# Privacy Does Matter!

Haojin Zhu

Professor

Computer Science & Engineering

Shanghai Jiao Tong University

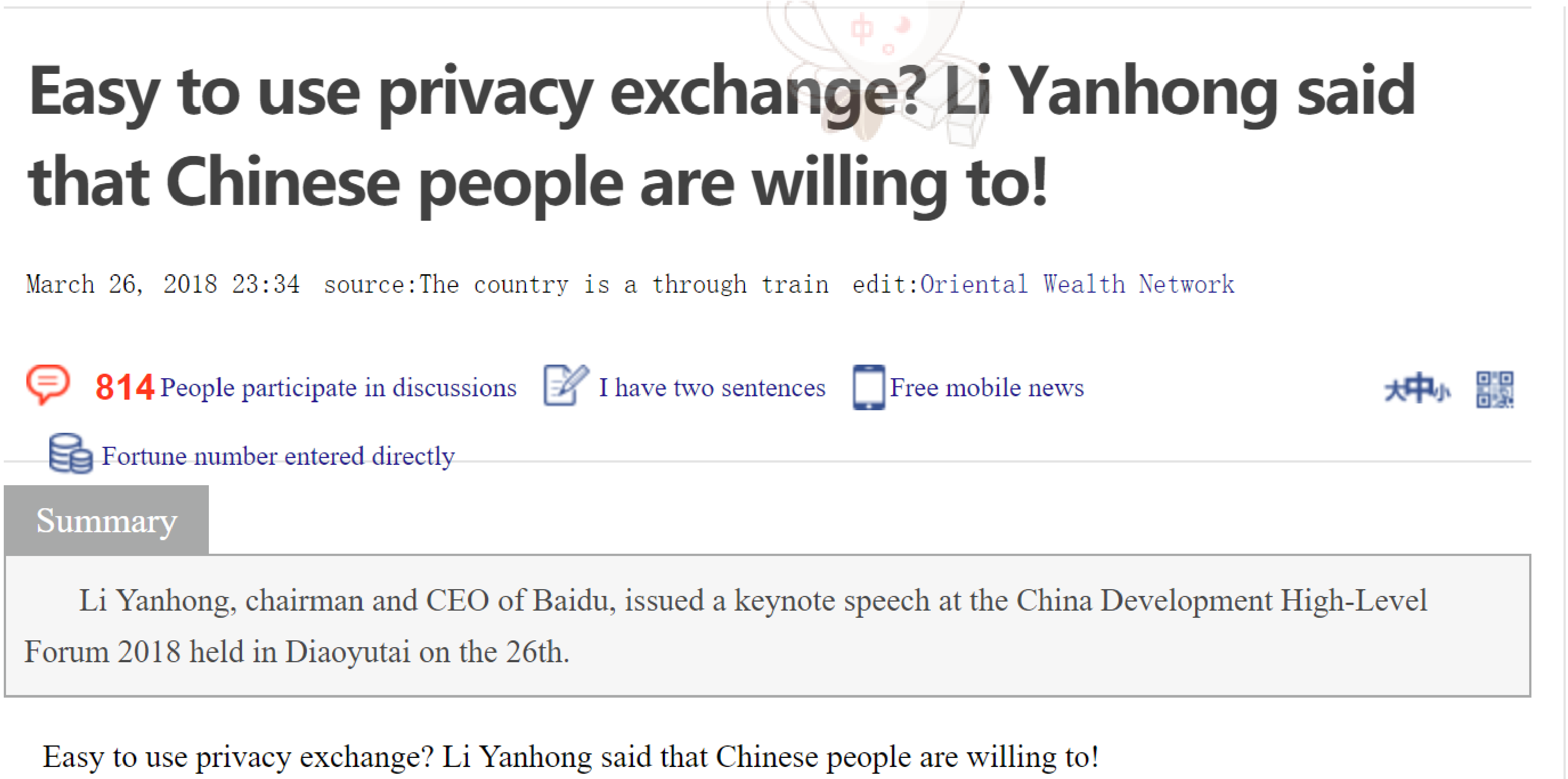
# Scope of Privacy in This Talk

- Data about individuals

- Collection, using, and sharing of such data

- Privacy is primarily a social, legal, and moral concept

# Let’s start from a recent news about baidu CEO’s talk on privacy



## Easy to use privacy exchange? Li Yanhong said that Chinese people are willing to!

March 26, 2018 23:34 source:The country is a through train edit:Oriental Wealth Network

814 People participate in discussions I have two sentences Free mobile news 大中小

Fortune number entered directly

Summary

Li Yanhong, chairman and CEO of Baidu, issued a keynote speech at the China Development High-Level Forum 2018 held in Diaoyutai on the 26th.

Easy to use privacy exchange? Li Yanhong said that Chinese people are willing to!

https://mp.weixin.qq.com/s/uhwph4gFvn0hDpLSCtR0ew

中国人对隐私问题的态度更加开放也相对来说没那么敏感

如果他们可以用隐私换取便利、安全或者效率

鲍达民
麦肯锡公司董事长、全球总裁

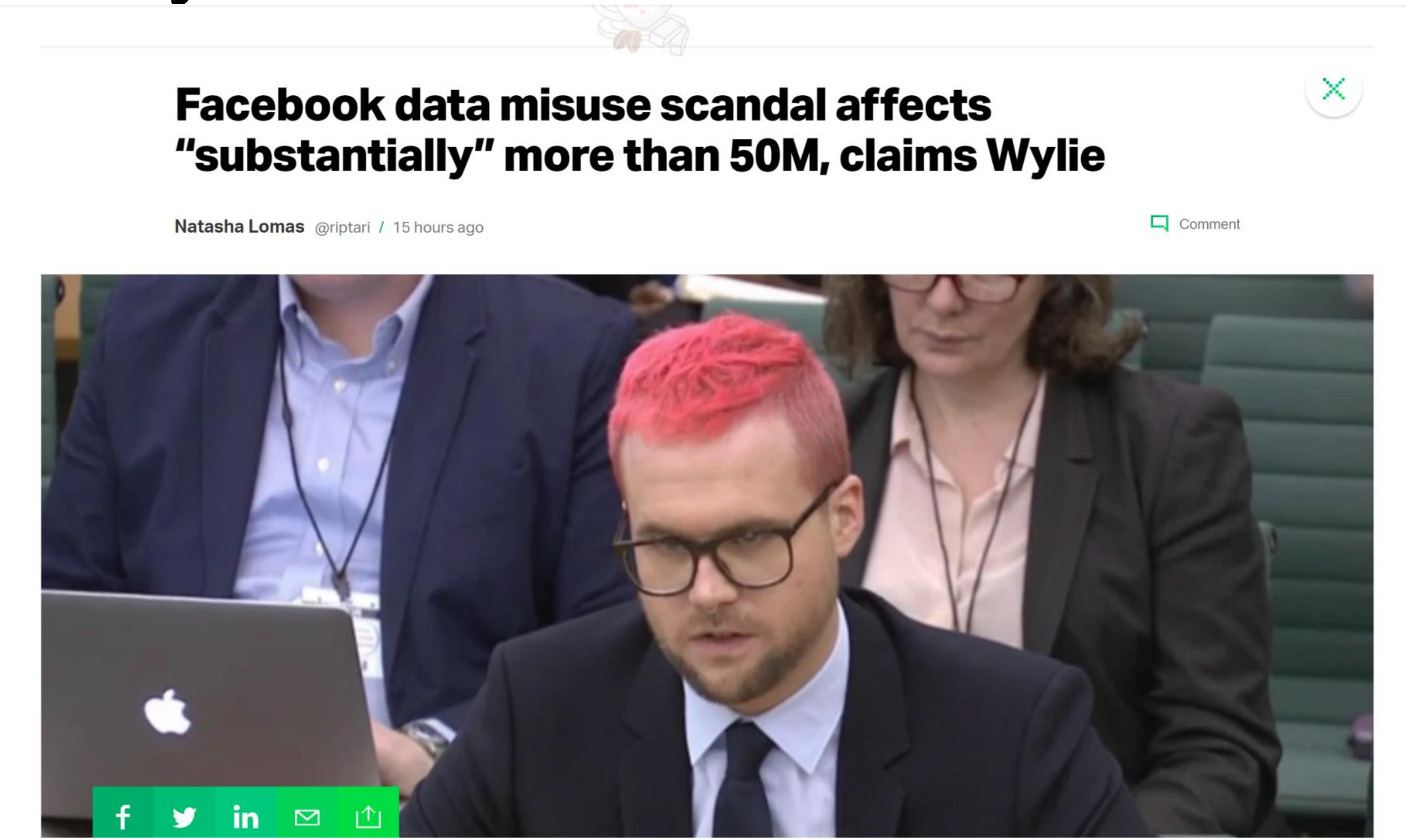
李彦宏
百度公司董事长兼首席执行官

在很多情况下他们就愿意这么做

# Let’s watch the full video

https://mp.weixin.qq.com/s/uhwph4gFvn0hDpLSCtR0ew

# On the other hand, when facebook data privacy leaks….



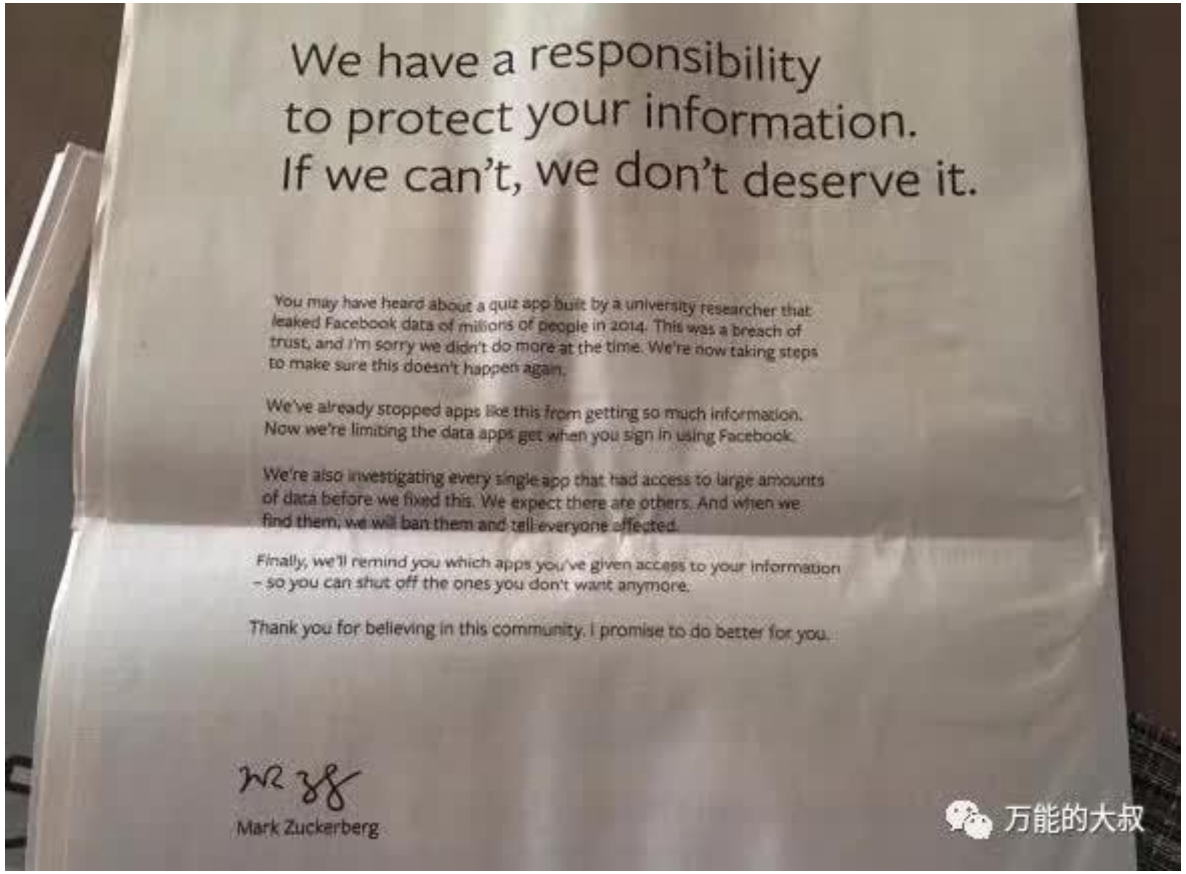
**Facebook data misuse scandal affects "substantially" more than 50M, claims Wylie**

**Natasha Lomas** @riptari / 15 hours ago

Comment

ıse

# “We have a responsibility to protect your data, and if we can‘t then we don’t deserve to serve you.” by Zuckerberg



We have a responsibility
to protect your information.
If we can’t, we don’t deserve it.

You may have heard about a quiz app built by a university researcher that leaked Facebook data of millions of people in 2014. This was a breach of trust, and I'm sorry we didn't do more at the time. We're now taking steps to make sure this doesn't happen again.

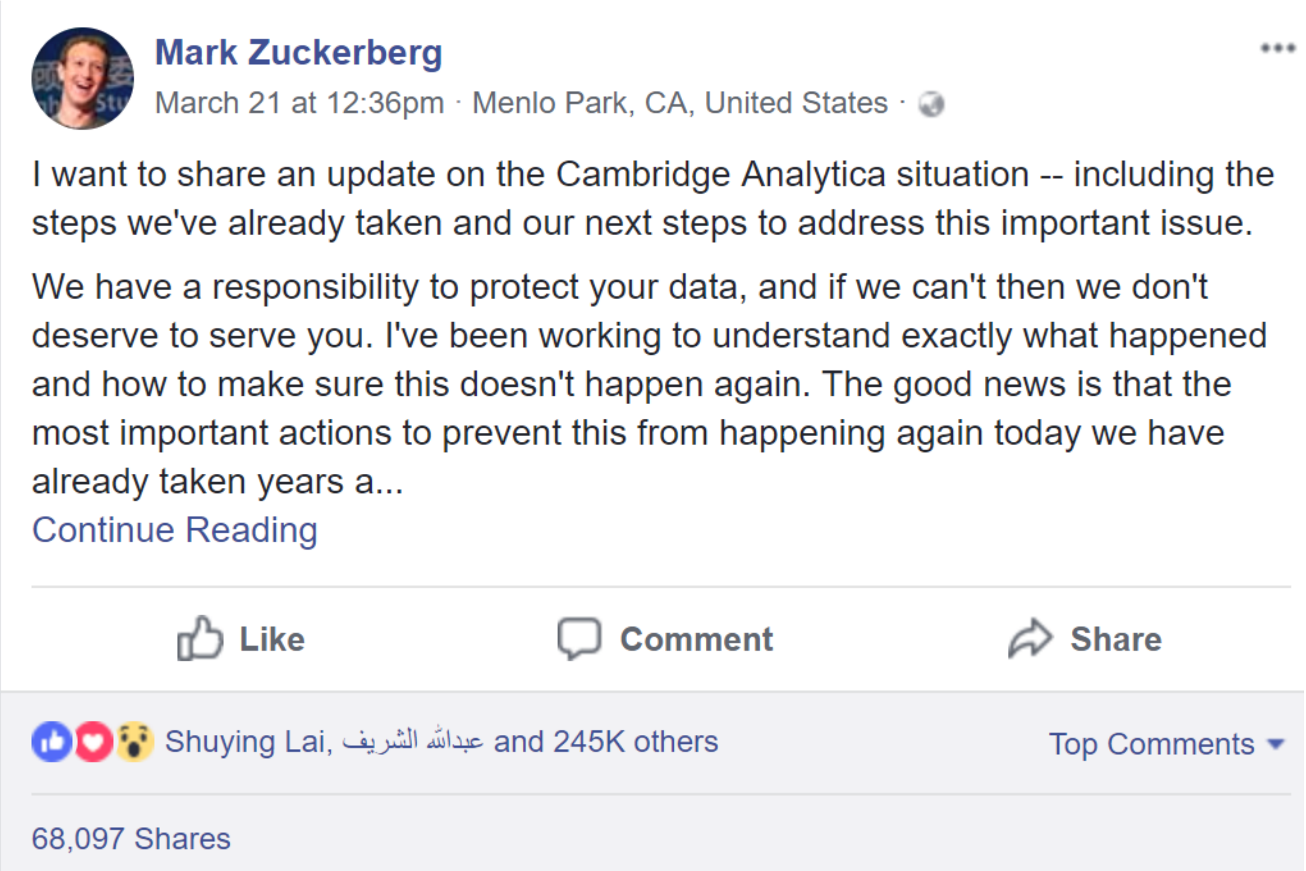
We've already stopped apps like this from getting so much information. Now we're limiting the data apps get when you sign in using Facebook.

We're also investigating every single app that had access to large amounts of data before we fixed this. We expect there are others. And when we find them, we will ban them and tell everyone affected.

Finally, we'll remind you which apps you've given access to your information – so you can shut off the ones you don't want anymore.

Thank you for believing in this community. I promise to do better for you.

Mark Zuckerberg



**Mark Zuckerberg**
March 21 at 12:36pm · Menlo Park, CA, United States ·

I want to share an update on the Cambridge Analytica situation -- including the steps we've already taken and our next steps to address this important issue.

We have a responsibility to protect your data, and if we can't then we don't deserve to serve you. I've been working to understand exactly what happened and how to make sure this doesn't happen again. The good news is that the most important actions to prevent this from happening again today we have already taken years a...
Continue Reading

Shuying Lai, عبدالله الشريف and 245K others

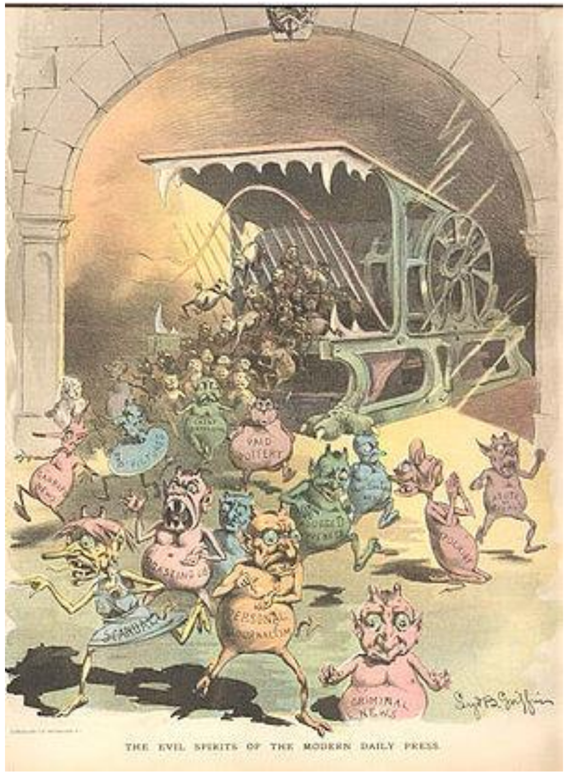
68,097 Shares

# Defining Privacy is Hard

- Lots of privacy notions
  - E.g., k anonymity, l diversity, t closeness, differential privacy, and many, many others
- Why defining privacy is hard?
  - Difficult  to agree on what should be protected from adversary.
  - Difficult to agree on adversary power.
  - Too strong , then not achievable.
  - Too weak, then not enough.
  - Information is correlated.

# Privacy

- Latin Privatus, meaning withdraw from public life
- In history
  - In 1086, William I of England commissioned the creation of the *Doomsday book*, a written record of major property holdings in England containing individual information collected for tax and draft purposes
  - 19th century, de-facto privacy was similarly threatened by *photographs* and *yellow journalism*.
  - one of the first publications advocating privacy in the U.S. in which Samuel Warren and Louis Brandeis argued that privacy law must evolve in response to technological changes [1]

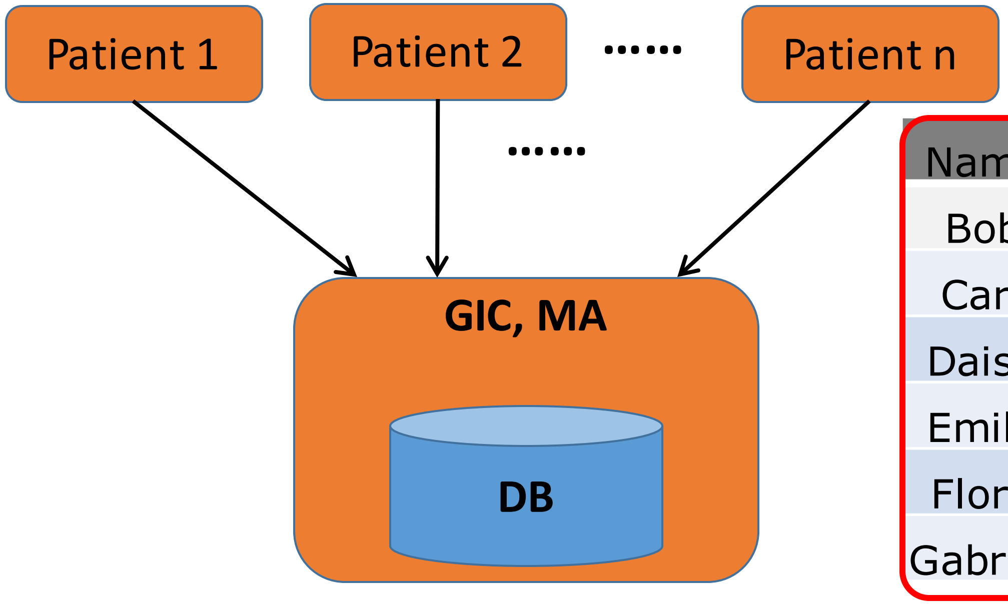


THE EVIL SPIRITS OF THE MODERN DAILY PRESS.

