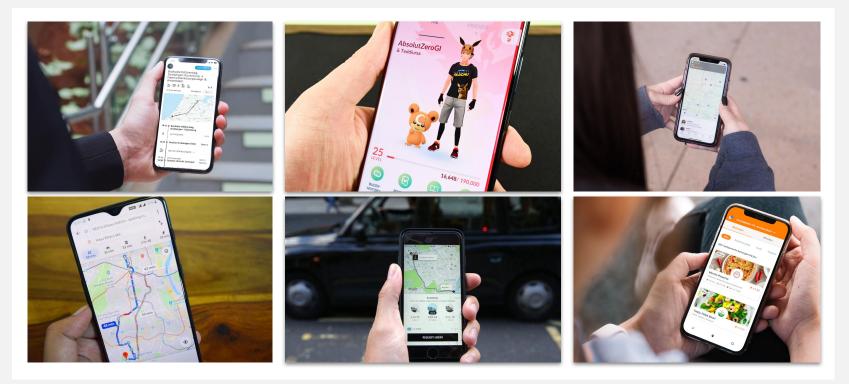
Privacy friendly location based services

Stef Vergeest - S1081227 Fouad Lamsettef - S1034545 Jesmer Logtenberg - S1107922



Location based services introduction





Questions

How much are you willing to sacrifice in terms of service for privacy?

Do you trust Google to responsibly handle you location data?



Contents

- Location based services
- Privacy Enhancing Technologies
 - Privacy policy mechanisms
 - Obfuscation mechanisms
 - Cryptographic mechanisms
 - Cooperation and cached mechanisms
- Practical case: Google Maps
- Discussion



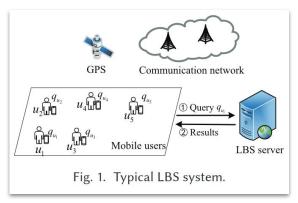


Location-based services



What are location-based services (LBS)?

- Services that integrate a mobile device's location or position with other information so as to provide added value to the user [Schiller, Jochen; Voisard, Agnès (2004)]
- **Mobile users**: the users of the LBS that send the query
- Location positioning system: infer location of the user
- **Communication network**: transfers the query from the user to the LBS server
- LBS server: processes the query and returns the result
- **Snapshot LBS**: one query to the server (restaurants, hotels)
- **Continuous LBS**: continuously report to server (navigation, sharing location)



[Location Privacy-preserving Mechanisms in Location-based Services: A Comprehensive Survey]



Applications of LBS





Value of location data

How sensitive is your location data to you?



Dangers of LBS

- LBS providers ...
 - ... selling data to data brokers and private investigators
 - ... selling data to governments and law enforcement
 - ... gathering more data than required or the user has agreed
 - ... being a single point of failure for malicious parties
- Location data ...
 - ... can be used alongside public data for de-anonymization
 - ... can reveal user's sensitive attributes







Dangers of LBS

TECH

The US military reportedly bought location data mined from a popular Muslim prayer app to track users for 'counterterrorism'

Aaron Holmes Nov 16, 2020, 7:45 PM CET

→ Share □ Save
 □
 Save

Fitness tracking app Strava gives away location of secret US army bases

Data about exercise routes shared online by soldiers can be used to pinpoint overseas facilities

Technologie • 10 jan 20:23 • Aangepast op 11 jan 07:44

Nederlandse telefoons online stiekem te volgen: 'Extreem veiligheidsrisico'

Auteur: Eric van den Berg

Locatiegegevens van Nederlandse mobiele telefoons zijn online gewoon te koop, blijkt uit onderzoek van BNR. Het gaan en staan van veel Nederlanders is hierdoor tegen betaling te volgen. Het aantal slachtoffers loopt mogelijk in de miljoenen.

Google to pay \$93m in settlement over deceptive location tracking

Tech giant 'continued to collect and store a user's location data' even if users turned off their location history, according to suit



California attorney general's office also alleged Google 'deceived users about their ability to opt out of advertisements targeted to location'. Photograph: Andre M Chang/ZUMA Wire/REX/ Shutterstock

'A Mass Invasion of Privacy' but No Penalties for Tim Hortons

A scathing report by four privacy commissioners found that the coffee and doughnut chain collected data on customers' daily lives.

