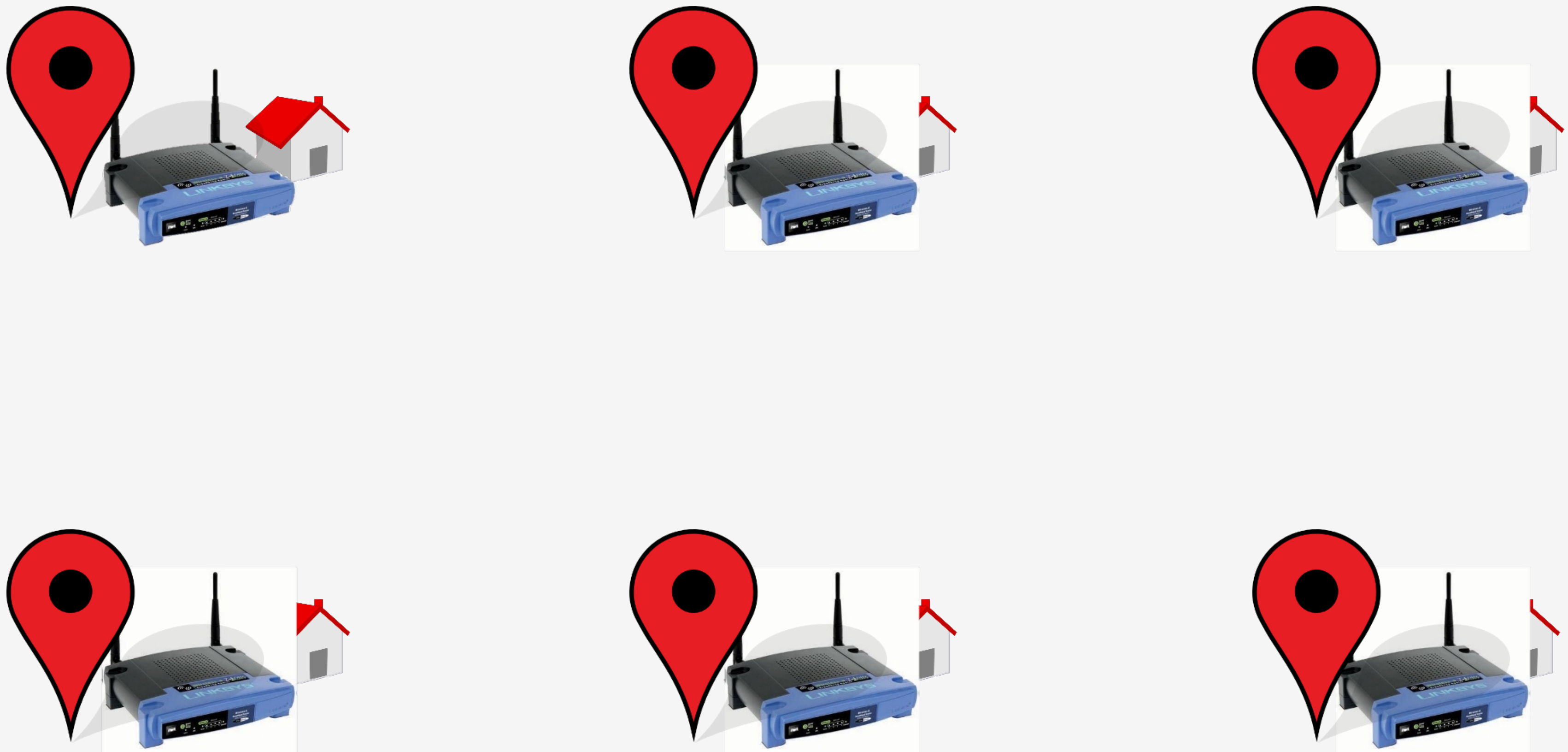
War driving

Search engines

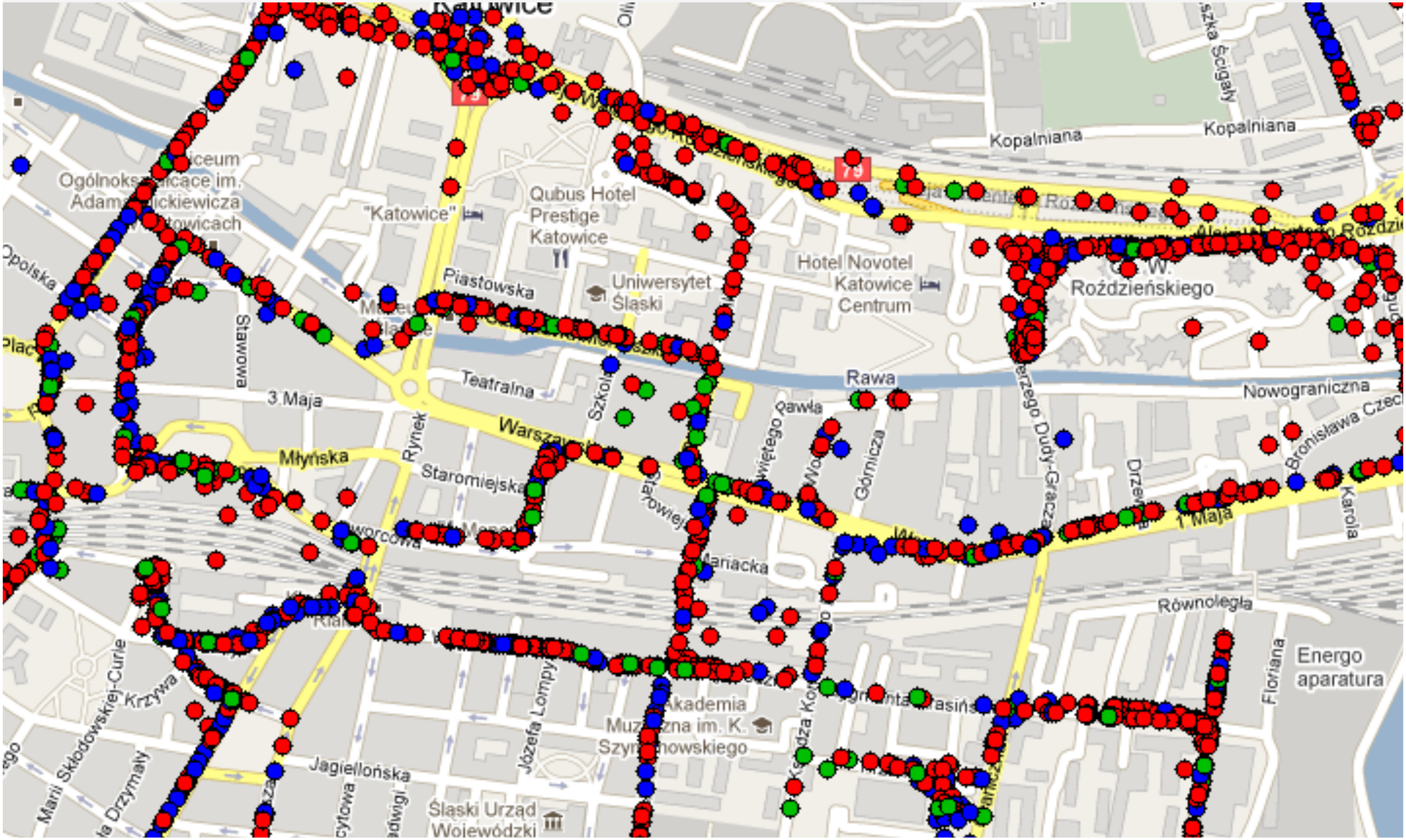
War driving



In a nutshell



WiFi security is getting better.



There's another wireless target out there.

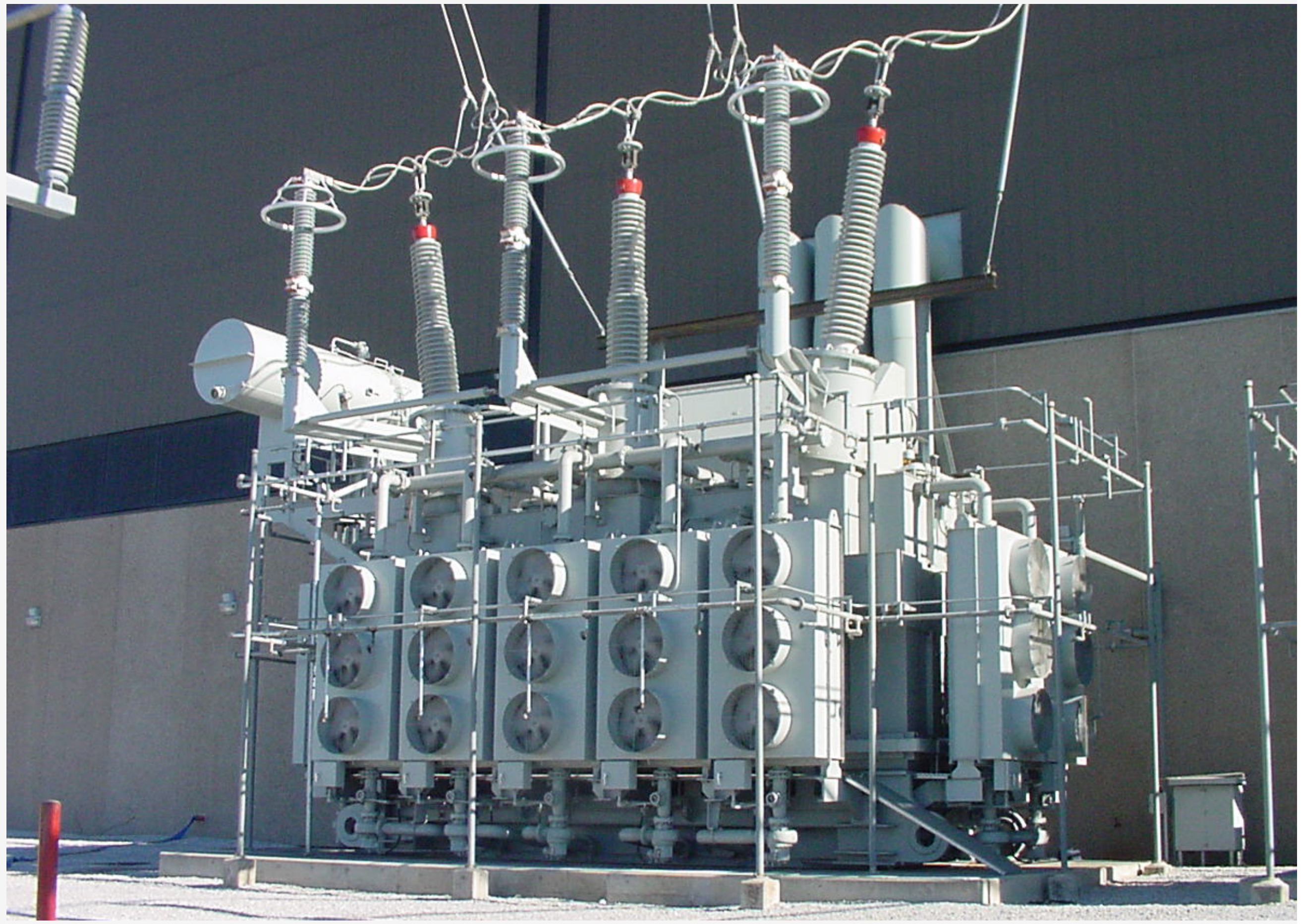
With

"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

he 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

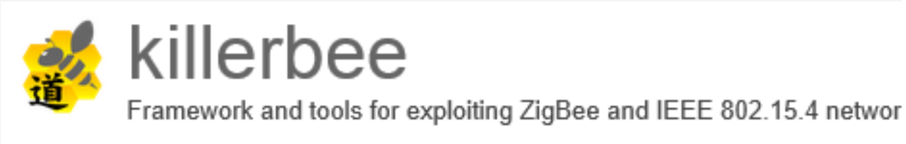
veryone uses different tech and different “standard