1. Warren, S. & Brandeis, L. The right to privacy. Harvard Law Review 193, 193–220 (1890).

# GIC Incidence [Sweeny 2002]

- Group Insurance Commissions (GIC, Massachusetts)
  - Collected patient data for ~135,000 state employees.
  - Gave to researchers and sold to industry.
  - Medical record of the former state governor is identified.



| Name | DoB | Gender | Zip code | Disease |
|---|---|---|---|---|
| Bob | 1/3/45 | M | 47906 | Cancer |
| Carl | 4/7/64 | M | 47907 | Cancer |
| Daisy | 9/3/69 | F | 47902 | Flu |
| Emily | 6/2/71 | F | 46204 | Gastritis |
| Flora | 2/7/80 | F | 46208 | Hepatitis |
| Gabriel | 5/5/68 | F | 46203 | Bronchitis |

Re-identification occurs!

# AOL Data Release [NYTimes 2006]

- In August 2006, AOL Released search keywords of 650,000 users over a 3-month period.
  - User IDs are replaced by random numbers.
  - 3 days later, pulled the data from public access.

AOL searcher # 4417749

*"landscapers in Lilburn, GA"*
*queries on last name "Arnold"*
*"homes sold in shadow lake subdivision Gwinnett County, GA"*
*"num fingers"*
*"60 single men"*
*"dog that urinates on everything"*

NYT →

Thelman Arnold, a 62 year old widow who lives in Liburn GA, has three dogs, frequently searches her friends' medical ailments.

Re-identification occurs!

# Genome-Wide Association Study (GWAS) [Homer et al. 2008]

- A typical study examines thousands of singe-nucleotide polymorphism locations (SNPs) in a given population of patients for statistical links to a disease.
- From aggregated statistics, one individual's genome, and knowledge of SNP frequency in background population, one can infer participation in the study.
  - The frequency of every SNP gives a very noisy signal of participation; combining thousands of such signals give high-confidence prediction

# GWAS Privacy Issue

**Published Data**

| | Disease Group Avg | Control Group Avg |
|---|---|---|
| SNP1=A | 43% | … |
| SNP2=A | 11% | … |
| SNP3=A | 58% | … |
| SNP4=A | 23% | … |
| … | | |

**Adv. Info & Inference**

| Population Avg | Target individual Info | Target in Disease Group |
|---|---|---|
| 42% | yes | + |
| 10% | no | - |
| 59% | no | + |
| 24% | yes | - |
| | | |

Membership disclosure occurs!

# Data Privacy Research Program

- Develop theory and techniques to anonymize data so that they can be beneficially used without privacy violations.

- How to define privacy for anonymized data?
- How to publish/anonymize data to satisfy privacy while providing utility?

# *k*-Anonymity [Sweeney, Samarati ]

**The Microdata**

| QID | | | SA |
|---|---|---|---|
| Zipcode | Age | Gen | Disease |
| 47677 | 29 | F | Ovarian Cancer |
| 47602 | 22 | F | Ovarian Cancer |
| 47678 | 27 | M | Prostate Cancer |
| 47905 | 43 | M | Flu |
| 47909 | 52 | F | Heart Disease |
| 47906 | 47 | M | Heart Disease |

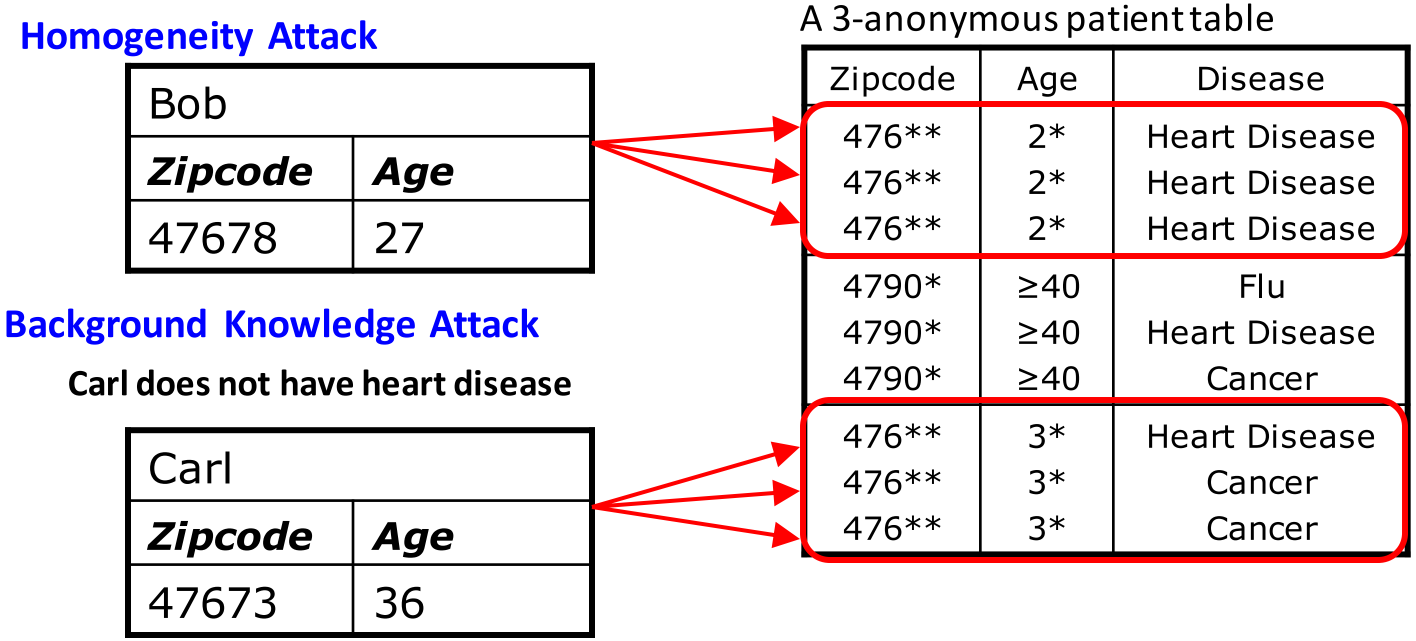
**A 3-Anonymous Table**

| QID | | | SA |
|---|---|---|---|
| Zipcode | Age | Gen | Disease |
| 476** | 2* | * | Ovarian Cancer |
| 476** | 2* | * | Ovarian Cancer |
| 476** | 2* | * | Prostate Cancer |
| 4790* | [43,52] | * | Flu |
| 4790* | [43,52] | * | Heart Disease |
| 4790* | [43,52] | * | Heart Disease |

- k-Anonymity
  - Each record is indistinguishable from ≥ k-1 other records when only “quasi-identifiers” are considered
  - These k records form an equivalence class

# Attacks on *k*-Anonymity

- k-anonymity does not protect against inference of sensitive attribute values:
  - Sensitive values lack diversity
  - The attacker has background knowledge



**Homogeneity Attack**

| Bob | |
|---|---|
| ***Zipcode*** | ***Age*** |
| 47678 | 27 |

**Background Knowledge Attack**

**Carl does not have heart disease**

| Carl | |
|---|---|
| ***Zipcode*** | ***Age*** |
| 47673 | 36 |

A 3-anonymous patient table

| Zipcode | Age | Disease |
|---|---|---|
| 476** | 2* | Heart Disease |
| 476** | 2* | Heart Disease |
| 476** | 2* | Heart Disease |
| 4790* | ≥40 | Flu |
| 4790* | ≥40 | Heart Disease |
| 4790* | ≥40 | Cancer |
| 476** | 3* | Heart Disease |
| 476** | 3* | Cancer |
| 476** | 3* | Cancer |

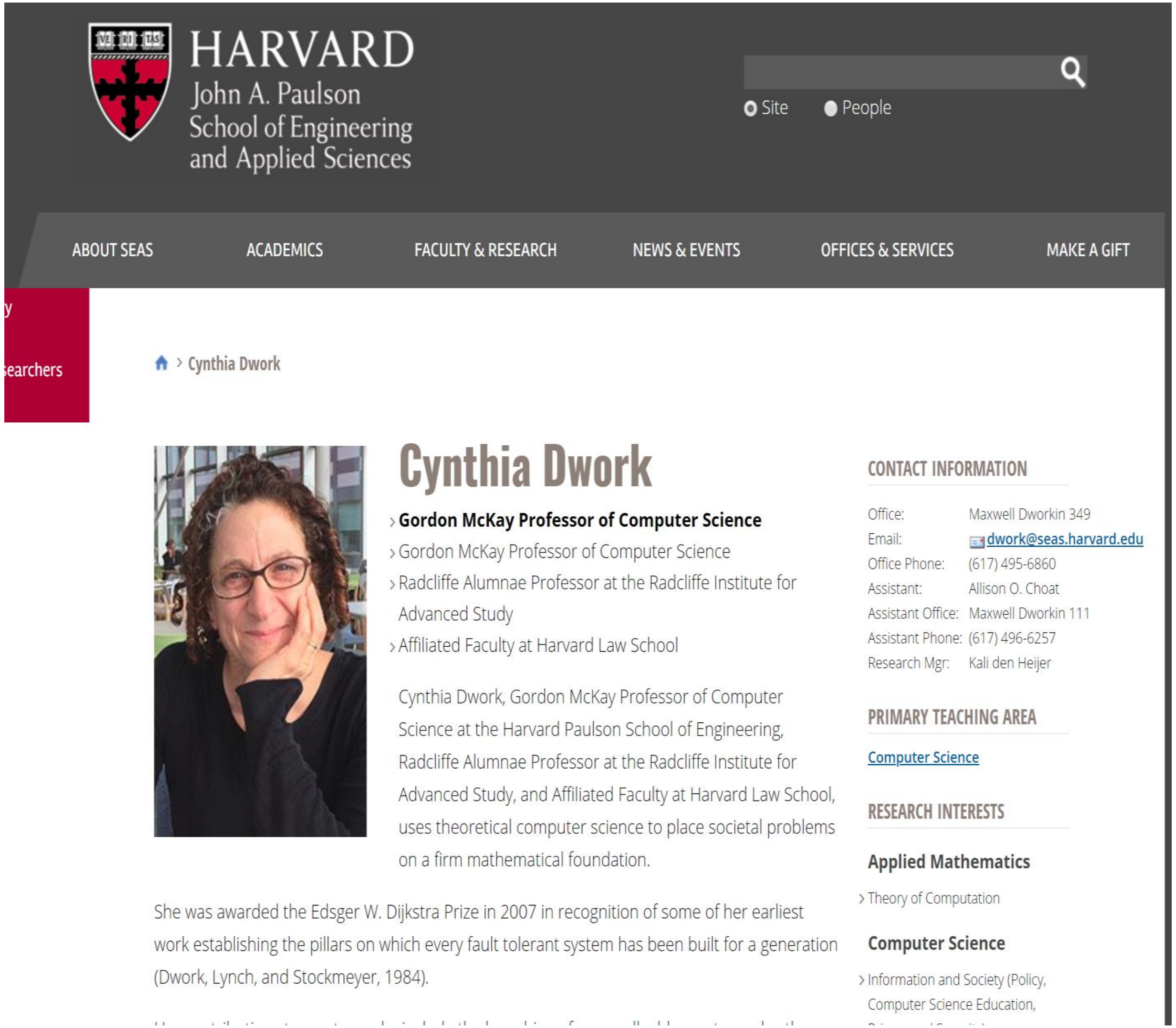
# $l$-diversity

- The $l$-diversity principle
  - Each equivalent class contains at least $l$ well-represented sensitive values
- Instantiation
  - Distinct $l$-diversity
    - Each equi-class contains $l$ distinct sensitive values
  - Entropy $l$-diversity
    - *entropy(equi-class)≥log$_2$($l$)*

$$H(X) = \mathrm{E}(I(X)) \;=\; -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$$

# Differential Privacy [Dwork et al. 2006]

- Definition: A mechanism *A* satisfies ε-Differential Privacy if and only if
  - for any neighboring datasets D and D’
  - and any possible transcript $t \in$ Range(A),
    $$\Pr[A(D) = t] \leq e^{\epsilon} \Pr[A(D') = t]$$
  - For relational datasets, typically, datasets are said to be neighboring if they differ by a single record.



## Cynthia Dwork

- **Gordon McKay Professor of Computer Science**
- Gordon McKay Professor of Computer Science
- Radcliffe Alumnae Professor at the Radcliffe Institute for Advanced Study
- Affiliated Faculty at Harvard Law School

Cynthia Dwork, Gordon McKay Professor of Computer Science at the Harvard Paulson School of Engineering, Radcliffe Alumnae Professor at the Radcliffe Institute for Advanced Study, and Affiliated Faculty at Harvard Law School, uses theoretical computer science to place societal problems on a firm mathematical foundation.

She was awarded the Edsger W. Dijkstra Prize in 2007 in recognition of some of her earliest work establishing the pillars on which every fault tolerant system has been built for a generation (Dwork, Lynch, and Stockmeyer, 1984).

### CONTACT INFORMATION

| | |
|---|---|
| Office: | Maxwell Dworkin 349 |
| Email: | dwork@seas.harvard.edu |
| Office Phone: | (617) 495-6860 |
| Assistant: | Allison O. Choat |
| Assistant Office: | Maxwell Dworkin 111 |
| Assistant Phone: | (617) 496-6257 |
| Research Mgr: | Kali den Heijer |

### PRIMARY TEACHING AREA

Computer Science

### RESEARCH INTERESTS

**Applied Mathematics**

- Theory of Computation

**Computer Science**

- Information and Society (Policy, Computer Science Education, Privacy and Security)

Cynthia Dwork (born 1958) is an American computer scientist at Harvard University, where she is Gordon McKay Professor of Computer Science, Radcliffe Alumnae Professor at the Radcliffe Institute for Advanced Study, and Affiliated Professor, Harvard Law School.

She was elected as a Fellow of the AAAS in 2008,[7][8] as a member of the National Academy of Engineering in 2008,[9] as a member of the National Academy of Sciences in 2014, as a fellow of the Association for Computing Machinery in 2015,[10], and as a member of the American Philosophical Society in 2016.[11] She received the Dijkstra Prize in 2007 for her work on consensus problems together with Nancy Lynch and Larry Stockmeyer.[12][13] In 2009 she won the PET Award for Outstanding Research in Privacy Enhancing Technologies.[14] 2017 Gödel Prize was awarded to Cynthia Dwork, Frank McSherry, Kobbi Nissim and Adam Smith for their seminal paper that introduced differential privacy.[15]

# Key Assumption Behind DP: The Personal Data Principle

- After removing one individual’s data, that individual’s privacy is protected perfectly.

- In other words, for each individual, the world after removing the individual’s data is an ideal world of privacy for that individual. Goal is to simulate all these ideal worlds.

# What Can Be Achieved Under DP?

- Publishing information of low-dimensional data
- Perform specific tasks for high-dimensional data

# Particular Data Mining Tasks

- K-means Clustering
- Classification
- Deep learning
- Frequent-itemset mining

- Solving genera problems for high-dimensional (and other complex) data remain an open problem
  - Appears possible with big data

# What Constitutes An Individual’s Data?

- Is the genome of my parents, children, sibling, cousins “my personal information”?

- Example: DeCode Genetics, based in Reykjavík, has collected full DNA sequences on 10,000 individuals. And because people on the island are closely related, DeCode says it can now also extrapolate to accurately guess the DNA makeup of nearly all other 320,000 citizens of that country, including those who never participated in its studies.

# Such legal and ethical questions still need to be resolved

- Evidences suggest that such privacy concerns will be recognized.
- In 2003, the supreme court of Iceland ruled that a daughter has the right to prohibit the transfer of her deceased father's health information to a Health Sector Database, not because her right acting as a substitute of her deceased father, but in the recognition that she might, on the basis of her right to protection of privacy, have an interest in preventing the transfer of health data concerning her father into the database, as information could be inferred from such data relating to the hereditary characteristics of her father which might also apply to herself.

https://epic.org/privacy/genetic/iceland_decision.pdf

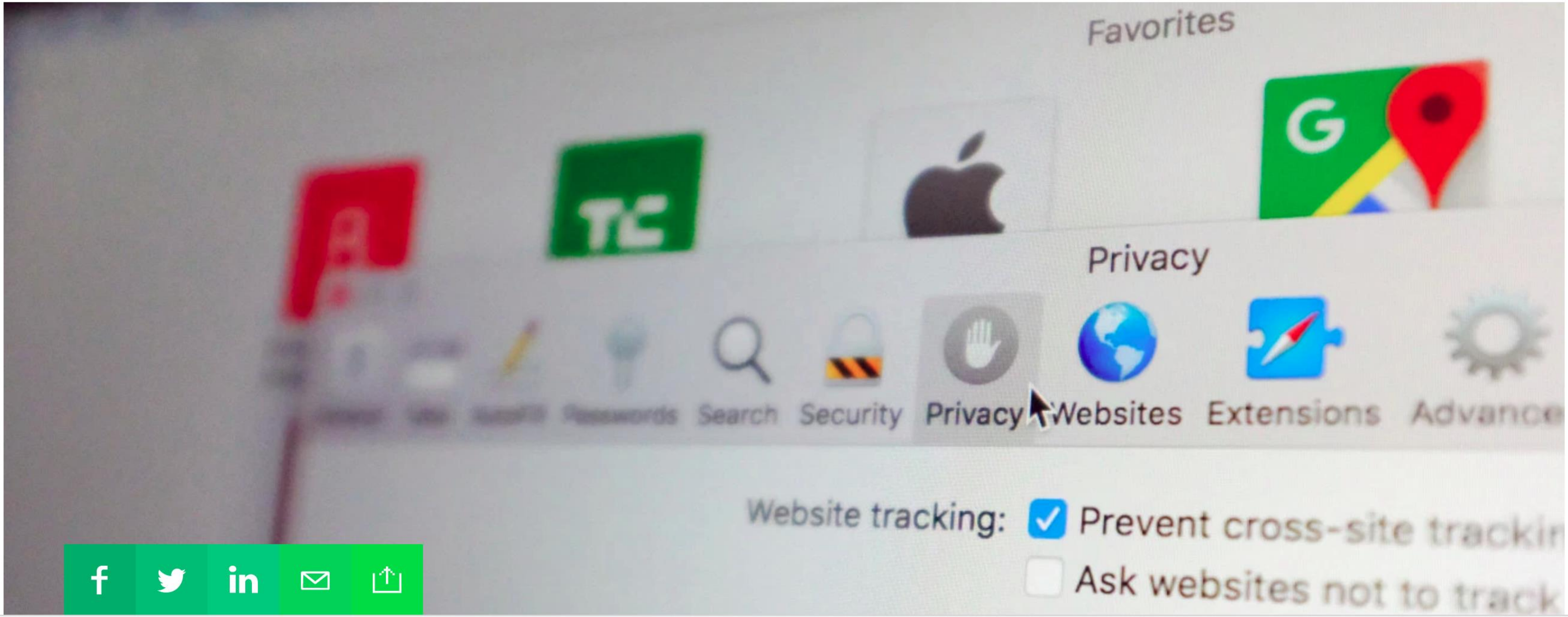
# Lesson

- When dealing with genomic and health data, one cannot simply say correlation doesn't matter because of Personal Data Principle, and may have to quantify and deal with such correlation.