The Popular Family Safety App Life360 Is Selling Precise Location Data on Its Tens of Millions of Users

The app is a major source of raw location data for a multibillion-dollar industry that buys, packages, and sells people's movements By Jon Keegan and Alfred Ng December 6, 2021 08:00 ET



Uber pulls U-turn on controversial tracking of users after trip has ended

Company will give users option to be tracked only while actively using the app, as it tries to clean up its act on privacy after unveiling of new CEO

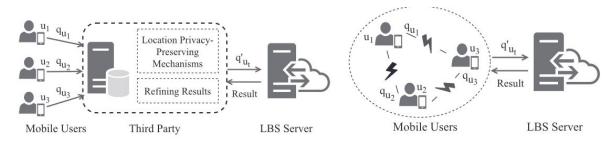






Privacy enhancing technologies (PET) in the context of LBS

- Term coined by the Dutch Data Protection Authority and the Ontario Information Commissioner.
 - Variety of technologies that safeguard personal privacy by minimizing or eliminating the collection of identifiable data [*Privacy-Enhancing Technologies: the path to anonymity (1995)*]
- **Third Party-based architecture**: mobile users communicate with third party (anonymizer), that protects sensitive location information in queries, and forwards these to the LBS server.
- Third Party-free architecture: mobile users communicate directly with the LBS server.



[Location Privacy-preserving Mechanisms in Location-based Services: A Comprehensive Survey]



Privacy enhancing techniques (PET)

- Privacy policy mechanisms
 - ePrivacy Directive
 - General Data Protection Directive
- Obfuscation mechanisms
 - Cloaking
 - Dummy locations
- Cryptographic mechanisms
 - Space Transformation
 - Secure Multiparty Computation
- Cooperation and cached mechanisms
 - Caching
 - MobiCache





Privacy policy mechanisms



Privacy policy mechanisms - Definition

common privacy management rules and trusted privacy agreements, constraining the service provider and the third party to fairly and securely <u>access</u>, <u>store</u>, and <u>use</u> the location information in LBS queries submitted by users.

(Hongbo Jiang, Jie Li, Ping Zhao, Fanzi Zeng, Zhu Xiao, and Arun Iyengar. 2021. Location Privacy-preserving Mechanisms in Location-based Services: A Comprehensive Survey.)



Privacy policy mechanisms - Question

Who has every fully read a privacy policy, before agreeing to it?



Privacy policy mechanisms - ePrivacy Directive (2002)

- Protects electronic communications from individuals inside the EU
- **location data** means any data processed in an electronic communications network, indicating the geographic position of the terminal equipment of a user of a publicly available electronic communications service (Article 2(c)).
- Processing of location data for a **value added service** is only allowed (Article 9(1)):
 - ... when the information is made **anonymous** or
 - ... with the **consent** of the users or subscribers to the extent and for the duration necessary for the provision of a value added service
- Users and subscribers are given the possibility to withdraw consent for processing at any time
- Prior to processing, users and subscribers have to be informed about:
 - the **type**, **purposes** and **duration** of processed location data
 - whether the location data will be transmitted to a **third party**





- Protects **personal data** from individuals inside the EU
- '**personal data**' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as (...) **location data**.
- Thus the location data of the data subject:
 - ...has to be processed according to the principles of the GDPR, described in Article 5
 - ... can only be processed once the controller can proof that one of the legitimate bases applies, described in Article 6



Article 5

Principles relating to processing of personal data

- 1. Personal data shall be:
- (a) processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');
- (b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');
- (c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');
- (d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy');
- (e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation');
- (f) processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').



Article 6

Lawfulness of processing

1. Processing shall be lawful only if and to the extent that at least one of the following applies:

(a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;

(b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;

- (c) processing is necessary for compliance with a legal obligation to which the controller is subject;
- (d) processing is necessary in order to protect the vital interests of the data subject or of another natural person;
- (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
- (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.