Lack of security mindset

Needs to survive for a very long time, slow updates

An example

+



$40

$0

LOW-tech
DATA MINING

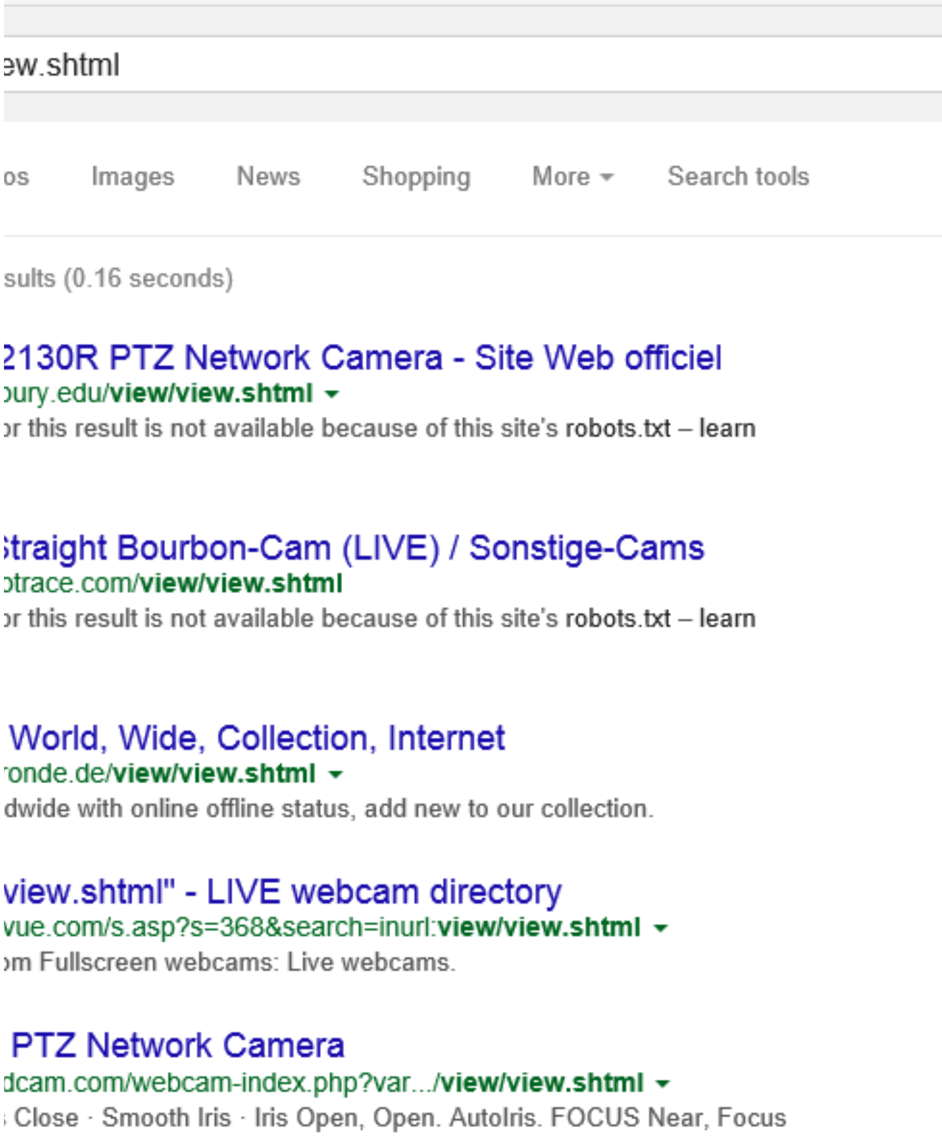
War dialing

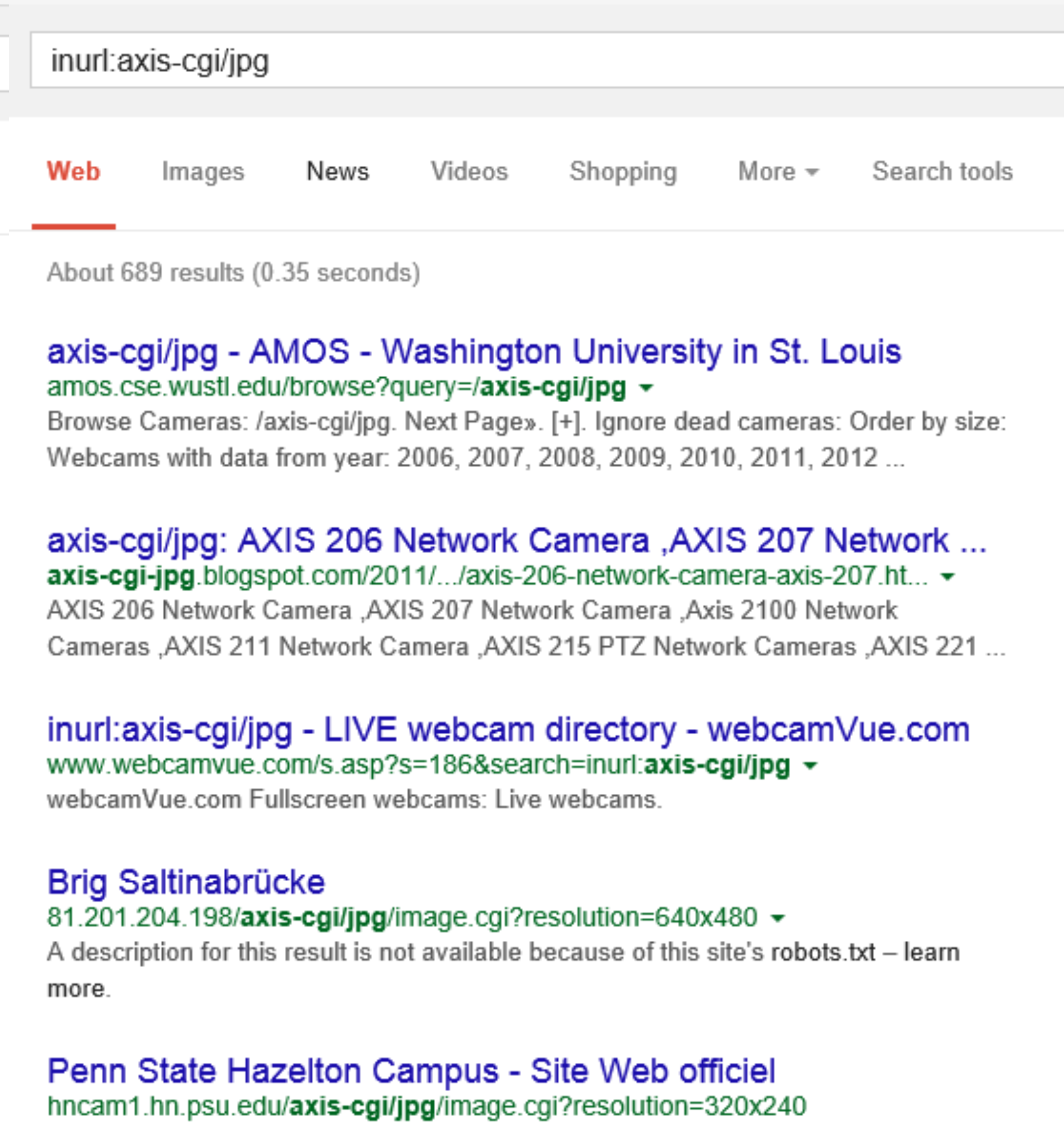
War driving

Search engines

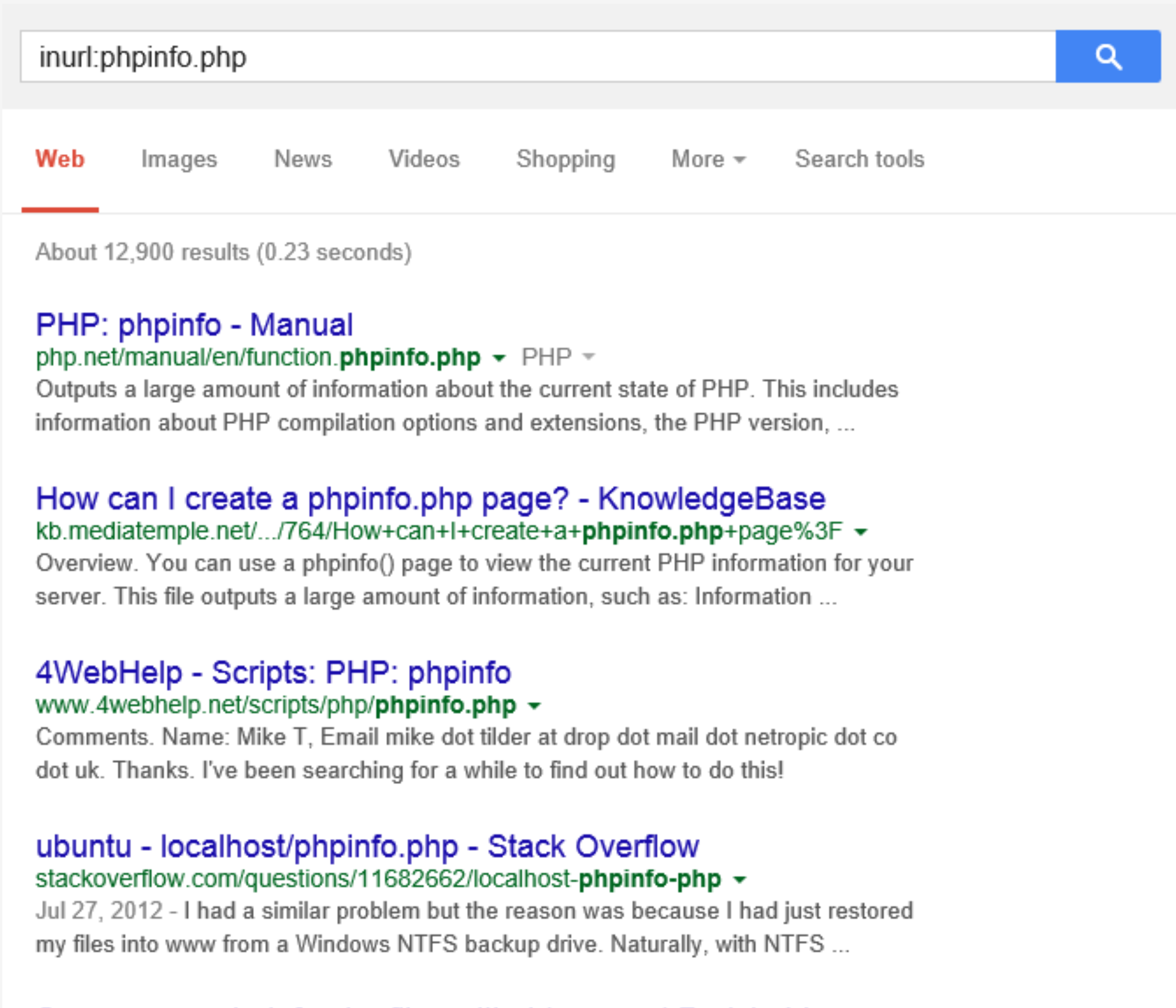
Search engines are really good at finding things.