# Apple starts collecting browsing data in Safari using its differential privacy tech

**Brian Heater** @bheater / Sep 26, 2017

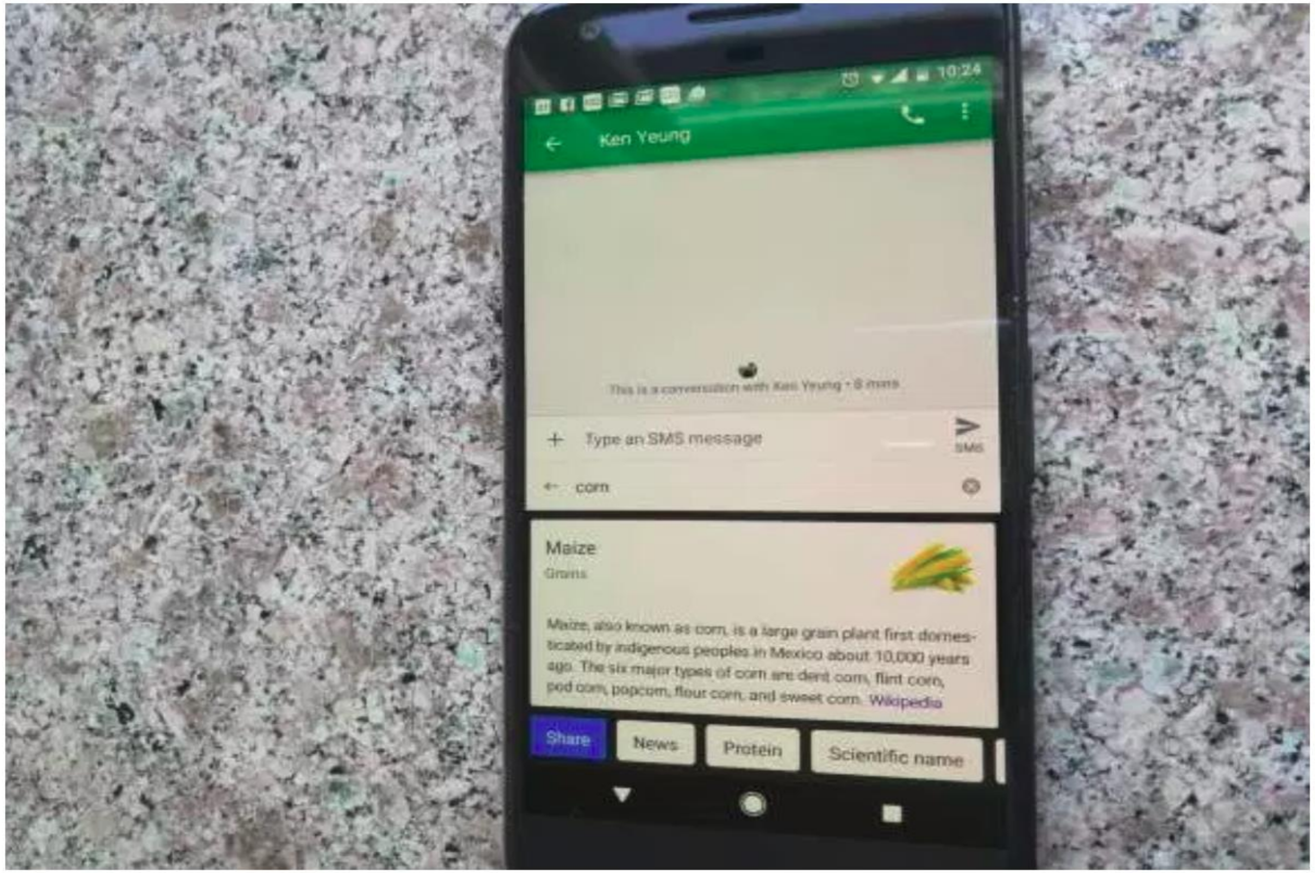
Comment



AI

# Following Apple, Google is exploring differential privacy in Gboard for Android

JORDAN NOVET @JORDANNOVET APRIL 6, 2017 6:50 PM



Maize, also known as corn, is a large grain plant first domesticated by indigenous peoples in Mexico about 10,000 years ago. The six major types of corn are dent corn, flint corn, pod corn, popcorn, flour corn, and sweet corn. Wikipedia

Above: A search result in Gboard for Android.

Image Credit: Jordan Novet/VentureBeat

## VB Recommendations

Nvidia reinvents workstations with real-time ray tracing

Atari pressed the reset button on its Atari VCS game console

When consumer packaged goods start acting like software

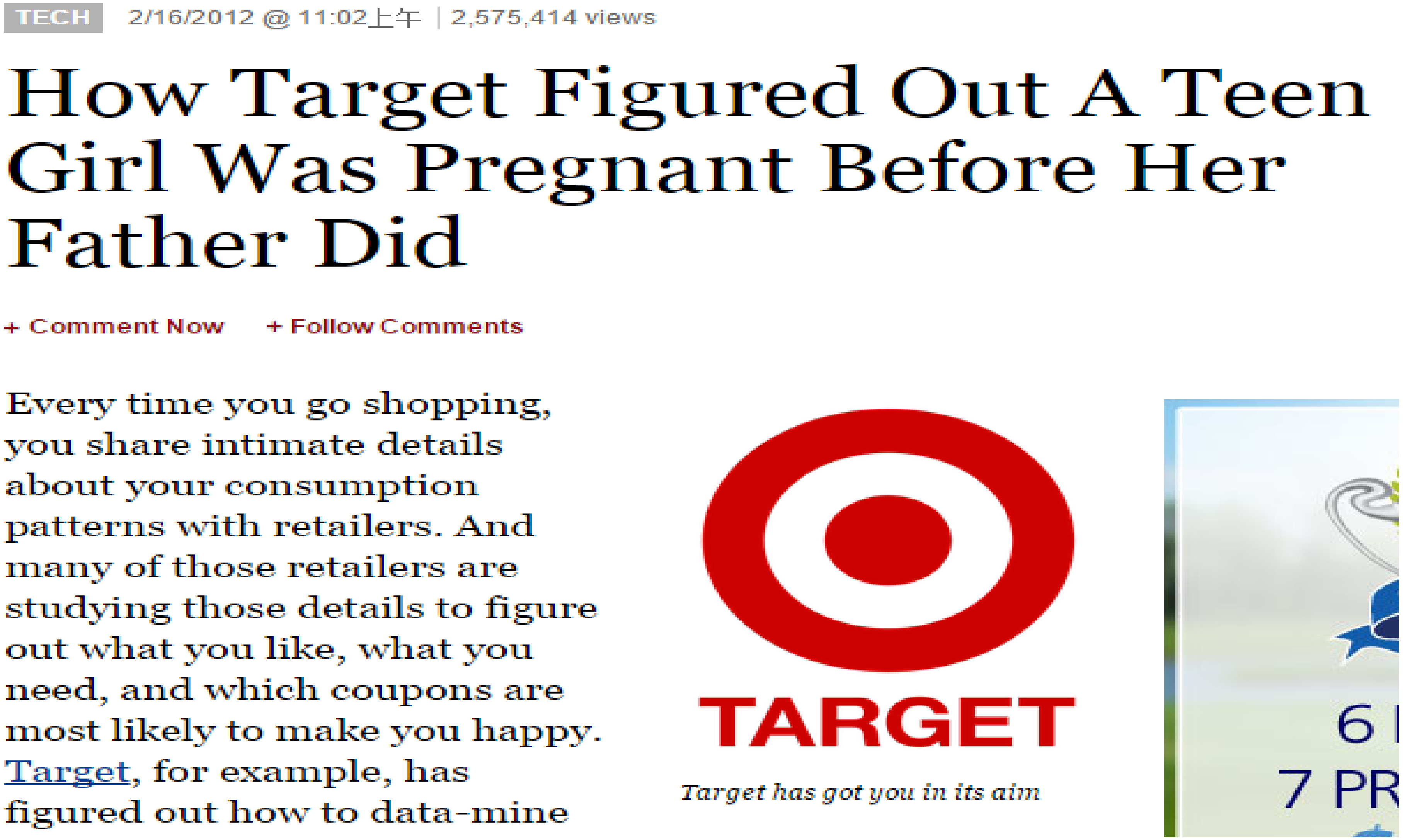
## Upcoming Events

GamesBeat Summit 2018 Apr 9 - 10

Transform 2018 Aug 21 - 22

VB Summit 2018 Oct 22 - 23

# Big Data Privacy



TECH 2/16/2012 @ 11:02上午 | 2,575,414 views

## How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

+ Comment Now + Follow Comments

Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying those details to figure out what you like, what you need, and which coupons are most likely to make you happy. Target, for example, has figured out how to data-mine

*Target has got you in its aim*

# Privacy and Discrimination

- What if one applies a classifier to public information (such as gender, age, race, nationality, etc.) and make decisions accordingly
- Is there privacy concern?
- Better privacy may cause more discrimination!
  - From Wheelan’s book “Naked Economics”
  - Hiring blacks with (and w/o) criminal background checks.

# The Legal Aspect of Privacy

# President Obama's Call for Review of Privacy (Jan 2014)

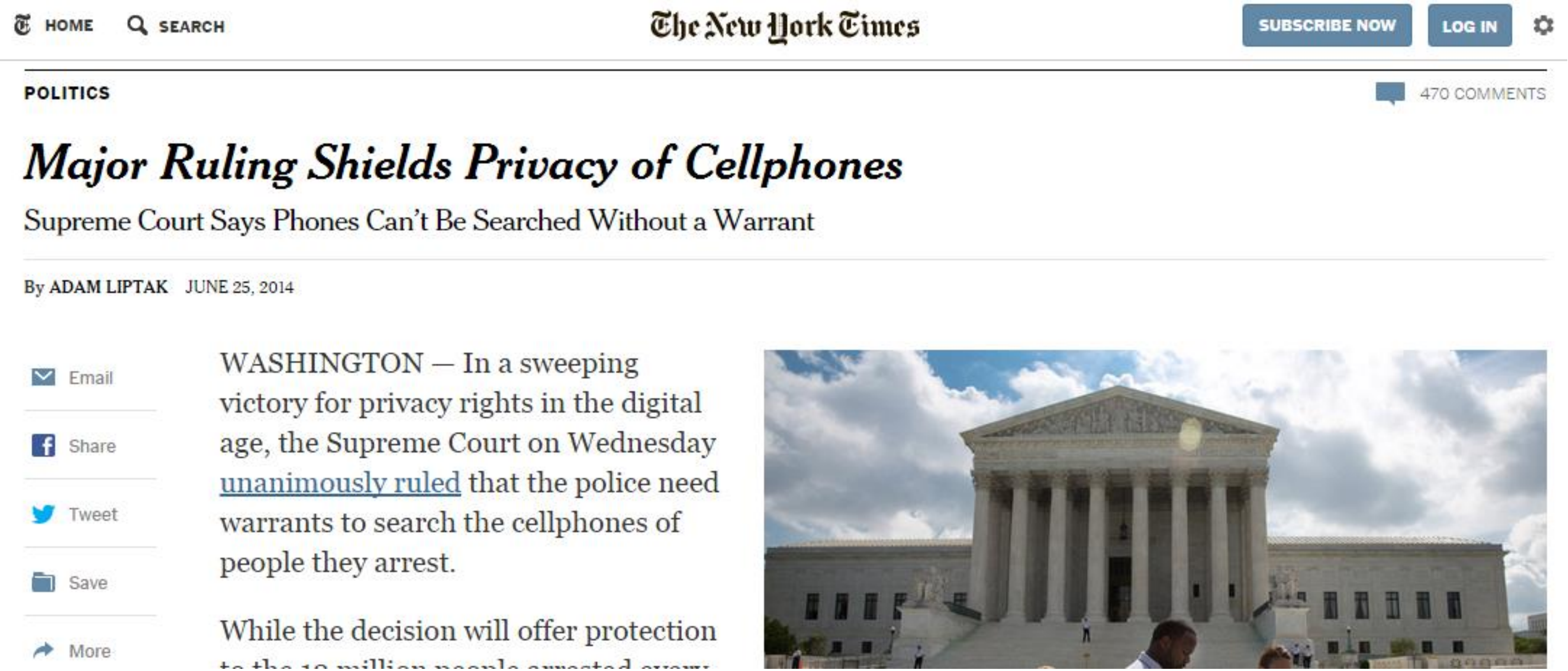
## President Obama Discusses U.S. Intelligence Programs at the Department of Justice

Megan Slack
January 17, 2014
06:44 PM EDT

Share This Post

President Barack Obama delivers remarks presenting the outcome of the Administration's review

# U.S. Supreme Court's Cellphone Ruling Is a Major Victory for Privacy (2014)



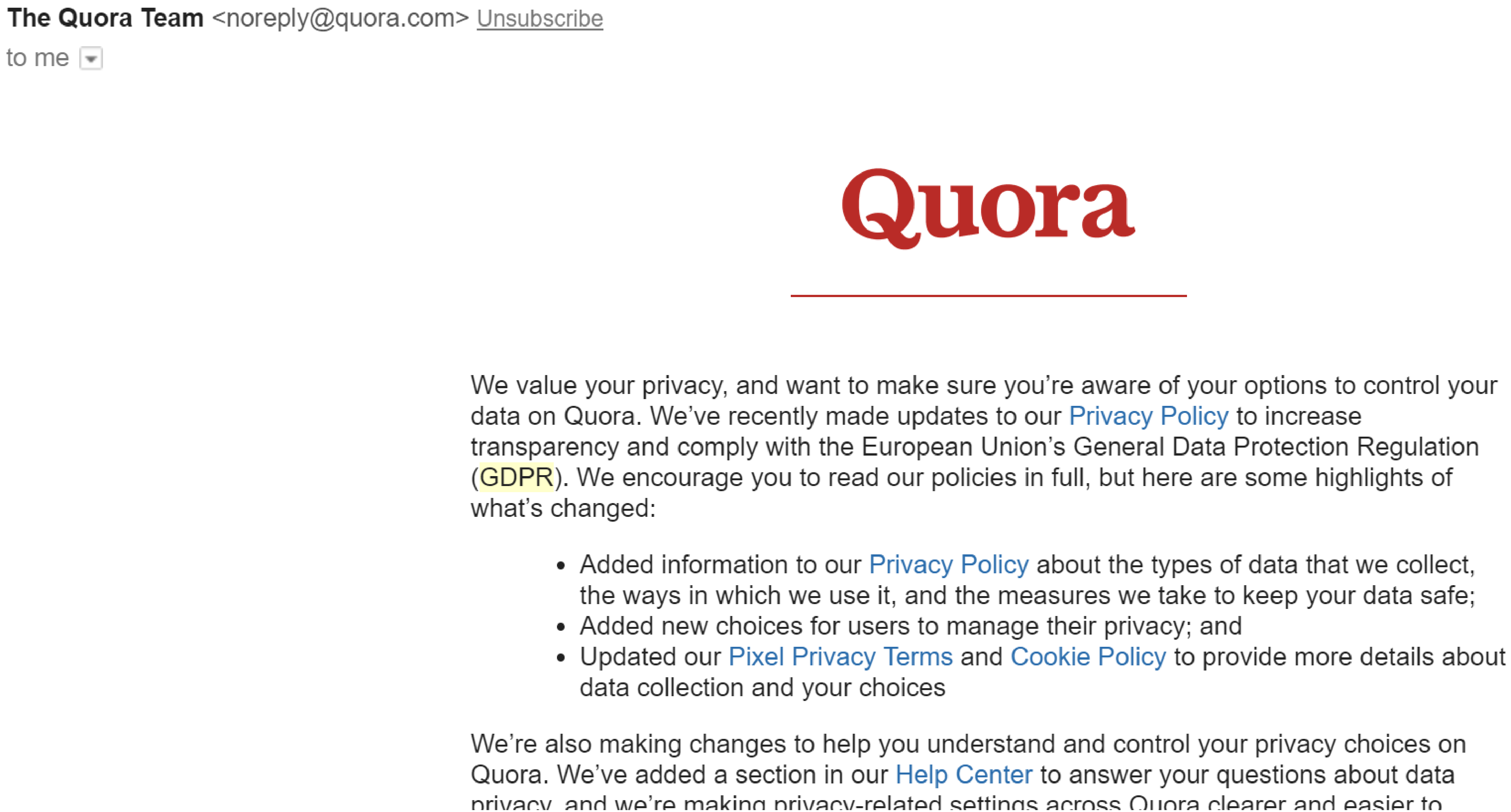
## *Major Ruling Shields Privacy of Cellphones*

Supreme Court Says Phones Can't Be Searched Without a Warrant

By ADAM LIPTAK JUNE 25, 2014

WASHINGTON — In a sweeping victory for privacy rights in the digital age, the Supreme Court on Wednesday unanimously ruled that the police need warrants to search the cellphones of people they arrest.

While the decision will offer protection to the 12 million people arrested every



**The Quora Team** <noreply@quora.com> Unsubscribe

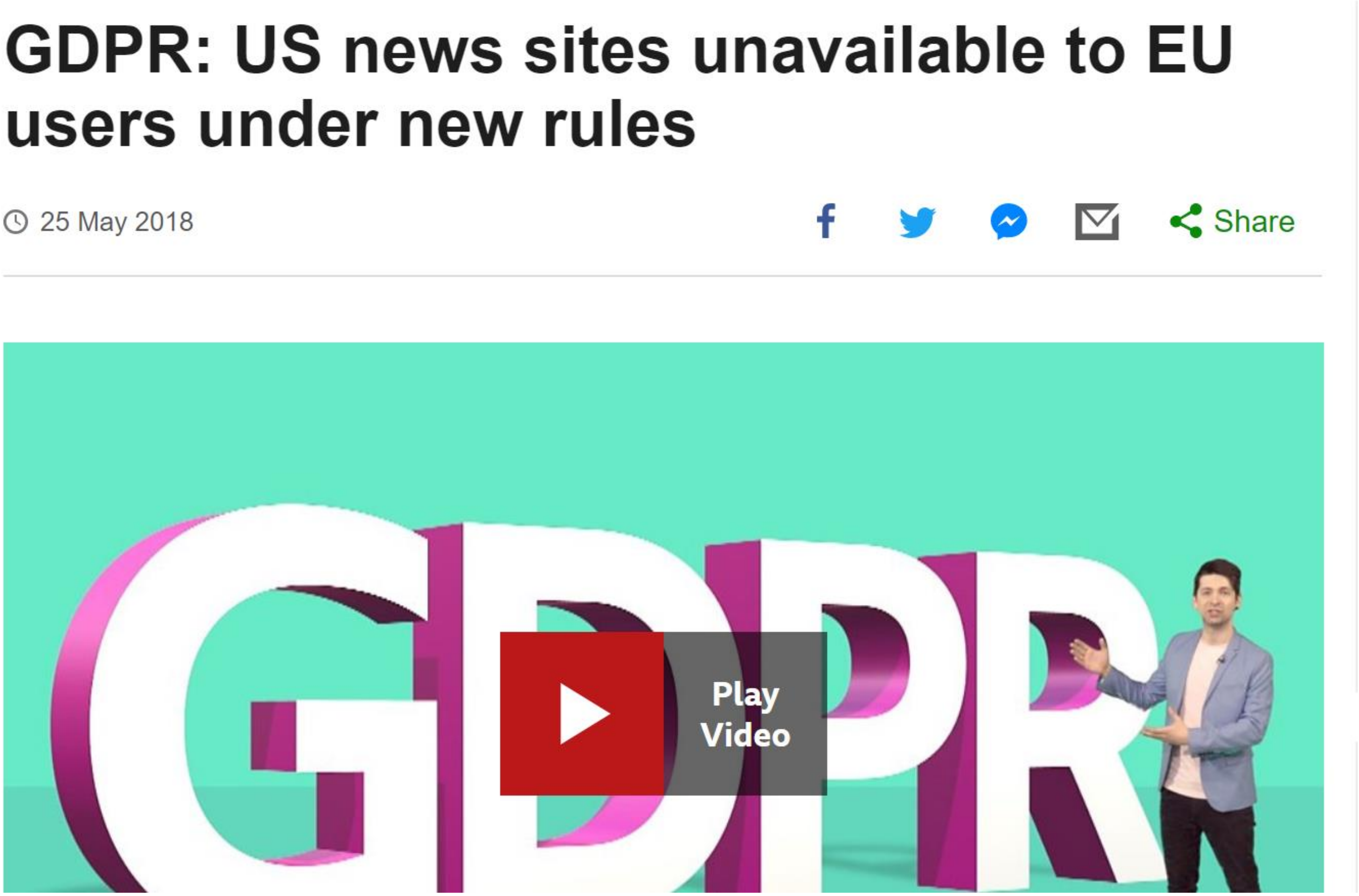
to me

Quora

We value your privacy, and want to make sure you're aware of your options to control your data on Quora. We've recently made updates to our Privacy Policy to increase transparency and comply with the European Union's General Data Protection Regulation (GDPR). We encourage you to read our policies in full, but here are some highlights of what's changed:

- Added information to our Privacy Policy about the types of data that we collect, the ways in which we use it, and the measures we take to keep your data safe;
- Added new choices for users to manage their privacy; and
- Updated our Pixel Privacy Terms and Cookie Policy to provide more details about data collection and your choices

We're also making changes to help you understand and control your privacy choices on Quora. We've added a section in our Help Center to answer your questions about data privacy, and we're making privacy-related settings across Quora clearer and easier to



# GDPR: US news sites unavailable to EU users under new rules

25 May 2018

## 关注点一：适用 GDPR 的场景

GDPR第三条规定，以下两类情形在其适用范围内：

**一是数据控制者或数据处理者在欧盟境内设有分支机构（establishment）。**在此情形中，只要个人数据处理活动发生在分支机构开展活动的场景中（in the context of the activities of an establishment），即使实际的数据处理活动不在欧盟境内发生，适用GDPR。

例如，在欧盟本地运营的A国（A国指某一非欧盟成员国）连锁酒店，直接将其收集的住客个人数据传输至A国总部进行处理，则需要履行GDPR中相关责任和义务。

**二是数据控制者或数据处理者在欧盟境内不设分支机构（establishment）的情形。**在此情形中，GDPR原则性地规定只要其面向欧盟境内的数据主体提供商品或服务（无论是否发生支付行为），或监控（monitor）欧盟境内数据主体的行为，适用GDPR。

例如，A国境内运营的某一电商平台，在欧盟不设分支机构，但提供专门

的法文、德文版本的页面，同时支持用欧元进行结算，支持向欧盟境内配送物流。该电商平台属于面向欧盟境内的数据主体提供商品或服务，需要适用GDPR。

例如，在A国运营的社交媒体平台，支持境外账户注册，且已有欧盟境内用户使用。该社交媒体平台根据用户的位置信息、浏览记录等行为信息，向用户推送个性化的信息和广告，有可能被欧盟的个人数据保护机构（Data Protection Authority）认定为监控（monitor）欧盟境内数据主体的行为，适用GDPR的可能性较高。