Requirements for legitimate interest:

- 1. The interest must be legitimate
- 2. The processing must be necessary for that purpose
- 3. The legitimate interest must outweigh the data subject's rights and interest:

Applied to selling location data for money:

- 1. Making money is a legitimate interest
- 2. Selling location data is not necessary for making money
- 3. Making money certainly does not outweigh the privacy rights of data subjects



Privacy policy mechanisms - Shortcomings

- ePrivacy Directive was originally meant for protection of **telecommunication services** and hard to incorporate new technologies into regulatory framework
 - ePrivacy Regulation is in the making to tackle this problem, but dead in the water due to lobbying
- ePrivacy Directive and GDPR restrictions do not apply to anonymous data, but creating truly anonymous data is hard
 - 'information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable'
- People do not read privacy policies, since they are hard to digest for the average user
- Enforcement of ePrivacy Directive and General Data Protection Regulation is lacking ٠
- High level of trust allocated to LBS providers to correctly implement and conform to rules





Obfuscation mechanisms



Obfuscation mechanisms

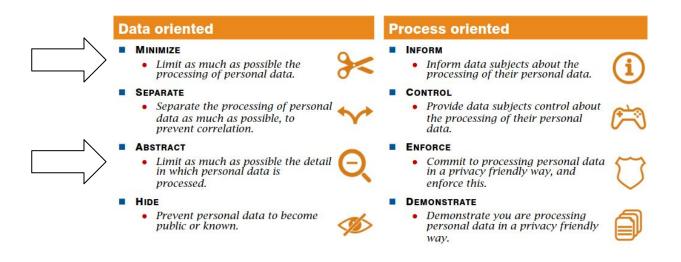
'The <u>act</u> of (intentionally) making something less <u>clear</u> and less <u>easy</u> to <u>understand</u>'





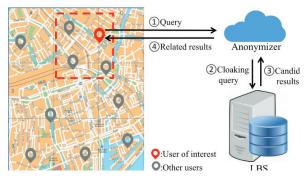
Obfuscation mechanisms

• In the context of LBSs: doing something with the location data in such a way that a service provider is no longer able to **identify** or **follow** you





- Conceals the precise position of an individual by using a **cloaking region**
- Cloaking can be performed *spatially* or *temporarily*
 - In spatial cloaking, the user sends a generalized region to the server, instead of a precise point
 - In temporal cloaking, the user intentionally delays their query
 - Note: this makes services needing real-time data unusable
- Being in a cloaking region doesn't ensure that you are not identifiable
 - To ensure this, we need some type of measurement: **k-anonymity**





K-anonymity: a way to 'measure' the degree of identification

- Each person possesses several identifying points, such as name, address and IP
- Next to these identifiers, we also possess **quasi-identifiers**
 - How many % of the United States population do you think is identifiable using solely these 3 quasi-identifiers?



• K-anonymity means that an individual's quasi identifiers have to be equivalent to at least k-1 other individuals

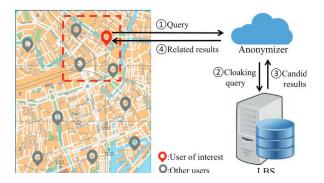


K-anonymity example





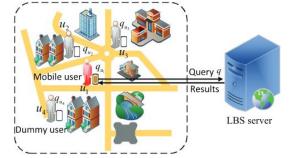
- K-anonymity does have some limitations:
 - If the data is not diverse, an individual can still be identified
 - The cloaking region may be too small, meaning the location of an individual can be recovered
 - The cloaking region may be too large, resulting in unusable restaurant recommendations
- These limitations can be further addressed using techniques like *l-diversity* (out of scope for this presentation)





