

*DSA 4263*

*Sense-Making in Fraud Analytics*

*Literature Review Presentations*

# Recap: Assignment

Imagine that you are a data scientist on a fraud team seeking to improve your product defences. Your managers have tasked you, the data science expert, to review vendors selling various anti-fraud solutions. Prior to signing a contract, you decide to research more about the product you are purchasing, in order to evaluate them accurately, and to evaluate the feasibility of replicating the very same service internally with your team. (10 min)

## Suggested outline

1. What is this paper's hypothesis and goals?
2. What were the methods and results?
3. What dataset did they use? How did they collect it?
4. What are metrics specific to this product, if any?

## Discussion

1. Discuss challenges and tradeoffs in implementation.
2. Discuss ways that this product can be spoofed, frauded, misled or any other weaknesses.
3. State your opinion on this method's overall feasibility, based on what you have read.

# Recap: Assignment

## Reminders

1. You don't have to include an explanation about what the product is, as the lecture will already include a section about this, and your group may wish to focus your limited presentation time on the lit review discussion instead.
2. Although helpful, you don't have to include a list of how the product can be used in different patterns of fraud. The list is non-exhaustive.
3. Many research papers and industry settings will show that this product is used in combination with other techniques to detect various kinds of fraud. That does not make it a weak method or a weak paper. Your focus is on implementation and development methods.

*Week 2*

*Device Fingerprinting*

# Device Fingerprinting

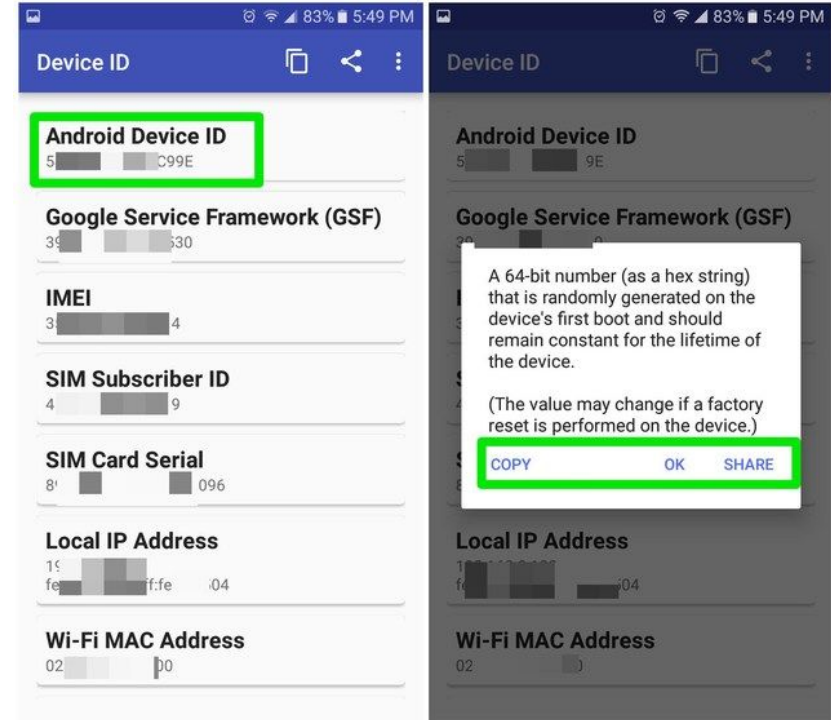
- Device fingerprinting is the technique of identifying a user by collecting information related to their hardware device, operating system, router and wifi signals, browser sessions, and user behaviour.
- Device fingerprints are terms used to refer to hardware details and information that uniquely identifies a single device amongst others.
- The technique is widely used to identify and authenticate the users of the application or website for security and advertisement purposes.



# Device Fingerprinting

Example of permissible signals collected in mobile device fingerprinting:

1. IP address (ie., 2001:4860:7:505::e4)
2. Mobile device type (ie., 'Xiaomi 12 Pro')
3. Operating System type (ie., Android, iOS)
4. Application version (ie., 3.3.1.2)
5. Storage size (ie., 16GB)
6. Device screen resolution (ie., 430x932 pixels)
7. International Mobile Equipment Identity (IMEI), android-only (ie., 350123451234560)
8. WiFi MAC Address
9. "Immutable" hardware identifiers:
  - a. iOS UUID (ie., 8753A44-4D6F-1226-9C60-0050E4C00067)
  - b. Android device\_id (ie., 2cbb8acf1c0a880da680b0a300c022b43b0647d4edf6e52246c62c8355270aaf)