Unsecured surveillance cameras

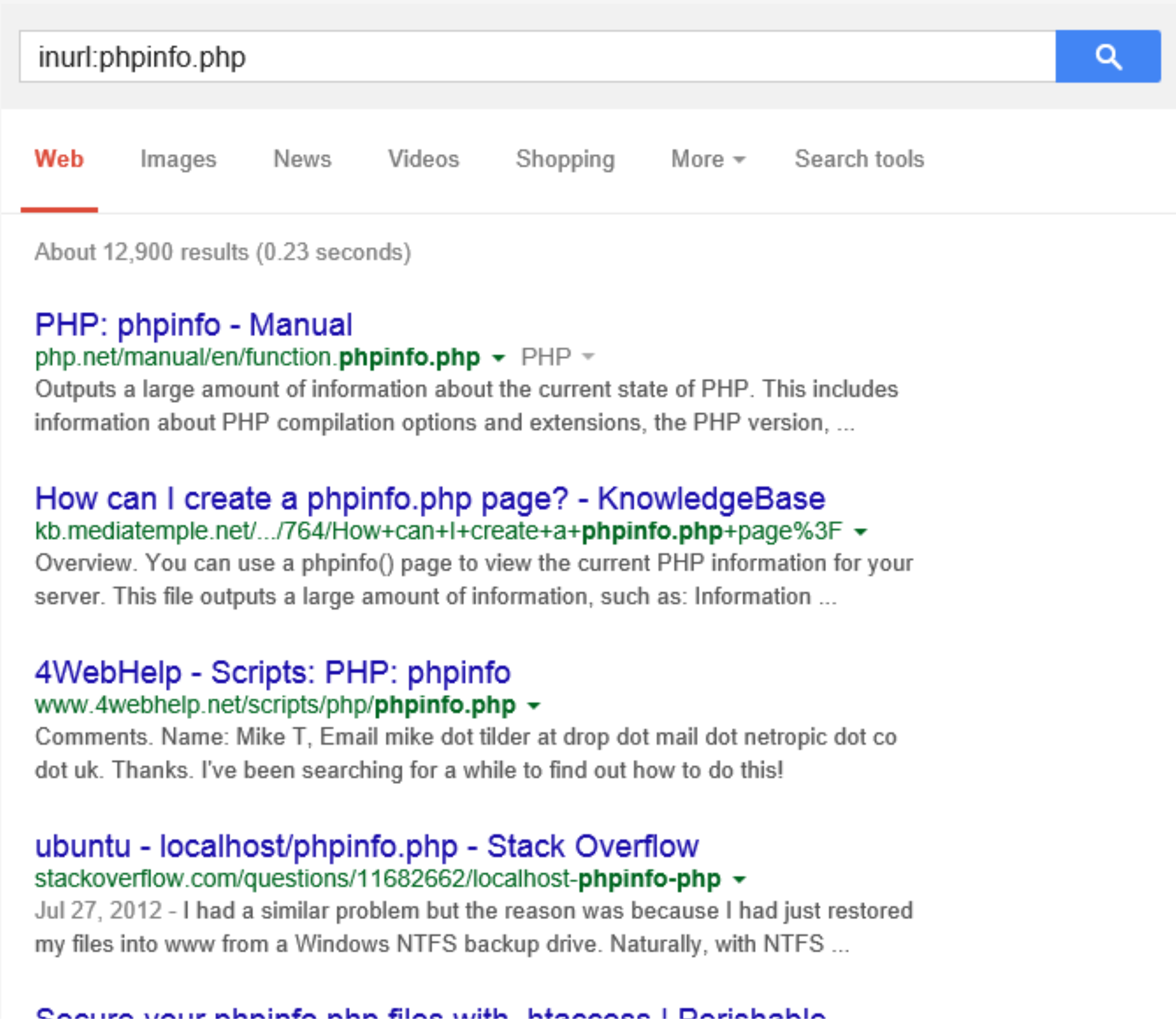




Exposed server info



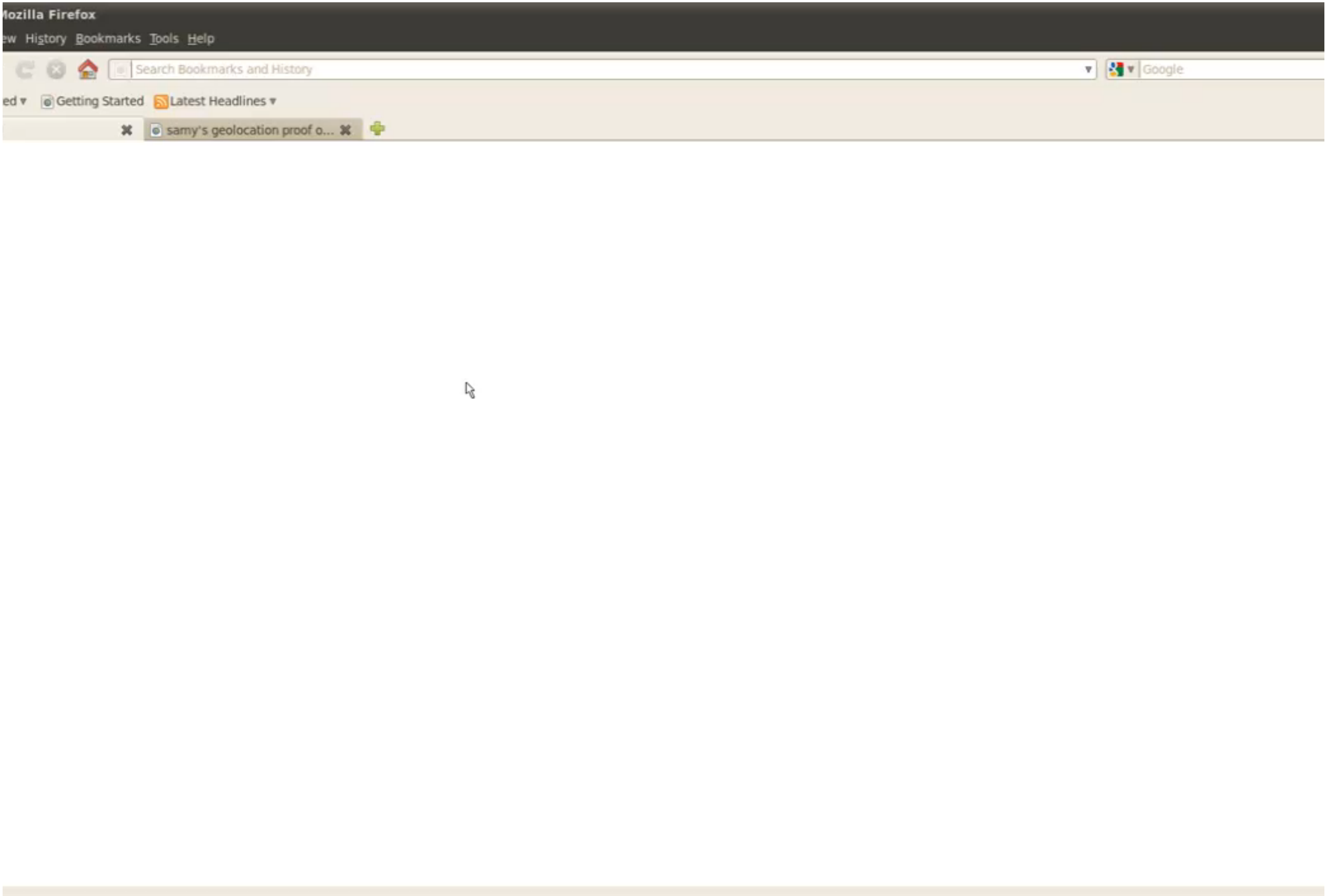
Exposed server info



Some search engines are tailor made for finding vulnerabilities.

SHODAN


SHODAN
Search
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.
The Register
It greatly lowers the technical bar needed to canvas the Internet...
threatpost
'Shodan for Penetration Testers' presented at DEF CON 18
DEFCON 18
It's a reminder to many to know what's on your network...
dark READING


Mozilla Firefox
ew History Bookmarks Tools Help
Search Bookmarks and History
Google
ed Getting Started Latest Headlines
samy's geolocation proof o...

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!