Obfuscation mechanisms - Dummy locations

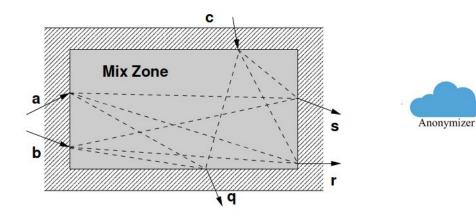
- Generalize location information by **spamming the LBS server with fake locations**, with one being (close to) your actual location
- Retrieve the correct information by filtering out the responses
- Dummy locations also imply certain limitations:
 - Dummy locations that are *too close* reveal the actual location of the user
 - Dummy locations that are too far again results in unusable query results
 - An increasing number of dummy locations increase the server overhead
- Dummy locations are nice, because
 - No third party is required (like the anonymizer in the previous PET)
 - Allows accurate query results (restaurant recommendations)
 - Follows Kerckhoffs principle





Obfuscation mechanisms - Mix zones

- Mix zones are a **spatial region where applications cannot access specific location** information of users therein
- Example spaces are hospital grounds, university spaces, etcetera
- The user identity is mixed with all other users in the mix zone
 - Third party only sees obfuscated version of location



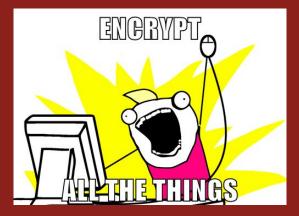
Obfuscation mechanisms - Path confusion

- Uses the **interaction between users** to achieve privacy protection
- When crossing another LBS user, send query to the anonymizer
 - The anonymizer delays user's queries to increase the probability of users crossing with more users
 - When anonymizer then releases query to LBS, data will be randomized









Cryptographic mechanisms



Cryptographic mechanisms

- CIA+P
 - Confidentiality
 - Integrity
 - Availability
 - Privacy
- Cryptography works well for Confidentiality and Integrity
- What about privacy?



Cryptographic mechanisms

- Only applicable for Snapshot LBS (restaurants)
- Limited application in Continuous LBS (traffic info)
 - Cryptography too much overhead
 - Fast response time is prioritized for situations like location sharing
- Thus limited use in LBS

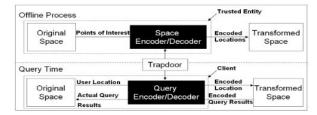






Cryptographic mechanisms - Space Transformation

- Applicable Snapshot LBS
- Third-party free LBS
- Protect the location data of the user
- Transform location data points to different space
 - Using one-way hash function
 - K-nearest-neighbor query
- User wants to know location of nearby restaurant
- User location is encoded and send with the query
- LBS server searches in encoded space the points of interests
 - Distance is preserved in the encoded space
- LBS server sends encoded points of interest back to user
- User can 'reverse'/decode the encoded data using the trapdoor to get the results
 - Trapdoor/ Key is given extra knowledge

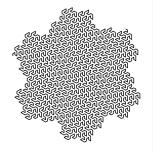


https://dl.acm.org/doi/proceedings/10.1145/ 3469830



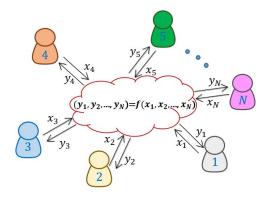
Space Transformation - decryption phase

- Space Filling Curves
 - Family of curves which pass through all points in space without crossing themselves
 - Retain proximity and neighboring aspects of data
- Hilbert Curves
 - Most used class due to superior clustering and distance preserving properties
 - We define H_d^N as the Nth order Hilbert curve for a d-dimensional space
 - H_d^N is a linear ordering with H-value: H = L(P) with P a coordinate in a d-dimensional space and L a mapping function
 - In our case we H = L(X, Y) for 2-D coordinates
 - Mapping function L becomes similar to a one-way function if curve-parameters are unknown
 - Curve starting point: (X_0, Y_0)
 - Curve orientation θ, order N and scale factor T
 - These parameters become our key called Space-Decryption-Key (SDK)
 - How we distribute the SDK to the user in order to decode the query can depend on implementation
 - Use of trusted entity which distributes keys based on indexes of POI
 - Embed keys on devices



Cryptographic mechanisms - Secure Multiparty Computation

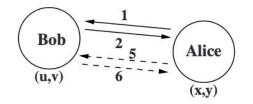
- Applicable for snapshot LBS
- Cryptography-based scheme used to confuse an untrusted third or LBS servers in LBS
- Multiple parties compute output of a function
- Each party does not know the input of the other party and does not learn it from doing the computation





Secure Multiparty Computation: approach 2 (Zhong et al.)