## 关注点二：适用的数据范围

GDPR规定，个人数据，是指与一个确定的或可识别的自然人相关的任何信息。可被识别的自然人,是指借助标识符，例如姓名、身份标识、位置数据、网上标识符，或借助与该个人生理、心理、基因、精神、经济、文化或社会身份特定相关的一个或多个因素，可被直接或间接识别出的个人。

GDPR规定，特殊类别（敏感）个人数据，是指揭示种

族或民族出身，政治观点、宗教或哲学信仰以及工会成员的个人数据，以及唯一识别自然人为目的的基因数据、生物特征数据、自然人的健康、性生活或性取向数据，还包括刑事定罪和犯罪相关的个人资料等。

**组织应识别其处理个人数据或特殊类别（敏感）个人数据的具体类型。**

## 关注点四：数据处理的合法正当性事由

GDPR规定，数据处理行为首先应具备合法性基础，GDPR规定的六种合法性情形包括：数据主体的同意、合同履行、履行法定义务、保护个人重要利益、维护公共利益以及追求正当利益。

GDPR强调同意是指“数据主体通过书面声明或经由一

个明确的肯定性动作，表示同意对其个人数据进行处理。该意愿表达应是自由给出的（freely given）、特定具体的（specific）、知情的（informed）、清晰明确的（unambiguous）”，且撤回同意的方式应该与表达同意同等便利。

**组织应保证其数据处理活动满足合法性要求。**

## 关注点五：对儿童的特别保护规定

GDPR认为，儿童可能不太了解有关个人数据处理的风险、后果以及他们在数据处理中所拥有的权利，因此在收集有关儿童的个人数据时和将其个人数据使用在对儿童提供营销或创设个人账户的服务时，应予以特别保护。GDPR在“信息社会服务中适用儿童同意的条件”规定，如果直接向儿童提供信息社会服务时，该儿童的年龄应当为 16 周岁以上。若儿童未满 16 周岁，只有在征得监护人同意或授权的范围内其处理才合法。

**组织如涉及儿童个人数据的处理，应予以特别保护。**

## 关注点六：数据主体权利

GDPR赋予了数据主体对其数据广泛的控制权，包括知情、访问、更正、删除、限制处理、可携带、反对等权利。比如：

在数据收集时，数据控制者应以简洁、透明、易懂及便于访问的方式向数据主体提供数据控制者身份及联系信息、

## 关注点七：对用户画像的规定

GDPR规定，用户画像是指通过自动化方式处理个人数据的活动，用于评估、分析以及预测个人的特定方面，可能包括工作表现、经济状况、位置、健康状况、个人偏

好、可信赖度或者行为表现等。如电商通过用户画像，开展广告、市场预测和推广工作。

GDPR规定，在数据主体明确同意、欧盟或者成员国法律的明确授权、履行合同所必需的情形下可以使用用户画像。同时还提出，在征得数据主体同意时，应对画像相关数据来源、算法原理及相应影响等予以充分告知，并赋予数据主体反对权、删除权、更正权和限制处理权等权利。

**组织如涉及对用户进行画像，需要关注如何获得合法性基础，以及向数据主体提供相应权利。**

## 关注点八：对数据处理者的规定

GDPR规定，数据处理者是指为数据控制者处理个人数据的自然人、法人、公共机构、行政机关或其他非法人组织。

GDPR规定，当数据控制者委托数据处理者具体处理数据时，数据控制者应选择采取了合适的技术和组织方面措施的数据处理者，以确保数据处理符合GDPR的要求，及保障数据主体的权利。在没有数据控制者事先或一般性的书面许可时，数据处理者不应再与另外的数据处理者合作。

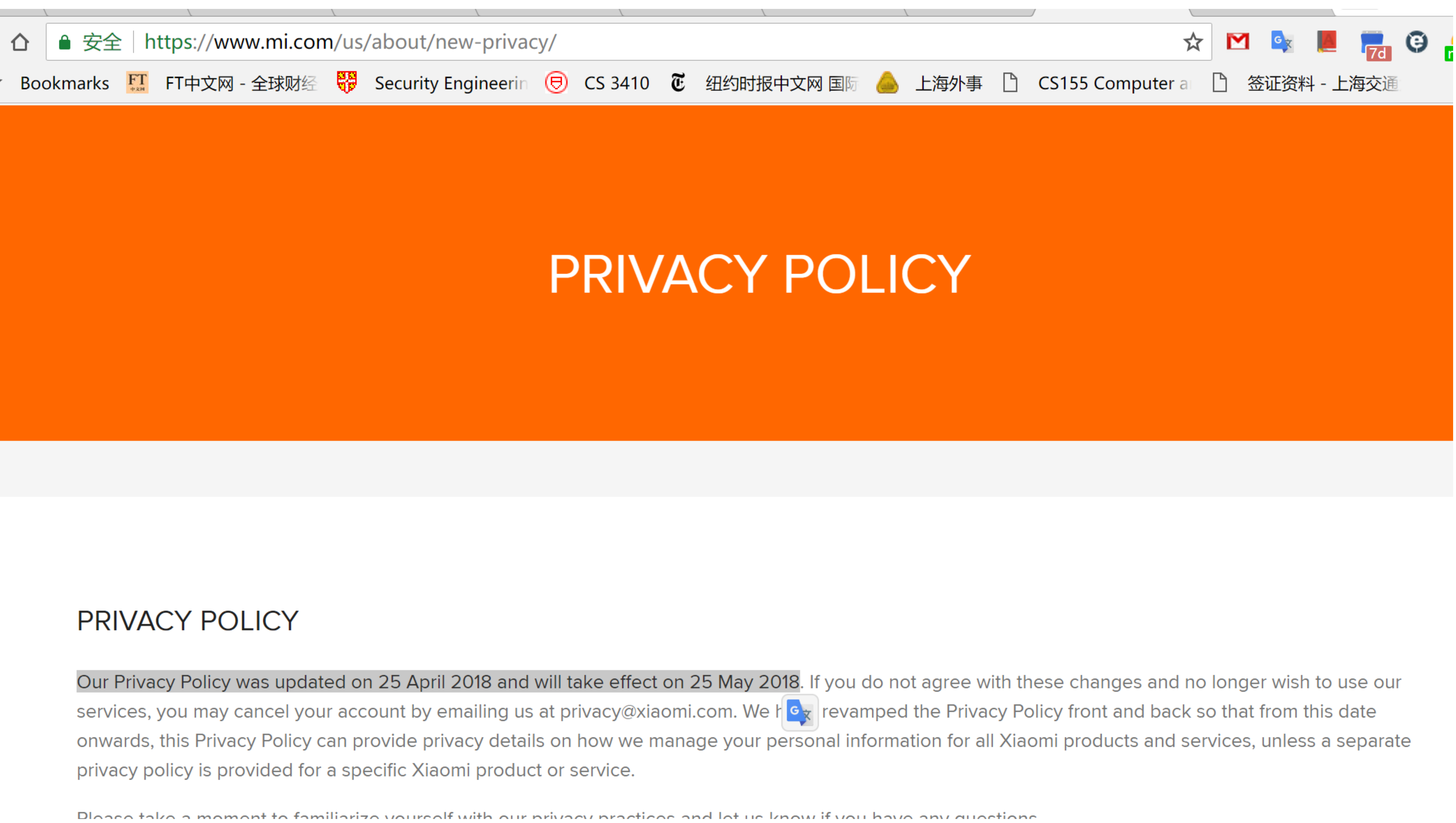
**组织如属于数据处理者，应根据数据控制者明确的指示处理个人数据，履行相应的保密义务，在数据处理服务结束时，删除或返还所有的个人数据，接受数据控制者的审计等。**

## 关注点十四：处罚规定

GDPR对违规组织采取根据情况分级处理的方法，并设定了最低一千万欧元的巨额罚款作为制裁。如果组织未按要求保护数据主体的权益、做好相关记录，或未将其违规行为通知监管机关和数据主体，或未进行数据保护影响评估或者未按照规定配合认证，或未委派数据保护官或欧盟境内代表，则可能被处以1000万欧元或其全球年营业额2%（两者取其高）的罚款。

如果发生了更为严重的侵犯个人数据安全的行为，如未获得客户同意处理数据，或核心理念违反“隐私设计”要求，或违反规定将个人数据跨境传输，或违反欧盟成员国法律规定的义务等，组织有可能面临最高 2000 万欧元或组织全球年营业额的 4%（两者取其高）的巨额罚款。

**组织可向其内部通报GDPR处罚规则，进一步提升安全意识。**



# PRIVACY POLICY

## PRIVACY POLICY

Our Privacy Policy was updated on 25 April 2018 and will take effect on 25 May 2018. If you do not agree with these changes and no longer wish to use our services, you may cancel your account by emailing us at privacy@xiaomi.com. We h revamped the Privacy Policy front and back so that from this date onwards, this Privacy Policy can provide privacy details on how we manage your personal information for all Xiaomi products and services, unless a separate privacy policy is provided for a specific Xiaomi product or service.

Please take a moment to familiarize yourself with our privacy practices and let us know if you have any questions.

# China's Cybersecurity Law

- effective on 1 June 2017

- Purposes: [Art 1]

    - guarantee cybersecurity
    - safeguard cyberspace sovereignty
    - safeguard national security and public interest
    - protect lawful rights and interests of citizens, legal persons and other organisations
    - promote sound development of economic and social informatisation (信息化)

# China’s Cybersecurity Law

## Scope of Application:

- apply to the construction, operation, maintenance and use of the network, and the supervision and administration of cybersecurity within China [Art 2]
- mainly regulate network operators, i.e. the owners and administrators of the network as well as network service providers. [Art 76(3)]
  - not limited to technology companies e.g. a financial institute which uses computer network in its operation is a network operator
- protect personal information

# China’s Cybersecurity Law - Requirements

## Data Collection & Use:

- where personal information is collected, notify users and obtain their consent  [Art 22]
- follow principles of legality, rightfulness and necessity during collection and use; explicitly indicate the purposes, means and scope of collection and use [Art 41]
- do not collect personal information irrelevant to services provided [Art 41]
- do not collect or use personal information in violation of any law or administrative regulation or agreement of both parties [Art 41]

# China's Cybersecurity Law - Requirements

## Data Accuracy & Record Retention:

- not tamper with personal information collected [Art 42]

- take technical measures to monitor and record the status of network operation and cybersecurity incidents, and preserve weblogs for not less than 6 months [Art 21(3)]

# China's Cybersecurity Law - Requirements

## Data Security & Breach Notification:

- strictly keep confidential users' personal information collected, and establish and improve the system for information protection [Art 40]

- do not damage personal information collected, and take technical measures and other necessary measures to ensure security of personal information collected, and prevent information leakage, damage and loss [Art 42]

- where personal information has been or is likely to be divulged, damaged or lost, take remedial measures, inform users, and report to regulatory authority [Art 42]

# China’s Cybersecurity Law - Requirements

## Data Localisation:

- personal information and important data collected and produced by operators of critical information infrastructure (CII) during their operations within China shall be stored within China [Art 37]

- if CII operators need to provide such information and data to overseas parties due to business requirements, they shall conduct security assessment according to the measures developed by the Cyberspace Administration of China (CAC) and relevant departments of State Council, unless otherwise prescribed [Art 37]

- other network operators are encouraged to voluntarily participate in the CII protection system [Art 31]

# China’s Cybersecurity Law - Sanctions

## Possible sanctions for a breach:

- corrective action
- warning
- confiscate illegal income,
- fine between 1 and 10 times of illegal income
- if no illegal income, impose fine < RMB 1 million
- impose fine between RMB 10,000 and 100,000 on directly responsible person
- in serious cases, suspend or cease business operation for rectification, or close down website, or revoke business permit or license [Arts 64 & 66]

*[Depending on the graveness of the case, the above sanctions may be applied simultaneously ]*

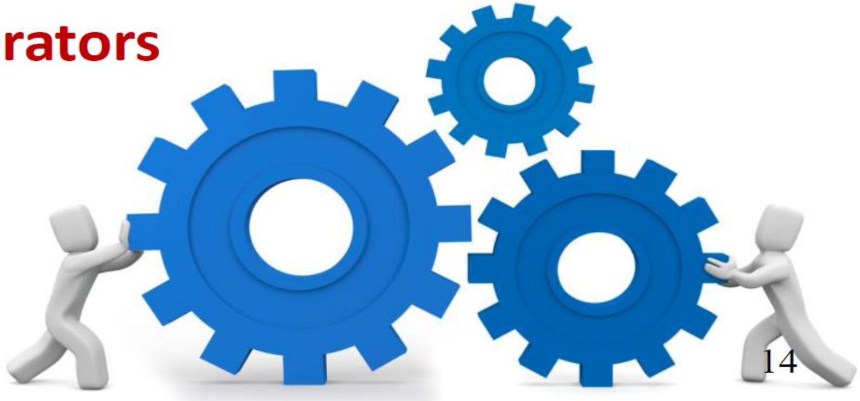
# “Measures for Security Assessment of Cross-Border Transfer of Personal Information and Important Data” 《個人信息和重要數據出境安全評估辦法 》

- Purposes: Elaborate the requirements of security assessment for cross-border data transfer under the Cybersecurity Law

- Draft Measures first issued in April 2017, and revised in May 2017

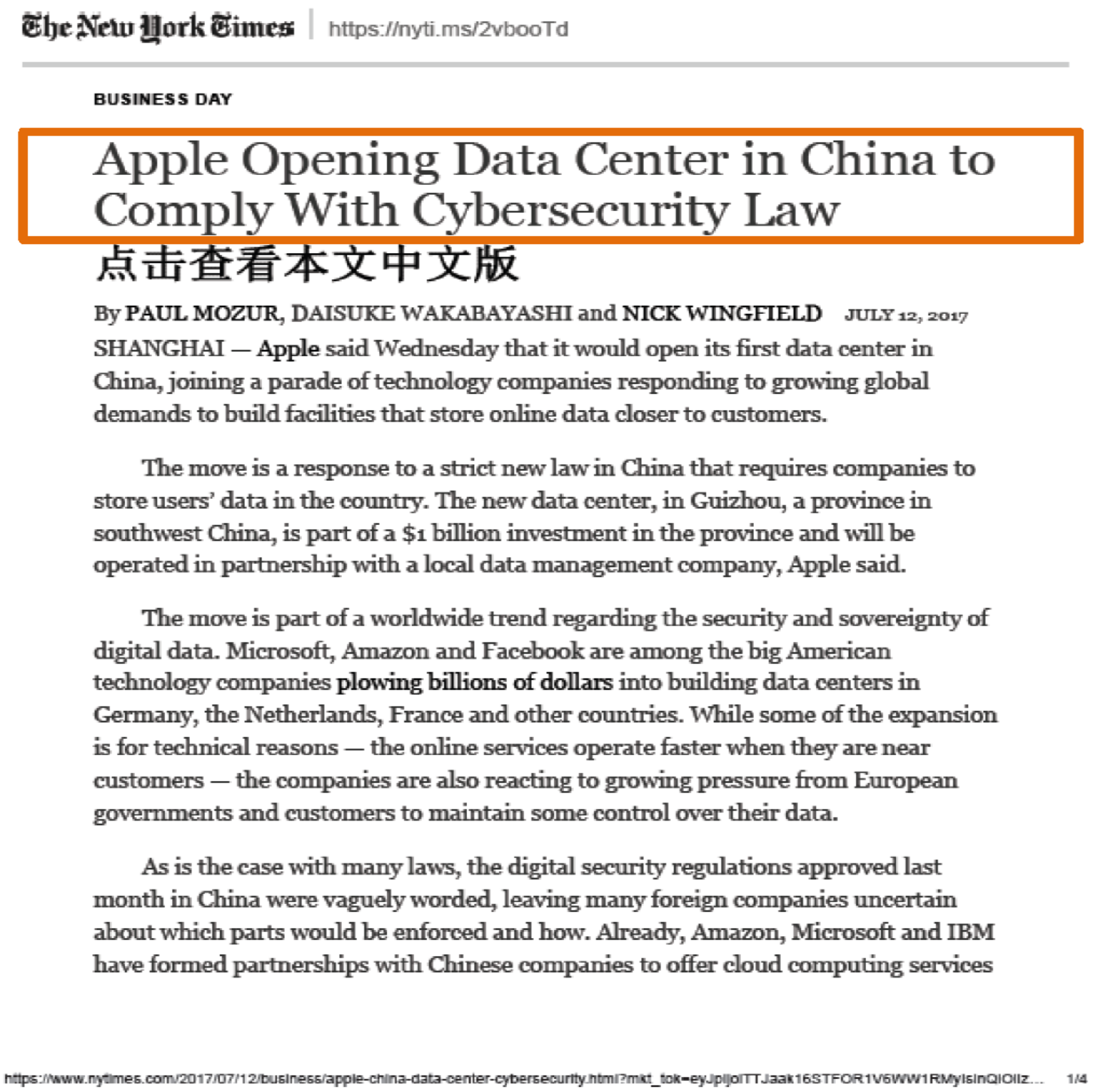
- Final version of the Measure is not yet available

- According to the latest draft, all network operators (not only CII operators) should conduct security assessment before transfer of personal information and important data to a place outside mainland of China



# “Guidelines for Data Cross-Border Transfer Security Assessment”
# 《數據出境安全評估指南》

- **Purposes: Provide substantive guidance on how to perform security assessment for transfer of personal information and important data to a place outside mainland of China**
- **Latest draft of the Guidelines issued on 31 August 2017**
- **Final version of the Guidelines is not yet available**
- **Clarify the operations by:**
  - providing definition of domestic operation and data cross-border transfer
  - elaborating self-security assessment process
  - specifying the conditions, process and requirements of government assessments



The New York Times | https://nyti.ms/2vbooTd

BUSINESS DAY

## Apple Opening Data Center in China to Comply With Cybersecurity Law

点击查看本文中文版

By PAUL MOZUR, DAISUKE WAKABAYASHI and NICK WINGFIELD JULY 12, 2017

SHANGHAI — Apple said Wednesday that it would open its first data center in China, joining a parade of technology companies responding to growing global demands to build facilities that store online data closer to customers.

The move is a response to a strict new law in China that requires companies to store users' data in the country. The new data center, in Guizhou, a province in southwest China, is part of a $1 billion investment in the province and will be operated in partnership with a local data management company, Apple said.

The move is part of a worldwide trend regarding the security and sovereignty of digital data. Microsoft, Amazon and Facebook are among the big American technology companies plowing billions of dollars into building data centers in Germany, the Netherlands, France and other countries. While some of the expansion is for technical reasons — the online services operate faster when they are near customers — the companies are also reacting to growing pressure from European governments and customers to maintain some control over their data.

As is the case with many laws, the digital security regulations approved last month in China were vaguely worded, leaving many foreign companies uncertain about which parts would be enforced and how. Already, Amazon, Microsoft and IBM have formed partnerships with Chinese companies to offer cloud computing services

https://www.nytimes.com/2017/07/12/business/apple-china-data-center-cybersecurity.html?mkt_tok=eyJpIjoiTTJaak16STFOR1V6WW1RMyIsInQiOiz... 1/4

**Apple Opening Data Centre in China to Comply With Cybersecurity Law**

**Source: The New York Times 12 July 2017**

# The Interpretation of Various Issues Concerning Application of Law in Handling Crimes of Infringing upon Citizen's Personal Data

# 《關於辦理侵犯公民個人信息刑事案件適用法律若干問題的解釋》

Jointly issued by the PRC Supreme People's Court and the Supreme People's Procuratorate in May 2017

## Clarifying and expanding the definition of personal information for the purpose of the Criminal Law

- personal information means any information, individually or combined with other information, that can identify a specific individual
- new examples include account password, financial position and geo-location data

## Make it clear that it is a crime to publicise personal information without consent

- publication of personal information through the Internet or other channels without consent fall within the crime of infringement of personal information under the Criminal Law, even if the information is legally obtained *(partly for addressing the increasing trend of "cyber manhunt" （人肉搜索））*

# Location Privacy: A Real-world Example

# Location Privacy is Gaining An Increasing Attention!

- A trace/location tells much about the individual’s habits, interests, activities, and relationships.

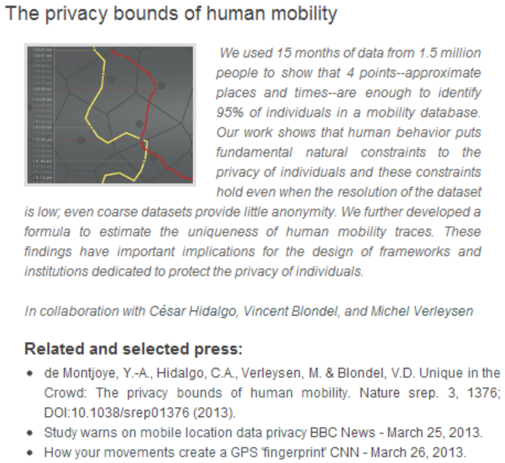
  --*Quantifying Location Privacy*, Oakland'11

- In a mobility database consisting of 1.5 million people, 4 temporal-spatio points are enough to identify 95% of individuals.