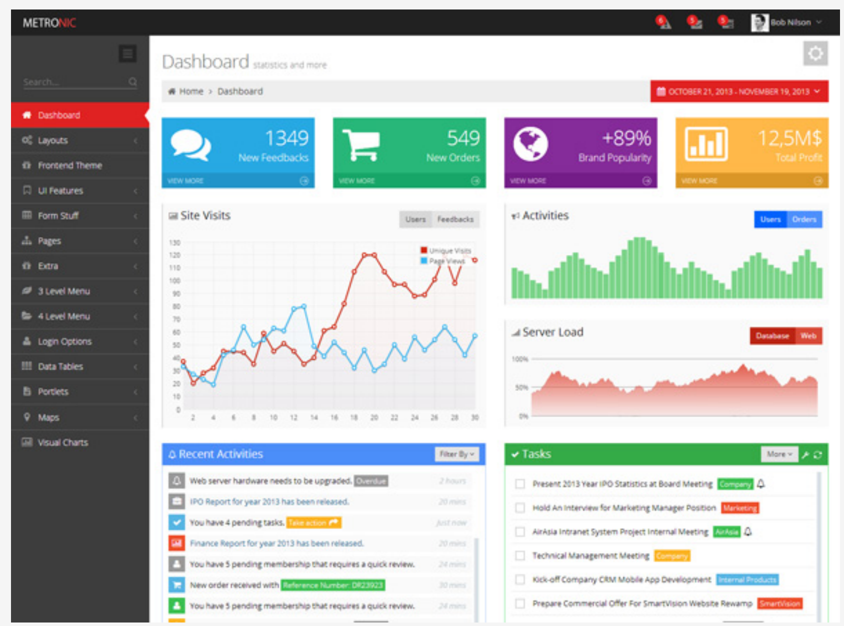
Use robots.txt

THANKS

Derrick Liu + Daniel Chiu


BANK


SENTRY
SURVEILLANCE


METRONIC
Dashboard statistics and more
Home > Dashboard
Dashboard
Layouts
Frontend Theme
UI Features
Form Stuff
Pages
Extra
3 Level Menu
4 Level Menu
Login Options
Data Tables
Portlets
Maps
Visual Charts
1349
New Feedbacks
549
New Orders
+89%
Brand Popularity
12,5M$
Total Profit
Site Visits
Activities
Server Load
Recent Activities
Tasks

Gyrophone: Recognizing Speech From Gyroscope Signals

Yan Michalevsky Dan Boneh
Computer Science Department
Stanford University

Gabi Nakibly
National Research & Simulation Center
Rafael Ltd.

Abstract

We show that the MEMS gyroscopes found on modern smart phones 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 gyro, 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 enable rich user experience. Being generally put to good use, they can sometimes unintentionally expose information 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 gyroscope and accelerometer 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 keystroke inference from nearby keyboards using accelerometers has been shown in [35]. In [21], the authors demonstrate the possibility of keystroke inference on a mobile device using accelerometers and mention the potential of using gyroscope measurements as well, while another study [19] points to the benefits of exploiting the gyroscope.

All of the above work focused on exploitation of motion events obtained from the sensors, utilizing the expected kinetic response of accelerometers and gyroscopes. 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. 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 from measurements of a single gyroscope. Therefore, we resort to automatic speech recognition. We extract features from the gyroscope measurements using various signal processing methods and train machine learning algorithms for recognition. We achieve about 50% success rate for speaker identification from a set of 10 speakers. We also show that while limiting ourselves to a small vocabulary consisting solely of digit pronunciations ("one", "two", "three", ...) and achieve speech recognition success rate of 65% for the speaker dependent case and up to 26% recognition rate for the speaker independent case. This capability allows an attacker to substantially leak information about numbers spoken over or next to a phone (i.e. credit card numbers, social security numbers and the like).

We also consider the setting of a conference room where two or more people are carrying smartphones or tablets. This setting allows an attacker to gain simultaneous measurements of speech from several gyroscopes. We show that by combining the signals from two or more phones we can increase the effective sampling rate of the acoustic signal while achieving better speech recognition rates. In our experiments we achieved 77% successful recognition rate in the speaker dependent case based on the digits vocabulary.

The paper structure is as follows: in Section 2 we provide a brief description of how a MEMS gyroscope works and present initial investigation of its properties as a microphone. In Section 3 we discuss speech analysis and describe our algorithms for speaker and speech recognition. 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. In particular, we argue that restricting the sampling rate is an effective and backwards compatible solution.

2 Gyroscope as a microphone

In this section we explain how MEMS gyroscopes operate and present an initial investigation of their susceptibility to acoustic signals.

2.1 How does a MEMS gyroscope work?

Standard-size (non-MEMS) gyroscopes are usually composed of a spinning wheel on an axle that is free to assume any orientation. Based on the principles of angular momentum the wheel resists to changes in orientation, thereby allowing to measure those changes. Nonetheless, all MEMS gyros take advantage of a different physical phenomenon – the Coriolis force. It is a fictitious force (d'Alembert force) that appears to act on an object while viewing it from a rotating reference frame (much like the centrifugal force). The Coriolis force acts in a direction perpendicular to the rotation axis of the reference frame and to the velocity of the viewed object. The Coriolis force is calculated by $F = 2m\vec{v} \times \vec{\omega}$ where m and v denote the object's mass and velocity, respectively, and ω denotes the angular rate of the reference frame.

Generally speaking, MEMS gyros measure their angular rate (ω) by sensing the magnitude of the Coriolis force acting on a moving proof mass within the gyro. Usually the moving proof mass constantly vibrates within the gyro. Its vibration frequency is also called the resonance frequency of the gyro. The Coriolis force is sensed by measuring its resulting vibration, which is orthogonal to the primary vibration movement. Some gyroscope designs use a single mass to measure the angular rate of different axes, while others use multiple masses. Such a general design is commonly called *vibrating structure gyroscope*.

There are two primary vendors of MEMS gyroscopes for mobile devices: STMicroelectronics [15] and InvenSense [7]. According to a recent survey [18] STMicroelectronics dominates with 80% market share. Teardown analyses show that this vendor's gyros can be found in Apple's iPhones and iPads [17, 8] and also in the latest generations of Samsung's Galaxy-line phones [5, 6]. The second vendor, InvenSense, has the remaining 20% market share [18]. InvenSense gyros can be found in Google's latest generations of Nexus-line phones and tablets [14, 13] as well as in Galaxy-line tablets [4, 3]. These two vendors' gyroscopes have different mechanical designs, but are both noticeably influenced by acoustic noise.

2.1.1 STMicroelectronics

The design of STMicroelectronics 3-axis gyros is based on a single driving (vibrating) mass (shown in Figure 1). The driving mass consists of 4 parts M_1, M_2, M_3 and M_4 (Figure 1(b)). They move inward and outward simultaneously at a certain frequency[1] in the horizontal plane. As shown in Figure 1(b), when an angular rate is applied on the Z-axis, due to the Coriolis effect, M_2 and M_4 will move in the same horizontal plane in opposite directions as shown by the red and yellow arrows. When an angular rate is applied on the X-axis, then M_1 and M_3 will move in opposite directions up and down out of the plane due to the Coriolis effect. When an angular rate is applied to the Y-axis, then M_2 and M_4 will move in opposite directions up and down out of the plane. The movement of the driving mass causes a capacitance change relative to stationary plates surrounding it. This change is sensed and translated into the measurement signal.

2.1.2 InvenSense

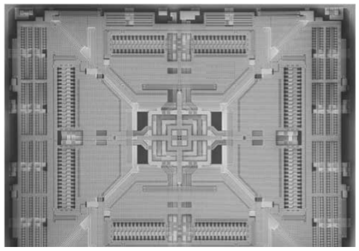
InvenSense's gyro design is based on the three separate driving (vibrating) masses[2]; each senses angular rate at a different axis (shown in Figure 2(a)). Each mass is a coupled dual-mass that move in opposite directions. The masses that sense the X and Y axes are driven out-of-plane (see Figure 2(b)), while the Z-axis mass is driven in-plane. As in the STMicroelectronics design the movement due to the Coriolis force is measures by capacitance changes.

2.2 Acoustic Effects

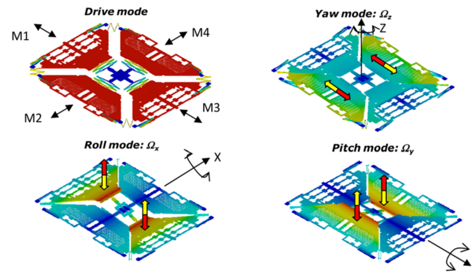
It is a well known fact in the MEMS community that MEMS gyros are susceptible to acoustic noise which degrades their accuracy [22, 24, 25]. An acoustic signal affects the gyroscope measurement by making the driving mass vibrate in the sensing axis (the axis which senses the Coriolis force). The acoustic signal can be transferred to the driving mass in one of two ways. First, it may induce mechanical vibrations to the gyros package. Additionally, the acoustic signal can travel through the gyroscope packaging and directly affect the driving mass in case it is suspended in air. The acoustic noise has the most substantial effect when it is near the resonance frequency of the vibrating mass. Such effects in some cases can render the gyro's measurements useless or even saturated. Therefore to reduce the noise effects vendors manufacture gyros with a high resonance frequency (above

[1] It is indicated in [1] that STMicroelectronics uses a driving frequency of over 20 KHz.

[2] According to [43] the driving frequency of the masses is between 25 KHz and 30 KHz.

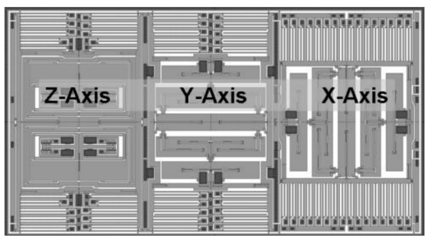


(a) MEMS structure

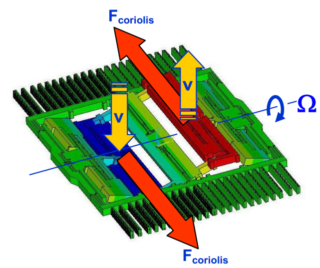


(b) Driving mass movement depending on the angular rate

Figure 1: STMicroelectronics 3-axis gyro design (Taken from [16]. Figure copyright of STMicroelectronics. Used with permission.)



(a) MEMS structure



(b) Driving mass movement depending on the angular rate

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

20 KHz) where acoustic signals are minimal. Nonetheless, in our experiments we found that acoustic signals at frequencies much lower than the resonance frequency still have a measurable effect on a gyro's measurements, allowing one to reconstruct the acoustic signal.

2.3 Characteristics of a gyro as a microphone

Due to the gyro's acoustic susceptibility one can treat gyroscope readings as if they were audio samples coming from a microphone. Note that the frequency of an audible signal is higher than 20 Hz, while in common cases the frequency of change of mobile device's angular velocity is lower than 20 cycles per second. Therefore, one can high-pass-filter the gyroscope readings in order to retain only the effects of an audio signal even if the mobile device is moving about. Nonetheless, it should be noted that this filtering may result in some loss of acoustic information since some aliased frequencies may be filtered out (see Section 2.3.2). In the following we explore the gyroscope characteristics from a standpoint of an acoustic sensor, i.e. a microphone. In this section we exemplify these characteristics by experimenting with Galaxy S III which has an STMicroelectronics gyro [6].

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 audio applications.