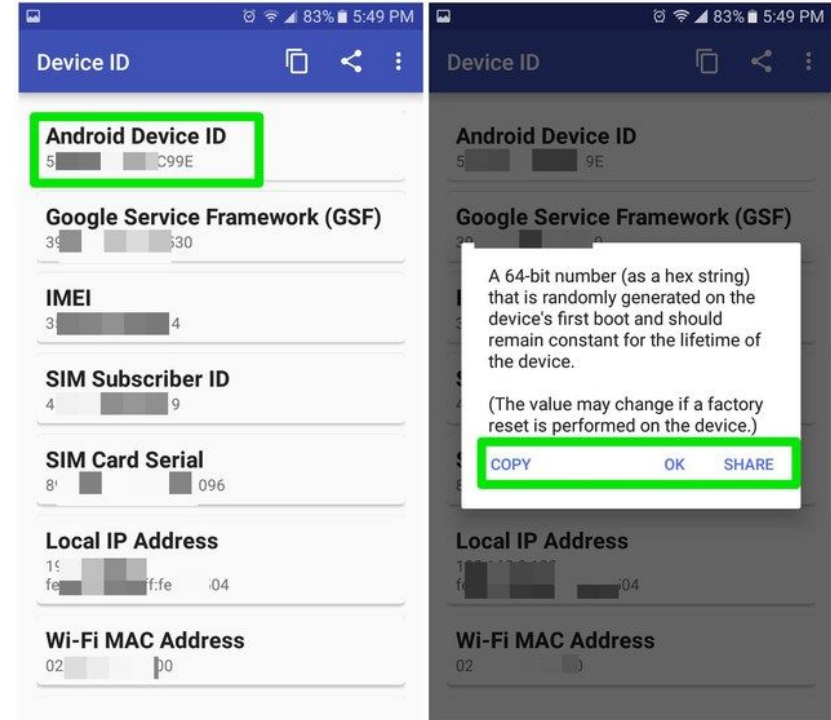
# Device Fingerprinting

Example of intrusive signals collected in mobile device fingerprinting:

- GPS location (always-on / active when app is open/ denied)
- List of applications downloaded

Examples of non-intrusive hardware signals:

- List of installed fonts and plugins
- Memory storage and usage
- Battery usage changes over time
- Clock skews
- RF frequency changes over time
- Device orientation
- System language
- System country



# Device Fingerprinting

Amongst normal users, identifying a computer or a phone by their `device_id` signal is expected to be sufficient.

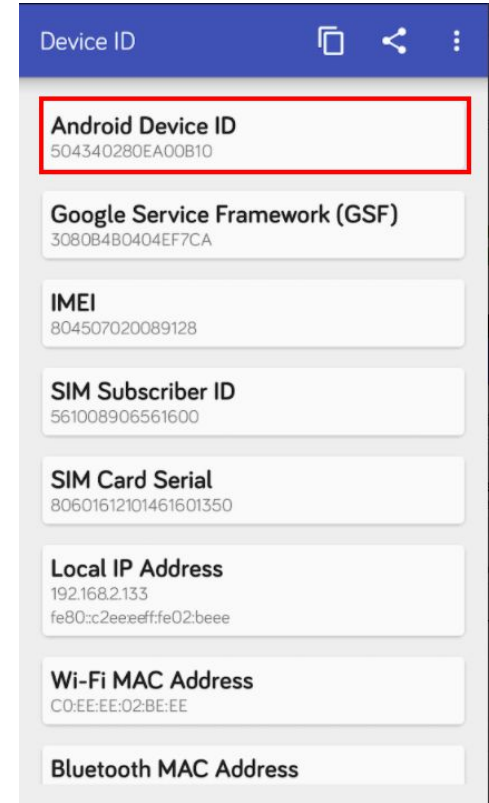
However... when tackling fraudsters...



# Device Fingerprinting

Challenges & known vulnerabilities around device spoofing (non-exhaustive):

1. When a device is re-formatted and reset, the device\_id header changes, as this is a unique hash key generated upon each system's initialisation
2. Android phones can be "rooted" and have device identifiers like IMEI tampered manually without resetting the phone
3. IP addresses and router information can easily be spoofed with numerous free apps.
4. IP addresses will also be changed when VPN is active
5. GDPR and PDPR policies require that advertising cookies and many other user-identifying cookies are reset every 30 days, preventing long-lasting identification
6. Use of virtual environments and online device emulators to create multiple virtual device fingerprints, often used in fraud farms
7. Widespread accessibility of legal and quasi-legal anti-fingerprinting no-code tools (FraudFox, AntiDetect, Kameleo, Linken sphere, MultiLogin)
8. Updated and improved privacy legislation require extensive user consent to collect any strong identifiers in device data, and reduce available features for fingerprinting (see next slide)



# Device Fingerprinting

## Broad Applications in Fraud

1. Incentive abuse: Data gathered through device fingerprinting can help you determine whether the bonus offer is going to a legitimate customer. You can spot users who share a similar device and password or even filter those who try to spoof their data using privacy-enhancing tools.
2. Account takeovers: Device fingerprinting is highly effective in detecting scenarios when there are login attempts to existing accounts from unknown devices, browsers or locations. Additional verifications and user authentication challenges can be issued.
3. Scripted account creation: Device fingerprinting is an essential tool in identifying the use of device emulators or virtual environments, which are commonly used during scripted account creation for fraudulent purposes. This technique can be used to verify when an invalid device is used, which is a strong signal of an inauthentic user.
4. Chargebacks & friendly fraud: Combined with digital profiling and IP analysis, device fingerprinting data can help companies verify customers' identities and intentions and spot chargeback & friendly fraud attempts.
5. Bot attacks: Device fingerprinting examines installed plugins, web browser version, browser window size, screen resolution, and more while also highlighting emulators and virtual machines for better bot management.