  -- *Unique in the Crowd: The privacy bounds of human mobility*. **Nature**. 2013

- Suggest offering a “Do Not Track” mechanism for smartphone users

  --*Mobile Privacy Disclosures: Building Trust Through Transparency*, **Federal Trade Commission (FTC), 2013**


The privacy bounds of human mobility

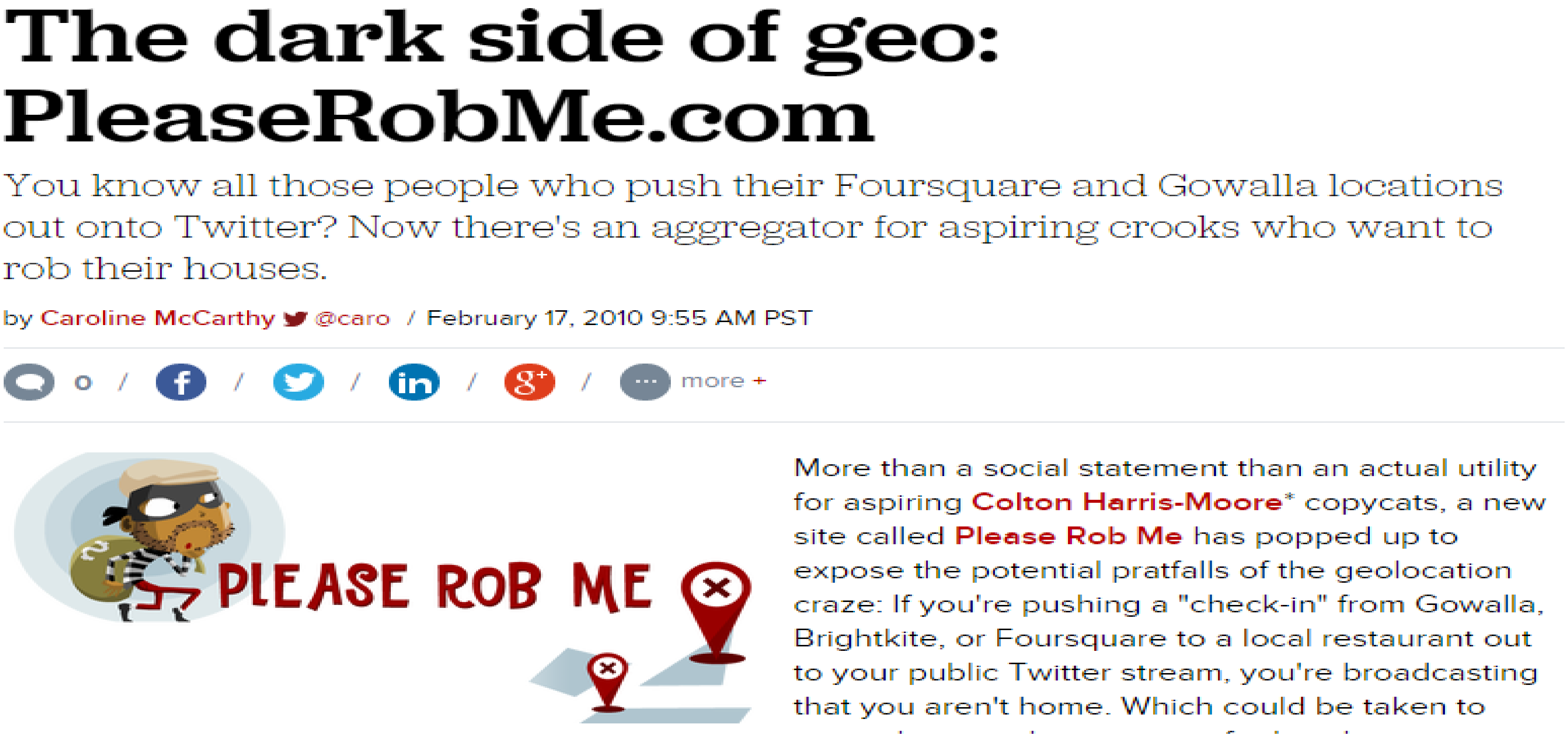
*We used 15 months of data from 1.5 million people to show that 4 points--approximate places and times--are enough to identify 95% of individuals in a mobility database. Our work shows that human behavior puts fundamental natural constraints to the privacy of individuals and these constraints hold even when the resolution of the dataset is low; even coarse datasets provide little anonymity. We further developed a formula to estimate the uniqueness of human mobility traces. These findings have important implications for the design of frameworks and institutions dedicated to protect the privacy of individuals.*

*In collaboration with César Hidalgo, Vincent Blondel, and Michel Verleysen*

**Related and selected press:**

- de Montjoye, Y.-A., Hidalgo, C.A., Verleysen, M. & Blondel, V.D. Unique in the Crowd: The privacy bounds of human mobility. Nature srep. 3, 1376; DOI:10.1038/srep01376 (2013).
- Study warns on mobile location data privacy BBC News - March 25, 2013.
- How your movements create a GPS 'fingerprint' CNN - March 26, 2013.

# Location Privacy: Rob Me



## The dark side of geo: PleaseRobMe.com

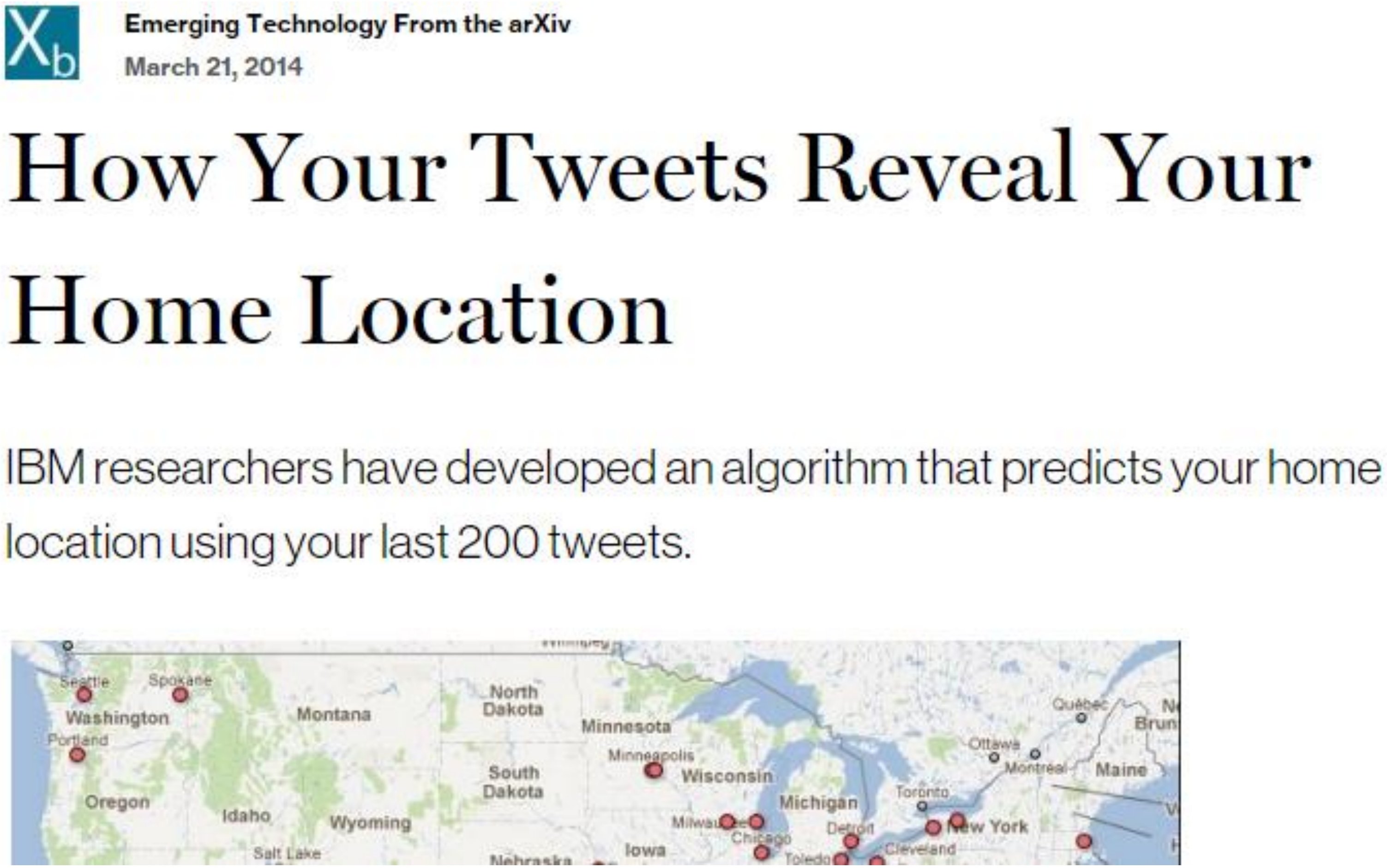
You know all those people who push their Foursquare and Gowalla locations out onto Twitter? Now there's an aggregator for aspiring crooks who want to rob their houses.

by Caroline McCarthy @caro / February 17, 2010 9:55 AM PST

0 / / / / / more +

More than a social statement than an actual utility for aspiring Colton Harris-Moore* copycats, a new site called Please Rob Me has popped up to expose the potential pratfalls of the geolocation craze: If you're pushing a "check-in" from Gowalla, Brightkite, or Foursquare to a local restaurant out to your public Twitter stream, you're broadcasting that you aren't home. Which could be taken to

# Location Privacy Leaking Risk (MIT Tech Review 2014)



Emerging Technology From the arXiv

March 21, 2014

## How Your Tweets Reveal Your Home Location

IBM researchers have developed an algorithm that predicts your home location using your last 200 tweets.

# Location Privacy In Emerging Wireless Networks

- While in the past, mobility traces were only available to mobile phone carriers, the advent of smartphones and other means of data collection has made these broadly available.

- For example, Apple recently updated its privacy policy to allow sharing the spatio-temporal location of their users with "partners and licensees'

- Skyhook wireless is resolving 400 M user's WiFi location every day

- a third of the 25B copies of applications available on Apple's App Store access a user's geographic location

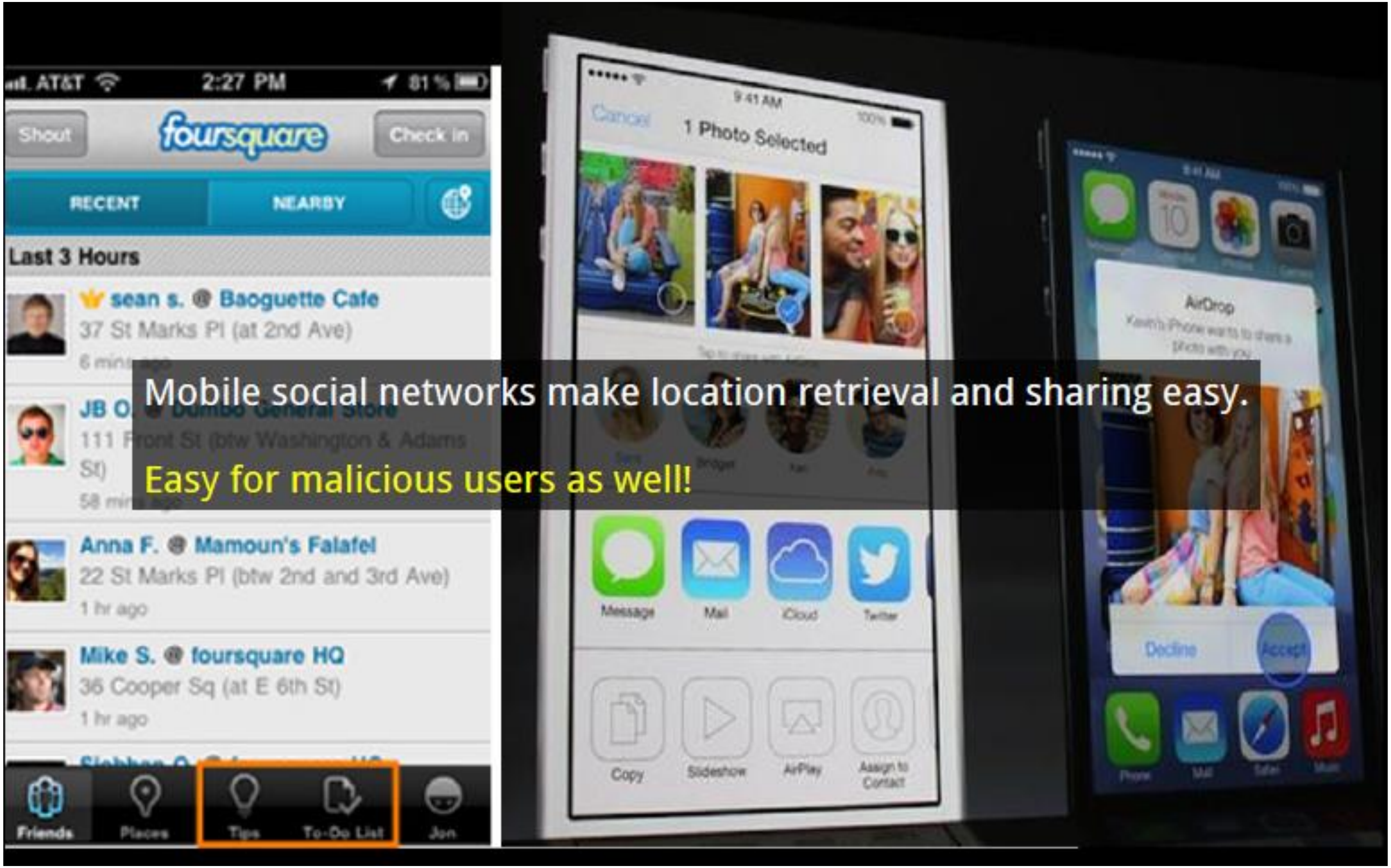
- the geo-location of, 50% of all iOS and Android traffic is available to ad networks

# All Your Location Are Belong to Us: Breaking Mobile Social Networks for Automated User Location Tracking

Muyuan Li, Haojin Zhu, Zhaoyu Gao, Si Chen, Kui Ren, Le Yu, Shangqian Hu, All Your Location are Belong to Us: Breaking Mobile Social Networks for Automated User Location Tracking, *ACM MobiHoc'14*, Main Conference, 2014.

# Outline

- **Introduction**
- Related Work
- Overview
- Location Privacy in LBSN
- FreeTrack
- Evaluation
- Performance Optimization
- A Demo
- Conclusion



sean s. @ Baoguette Cafe
37 St Marks Pl (at 2nd Ave)
6 min ago

JB O. @ Dumbo General Store
111 Front St (btw Washington & Adams St)
58 min ago

Anna F. @ Mamoun's Falafel
22 St Marks Pl (btw 2nd and 3rd Ave)
1 hr ago

Mike S. @ foursquare HQ
36 Cooper Sq (at E 6th St)
1 hr ago

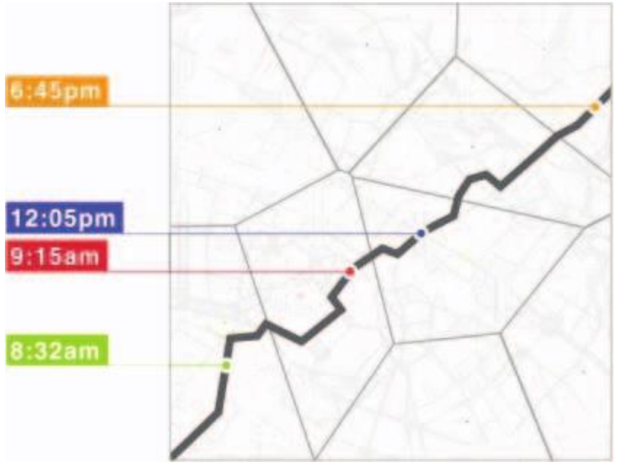
Mobile social networks make location retrieval and sharing easy.

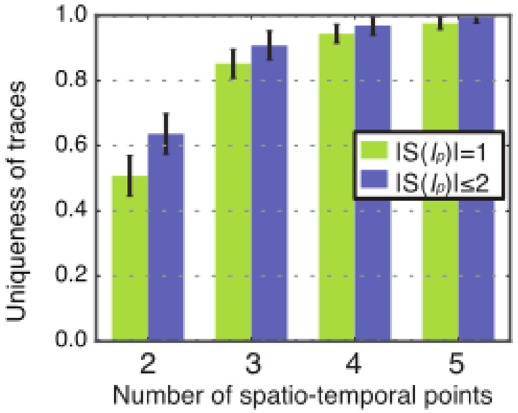
Easy for malicious users as well!

# Location can even reveal your identity

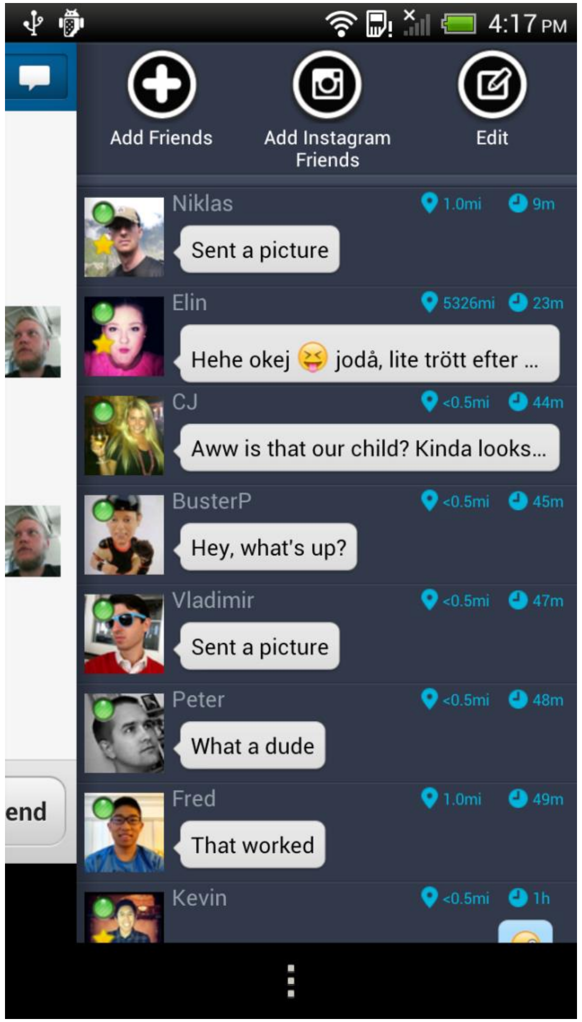
Unique in the Crowd: The privacy bounds of human mobility (Nature, 2013)

- Analyzed millions of traces
- Make re-identifications
- Amazingly, 95% of people can be re-identified with 4 or less points!





# Location-based Mobile Social Networks



Niklas 1.0mi 9m
Sent a picture

Elin 5326mi 23m
Hehe okej 😝 jodå, lite trött efter ...

CJ <0.5mi 44m
Aww is that our child? Kinda looks...

BusterP <0.5mi 45m
Hey, what's up?

Vladimir <0.5mi 47m
Sent a picture

Peter <0.5mi 48m
What a dude

Fred 1.0mi 49m
That worked

Kevin <0.5mi 1h

**Typical examples:** Wechat, Skout, Momo

**Common Feature**

- Enable location-based social discovery
- Display the relative distance with your neighbors

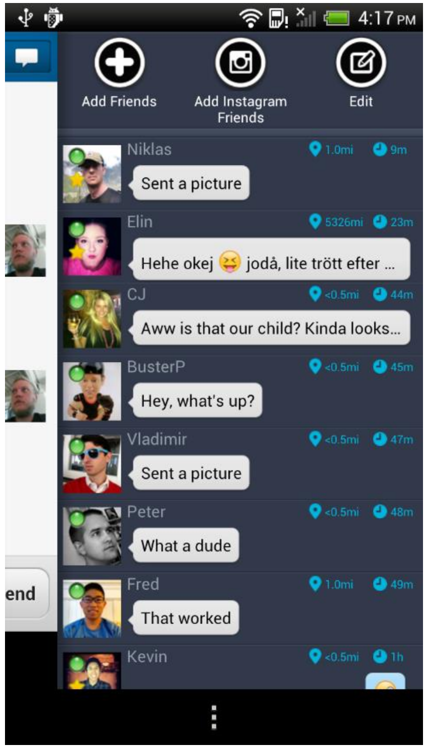
**Super Popularity of LBSN**

- **Wechat**: 300 millions in China
  (60 million international users)
- **Momo**: 30 millions
- **Skout**: 5 millions in north America
- **MiTalk**: 20 millions

# Best Practice Location Protection in LBSNs

- **Industry standard method to protect users location privacy**

  (It is claimed that NEVER reveal users' exact locations)



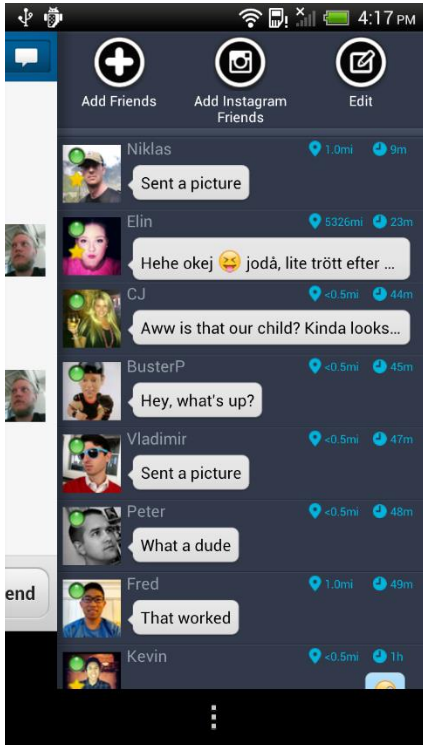
| | | | |
|---|---|---|---|
| Niklas | 1.0mi | 9m | Sent a picture |
| Elin | 5326mi | 23m | Hehe okej 😝 jodå, lite trött efter ... |
| CJ | <0.5mi | 44m | Aww is that our child? Kinda looks... |
| BusterP | <0.5mi | 45m | Hey, what's up? |
| Vladimir | <0.5mi | 47m | Sent a picture |
| Peter | <0.5mi | 48m | What a dude |
| Fred | 1.0mi | 49m | That worked |
| Kevin | <0.5mi | 1h | |

1. Relative Distance Only: showing the distance rather than your exact locations

# Best Practice Location Protection in LBSNs

- **Industry standard method to protect users location privacy**

  (It is claimed that NEVER reveal users' exact locations)



Niklas 1.0mi 9m
Sent a picture

Elin 5326mi 23m
Hehe okej 😝 jodå, lite trött efter ...