Sampling frequency is the rate at which a signal is sampled. According to the Nyquist sampling theorem a sampling frequency f enables us to reconstruct signals at frequencies of up to $f/2$. Hence, a higher sampling frequency allows us to more accurately reconstruct the audio signal. In most mobile devices and operating systems an application is able to sample the output of a microphone at up to 44.1 KHz. A telephone system (POTS) samples an audio signal at 8000 Hz. However, STMicroelectronics' gyroscope hardware supports sampling frequencies of up to 800 Hz [9], while InvenSense gyros' hardware support sampling frequency up to 8000 Hz [12]. Moreover, all mobile operating systems bound the sampling frequency even further – up to 200 Hz – to limit power consumption. On top of that, it appears that some browser toolkits limit the sampling frequency even further. Table 1 summarizes the results of our experi-

		Sampling Freq. [Hz]
Android 4.4	application	200
	Chrome	25
	Firefox	200
	Opera	20
iOS 7	application	100 [2]
	Safari	20
	Chrome	20

Table 1: Maximum sampling frequencies on different platforms

ments measuring the maximum sampling frequencies allowed in the latest versions of Android and iOS both for application and for web application running on common browsers. The code we used to sample the gyro via a web page can be found in Appendix B. The results indicate that 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.

2.3.2 Aliasing

As noted above, the sampling frequency of a gyro is uniform and can be at most 200 Hz. This allows us to directly sense audio signals of up to 100 Hz. Aliasing is a phenomenon where for a sinusoid of frequency f, sampled with frequency f_s, the resulting samples are indistinguishable from those of another sinusoid of frequency $|f - N \cdot f_s|$, for any integer N. The values corresponding to $N \neq 0$ are called images or aliases of frequency f. An undesirable phenomenon in general, here aliasing allows us to sense audio signals having frequencies which are higher than 100 Hz, thereby extracting more information from the gyroscope readings. This is illustrated in Figure 3.

Using the gyro, we recorded a single 280 Hz tone. Figure 3(a) depicts the recorded signal in the frequency domain (x-axis) over time (y-axis). A lighter shade in the spectrogram indicates a stronger signal at the corresponding frequency and time values. It can be clearly seen that there is a strong signal sensed at frequency 80 Hz starting around 1.5 sec. This is an alias of the 280 Hz-tone. Note that the aliased tone is indistinguishable from an actual tone at the aliased frequency. Figure 3(b) depicts a recording of multiple short tones between 130 Hz and 200 Hz. Again, a strong signal can be seen at the aliased frequencies corresponding to 130 - 170 Hz[3]. We also observe some weaker aliases that do not correspond to the base frequencies of the recorded tones, and perhaps correspond to their harmonics. Figure 3(c) depicts the recording of a chirp in the range of 420 - 480 Hz. The aliased chirp is detectable in the range of 20 - 80 Hz; however it is a rather weak signal.

2.3.3 Self noise

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. To measure the gyroscope's self noise we played 80 Hz tones for 10 seconds at different volumes while measuring it using a decibel meter. Each tone was recorded by the Galaxy S III gyroscope. While analyzing the gyro recordings we realized that the gyro readings have a noticeable increase in amplitude when playing tones with volume of 75 dB or higher which is comparable to the volume of a loud conversation. Moreover, a FFT plot of the gyroscope recordings gives a noticeable peak at the tone's frequency when playing tone with a volume as low as 57 dB which is below the sound level of a normal conversation. These findings indicate that a gyro can pick up audio signals which are lower than 100 HZ during most conversations made over or next to the phone. To test 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 much higher values that are comparable to a loud conversation.

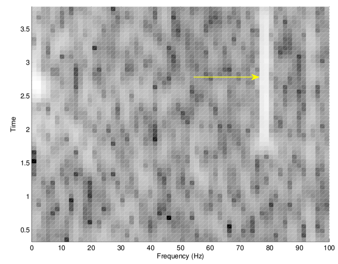
2.3.4 Directionality

We now measure how the angle at which the audio signal hits the phone affects the gyro. For this experiment we played an 80 Hz tone at the same volume three times. The tone was recorded at each time by the Galaxy S III gyro while the phone rested at a different orientation allowing the signal to hit it parallel to one of its three axes (see Figure 4). The gyroscope senses in three axes, hence for each measurement the gyro actually outputs three readings – one per axis. As we show next this property benefits the gyro's ability to pick up audio signals from every direction. For each recording we calculated the FFT magnitude at 80 Hz. Table 2 summarizes the results.

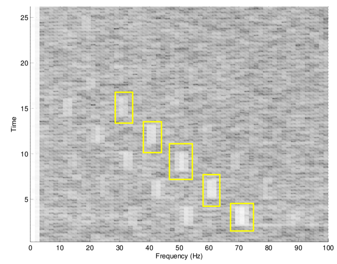
It is obvious from the table that for each direction the audio hit the gyro, there is at least one axis whose readings are dominant by an order of magnitude compared to the rest. This can be explained by STMicroelectronics gyroscope design as depicted in Figure 1[4]. When the signal travels in parallel to the phone's x or y axes, the sound pressure vibrates mostly masses laid along the respective axis, i.e. M_2 and M_4 for x axis and M_1 and M_3

[3]We do not see the aliases corresponding to 180 - 200 Hz, which might be masked by the noise at low frequencies, i.e., under 20 Hz.

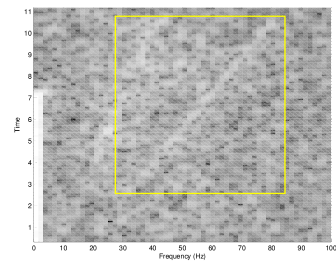
[4]This is the design of the gyro built into Galaxy S III.



(a) A single 280 Hz tone



(b) Multiple tones in the range of 130 – 170 Hz



(c) A chirp in the range of 420 – 480 Hz

Figure 3: Example of aliasing on a mobile device. Nexus 4 (a,c) and Galaxy SII (b).

Tone direction:	X			Y			Z		
Recording direction:	x	y	z	x	y	z	x	y	z
Amplitude:	0.002	0.012	0.0024	0.01	0.007	0.004	0.007	0.0036	0.0003

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.

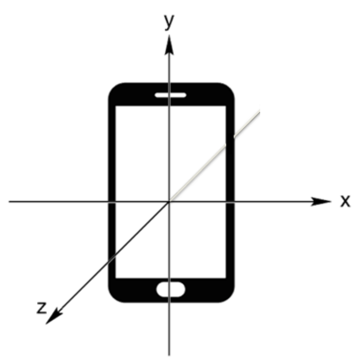


Figure 4: Coordinate system of Android and iOS.

for the *y* axis; therefore, the gyro primarily senses a rotation at the *y* or *x* axes, respectively (see Section 2.1.1). When the signal travels in parallel to the phone's *z* axis then the sound pressure vibrates all the 4 masses up and down, hence the gyro primarily senses a rotation at both *x* and *y* axes.

These findings indicate that the gyro is an omnidirectional audio sensor allowing it to pick up audio signal from every direction.

3 Speech analysis based on a single gyroscope

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, and even recognize the spoken words or phrases. We do so by leveraging the fact that aliasing causes information leaks from higher frequency bands into the sub-Nyquist range.

Since the fundamentals of human voices are roughly in the range of 80 – 1100 Hz [20], we can capture a large fraction of the interesting frequencies, considering the results we observe in 2.3.2. Although we do not delve into comparing performance for different types of speakers, one might expect that given a stronger gyroscope response for low frequencies, typical adult male speech (Bass, Baritone, Tenor) could be better analyzed than typical female or child speech (Alto, Mezzo-Soprano, Soprano) [5], however our tests show that this is not necessarily the case.

The signal recording, as captured by the gyroscope, is not comprehensible to a human ear, and exhibits a mixture of low frequencies and aliases of frequencies beyond the Nyquist sampling frequency (which is 1/2 the sampling rate of the Gyroscope, i.e. 100 Hz). While the signal recorded by a single device does not resemble speech, it is possible to train a machine to transcribe the signal with significant success.

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[6]; It can also be speaker dependent (in which case the recognizer is trained per speaker) or speaker in-

[5] For more information about vocal range see `http://www.wikipedia.org/wiki/Vocal_range`

[6] For example by identifying phonemes and combining them to words.

dependent (in which case the recognizer is expected to identify phrases pronounced by different speakers and possibly ones that were not encountered in the training set). Additionally, speech analysis may be also used to identify the speaker.

We focused on speaker identification (including gender identification of the speaker) and isolated words recognition while attempting both speaker independent and speaker dependent recognition. Although we do not demonstrate fluent speech transcription, we suggest that successful isolated words recognition could be fairly easily transformed into a transcription algorithm by incorporating word slicing and HMM [40]. We did not aim to implement a state-of-the-art speech recognition algorithm, nor to thoroughly evaluate or do a comparative analysis of the classification tests. Instead, we tried to indicate the potential risk by showing significant success rates of our speech analysis algorithms compared to randomly guessing. This section describes speech analysis techniques that are common in practice, our approach, and suggestions for further improvements upon it.

3.1 Speech processing: features and algorithms

3.1.1 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 to short (10 – 30 ms) overlapping or non-overlapping windows and operation on them. This results in a time-series of features that characterize 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 basically compensates for the non-linear frequency response of the human ear[7]. The Cepstrum transformation is an attempt to separate the excitation signal originated by air passing through the vocal tract from the effect of the vocal tract (acting as a filter shaping that excitation signal). The latter is more important for 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 [30].

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 to groups. We use SVM to distinguish male and female speakers. Multi-class SVMs can be constructed using multiple binary SVMs, to distinguish between multiple groups. We used a multi-class SVM to distinguish between multiple speakers, and to recognize words from a limited dictionary.

Gaussian Mixture Model (GMM) has been successfully used for speaker identification [41]. We can train a GMM for each group in the training stage. In the testing stage we can obtain a match score for the sample using each one of the GMMs and classify the sample according to the group corresponding to the GMM that yields the maximum score.

Dynamic Time Warping (DTW) is a time-series matching and alignment technique [37]. 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[8]. We applied silence removal to include only relevant information and minimize noise. The silence removal algorithm was based on the implementation in [29], which classifies the speech into voiced and unvoiced segments (filtering out the unvoiced) according to dynamically set thresholds for Short-Time Energy 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.

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

[7] Approximated as logarithmic by the Mel-scale

[8] 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 WAV file at higher sampling rate with standard speech processing tools.

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 an 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_i^l corresponding to that guess in the training set. Let us denote this similarity function by $D(y, X_i^l)$. 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

$$S^l = \sum_i D(y, X_i^l)$$

After obtaining a total score for all possible words, the sample is classified according to the maximum total score

$$C(y) = \underset{l}{\operatorname{argmax}}\, S^l$$

3.4 Experiment setup

Our setup consisted of a set of loudspeakers that included 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 to obtain as high SNR as possible for our experiments. This means that for more restrictive attack scenarios (farther source, lower volume) there will be a need to handle low SNR, perhaps by filtering out the noise or applying some other preprocessing for emphasizing the speech signal. [9]

Figure 5: Experimental setup

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 set out to recognize speech of a limited dictionary, the recognition of which would still leak substantial private information. For this work we chose to focus on the digits dictionary, which includes the words: zero, one, two..., nine, and ”oh”. Recognition of such words would enable an attacker to eavesdrop on private information, such as credit card numbers, telephone numbers, social security numbers and the like. This information may be eavesdropped when the victim speaks over or next to the phone.