# Device Fingerprinting

Research problem:

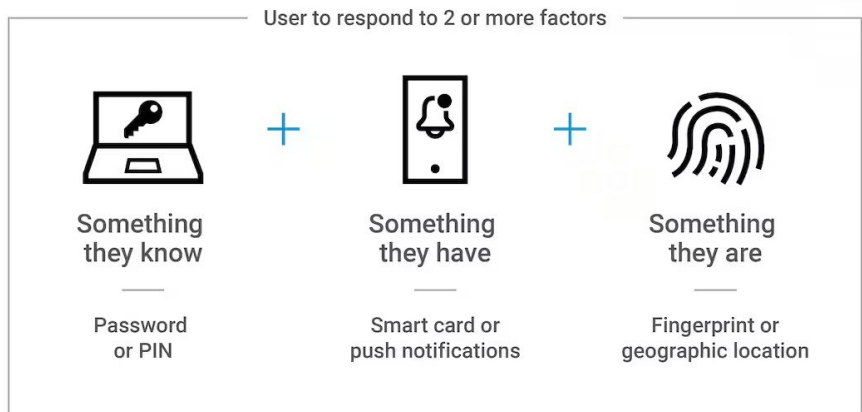
Given a list of strong and weak device identifiers, each with varying degrees of intrusiveness and mutability, how can we build a consistent and accurate means of device identification that is robust to spoofing attempts?

*Week 3*

*Multi-Factor Authentication*

# Multi-Factor Authentication

- Usernames and passwords are vulnerable to theft through brute force attacks, keyloggers, phishing, malware and social engineering techniques.
- Multi-factor Authentication (MFA) is an authentication method that requires the user to provide two or more verification factors to gain access to a resource such as an application, online account, or a VPN.



# Multi-Factor Authentication

There are 3 categories of authentication factors, each with their own examples:

Knowledge ('something you know'):

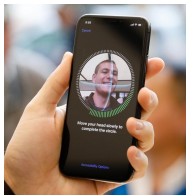
- Username
- Password
- Security question
- PIN number
- Credit card CVV number

Possession ('something you have'):

- Smart cards for office entry badges
- Physical hardware tokens like YubiKey
- Mobile phone to receive SMS OTP
- Authenticated mobile application that generates a PIN, like Google Authenticator

Inherence ('something you are'):

- Apple's FaceID system
- Google's FaceUnlock system
- Android BiometricPrompt API fingerprint scanner



**Security Questions**

Select a security question or create one of your own. This question will help us verify your identity should you forget your password.

Security Question: What is the first name of your best friend in high school?

Please select

Answer: What is the first name of your best friend in high school?

What was the name of your first pet?

What was the first thing you learned to cook?

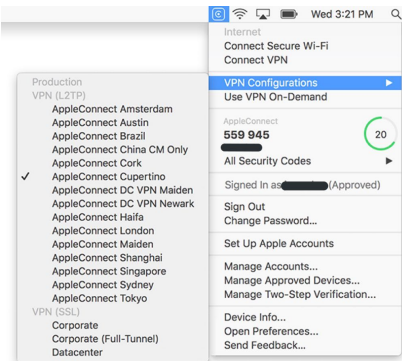
What was the first film you saw in a theater?

Where did you go the first time you flew on a plane?

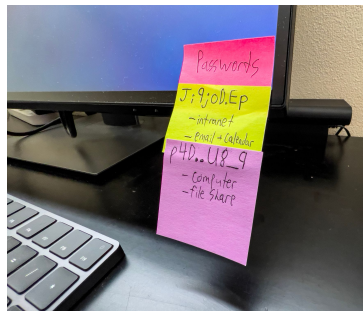
What is the last name of your favorite elementary school teacher?

Answer: \*\*\*\*\*

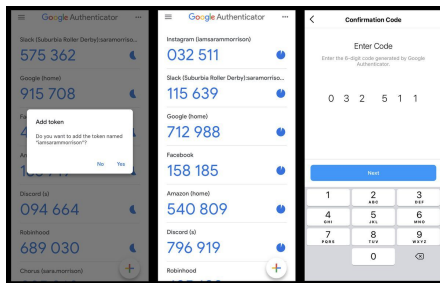
Save answers Cancel



AppleConnect is the Apple-specific single sign-on (SSO) and authentication system that allows employees to access certain applications inside Apple's network.



Yubikey tokens are portable hardware authentication devices that require users to physically touch the key to activate it.



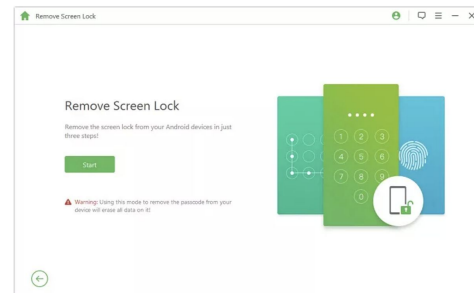
# Multi-Factor Authentication

Challenges & known vulnerabilities (non-exhaustive):