- Scenario: Alice wants to know if Bob is nearby
- Third party free LBS
- Alice and Bob encrypt their location data using public key crypto
 - Alice and Bob each calculate C= A^b = B^a
 - Alice sends her location $\epsilon_A(x^2 + y^2)$ to Bob
 - Bob computes the distance D and sends it back to Alice
 - Alice can learn the distance using her private key
- Alice reveals her location to Bob
 - For mutual exchange of information, protocol needs be done twice.
- Based on trust of both parties.
 - Bob can send fake distances by computing D with a fake (u, v)







Cooperation & Caching mechanisms



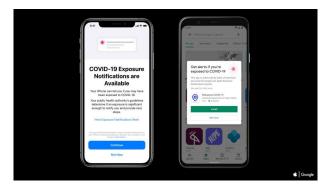
Cooperation & Caching mechanisms: (Amini et al.)

- Idea: reduce the amount of trust to third parties
- Cache important/relevant location data periodically before it is needed
- User gives preferred areas which are then cached for later
- Possible challenges:
 - Storage requirements
 - Data freshness
 - Data consistency
 - User moving outside the boundary of the cached content
 - Cache misses
 - No cache for specific area of interest at the moment
- Not applicable for real-time applications like traffic flow or nearby-friends



Cooperation & Caching mechanisms: MobiCrowd scheme (Shokri et al)

- Idea: let other nearby users answer the location query
- Informed users: users with valid information of a certain regions
- Seekers: users interested in getting information of that region
- Send information through an ad-hoc wireless interface of the device
- Analogous to the epidemic, where information of "infected" users are spread to nearby users
- Disadvantage
 - User still need to query to LBS if other users fail to answer
 - Freshness of cached data









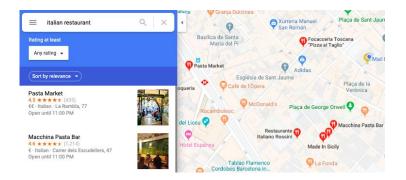
Google Maps

- One of the most popular maps application on mobile devices
- Pre-installed on Android devices
- How privacy friendly is Google Maps already?
- How can we make Google Maps more privacy friendly using PETS?



Google Maps: how is data anonymized

- Google Maps already implements multiple PETs: obfuscation via cloaking
 - K-anonymity / L-diversity
 - Adding noise by means of differential privacy
- Mostly applicable for snapshot-based services in Google Maps:
 - Queries like "Closest restaurant in Nijmegen"





Google Maps: how is location data used?

How does Google use location information that is pseudonymous or anonymous?

Google uses anonymized and pseudonymized location information to help enhance people's privacy. Anonymized information generally cannot be associated with any individual. Pseudonymized information may be tied to a unique identifier, such as a string of numbers, rather than more personally identifiable information such as a person's account, name, or email address. Anonymized and pseudonymized location information may be used by Google in its products and services for purposes such as advertising or trends.

Users may be able to reset certain pseudonymous identifiers linked to location information. For example, people can reset certain pseudonymous identifiers by resetting advertising IDs on their Android devices. In addition, Google automatically resets certain pseudonymous identifiers to enhance user privacy, including for GLA, the device setting that users can control to improve location-based service and accuracy on their devices.

Separately, Google may use anonymized location information. For example, people can tap on places in Google Maps, e.g., a restaurant or a park, and see trends from those places in an area. Location information used to build trends, like popular times, cannot be used to identify an individual. If Google does not have enough information to provide accurate and anonymous busyness information, it doesn't appear on Google.

Google also offers people who are signed-out other ways to manage information associated with their browser or device, including the Search customization setting, YouTube settings, and ads settings. Learn more 2

- Make experiences useful
- Help people remember places
- Get queries faster
- Show relevant ads
- Show trends in certain communities



- Google provides many services
 - Youtube (What you watch)
 - Search (What you search)
 - Office Suite (What you create)
 - Maps (Where you want to go)
 - Android Activity

How does Google know my location?