In our experiments, we use the following corpus of audio signals on which we tested our recognition algorithms.

TIDIGITS This is a subset 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. The corpus is digitized at 20 kHz.

3.4.2 Mobile devices

We primarily conducted our experiments using the following mobile devices:

[9] We tried recording in an anechoic chamber, but it didn't seem to provide better recognition results compared to a regular room. We therefore did not proceed with the anechoic chamber experiments. Yet, further testing is needed to understand whether we can benefit significantly from an anechoic environment.

1. Nexus 4 phone which according to a teardown analysis [13] is equipped with an InvenSense MPU-6050 [12] gyroscope and accelerometer chip.

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

3. Samsung Galaxy S III phone which according to a teardown analysis [6] is equipped with an STMicroelectronics LSM330DLC [10] gyroscope and accelerometer chip.

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 TIDITS 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 readings in our discussion. We also included in the tables some results obtained using a Galaxy III device, for comparison.

Results for gender identification are presented in Table 3. As we see, using DTW scoring for STFT features yielded a much better success rate.

Results for speaker identification are presented in Table 4. Since the results for a mixed female-male set of speakers may be partially attributed to successful gender identification, we tested classification for speakers of the same gender. In this setup we have 5 different speakers. The improved classification rate (except for DTW for female speaker set) can be partially attributed to a smaller number of speakers.

The results for speaker-independent isolated word recognition are summarized in Table 5. We had correct classification rate of $\sim 10\%$ 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, explains the not so high recognition rate, exhibiting many false positives for the words ”6” and ”9”. At the same time the recognition rate for

	SVM	GMM	DTW
Nexus 4	80%	72%	84%
Galaxy S III	82%	68%	58%

Table 3: Speaker's gender identification results

		SVM	GMM	DTW
Nexus 4	Mixed female/male	23%	21%	50%
	Female speakers	33%	32%	45%
	Male speakers	38%	26%	65%
Galaxy S III	Mixed female/male	20%	19%	17%
	Female speakers	30%	20%	29%
	Male speakers	32%	21%	25%

Table 4: Speaker identification results

		SVM	GMM	DTW
Nexus 4	Mixed female/male	10%	9%	17%
	Female speakers	10%	9%	26%
	Male speakers	10%	10%	23%
Galaxy S III	Mixed female/male	7%	12%	7%
	Female speakers	10%	10%	12%
	Male speakers	10%	6%	7%

Table 5: Speaker-independent case – isolated words recognition results

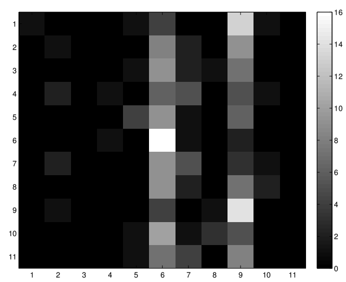


Figure 6: Speaker independent word recognition using DTW: confusion matrix as a heat map. $c_{(i,j)}$ corresponds to the number of samples from group i that were classified as j, where i, j are the row and column indices respectively.

SVM	GMM	DTW
15%	5%	65%

Table 6: Speaker-dependent case – isolated words recognition for a single speaker. Results obtained via ”leave-one-out” cross-validation on 44 recorded words pronounced by a single speaker. Recorded using a Nexus 4 device.

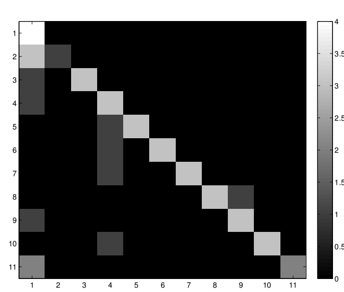


Figure 7: Speaker dependent word recognition using DTW: confusion matrix as a heat map.

these particular words is high, contributing to the correct identification rate.

For a speaker-dependent case one may expect to get better recognition results. We recorded a set of 44 digit pronunciations, where each digit was pronounced 4 times. We tested the performance of our classifiers using ”leave-one-out” cross-validation. The results are presented in Table 6, and as we expected exhibit an improvement compared to the speaker independent recognition[10] (except for GMM performance that is equivalent to randomly guessing). The confusion matrix corresponding to speaker-dependent word recognition using DTW is presented in Figure 7.

DTW method outperforms SVM and GMM in most cases. One would expect that DTW would perform better for word recognition since the changing in time of the spectral features is taken into account. While true for Nexus 4 devices it did not hold for measurements taken with Galaxy III. possible explanation to that is that the low-pass filtering on the Galaxy III device renders all methods quite ineffective resulting in a success rate equivalent to a random guess. For gender and speaker identification, we would expect statistical spectral features based methods (SVM and GMM) to perform at least as good as DTW. It is only true for the Galaxy S III mixed speaker set and gender identification cases, but not for the other experiments. Specifically for gender identification, capturng the temporal development of the spectral feature wouldn’t seem like a clear advantage and is therefore somewhat surprising. One comparative study that supports the advantage of DTW over SVM for speaker recognition is [48]. It doesn’t explain though why it outperforms GMM which is a well established method for speaker identification. More experimentation is required to confirm whether this phenomenon is consistent and whether it is related to capturing the high frequencies.

3.7 Further improvement

We suggest several possible future improvements on our recognition algorithms. Phoneme recognition instead of whole words, in combination with an HMM could improve the recognition results. This could be more suitable since different pronunciations have different lengths, while an HMM could introduce a better probabilistic recognition of the words. Pre-filtering of the signal could be beneficial and reduce irrelevant noise. It is not clear which frequencies should be filtered and therefore some experimentation is needed to determine it.

[10] It is the place to mention that a larger training set for speaker independent word recognition is likely to yield better results. For our tests we used relatively small training and evaluation sets.

For our experiments, we used samples recorded by the gyroscope for training. For speaker-dependent speech recognition we can imagine it may be easier to obtain regular speech samples for a particular speaker than a transcribed recording of gyroscope samples. Even for speaker independent 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 when the training set is based on normal audio recordings, downsampled to 200 Hz to simulate a gyroscope recording.

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, one may take advantage of this to get a composed signal of the three axes to get a better signal-to-noise ratio.

While we suggested that the signal components related to speech, and those related to motion lie in separate frequency bands, the performance of speech analysis in the presence of such noise is yet to be evaluated.

4 Reconstruction using multiple devices

In this section we suggest that isolated word recognition can be improved if we sample the gyroscopes of multiple devices that are in close proximity, such that they exhibit a similar response to the acoustic signals around them. This can happen for instance in a conference room where two mobile devices are running malicious applications or, having a browser supporting high-rate sampling of the gyroscope, are tricked into browsing to a malicious website.

We do not refer here to the possibility of using several different gyroscope readings to effectively obtain a larger feature vector, or have the classification algorithm take into account the score obtained for all readings. While such methods to exploit the presence of more than one acoustic side-channel may prove very efficient we leave them outside the scope of this study. It also makes sense to look into existing methods for enhancing speech recognition using multiple microphones, covered in signal processing and machine learning literature (e.g., [23]).

Instead, we look at the possibility of obtaining an enhanced signal by using all of the samples for reconstruction, thus effectively obtaining higher sampling rate. Moreover, we hint at the more ambitious task of reconstructing a signal adequate enough to be comprehensible by a human listener, in a case where we gain access to readings from several compromised devices. While there are several practical obstacles to it, we outline the idea, and demonstrate how partial implementation of it facilitates the automatic speech recognition task.

We can look at our system as an array of time-interleaved data converters (interleaved ADCs). Interleaved ADCs are multiple sampling devices where each samples the signal with a sub-Nyquist frequency. While the ADCs should ideally have time offsets corresponding to a uniform sampling grid (which would allow to simply interleave the samples and reconstruct according to the Whittaker-Shannon interpolation formula [44]), usually there will be small time skews. Also, DC offsets and different input gains can affect the result and must all be compensated.

This problem is studied in a context of analog design and motivated by the need to sample high-frequency signals using low-cost and energy-efficient low-frequency A/D converters. While many papers on the subject exist, such as [27], the proposed algorithms are usually very hardware centric, oriented towards real-time processing at high-speed, and mostly capable of compensating for very small skews. Some of them require one ADC that samples the signal above the Nyquist rate, which is not available in our case. At the same time, we do not aim for a very efficient, real-time algorithm. Utilizing recordings from multiple devices implies offline processing of the recordings, and we can afford a long run-time for the task.

The ADCs in our case have the same sampling rate $F_s = 1/T = 200$. We assume the time-skews between them are random in the range $[0, T_Q]$ where for N ADCs $T_Q = \frac{T}{N}$ is the Nyquist sampling period. Being located at different distances from the acoustic source they are likely to exhibit considerably different input gains, and possibly have some DC offset. [26] provides background for understanding the problems arising in this configuration and covers some possible solutions.

4.1 Reconstruction algorithm

4.1.1 Signal offset correction

To correct a constant offset we can take the mean of the Gyro samples and compare it to 0 to get the constant offset. It is essentially a simple DC component removal.

4.1.2 Gain mismatch correction

Gain mismatch correction is crucial for a successful signal reconstruction. We correct the gain by normalizing the signal to have standard deviation equal to 1. In case we are provided with some reference signal with a known peak, we can adjust the gains of the recordings so that the amplitude at this peak is equal for all of them.

4.1.3 Time mismatch correction

While gyroscope motion events are provided with precise timestamps set by the hardware, which theoretically could have been used for aligning the recordings, in practice, we cannot rely on the clocks of the mobile devices to be synchronized. Even if we take the trouble of synchronizing the mobile device clock via NTP, or even better, a GPS clock, the delays introduced by the network, operating system and further clock-drift will stand in the way of having clock accuracy on the order of a millisecond[11]. While not enough by itself, such synchronization is still useful for coarse alignment of the samples.

El-Manar describes foreground and background time-mismatch calibration techniques in his thesis [27]. Foreground calibration means there is a known signal used to synchronized all the ADCs. While for the purpose of testing we can align the recordings by maximizing the cross-correlation with a known signal, played before we start recording, in an actual attack scenario we probably won't be able to use such a marker[12]. Nevertheless, in our tests we attempted aligning using a reference signal as well. It did not exhibit a clear advantage over obtaining coarse alignment by finding the maximum of the cross-correlation between the signals. One can also exhaustively search a certain range of possible offsets, choosing the one that results in a reconstruction of a sensible audio signal.

Since this only yields alignment on the order of a sampling period of a single gyroscope (T), we still need to find the more precise time-skews in the range $[0,T]$. We can scan a range of possible time-skews, choosing the one that yields a sensible audio signal. We can think of an automated evaluation of the result by a speech recognition engine or scoring according to features that would indicate human speech, suggesting a successful reconstruction.