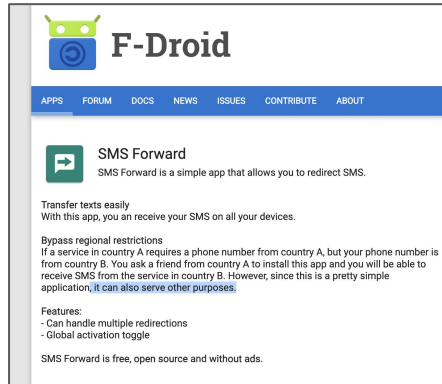
1. Authentication via possession (ie., badge, mobile phone, hardware key) is entirely based on a physical device. If a malicious actor gains access to your device, the factor is as good as broken.
2. Authentication via inherence (ie., face recognition, fingerprint IDs) can be broken through wide availability of free applications (eg., iMyFone LockWiper), and various well-intended procedures to bypass biometric checks should hardware device features fail or have false positives.
3. Authentication via mobile app PIN push notifications are vulnerable to Man-in-the-middle (MITM) attacks, which allow hackers to steal a user's identity by eavesdropping on the communication between the user and the security system.
4. Authentication via SMS OTP assumes that the identity of the account matches the identity of the owner of the SIM card. However:
  - a. SIM cards can be stolen, cloned and spoofed. SIM card ownership is not a strong identity or representation of an authentic user.
  - b. Virtual phone number generators can be used to create multiple fake accounts that do not match any offline identity, used in fraud farms and other scripted attacks.
  - c. *Silent call-forwarding apps, previously created for remote disability support or account management functions, can also be used to steal SMS OTPs away from the true owner.*

## How to Unlock Android Fingerprint Lock in 5 Minutes

**Step 1.** Download the iMyFone LockWiper (Android) program and launch it on your computer. Using a USB cable, connect your Android phone to your computer. The program will automatically detect it and load up your device's information.



A cursory search reveals many publicly available programs and processes to bypass biometric IDs, such as iMyFone LockWiper (above).



Unsuspecting users can be manipulated to download an SMS-forwarding app disguised as another app. This silently forwards all SMSes, including OTPs, to an unknown third party number.

# Multi-Factor Authentication

Case example: Indonesian celebrity falls for call-forwarding scam, gets credit card, Grab and Gopay wallets hacked.

1. Scammers use various social engineering or manipulation to get users to key ## 21# <phone number> \* 1 into phone.
2. Subsequent phone calls are forwarded to a scammer's number without the user knowing.
3. Grab, Gojek, other apps have fallback option that calls a user when SMS OTP retries exceeded a given count.
4. Scammers trigger SMS OTP requests until retry limit is reached, then wait for the (forwarded) call with OTP, then use it to login to accounts.

Call Forwarding or SMS Forwarding is the activity of transferring communications from one cellphone number to another cellphone number. When someone makes a call forwarding by pressing "21" (followed by the cellphone number to which the forward is sent) #, all calls will go to that destination number.

According to Alfons Tanujaya, a cyber security expert from Akuncom, the hack that Maia experienced was caused by an SMS forward. He suspects that Maia's activation of call forward unknowingly meant that SMS activity on her cellphone was also diverted automatically and could be accessed by fraudsters.

"Just calling forward will not result in account takeover. The possibility is that the call forward feature automatically activates SMS forward," he continued.

**Bandung, CNN Indonesia** -- Celebrity Maia Estianty's **Gopay** balance was recently hacked. Ahmad Dhani's ex-wife admitted that hackers managed to drain the payment balance in the **Gojek** application using the USSD call forwarding code method.

The line of code called Maia also activates SMS forwarding. Maia was tricked by a Gojek driver who said her motorbike had broken down and asked her to change drivers by typing in the USSD code.

The USSD code requested is "21" followed by the foreign cell phone number ("21\*082178912261).

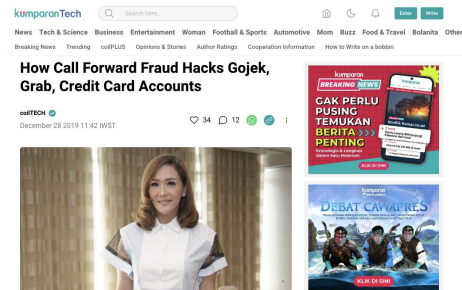
See also: Revealing the Gopay Hacking Method that Affected Maia Estianty

Responding to this, Telkomsel said that the USSD code is used to activate and deactivate call forwarding.

Sources:

<https://www.cnnindonesia.com/teknologi/20191230163700-185-461091/tips-cegah-saldo-gopay-dibobol-seperti-kasus-maia>

<https://kumparan.com/kumparantech/bagaimana-penipuan-call-forward-bobol-akun-gojek-grab-kartukredit-1sWzpfNIAjk>



Maia herself admitted that she had received an OTP code via SMS on her cellphone. However, he didn't give the code to anyone.

Telkomsel also provides tips on how to activate and deactivate this call forwarding feature, namely via the following USSD code:

1. Divert for all calls
  - press \* 21 \* diversion destination number # Yes/OK
  - to cancel press ##21# Yes/OK
2. Missed call (no reply)
  - dial \* 61 \* diversion destination number # Yes/OK
  - to cancel press ##61# Yes/OK
3. Number is busy (on busy)
  - dial \* 67 \* diversion destination number # Yes /OK
  - to cancel press ##67# Yes/OK
4. The number is inactive or out of reach (not reachable)
  - press \* 62 \* diversion destination number # Yes/OK

All of the above features can also be deactivated via USSD as follows:

- press ##002# Yes/OK or from the "Divert" Menu, select "Clear All Divert" then press "Yes/OK" (jnp/eks)