Depending on the products you're using and settings you choose, Google may use different types of location information to help make some services and products you use more helpful.

This location information can come from real-time signals, like your IP address or from your device, and also your saved activity on Google sites and services. Here are the main ways Google may get information about your location.

From your IP address

An IP address, also called an Internet Protocol address, is a number that is assigned to your computer or device by your Internet Service Provider. IP addresses are used to make the connection between your devices and the websites and services you use.

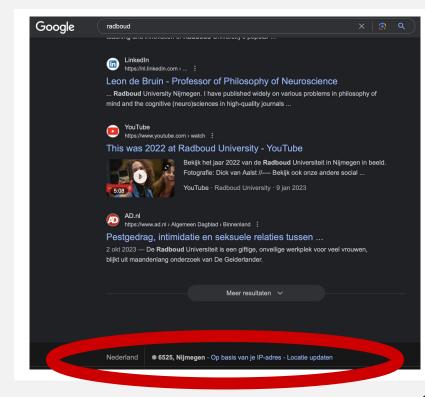
Like many other internet services, Google may use information about the general area that you're in to provide some basic services—relevant results, such as when someone does a search asking what time it is, or keeping your account safe by detecting unusual activity, such as a sign-in from a new city.

Keep in mind: Devices need an IP address in order to send and receive internet traffic. IP addresses are roughly based on geography. This means that any apps, services, or websites you use, including google.com, may be able to infer and use some information about your general area from your IP address.

• They get your IP-address and infer the general location

• Can be used for detection of unauthorized use of Google account





Huygensgebouw





From your saved activity

If you're signed in to your Google Account and have Web & App Activity turned on, your activity data on Google sites, apps, and services may be saved in your account's Web & App Activity. Some activity may include information about the general area you were in when using the Google service. When you search for something using a general area, your search will use an area of at least 3 sq km, or expand until the area represents the locations of at least 1,000 people. This helps protect your privacy.

In some cases, areas that you have searched from in the past may be used to estimate a relevant location for your search. For example, if you search for coffee shops while in Chelsea, Google might show results for Chelsea in future searches.

You can view and control your Web & App Activity at My Activity 2.

If you're not signed in to your Google Account, Google may store some location information for previous searches from the device you're using to help provide more relevant results and recommendations. If you turn off Search customization, Google won't use previous search activity to estimate your location. Learn more about how to search and browse privately.

- Everything you do with Google services is saved
- Web & App Activity can be disabled



From home or work addresses you saved

You might choose to save places to your Google Account that are important to you, such as your home or your work. If you set your home or work addresses, they can be used to help you do things more easily, such as getting directions or finding results closer to your home or work, and to show you more useful ads.

You can edit or delete your home or work addresses anytime in your Google Account 2.

• You can't save you work or home address on Google Maps without enabling location history.



- Google can still infer your location if you have Location History disabled
- Users must remember to delete the history or set an earlier deletion period

Keep in mind

If you turn off Location History

- Google will continue to store any past Location History data you've saved until you
 delete it, or it will be deleted after a period of time that you've chosen as part of your
 auto-delete settings.
- Turning off Location History doesn't impact how location information is saved or used by Web & App Activity or other Google products, e.g., based on your IP address. You may still have other settings that save location information.

To see if you've turned on Location History, visit your Activity Controls 2. Learn more 2.