CJ <0.5mi 44m
Aww is that our child? Kinda looks...

BusterP <0.5mi 45m
Hey, what's up?

Vladimir <0.5mi 47m
Sent a picture

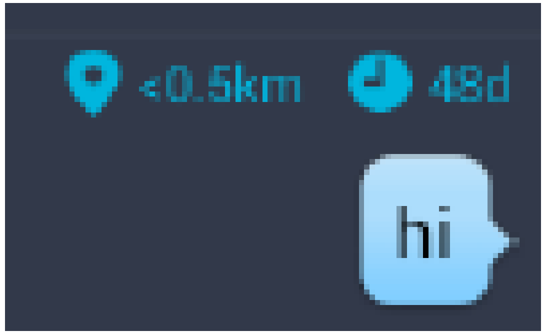
Peter <0.5mi 48m
What a dude

Fred 1.0mi 49m
That worked

Kevin <0.5mi 1h

1. Relative Distance Only :

2. **Setting the Minimum Accuracy Limit:**
   (0.5 mile for Skout, 100 m for Wechat)



<0.5km 48d

hi

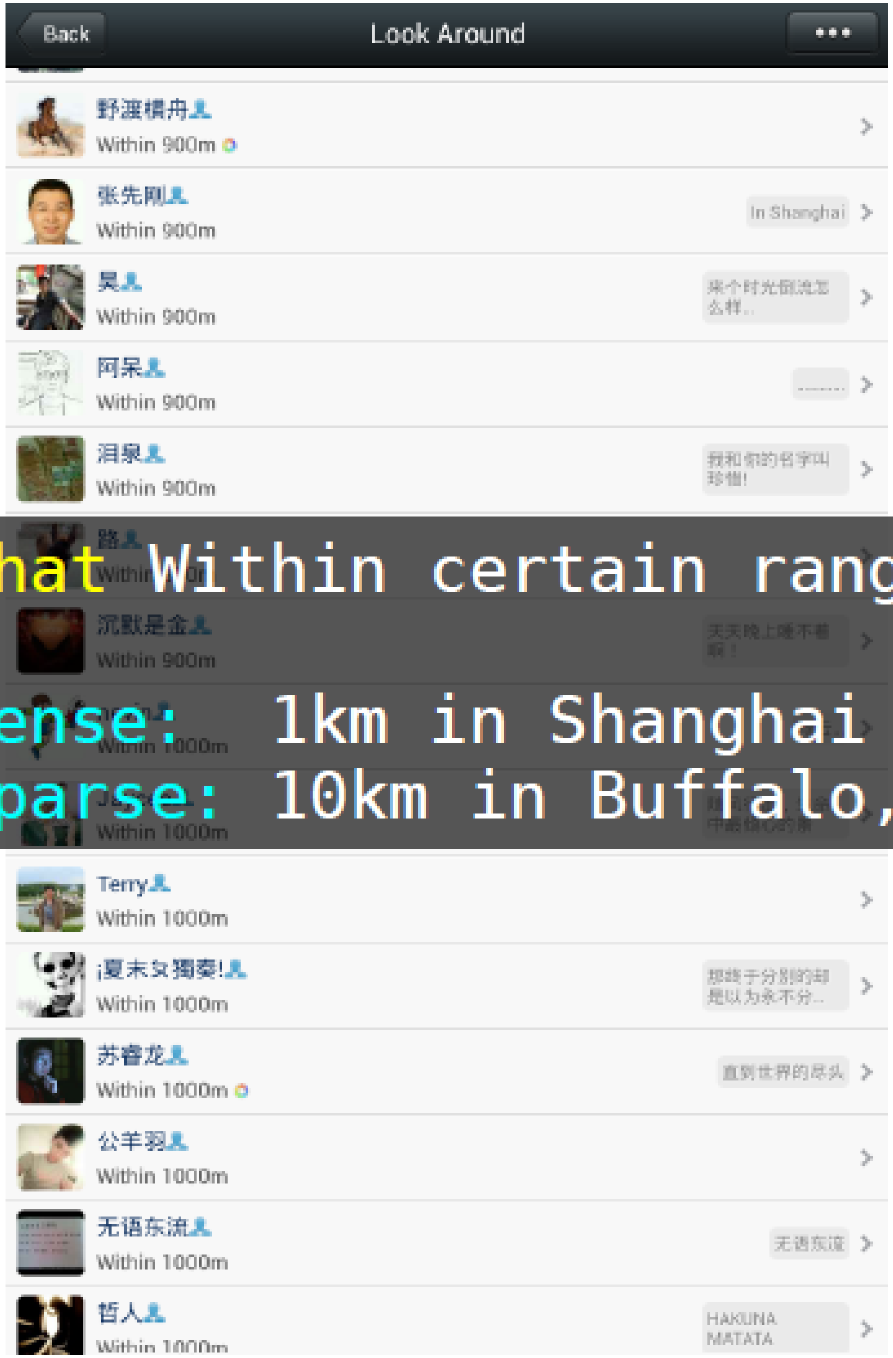
# Best Practice Location Protection in LBSNs

- **Industry standard method to protect users location privacy**

  (It is claimed that NEVER reveal users' exact locations)

1. Relative Distance Only :
2. Setting the Minimum Accuracy Limit:
3. **Setting the Localization Coverage Limits:**

   (restrict the users' localization capability to a specific region.)



Back — Look Around — •••

野渡横舟
Within 900m

张先刚
Within 900m
In Shanghai

吴
Within 900m
来个时光倒流怎么样..

阿呆
Within 900m
.......

泪泉
Within 900m
我和你的名字叫珍惜!

路
Within 900m

沉默是金
Within 900m
天天晚上睡不着啊！

Within 1000m

Within 1000m

Terry
Within 1000m

¡夏末女獨奏!
Within 1000m
那终于分别的却是以为永不分...

苏睿龙
Within 1000m
直到世界的尽头

公羊羽
Within 1000m

无语东流
Within 1000m
无语东流

哲人
Within 1000m
HAKUNA MATATA

Wechat Within certain range

* Dense:  1km in Shanghai
* Sparse: 10km in Buffalo, NY

# Summary of Location Privacy Protection Approaches in LBSNs

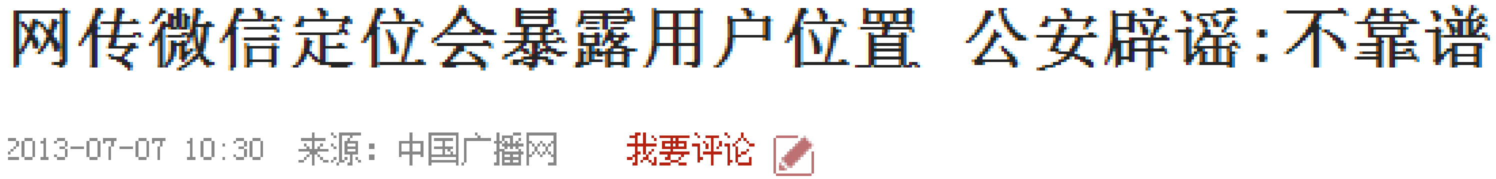
- Momo: (Strategy I: only showing the relative distances)
- Skout: (Strategy I & II: shows the distance & enforces the minimum localization limit)
- Wechat: (Strategy I & II & III)

Are these "seemly" safe privacy protection approaches really safe in reality?

# Misunderstanding of the Public

- LBSN users are willing to share their locations because they trust these privacy protection method.
- A recent news about location privacy issue of Wechat

  Chinese police states that: it is impossible to figure out users exact locations by Wechat



网传微信定位会暴露用户位置 公安辟谣:不靠谱

2013-07-07 10:30 来源：中国广播网 我要评论

Our work shows that LBSN users are facing a big risk of leaking their very sensitive location information

# Our Contributions

- We identify new location privacy issues in mobile social networks (LBSNs).

- Targeting at 3 popular location-based social network applications: Wechat, Skout and Momo and performing evaluations with 30 volunteers from China, Japan and United States for 3 weeks.

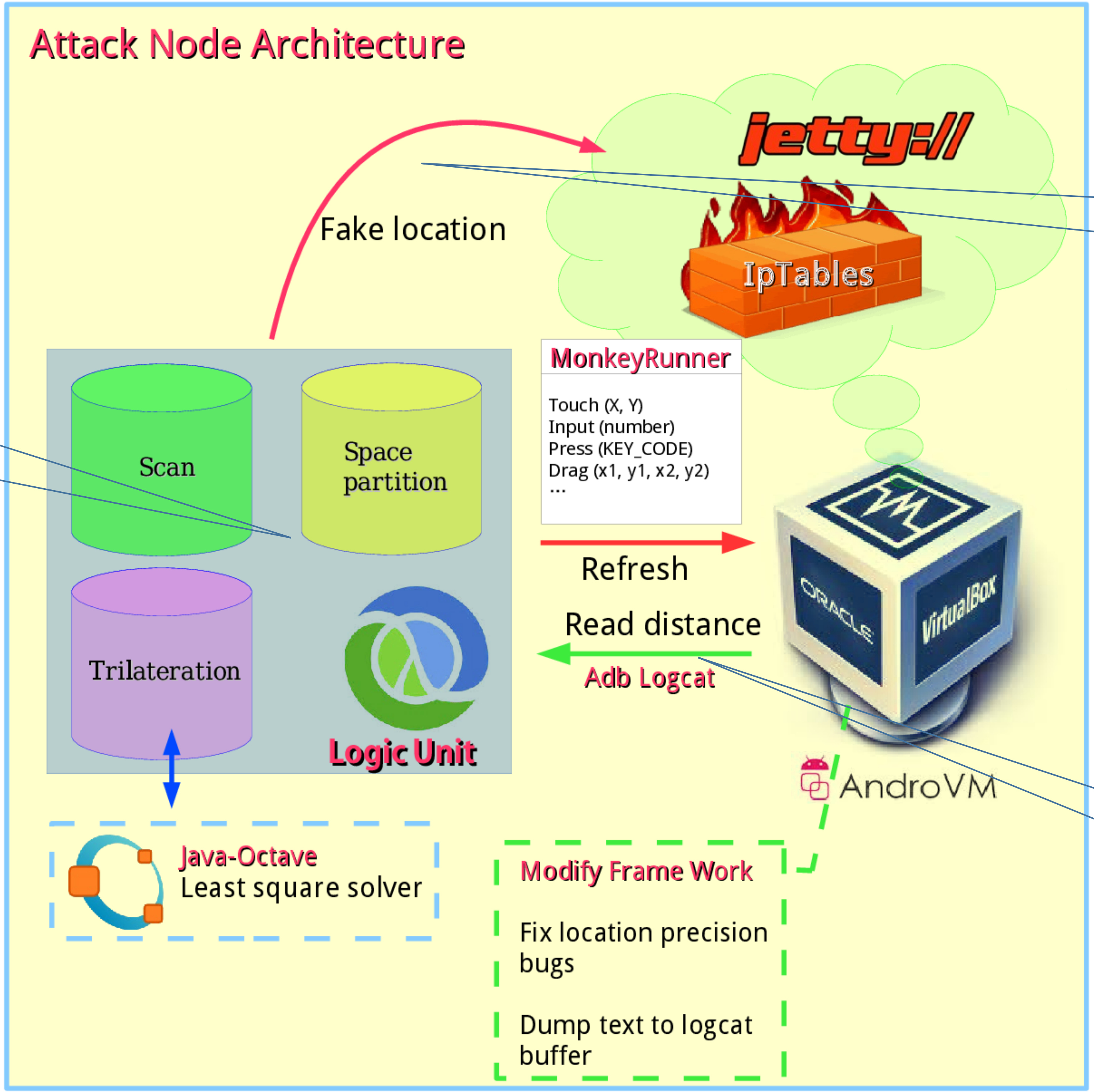
- We show that:
  - Users' location privacy is totally compromised
  - locate users with a very high accuracy
  - long-term tracking is easy to achieve
  - high possibility to reveal top locations

# FreeTrack: Automated User Location Tracking System

- Target: Obliviously obtain user location

- Attacker capability:
  - Need user ID (Not necessarily being friend, nor require your approval)
  - Only exploiting public available information
  - Conventional hardware
  - Do not need to modify applications

- Features:
  - Large coverage (Global tracking)
  - High accuracy
  - Recover top locations



## Attack Node Architecture

MonkeyRunner

Touch (X, Y)
Input (number)
Press (KEY_CODE)
Drag (x1, y1, x2, y2)
...

Java-Octave
Least square solver

Modify Frame Work

Fix location precision bugs

Dump text to logcat buffer

Three attack methodologies

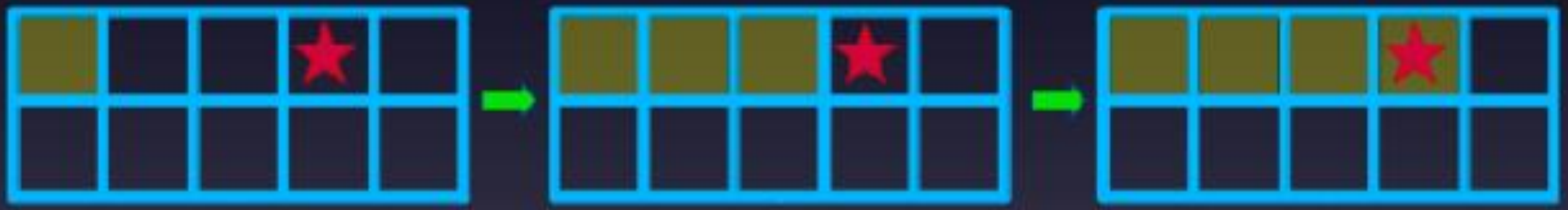
Setting the bogus anchor points

Automatic input/output fetch

# Attack Methodology

We utilize 3 types of attack

- Iterative trilateration attack
- Space partition attack
- Scan



# Trilateration

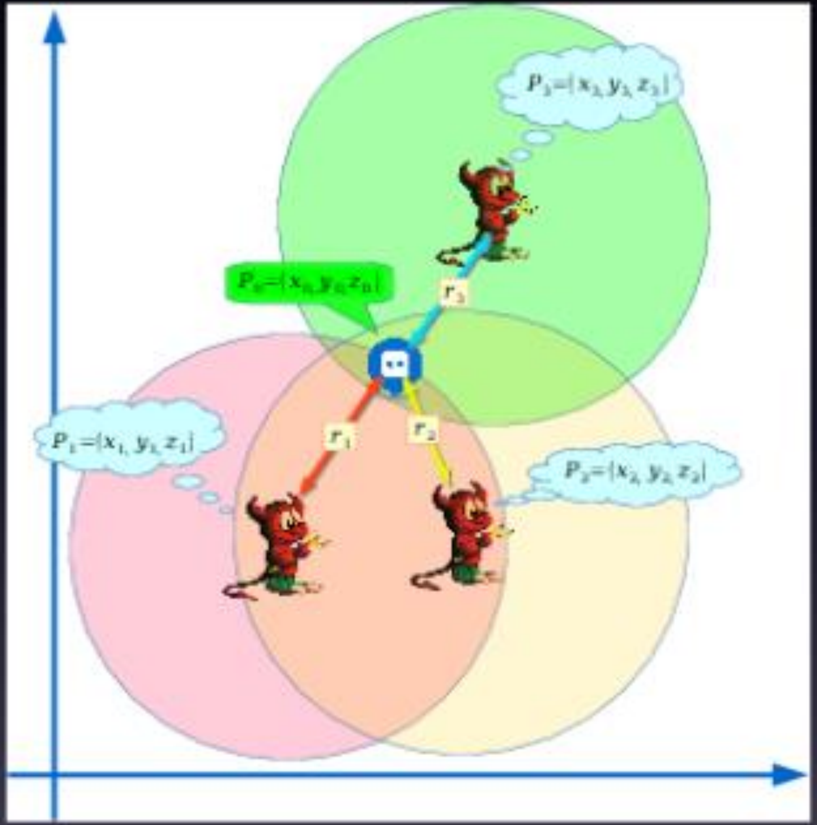
Relative distance
Minimal bounding
Within certain range

Givien 3 points $(x_i, y_i, z_i)$, trilateration finds a point $(x_0, y_0, z_0)$ that satisfy:

$$(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2 = R_i^2$$

Least square solution:

$$\sum_{i=1,2,3} \{(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2 - R_i^2\}^2$$