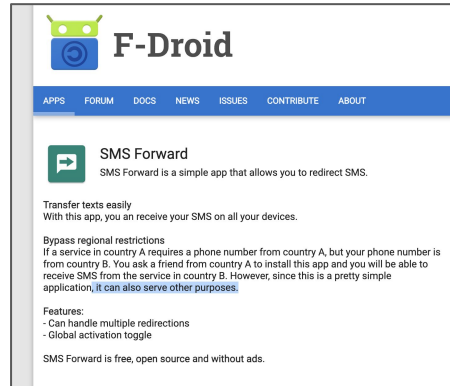
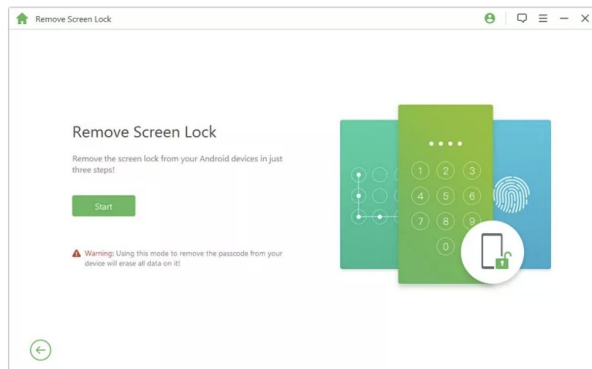
# Multi-Factor Authentication

## Additional Challenges:

- User experience is often negatively impacted due to inconvenience. Users may find additional steps tedious, leading to non-compliance in adopting MFA.
- Not all systems and services support MFA. Integration of MFA services into existing networks and databases may require significant modifications, making it difficult for large companies and organisations to integrate them, and to do so with full coverage.
- Companies also face challenges making MFA scalable across thousands and millions of users without compromising response time requirements.
- MFA solutions are costly. Yubikey hardware tokens are not cheap (estimated cost USD\$89/key). SMS OTPs that cost \$0.64 (a conservative estimate) quickly add up to millions per month as users make repeated login attempts.
- Mobile phones and laptops that have biometric identification hardware tools are neither cheap nor common in most parts of the world.

## How to Unlock Android Fingerprint Lock in 5 Minutes

**Step 1.** Download the iMyFone LockWiper (Android) program and launch it on your computer. Using a USB cable, connect your Android phone to your computer. The program will automatically detect it and load up your device's information.



Example: Unsuspecting users can be manipulated to download an SMS-forwarding app disguised as another app (eg., F-Droid under a cloned skin). This silently forwards all SMSes, including OTPs, to an unknown third party number.

# Multi-Factor Authentication

High priority areas that may be worth the trade-off between security and convenience:

1. Access to Sensitive Data: MFA should be employed when accessing sensitive or confidential information, such as financial records, personal identification details, healthcare records, or proprietary business data.
2. Financial Transactions: Online banking, financial transactions, and any activities involving the transfer of funds or payments should utilize MFA to prevent unauthorized access and fraudulent activities.
3. Enterprise remote Access and VPNs: When employees or users access company networks or resources remotely, especially through Virtual Private Networks (VPNs), MFA can add an extra layer of security to protect against unauthorized access.
4. Critical government Infrastructure and Systems: Industries such as healthcare, utilities, and government sectors that operate critical infrastructure and systems should employ MFA to safeguard against unauthorized access that could lead to disruptions or compromises.
5. Compliance Requirements: Compliance standards and regulations often mandate the use of MFA in specific industries to ensure the protection of sensitive information and to meet regulatory requirements.
6. User Account Management: When managing user accounts, especially for *administrators* or privileged users with elevated access rights, MFA is crucial to prevent unauthorized individuals from gaining control over critical systems.
7. Healthcare and Personal Records: Healthcare organizations dealing with patient records and personal information should implement MFA to protect sensitive data from unauthorized access and potential breaches.

# Multi-Factor Authentication

Research problem:

Given a list of various authentication factors, and an understanding of how each category of factor can be spoofed or stolen, how can we design a robust and accurate system of authenticating one's virtual identity to an offline one?

# Multi-factor Authentication

Some alternative MFA methods instead of SMS OTPs:

- Use of login links instead of PINs, which are harder and more suspicious to share
- Use of graphical methods (ie., choose familiar images from set of options as passwords)
- Use of 'SMS Magic' MagicKeys (ie., instead of an OTP that can easily be shared, calling the authentication API directly triggers a call from the user's application, bypassing the need for OTP confirmation)
- Use of authentication via host-based characteristics (ie., device fingerprinting of formerly-authenticated 'trustworthy' devices, user behavior, systems settings)

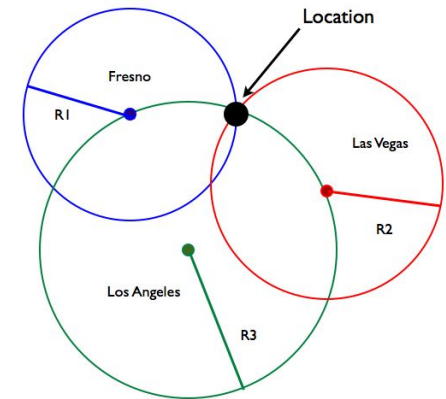
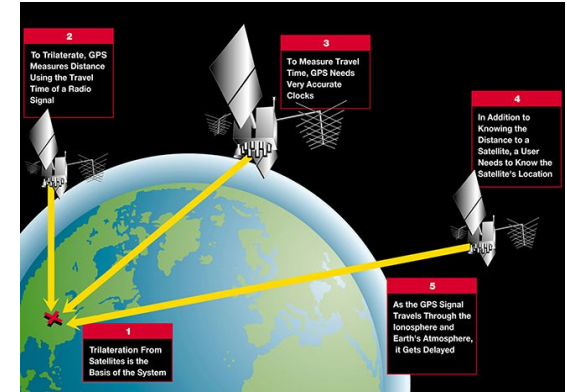
Considerations: Cost, accessibility, friction