Google to pay \$93m in settlement over deceptive location tracking

Tech giant 'continued to collect and store a user's location data' even if users turned off their location history, according to suit

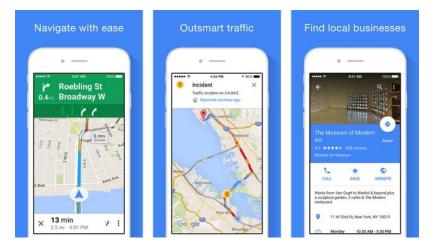


■ California attorney general's office also alleged Google 'deceived users about their ability to opt out of advertisements targeted to location'. Photograph: Andre M Chang/ZUMA Wire/REX/ Shutterstock



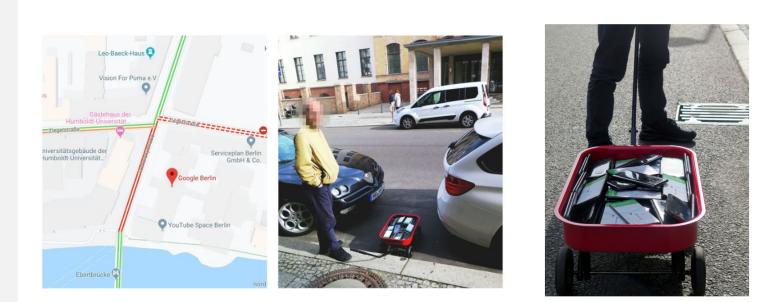
Google Maps: traffic congestion

- Continuous LBS application
- Multiple users share anonymous bits of location data
 - User's location
 - Speed of the vehicle
 - Speed Limit
- PET: mixed zones
 - Google Maps only knows that a pool of people are not driving very fast





Google Maps: traffic congestion





Approach for a more privacy friendly Google Maps: Caching & Cooperation

- User Approach
 - User Alice wants to go on vacation to New York
 - Alice downloads the entire region of New York on Google Maps and only uses Maps offline.
 - Alice deletes the cache when she is back home
 - Still limitations:
 - Data freshness
 - Cache misses
- What Google can implement?
 - Users get the option to only use location data using Cooperation if other cached users are nearby
 - For snapshot applications





Approach for a more privacy friendly Google Maps

- Use dummy locations to send fake locations to Google Maps and still get accurate query results
- Use a VPN and query Google Maps before visiting certain region
 - Does not help when Location History & Web App Activity enabled.

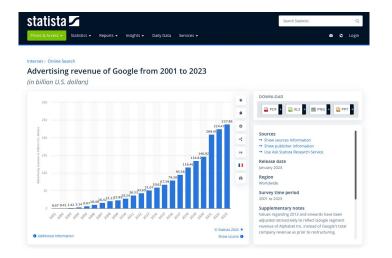




Why should Google protect your privacy?

- Besides requirements by law, why would Google be more privacy friendly?
- Where is the competition?
 - Apple Maps?





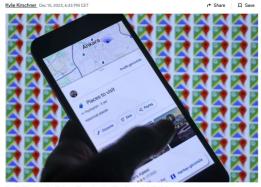




Why should Google protect your privacy?

- Users find their location data more valuable due to law

Google will no longer hold onto people's location data in Google Maps — meaning it can't turn that info over to the police



Google Maps location history will soon be stored locally. Anadolu/Getty Images

- Location history data in <u>Google Maps</u> will soon be stored directly on user devices.
- Google itself will no longer have access to the data.
- This also means law enforcement won't be able to request it from Google anymore.

Google has come under increasing pressure to stop collecting user location data, especially since <u>Roe v. Wade was overturned</u>. Location data, along with internet search history and even <u>messaging history</u> <u>can be used as criminal evidence</u> against individuals who get an abortion in states where abortion is illegal.

42 Democrats from the US House and Senate signed a letter last May addressed to Google CEO Sundar Pichai urging the company to <u>stop</u> collecting and retaining user location information.

"Google's current practice of collecting and retaining extensive records of cell phone location data will allow it to become a tool for far-right extremists looking to crack down on people seeking reproductive health care," the letter read.

Last July, Google announced it would <u>delete the location history</u> <u>data of users who visited abortion clinics</u>, drug treatment centers, domestic violence shelters, weight loss clinics, and other sensitive health-related locations. The company said that if its systems identified that a user had visited one of these sensitive locations, it would then delete the entry from that user's location history "soon after they visit."



Discussion



Questions?

How much are you willing to sacrifice in terms of service for privacy?

Do you trust Google to responsibly handle you location data?