This scanning is obviously time consuming. If we have n sources, we set one of the time skews (arbitrary) to 0, and have $n-1$ degrees of freedom to play with, and the complexity grows exponentially with the number of sources. Nevertheless, in an attack scenario, it is not impossible to manually scan all possibilities looking for the best signal reconstruction, provided the information is valuable to the eavesdropper.

[11]Each device samples with a period of 5 ms, therefore even 1 ms clock accuracy would be quite coarse.

[12]While an attacker may be able to play using one of the phones' speakers a known tone/chirp (no special permissions are needed), it is unlikely to be loud enough to be picked up well by the other device, and definitely depends on many factors such as distance, position etc.

4.1.4 Signal reconstruction from non-uniform samples

Assuming we have compensated for offset, gain mismatch and found the precise time-skews between the sampling devices, we are dealing with the problem of signal reconstruction from periodic, non-uniform samples. A seminal paper on the subject is [28] by Eldar et al. Among other works in the field are [39, 46] and [31]. Sindhi et al. [45] propose a discrete time implementation of [28] using digital filterbanks. The general goal is, given samples on a non-uniform periodic grid, to obtain estimation of the values on a uniform sampling grid, as close as possible to the original signal.

A theoretic feasibility justification lies in Papoulis' Generalized Sampling theorem [38]. Its corollary is that a signal bandlimited to π/T_Q can be recovered from the samples of N filters with sampling periods $T = NT_Q$. [13] We suggest using one of the proposed methods for signal reconstruction from periodic non-uniform samples. With only several devices the reconstructed speech will still be narrow-band. While it won't necessarily be easily understandable by a human listener, it could be used for better automated identification. Applying narrowband to wideband speech extension algorithms [36] might provide audio signals understandable to a human listener.

We suggest using one of the methods for signal reconstruction from periodic non-uniform samples mentioned above. With only several devices the reconstructed speech will still be narrow-band. For example, using readings from two devices operating at 200 Hz and given their relative time-skew we obtain an effective sampling rate of 400 Hz. For four devices we obtain a sampling rate of 800 Hz, and so on. While a signal reconstructed using two devices still won't be easily understandable by a human listener, it could be used to improve automatic identification.

We used [28] as a basis for our reconstruction algorithm. The discussion of *recurrent non-uniform sampling* directly pertains to our task. It proposes a filterbank scheme to interpolate the samples such that an approximation of the values on the uniform grid is obtained. The derivation of the discrete-time interpolation filters is provided in Appendix A.

This method allows us to perform reconstruction with arbitrary time-skews; however we do not have at the time a good method for either a very precise estimation

[13]It is important to note that in our case the signal is not necessarily bandlimited as required. While the base pitch of the speech can lie in the range $[0, 200 \cdot N]$, it can contain higher frequencies that are captured in the recording due to aliasing, and may interfere with the reconstruction. It depends mainly on the low-pass filtering applied by the gyroscope. In InvenSense's MPU-6050, Digital Low-Pass Filtering (DLPF) is configurable through hardware registers [11], so the conditions depend to some extent on the particular driver implementation.

SVM	GMM	DTW
18%	14%	77%

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.

of the time-skews or automatic evaluation of the reconstruction outcome (which would enable searching over a range of possible values). For our experiment we applied this method to the same set of samples used for speaker-dependent speech recognition evaluation, which was recorded simultaneously by two devices. We used the same value for τ, the time-skew for all samples, and therefore chose the expected value $\tau = T/2$ which is equivalent to the particular case of sampling on a uniform grid (resulting in all-pass interpolation filters). It is essentially the same as interleaving the samples from the two readings, and we ended up implementing this trivial method as well, in order to avoid the adverse effects of applying finite non-ideal filters.

It is important to note that while we propose a method rooted in signal processing theory, we cannot confidently attribute the improved performance to obtaining a signal that better resembles the original, until we take full advantage of the method by estimating the precise time-skew for each recording, and applying true non-uniform reconstruction. It is currently left as an interesting future improvement, for which the outlined method can serve as a starting point. In this sense, our actual experiment can be seen as taking advantage of better feature vectors, comprised of data from multiple sources.

4.1.5 Evaluation

We evaluated this approach by repeating the speaker-dependent word recognition experiment on signals reconstructed from readings of two Nexus 4 devices. Table 7 summarizes the final results obtained using the sample interleaving method[14].

There was a consistent noticeable improvement compared to the results obtained using readings from a single device, which supports the value of utilizing multiple gyroscopes. We can expect that adding more devices to the setup would further improve the speech recognition.

[14] We also compared the performance of the DTW classifier on samples reconstructed using the filterbank approach. It yielded a slightly lower correct classification rate of 75% which we attribute to the mentioned effects of applying non-ideal finite filters.

5 Further Attacks

In this section we suggest directions for further exploitation of the gyroscopes:

Increasing the gyro's sampling rate. One possible attack is related to the hardware characteristics of the gyro devices. The hardware upper bound on sampling frequency is higher than that imposed by the operating system or by applications[15]. InvenSense MPU-6000/MPU-6050 gyroscopes can provide a sampling rate of up to 8000 Hz. That is the equivalent of a POTS (telephony) line. STMicroelectronics gyroscopes only allow up to 800 Hz sampling rate, which is still considerably higher than the 200 Hz allowed by the operating system (see Appendix C). If the attacker can gain a one-time privileged access to the device, she could patch an application, or a kernel driver, thus increasing this upper bound. The next steps of the attack are similar: obtaining gyroscope measurements using an application or tricking the user into leaving the browser open on some website. Obtaining such a high sampling rate would enable using the gyroscope as a microphone in the full sense of hearing the surrounding sounds.

Source separation. Based on experimental results presented in Section 2.3.4 it is obvious that the gyroscope measurements are sensitive to the relative direction from which the acoustic signal arrives. This may give rise to the possibility to detect the angle of arrival (AoA) at which the audio signal hits the phone. Using AoA detection one may be able to better separate and process multiple sources of audio, e.g. multiple speakers near the phone.

Ambient sound recognition. There are works (e.g. [42]) which aim to identify a user's context and whereabouts based on the ambient noise detected by his smart phone, e.g restaurant, street, office, and so on. Some contexts are loud enough and may have distinct fingerprint in the low frequency range to be able to detect them using a gyroscope, for example railway station, shopping mall, highway, and bus. This may allow an attacker to leak more information on the victim user by gaining indications of the user's whereabouts.

6 Defenses

Let us discuss some ways to mitigate the potential risks. As it is often the case, a secure design would require an

[15] As we have shown, the sampling rate available on certain browsers is much lower than the maximum sampling rate enabled by the OS. However, this is an application level constraint.

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 website), 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 hardware vendors to block this threat.

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 greatful to 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.

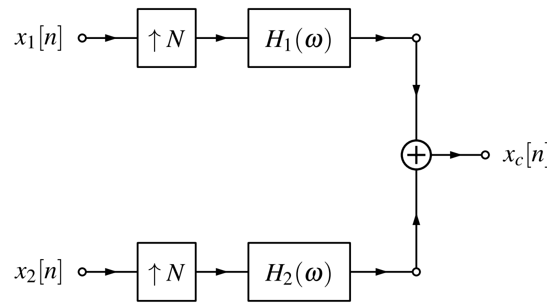


Figure 8: Filterbank reconstruction scheme

A Signal reconstruction from Recurrent Non-Uniform Samples

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

$$h_p(t) = a_p sinc\left(\frac{t}{T}\right) \prod_{q=0, q\neq p}^{N-1} \sin\left(\frac{\pi\left(t+t_p-t_q\right)}{T}\right)$$

We then sample this expression at times $t = nT_Q - t_p$ and calculate the filter coefficients for 48 taps. Given these filters, the reconstruction process consists of up-sampling the input signals by factor N, where $N = T/T_Q$ is the number of ADCs, 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:

```
if(window.DeviceMotionEvent) {
  window.addEventListener('devicemotion', function(
     event) {
    var r = event.rotationRate;
    if ( r!=null ) {
      console.log('Rotation at [x,y,z] is: [' +
        r.alpha+','+r.beta+','+r.gamma+']\n');
    }
  }
}
```

Figure 9 depicts measurements of the above code running on Firefox (Android) while sampling an audio chirp 50 – 100 Hz.

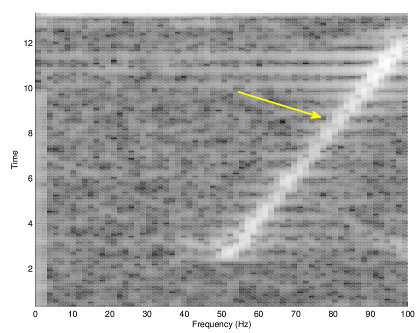


Figure 9: Recording audio at 200 Hz using JavaScript code on a web-page accessed from the Firefox browser for Android.

C Gyroscope rate limitation on Android

Here we see a code snippet from the Invensense driver for Android, taken from *hardware/invensense/65xx/libsensors_iio/MPLSensor.cpp*.
The OS is enforcing a rate of 200 Hz.

```
static int hertz_request = 200;
#define DEFAULT_MPL_GYRO_RATE          (20000L)      //us
...
#define DEFAULT_HW_GYRO_RATE           (100)         //Hz
#define DEFAULT_HW_ACCEL_RATE          (20)          //ms
...
/* convert ns to hardware units */
#define HW_GYRO_RATE_NS                (1000000000LL / rate_request) // to Hz
#define HW_ACCEL_RATE_NS               (rate_request / (1000000L))   // to ms
...
/* convert Hz to hardware units */
#define HW_GYRO_RATE_HZ                (hertz_request)
#define HW_ACCEL_RATE_HZ               (1000 / hertz_request)
```

D Code Release

We provide the source code of the Android application we used for recording the sensor measurements, as well as the Matlab code we used for analyzing the data and training and testing of the speech recognition algorithms. We also provide the gyroscope recordings used for the evaluation of our method. The code and data can be downloaded from the project website at http://crypto.stanford.edu/gyrophone. In addition, we provide a web page that records gyroscope measurements if accessed from a device that supports it.

References

[1] 3-axis digital gyroscopes. http://www.st.com/st-web-ui/static/active/en/resource/sales_and_marketing/promotional_material/flyer/fl3axdigitalgyro.pdf.

[2] Corona SDK API reference. http://docs.coronalabs.com/api/library/system/setGyroscopeInterval.html.

[3] Galaxy Tab 7.7. http://www.techrepublic.com/blog/cracking-open/galaxy-tab-77-teardown-reveals-lots-of-samsungs-homegrown-hardware/588/.