*Week 4*

*GPS spoofing*

# GPS Spoofing

- Global Positioning System (GPS) is a worldwide navigational and surveying facility based on the reception of signals from an array of orbiting satellites.
- Satellites are continually broadcasting their orbital position and exact time at that position on radio frequencies. GPS receivers receive these by antennas, and use at least 3 of such satellite signals to compute a user's location.
- GPS signal spoofing (technically not illegal) happens when a user intentionally transmits counterfeit GPS signals to interfere or override legitimate signals. This can be used to project a false location for themselves or someone else.
- GPS jamming (strictly illegal) happens when cyber criminals utilise tools to block GPS signals altogether.



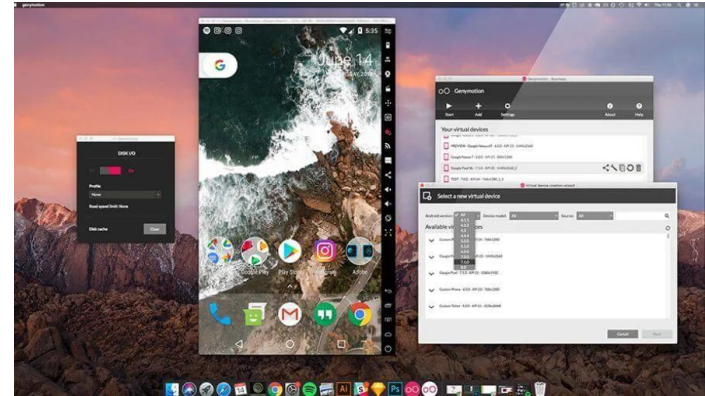
Example: If a user knows their distance ( $R_1$ ,  $R_2$ ,  $R_3$ ) from 3 transmission satellite locations in Los Angeles, Fresno, and Las Vegas, that user can calculate their GPS location.

# GPS Spoofing

There are many methods of location spoofing:

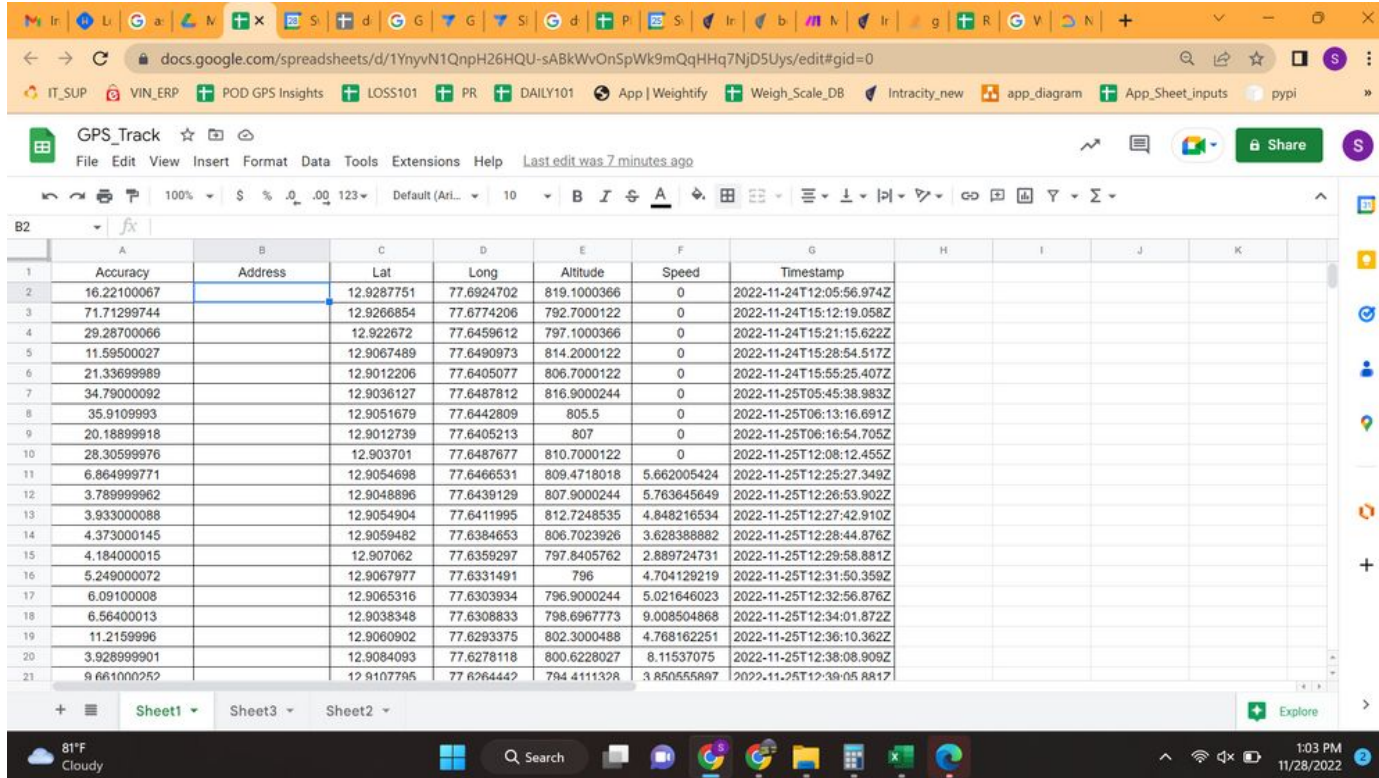
1. VPN & Proxies that hide IP addresses
2. Emulators (ie., using VMs to spin up mobile app copies on laptops)
3. Instrumentation tools (e.g., Frida)
4. App tampering (ie., modifying the compiled code of the application to send custom GPS signals)
5. GPS location spoofing mobile applications (eg., FakeGPS)

We will focus on the last option: GPS spoofing via mobile applications.