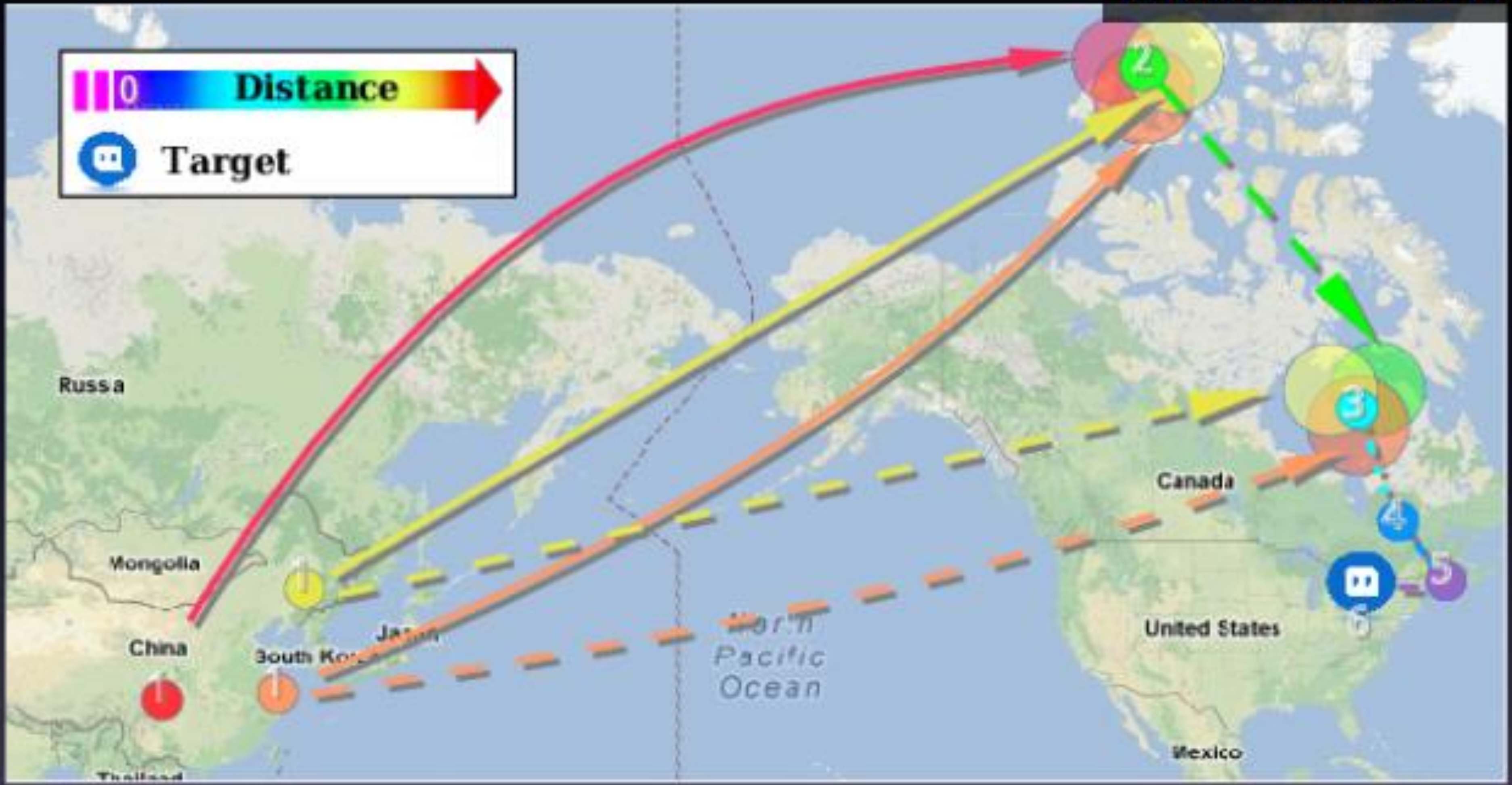
Trilateration

# Real Example of Trilateration Attack

Relative distance

Minimal bounding

Within certain range



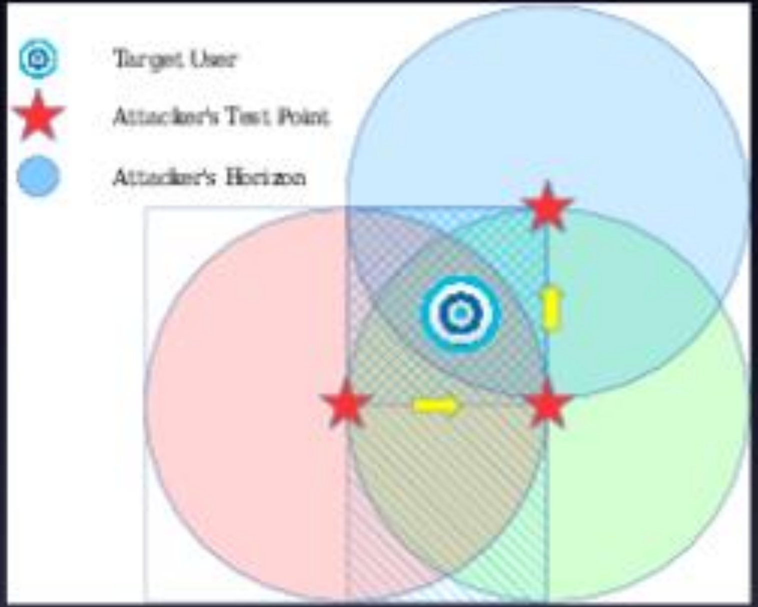
Trilateration Attack on Global Scale

# Space Partition Attack

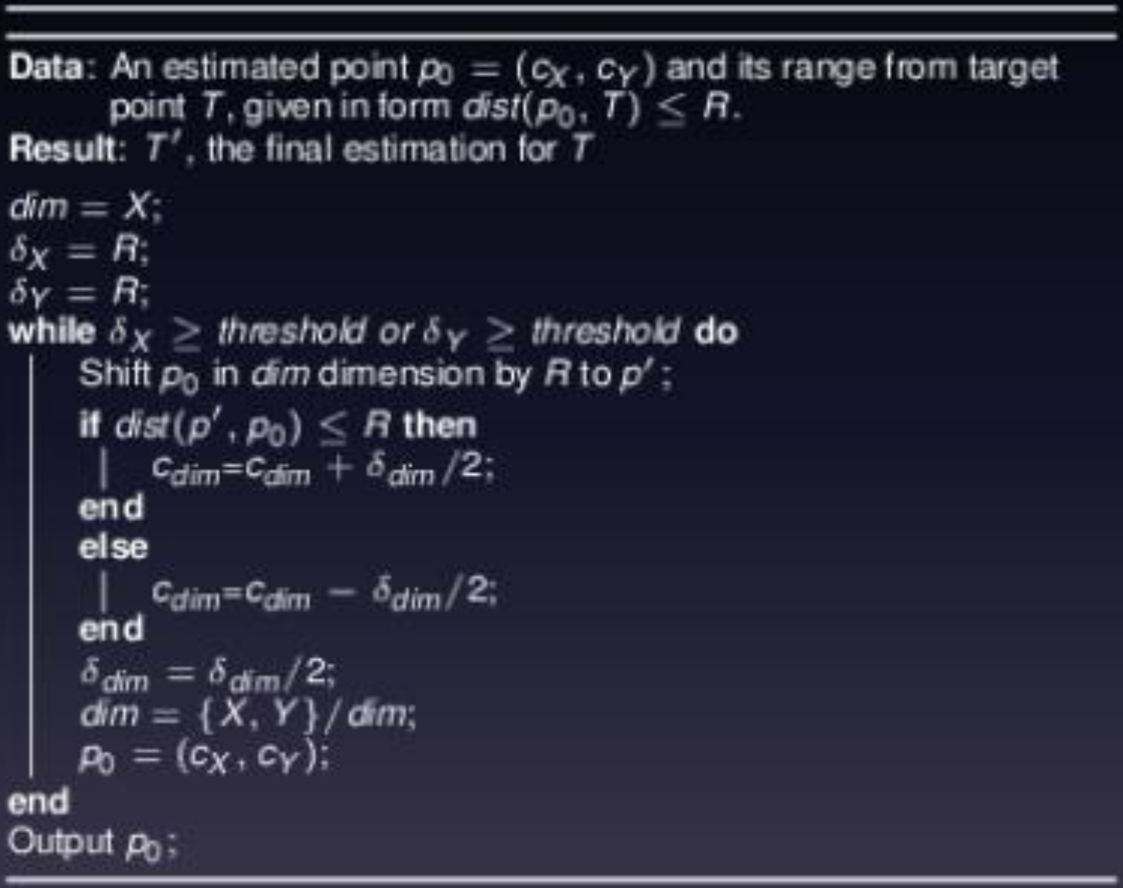
Relative distance

Minimal bounding

Within certain range

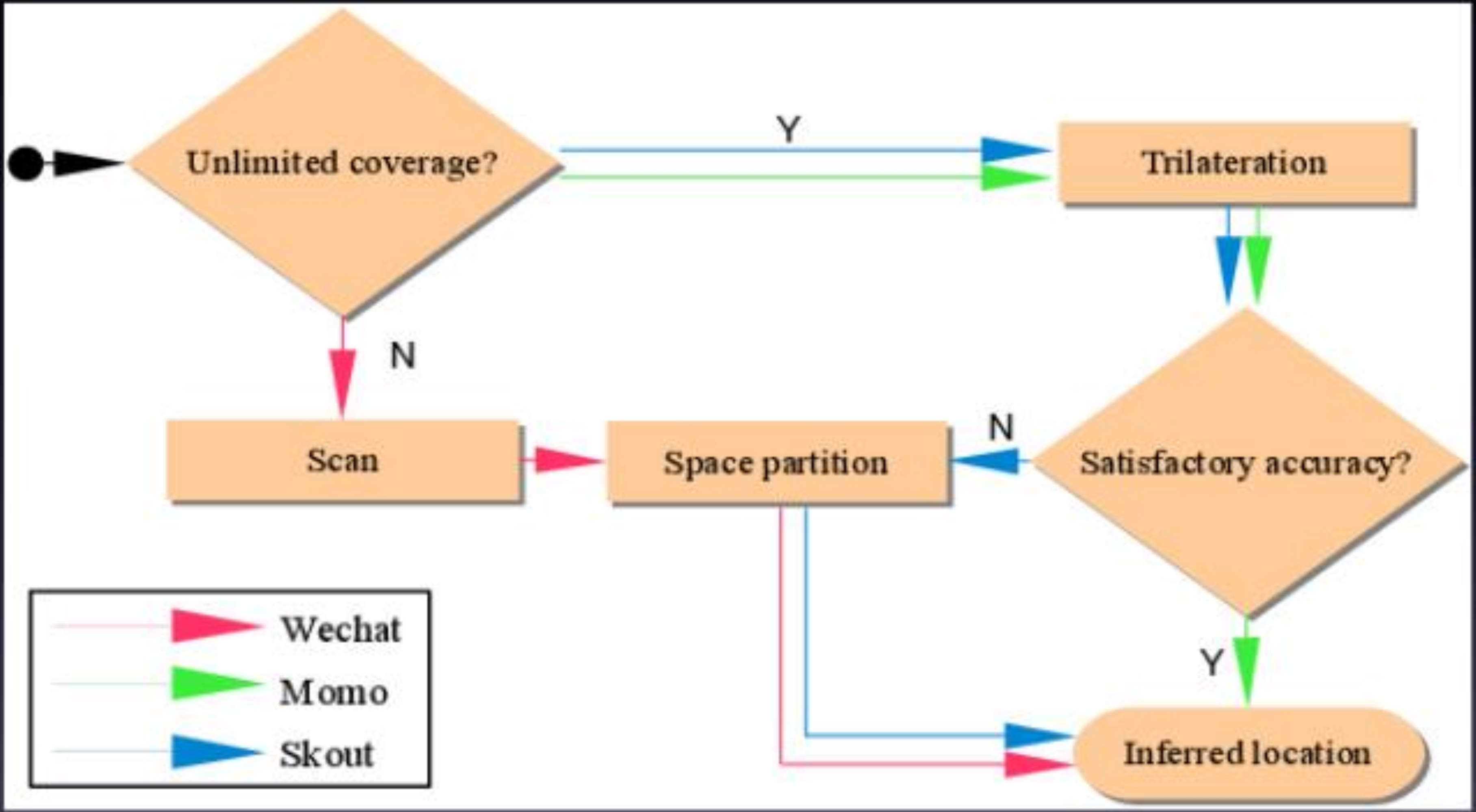


Space Partition Attack



```
Data: An estimated point $p_0 = (c_X, c_Y)$ and its range from target
      point $T$, given in form $dist(p_0, T) \leq R$.
Result: $T'$, the final estimation for $T$
$dim = X$;
$\delta_X = R$;
$\delta_Y = R$;
while $\delta_X \geq threshold$ or $\delta_Y \geq threshold$ do
    Shift $p_0$ in $dim$ dimension by $R$ to $p'$;
    if $dist(p', p_0) \leq R$ then
        $c_{dim}=c_{dim} + \delta_{dim}/2$;
    end
    else
        $c_{dim}=c_{dim} - \delta_{dim}/2$;
    end
    $\delta_{dim} = \delta_{dim}/2$;
    $dim = \{X, Y\}/dim$;
    $p_0 = (c_X, c_Y)$;
end
Output $p_0$;
```

# Put It Altogether

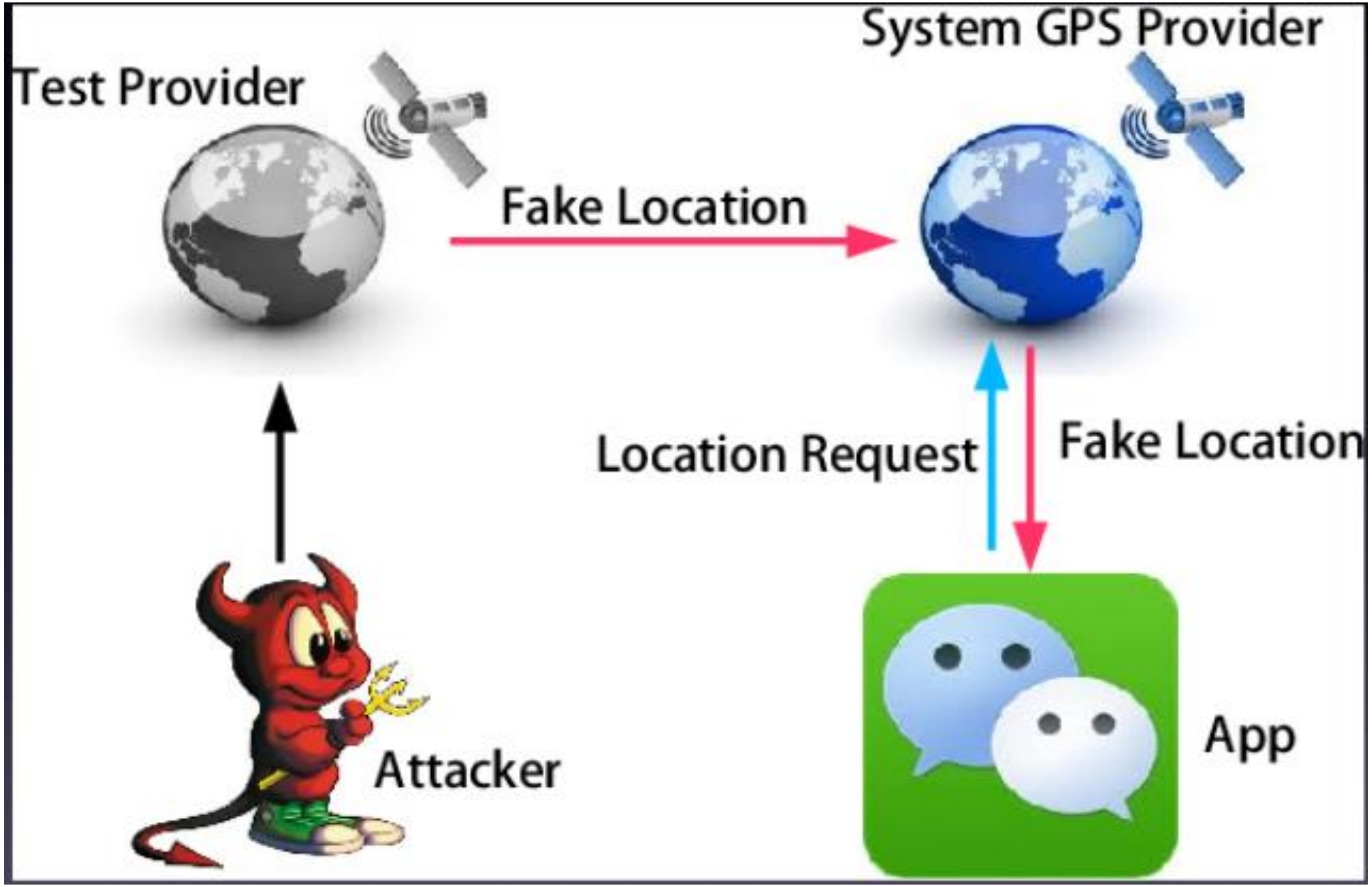


Logic Flow of the Attack

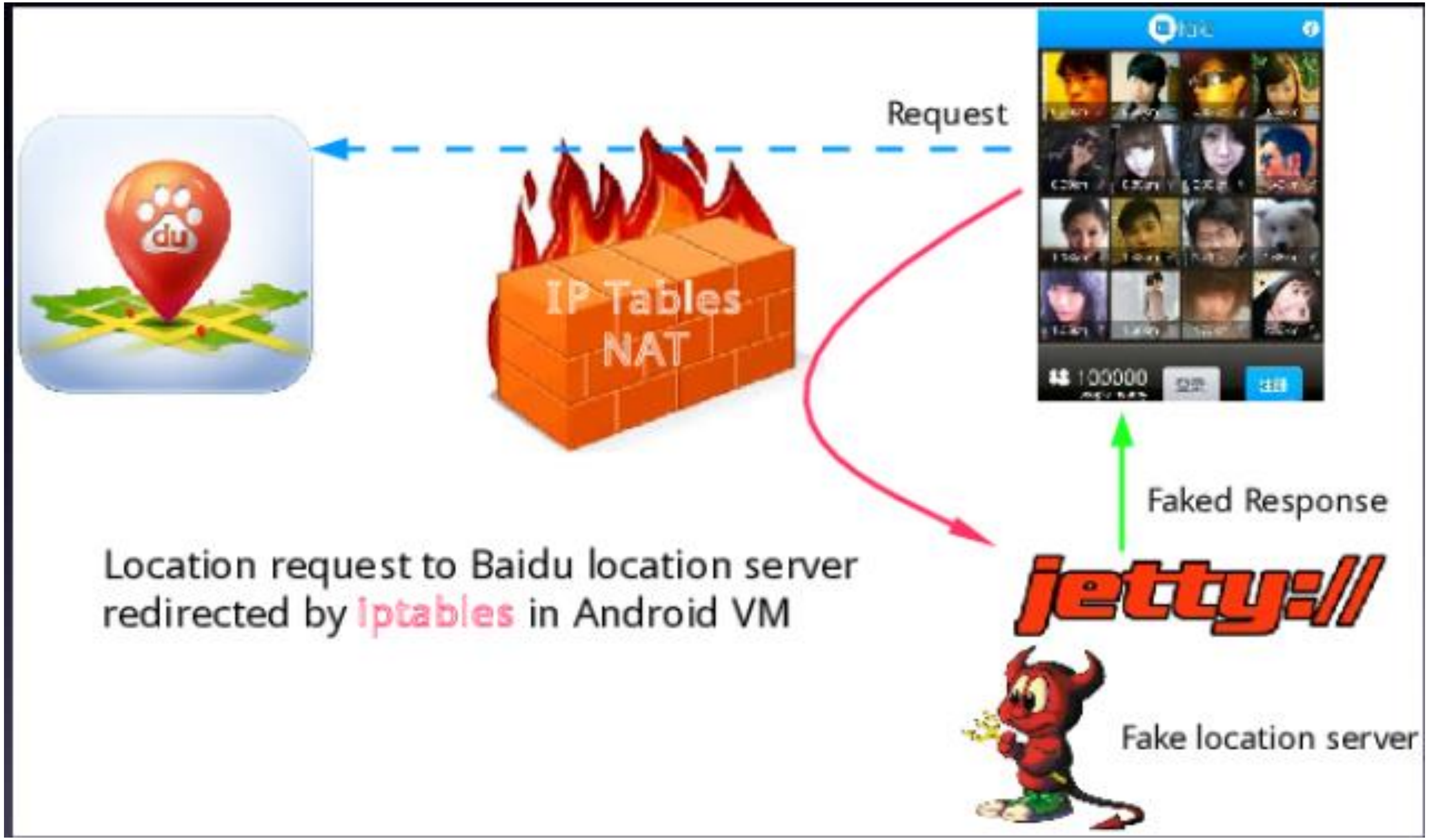
# Key Component: Generate fake GPS location to set bogus anchor points

- Way 1: Intercept network traffic
- Way 2: Utilize test location provider

# Add Test Location Provider



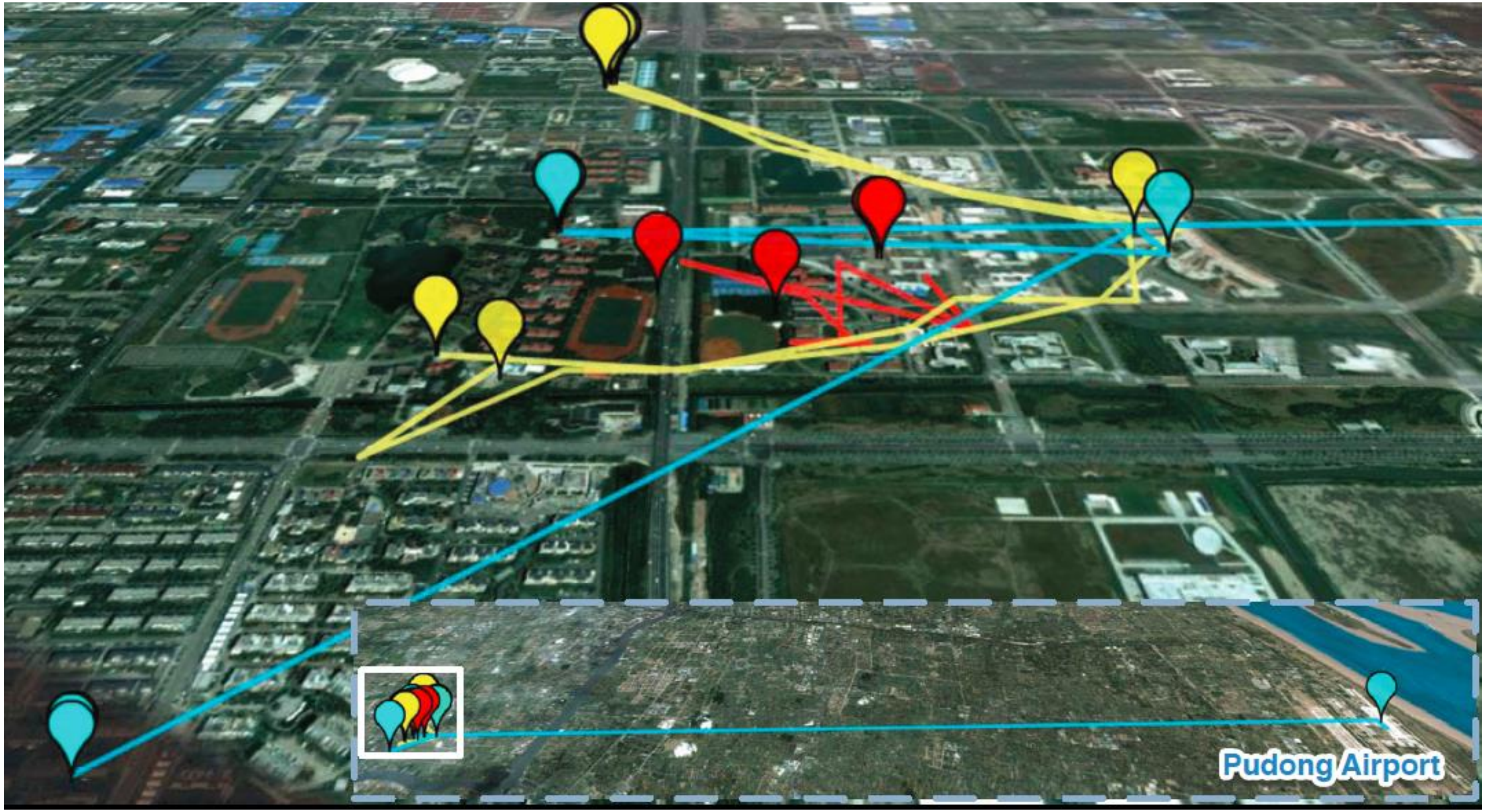
# Redirect Network Traffic


Request
IP Tables
NAT
Faked Response
jetty://
Location request to Baidu location server
redirected by iptables in Android VM
Fake location server
100000