[4] Inside the Latest Galaxy Note 3. http://www.chipworks.com/en/technical-competitive-analysis/resources/blog/inside-the-galaxy-note-3/.

[5] Inside the Samsung Galaxy S4. http://www.chipworks.com/en/technical-competitive-analysis/resources/blog/inside-the-samsung-galaxy-s4/.

[6] Inside the Samsung Galaxy SIII. http://www.chipworks.com/en/technical-competitive-analysis/resources/blog/inside-the-samsung-galaxy-siii/.

[7] InvenSense Inc. http://www.invensense.com/.

[8] iPad Mini Retina Display Teardown. http://www.ifixit.com/Teardown/iPad+Mini+Retina+Display+Teardown/19374.

[9] L3G4200D data sheet. http://www.st.com/st-web-ui/static/active/en/resource/technical/document/datasheet/CD00265057.pdf.

[10] LSM330DLC data sheet. http://www.st.com/st-web-ui/static/active/en/resource/technical/document/datasheet/DM00037200.pdf.

[11] MPU-6000 and MPU-6050 Register Map and Descriptions. http://invensense.com/mems/gyro/documents/RM-MPU-6000A.pdf.

[12] MPU-6050 product specification. http://www.invensense.com/mems/gyro/documents/PS-MPU-6000A-00v3.4.pdf.

[13] Nexus 4 Teardown. http://www.ifixit.com/Teardown/Nexus+4+Teardown/11781.

[14] Nexus 7 Teardown. http://www.ifixit.com/Teardown/Nexus+7+Teardown/9623.

[15] STMicroelectronics Inc. http://www.st.com/.

[16] Everything about STMicroelectronics 3-axis digital MEMS gyroscopes. http://www.st.com/web/en/resource/technical/document/technical_article/DM00034730.pdf, July 2011.

[17] iPhone 5S MEMS Gyroscope STMicroelectronics 3x3mm - Reverse Costing Analysis. http://www.researchandmarkets.com/research/lxrnrn/iphone_5s_mems, October 2013.

[18] MEMS for Cell Phones and Tablets. http://www.i-micronews.com/upload/Rapports/Yole_MEMS_for_Mobile_June_2013_Report_Sample.pdf, July 2013.

[19] AL-HAIQI, A., ISMAIL, M., AND NORDIN, R. On the best sensor for keystrokes inference attack on android. *Procedia Technology 11* (2013), 989–995.

[20] APPELMAN, D. *The Science of Vocal Pedagogy: Theory and Application.* Midland book. Indiana University Press, 1967.

[21] CAI, L., AND CHEN, H. Touchlogger: inferring keystrokes on touch screen from smartphone motion. In *Proceedings of the 6th USENIX conference on Hot topics in security* (2011), USENIX Association, pp. 9–9.

[22] CASTRO, S., DEAN, R., ROTH, G., FLOWERS, G. T., AND GRANTHAM, B. Influence of acoustic noise on the dynamic performance of mems gyroscopes. In *ASME 2007 International Mechanical Engineering Congress and Exposition* (2007), pp. 1825–1831.

[23] CRAMMER, K., KEARNS, M., AND WORTMAN, J. Learning from multiple sources. *The Journal of Machine Learning Research 9* (2008), 1757–1774.

[24] DEAN, R. N., CASTRO, S. T., FLOWERS, G. T., ROTH, G., AHMED, A., HODEL, A. S., GRANTHAM, B. E., BITTLE, D. A., AND BRUNSCH, J. P. A characterization of the performance of a mems gyroscope in acoustically harsh environments. *Industrial Electronics, IEEE Transactions on 58*, 7 (2011), 2591–2596.

[25] DEAN, R. N., FLOWERS, G. T., HODEL, A. S., ROTH, G., CASTRO, S., ZHOU, R., MOREIRA, A., AHMED, A., RIFKI, R., GRANTHAM, B. E., ET AL. On the degradation of mems gyroscope performance in the presence of high power acoustic noise. In *Industrial Electronics, 2007. ISIE 2007. IEEE International Symposium on* (2007), IEEE, pp. 1435–1440.

[26] EL-CHAMMAS, M., AND MURMANN, B. *Background calibration of time-interleaved data converters*. Springer, 2012.

[27] EL-CHAMMAS, M. I. *Background Calibration of Timing Skew in Time-Interleaved A/D Converters*. Stanford University, 2010.

[28] ELDAR, Y. C., AND OPPENHEIM, A. V. Filterbank reconstruction of bandlimited signals from nonuniform and generalized samples. *Signal Processing, IEEE Transactions on 48*, 10 (2000), 2864–2875.

[29] GIANNAKOPOULOS, T. A method for silence removal and segmentation of speech signals, implemented in matlab.

[30] HASAN, M. R., JAMIL, M., AND RAHMAN, M. G. R. M. S. Speaker identification using mel frequency cepstral coefficients. *variations 1* (2004), 4.

[31] JOHANSSON, H. K., AND LÖWENBORG, P. *Reconstruction of periodically nonuniformly sampled bandlimited signals using time-varying FIR filters*. 2005.

[32] LARTILLOT, O., TOIVIAINEN, P., AND EEROLA, T. A matlab toolbox for music information retrieval. In *Data analysis, machine learning and applications*. Springer, 2008, pp. 261–268.

[33] LEONARD, R. G., AND DODDINGTON, G. TIDIGITS. `http://catalog.ldc.upenn.edu/LDC93S10`, 1993.

[34] MANTYJARVI, J., LINDHOLM, M., VILDJIOUNAITE, E., MAKELA, S.-M., AND AILISTO, H. Identifying users of portable devices from gait pattern with accelerometers. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on* (2005), vol. 2, IEEE, pp. ii–973.

[35] MARQUARDT, P., VERMA, A., CARTER, H., AND TRAYNOR, P. (sp) iphone: decoding vibrations from nearby keyboards using mobile phone accelerometers. In *Proceedings of the 18th ACM conference on Computer and communications security* (2011), ACM, pp. 551–562.

[36] MIET, G. Towards wideband speech by narrowband speech bandwidth extension: magic effect or wideband recovery? *These de doctorat, University of Maine* (2001).

[37] MÜLLER, M. Dynamic time warping. *Information retrieval for music and motion* (2007), 69–84.

[38] PAPOULIS, A. Generalized sampling expansion. *Circuits and Systems, IEEE Transactions on 24*, 11 (1977), 652–654.

[39] PRENDERGAST, R. S., LEVY, B. C., AND HURST, P. J. Reconstruction of band-limited periodic nonuniformly sampled signals through multirate filter banks. *Circuits and Systems I: Regular Papers, IEEE Transactions on 51*, 8 (2004), 1612–1622.

[40] RABINER, L. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* (1989).

[41] REYNOLDS, D. A., AND ROSE, R. C. Robust text-independent speaker identification using gaussian mixture speaker models. *Speech and Audio Processing, IEEE Transactions on 3*, 1 (1995), 72–83.

[42] ROSSI, M., FEESE, S., AMFT, O., BRAUNE, N., MARTIS, S., AND TROSTER, G. Ambientsense: A real-time ambient sound recognition system for smartphones. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on* (2013), IEEE, pp. 230–235.

[43] SEEGER, J., LIM, M., AND NASIRI, S. Development of high-performance high-volume consumer MEMS gyroscopes. `http://www.invensense.com/mems/gyro/documents/whitepapers/Development-of-High-Performance-High-Volume-Consumer-MEMS-Gyroscopes.pdf`.

[44] SHANNON, C. E. Communication in the presence of noise. *Proceedings of the IRE 37*, 1 (1949), 10–21.

[45] SINDHI, S., AND PRABHU, K. Reconstruction of N-th Order Nonuniformly Sampled Signals Using Digital Filter Banks. *commsp.ee.ic.ac.uk* (2012).

[46] STROHMER, T., AND TANNER, J. Fast reconstruction algorithms for periodic nonuniform sampling with applications to time-interleaved adcs. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on* (2007), vol. 3, IEEE, pp. III–881.

[47] WALKER, W., LAMERE, P., KWOK, P., RAJ, B., SINGH, R., GOUVEA, E., WOLF, P., AND WOELFEL, J. Sphinx-4: A flexible open source framework for speech recognition. Tech. rep., 2004.

[48] YEE, L., AND AHMAD, A. Comparative Study of Speaker Recognition Methods: DTW, GMM and SVM. *comp.utm.my*.

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

State of the art

- 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

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]

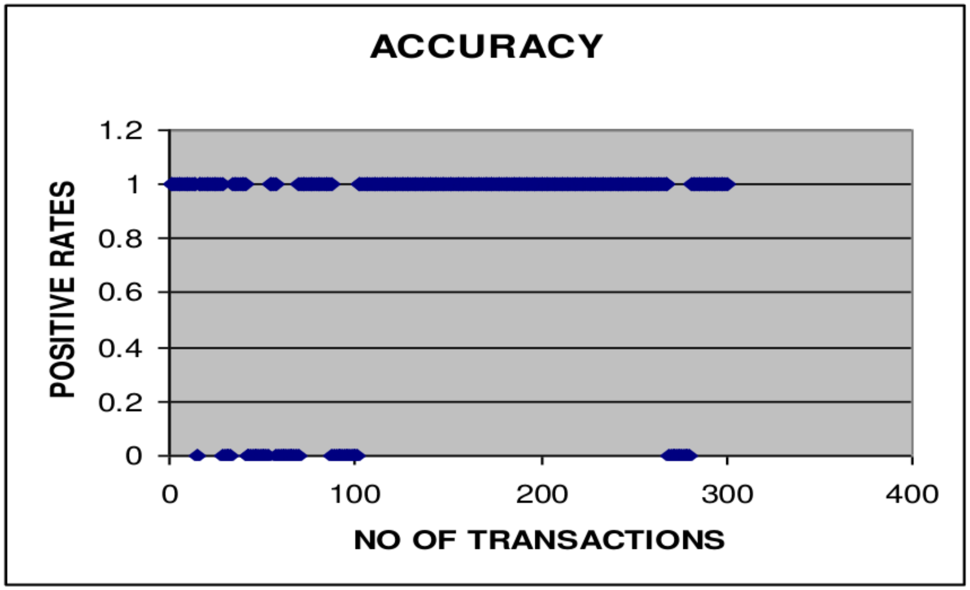
- Online method for fraud detection
- Low penalty of false negative, assumed patrons would be asked a credit question

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 Δ's in probabilities of emission between running window updates
- 5) Threshold on large Δ'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

References

1 Ngai E.W.T, et al. The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of Literature, Decision Support Systems, Vol 50, 2011

2 Fletcher H, et al. A Computational Model for Financial Reporting Fraud Detection, Decision Support Systems, Vol 50, 2011

3 S. Esakkiraj, A Predictive Approach for Fraud Detection Using Hidden Markov Model, International Journal of Engineering Research & Technology, Vol. 2 Issue 1 2013