# GPS Spoofing

Image: A user submitted their location data below in a forum discussing how these latitude and longitude points might be converted to a visualisation of their traveling path. This location data set reflects what existing ride-hailing, food delivery, logistics and online marketplace services collect.



|    | A           | B       | C          | D          | E           | F           | G                        | H | I | J | K |
|----|-------------|---------|------------|------------|-------------|-------------|--------------------------|---|---|---|---|
|    | Accuracy    | Address | Lat        | Long       | Altitude    | Speed       | Timestamp                |   |   |   |   |
| 1  | 16.22100067 |         | 12.9287751 | 77.6924702 | 819.1000366 | 0           | 2022-11-24T12:05:56.974Z |   |   |   |   |
| 2  | 71.71299744 |         | 12.9266854 | 77.6774206 | 792.7000122 | 0           | 2022-11-24T15:12:19.058Z |   |   |   |   |
| 3  | 29.28700066 |         | 12.922672  | 77.6459612 | 797.1000366 | 0           | 2022-11-24T15:21:15.622Z |   |   |   |   |
| 4  | 11.59500027 |         | 12.9067489 | 77.6490973 | 814.2000122 | 0           | 2022-11-24T15:28:54.517Z |   |   |   |   |
| 5  | 21.33699989 |         | 12.9012206 | 77.6405077 | 806.7000122 | 0           | 2022-11-24T15:55:25.407Z |   |   |   |   |
| 6  | 34.79000092 |         | 12.9036127 | 77.6487812 | 816.9000244 | 0           | 2022-11-25T05:45:38.983Z |   |   |   |   |
| 7  | 35.9109993  |         | 12.9051679 | 77.6442809 | 805.5       | 0           | 2022-11-25T06:13:16.691Z |   |   |   |   |
| 8  | 20.18899918 |         | 12.9012739 | 77.6405213 | 807         | 0           | 2022-11-25T06:16:54.705Z |   |   |   |   |
| 9  | 28.30599976 |         | 12.903701  | 77.6487677 | 810.7000122 | 0           | 2022-11-25T12:08:12.455Z |   |   |   |   |
| 10 | 6.864999771 |         | 12.9054698 | 77.6466531 | 809.4718018 | 5.662005424 | 2022-11-25T12:25:27.349Z |   |   |   |   |
| 11 | 3.789999962 |         | 12.9048896 | 77.6439129 | 807.9000244 | 5.763645649 | 2022-11-25T12:26:53.902Z |   |   |   |   |
| 12 | 3.933000088 |         | 12.9054904 | 77.6411995 | 812.7248535 | 4.848216534 | 2022-11-25T12:27:42.910Z |   |   |   |   |
| 13 | 4.373000145 |         | 12.9059482 | 77.6384653 | 806.7023926 | 3.628388882 | 2022-11-25T12:28:44.876Z |   |   |   |   |
| 14 | 4.184000015 |         | 12.907062  | 77.6359297 | 797.8405762 | 2.889724731 | 2022-11-25T12:29:58.881Z |   |   |   |   |
| 15 | 5.249000072 |         | 12.9067977 | 77.6331491 | 796         | 4.704129219 | 2022-11-25T12:31:50.359Z |   |   |   |   |
| 16 | 6.09100008  |         | 12.9065316 | 77.6303934 | 796.9000244 | 5.021646023 | 2022-11-25T12:32:56.876Z |   |   |   |   |
| 17 | 6.56400013  |         | 12.9038348 | 77.6308833 | 798.6967773 | 9.008504868 | 2022-11-25T12:34:01.872Z |   |   |   |   |
| 18 | 11.2159996  |         | 12.9060902 | 77.6293375 | 802.3000488 | 4.768162251 | 2022-11-25T12:36:10.362Z |   |   |   |   |
| 19 | 3.928999901 |         | 12.9084093 | 77.6278118 | 800.6228027 | 8.11537075  | 2022-11-25T12:38:08.909Z |   |   |   |   |
| 20 | 9.661000252 |         | 12.9107795 | 77.6264442 | 794.4111328 | 3.850555867 | 2022-11-25T12:39:05.881Z |   |   |   |   |

Sample of realistic location data collected by mobile applications with the following headers:

Latitude, Longitude, Altitude, Speed, Accuracy.

These must be organised by user\_id, timestamp, etc.

This data can be used to visualise a user's location and traveling path along a map.



# GPS Spoofing

Fraud applications of GPS spoofing in various industries:

## 1. Commercial shipping

- a. Logistics companies (via land, air, sea) all use GPS navigation systems to transport goods to numerous destinations. Hijackers can use GPS spoofing to misdirect a ship or lorry to a location where they will be robbed.
- b. With sensitive goods that are GPS-enabled locks, that only release upon arrival, hijackers can use GPS spoofing to gain access to that cargo as well.