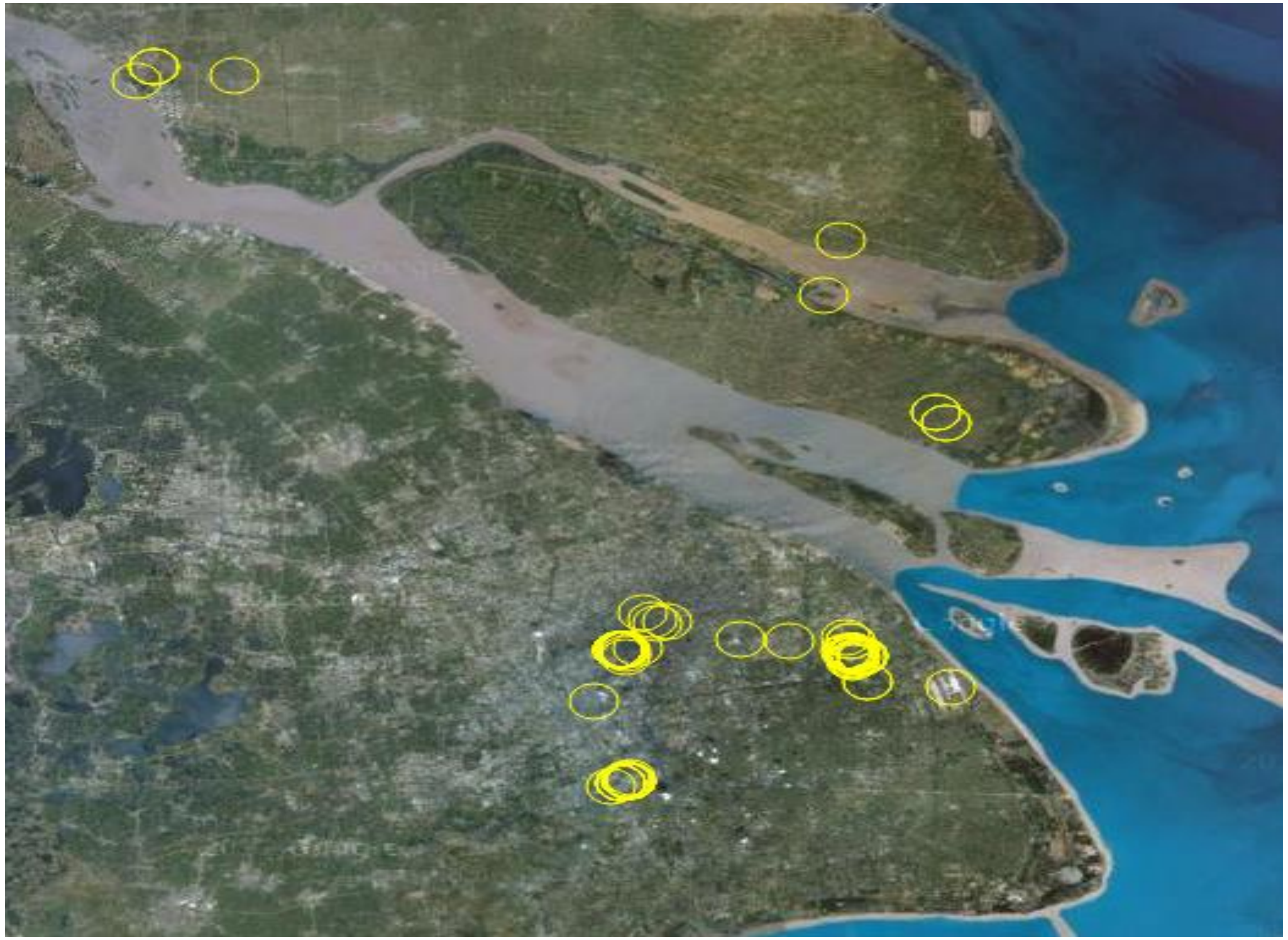
# Real-world Tracking

- Experiment Setup:
  - 30 volunteers from United States, China, and Japan
  - 3 apps: Wechat, Momo and Skout
  - Global Tracking (Momo and Skout ), covering SJTU campus (wechat)
  - 3 weeks tracking

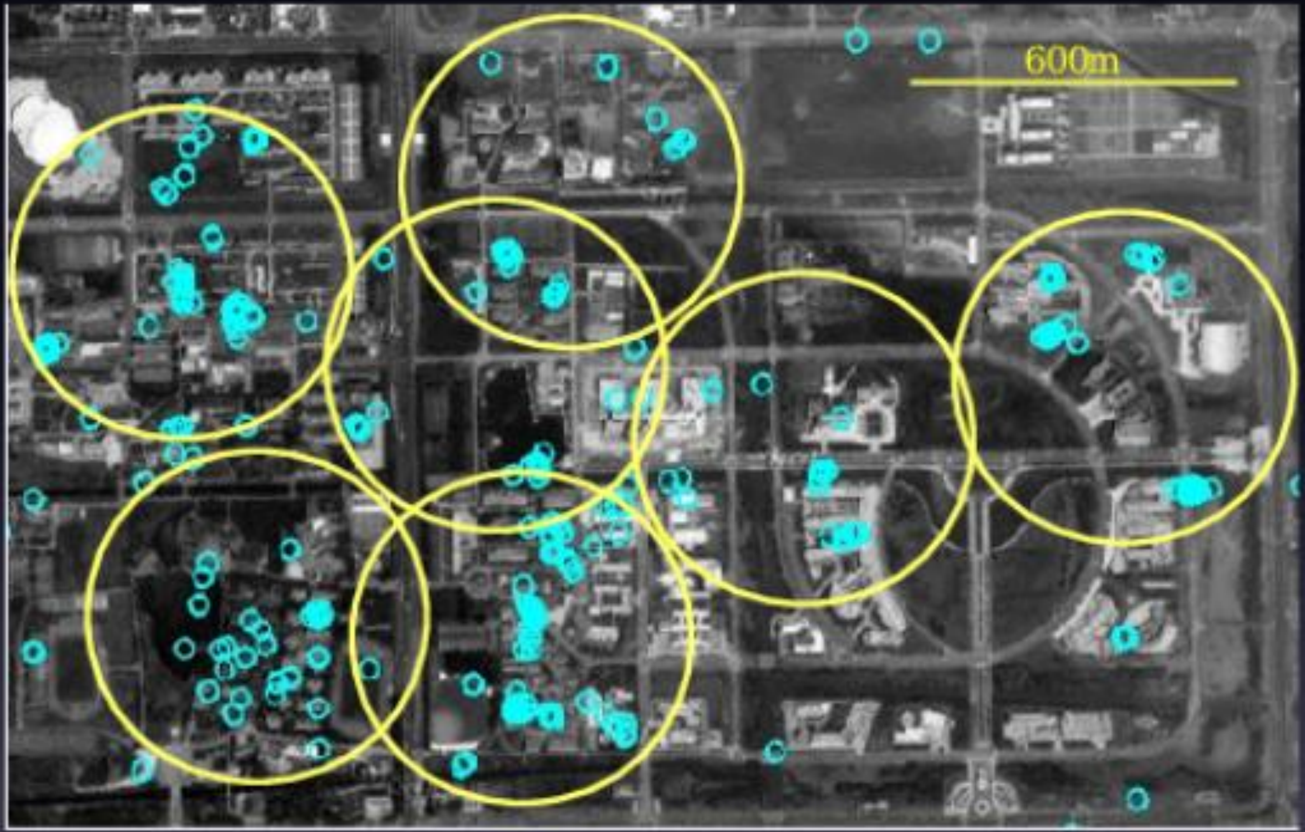
# Three Real-world Traces and Inferred Locations



# One volunteer’s three weeks’ trace

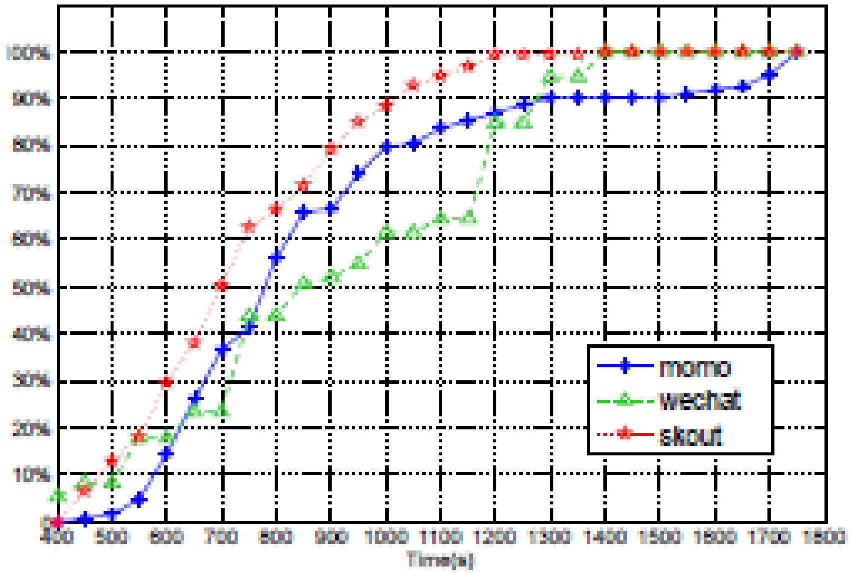


# Optimize Efficiency with Side Information

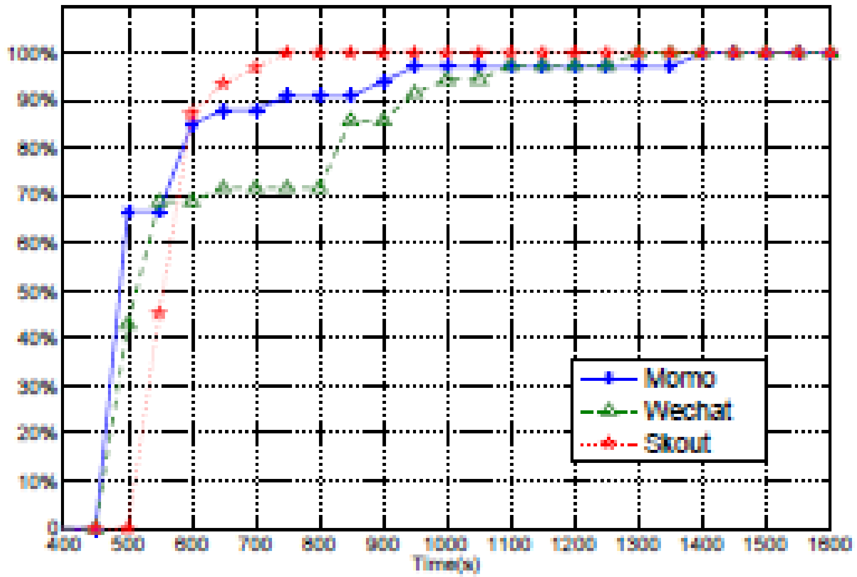


Location popularity as index
(Anonymous trace for 40 random users in 1 week)

# Attack Performance Enhancement by Using Side Information

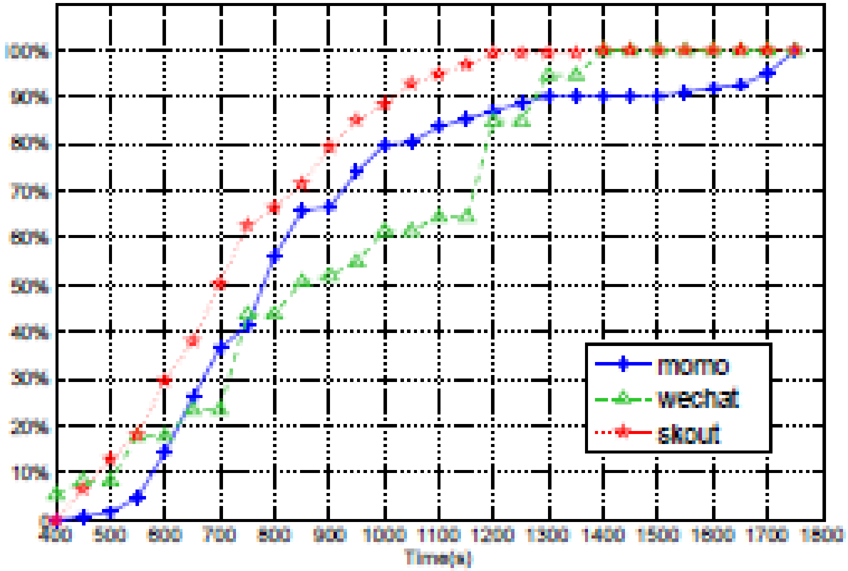


(a) Localization inference time on different apps

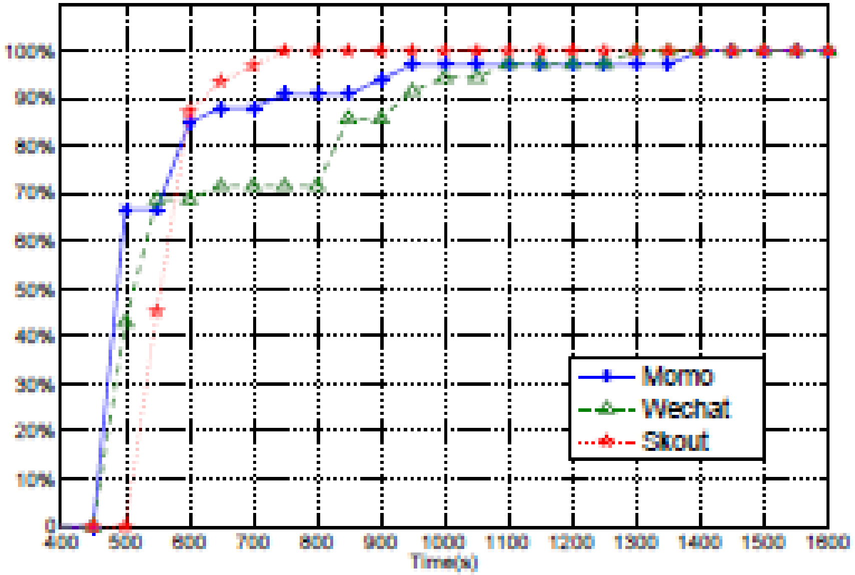


(b) Improved localization inference time on different apps

# Attack Performance Enhancement by Using Side Information

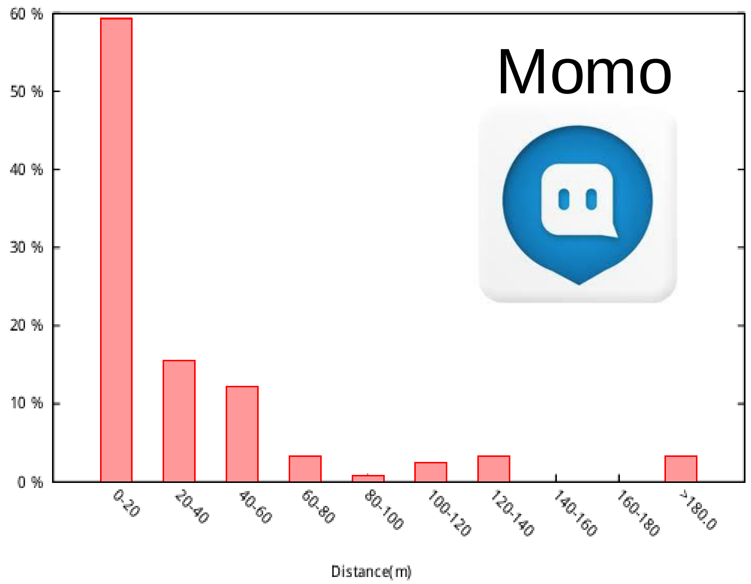


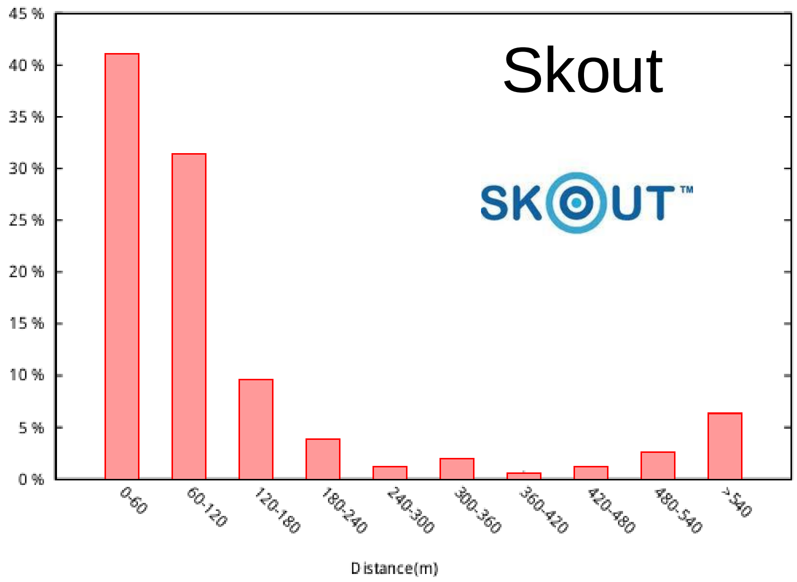
(a) Localization inference time on different apps

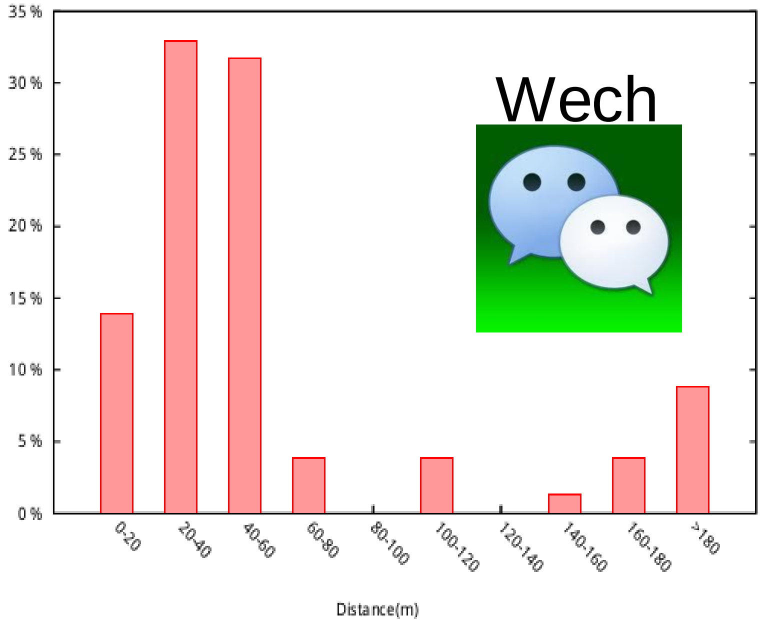


(b) Improved localization inference time on different apps

# Accuracy Evaluations

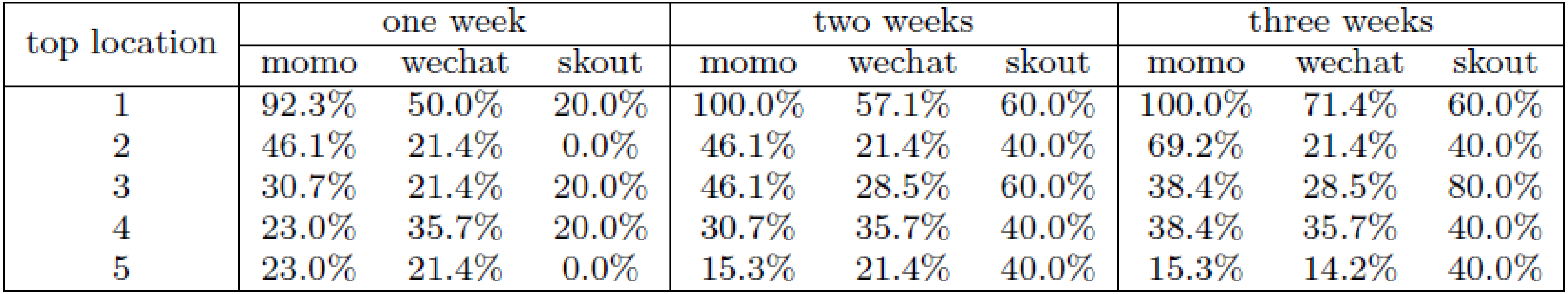






# Recover Top-5 Locations

TOP locations: locations that are most correlated to users' identities.
E.g., top 2 locations likely correspond to home and work locations



| top location | one week | | | two weeks | | | three weeks | | |
|---|---|---|---|---|---|---|---|---|---|
| | momo | wechat | skout | momo | wechat | skout | momo | wechat | skout |
| 1 | 92.3% | 50.0% | 20.0% | 100.0% | 57.1% | 60.0% | 100.0% | 71.4% | 60.0% |
| 2 | 46.1% | 21.4% | 0.0% | 46.1% | 21.4% | 40.0% | 69.2% | 21.4% | 40.0% |
| 3 | 30.7% | 21.4% | 20.0% | 46.1% | 28.5% | 60.0% | 38.4% | 28.5% | 80.0% |
| 4 | 23.0% | 35.7% | 20.0% | 30.7% | 35.7% | 40.0% | 38.4% | 35.7% | 40.0% |
| 5 | 23.0% | 21.4% | 0.0% | 15.3% | 21.4% | 40.0% | 15.3% | 14.2% | 40.0% |

Table 1: Top 5 Location Coverage Result for 3 Weeks

# A Demo