## 2. Location-based Services

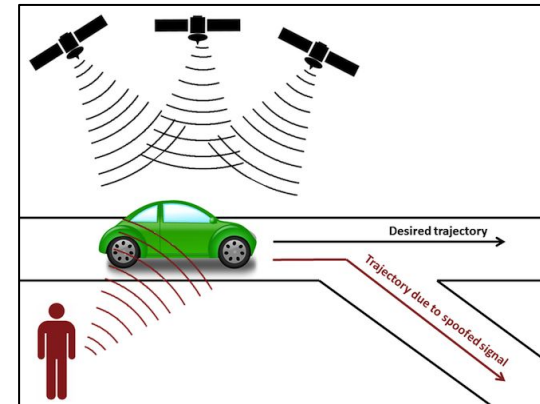
- a. Tinder charges a premium for a “passport” features that allow users to match with other users outside their current location. Users may use GPS spoofing to bypass this additional cost by simply spoofing their desired location to meet matches.
- b. PokemonGo is an Augmented Reality (AR) game that incentivise users to physically travel to specific locations to discover attractive items and events. Users may use GPS spoofing to rapidly travel to multiple location to unlock such events, giving them an unfair advantage over users who must physically travel to these locations instead.

# GPS Spoofing

Fraud applications of GPS spoofing in various industries:

## 3. Ride Hailing

- a. Drivers can use GPS spoofing to place themselves in surge areas to get more expensive orders, while they are actually much further away, affecting customer experience and unfairly gaining an advantage.
- b. Fraudsters can use GPS spoofing to create fake orders for incentive abuse. They can coordinate fake driver and customer accounts to place themselves in isolated areas in order to increase their odds of matching with each other, to conduct incentive abuse.
- c. Drivers using a false location is potentially dangerous for customers, especially women and children traveling alone. This hinders the company from monitoring an order for prolonged stops or unexpected detours that indicate a safety violation.
- d. Companies may create driver incentives to guarantee a certain minimum earnings per hour, to encourage drivers to stay online and improve service reliability. Fraudsters can use GPS spoofing to have fake driver accounts appear online and active, to cheat these incentives. These unfairly disadvantage other drivers, and reduce the available share of funds for these remaining good drivers.



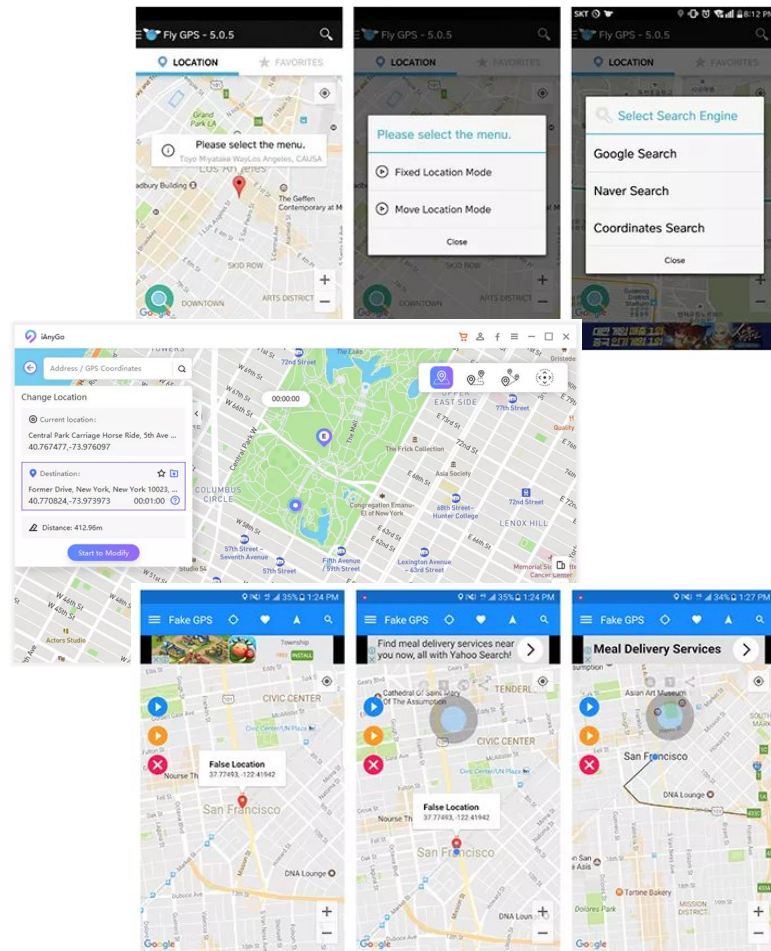
# GPS Spoofing

Complexity and services (from basic to challenging):

1. Single-point fixed location spoofing
2. Point-to-point journey spoofing
3. Joystick control, free navigation spoofing
4. Spoofing custom routes (ie., does not follow roads or public transport routes or existing map API-provided routes)
5. Concurrent multiple-location spoofing for different apps

Additional features that determine the quality of a “fake GPS” app and service:

- Accurate altitude signals
- Realistic noise and jitters
- No jailbreak required
- No intrusive device permissions



# GPS Spoofing

Research problem:

What are some inexpensive, quick, common, lightweight, non-intrusive data signals can we collect, and which machine-learning and analytical techniques can we apply, that will help us determine whether a GPS signal (or series of GPS signals along a journey) is accurate and authentic?