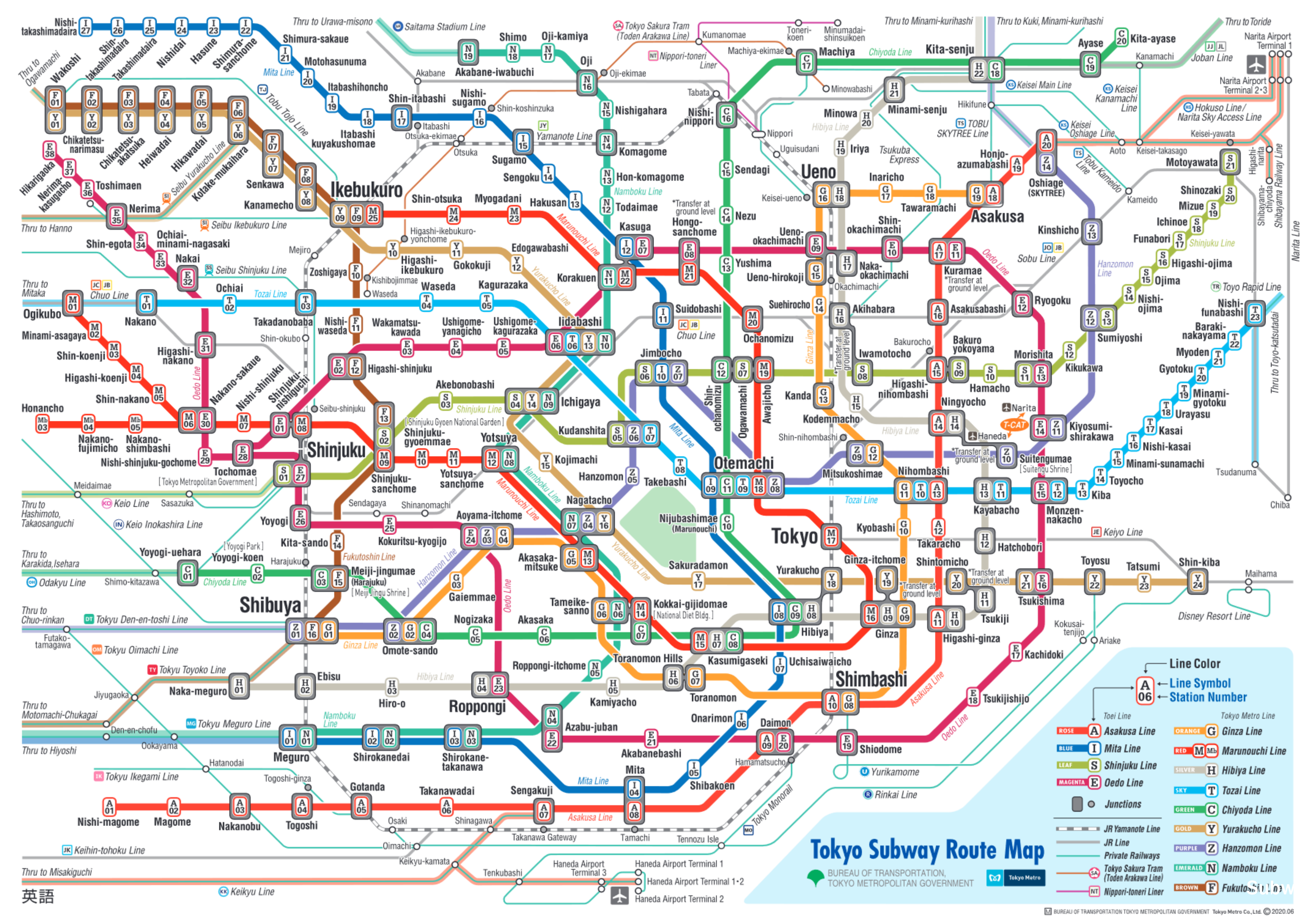
Private Attribution

PEARG, IETF 116

Martin Thomson

Trains


Tokyo Subway Route Map
BUREAU OF TRANSPORTATION, TOKYO METROPOLITAN GOVERNMENT
Tokyo Metro
Line Color
Line Symbol
Station Number
Toei Line
Tokyo Metro Line
A Asakusa Line
I Mita Line
S Shinjuku Line
E Oedo Line
G Ginza Line
M Mb Marunouchi Line
H Hibiya Line
T Tozai Line
C Chiyoda Line
Y Yurakucho Line
Z Hanzomon Line
N Namboku Line
F Fukutoshin Line
Junctions
JR Yamanote Line
JR Line
Private Railways
Tokyo Sakura Tram (Toden Arakawa Line)
Nippori-toneri Liner
英語
Nishi-takashimadaira
Shimura-sakaue
Motohasunuma
Itabashihoncho
Itabashi kuyakushomae
Shin-itabashi
Nishi-sugamo
Sugamo
Sengoku
Hakusan
Kasuga
Suidobashi
Jimbocho
Otemachi
Hibiya
Uchisaiwaicho
Onarimon
Shibakoen
Mita
Shirokane-takanawa
Shirokanedai
Meguro
Ikebukuro
Shinjuku
Shibuya
Ueno
Asakusa
Tokyo
Shimbashi
Roppongi
Otemachi
Akabane-iwabuchi
Oji-kamiya
Oji
Nishigahara
Komagome
Hon-komagome
Todaimae
Korakuen
Iidabashi
Ichigaya
Yotsuya
Kojimachi
Nagatacho
Tameike-sanno
Roppongi-itchome
Azabu-juban
Shirokane-takanawa
Kita-ayase
Ayase
Kita-senju
Machiya
Nishi-nippori
Sendagi
Nezu
Yushima
Shin-ochanomizu
Ochanomizu
Awajicho
Ogawamachi
Kanda
Mitsukoshimae
Nihombashi
Kayabacho
Monzen-nakacho
Kiba
Toyocho
Minami-sunamachi
Nishi-kasai
Kasai
Urayasu
Minami-gyotoku
Gyotoku
Myoden
Baraki-nakayama
Nishi-funabashi
Wakoshi
Chikatetsu-narimasu
Chikatetsu-akatsuka
Heiwadai
Hikawadai
Kotake-mukaihara
Senkawa
Kanamecho
Zoshigaya
Nishi-waseda
Higashi-shinjuku
Shinjuku-sanchome
Meiji-jingumae (Harajuku)
Kita-sando
Ogikubo
Minami-asagaya
Shin-koenji
Higashi-koenji
Shin-nakano
Nakano-sakaue
Nishi-shinjuku
Honancho
Nakano-fujimicho
Nakano-shimbashi
Ochiai
Nakano
Takadanobaba
Waseda
Kagurazaka
Ginza
Yurakucho
Kyobashi
Ginza-itchome
Shintomicho
Tsukishima
Toyosu
Tatsumi
Shin-kiba
Higashi-ginza
Tsukiji
Hatchobori
Kasumigaseki
Kokkai-gijidomae
Akasaka
Akasaka-mitsuke
Nogizaka
Omote-sando
Gaiemmae
Aoyama-itchome
Hanzomon
Kudanshita
Jimbocho
Suitengumae
Kiyosumi-shirakawa
Sumiyoshi
Kinshicho
Oshiage (SKYTREE)
Honjo-azumabashi
Kuramae
Asakusabashi
Bakuro-yokoyama
Higashi-nihombashi
Ningyocho
Hamacho
Morishita
Kikukawa
Ryogoku
Shin-okachimachi
Ueno-okachimachi
Naka-okachimachi
Ueno-hirokoji
Suehirocho
Iriya
Minowa
Minami-senju
Inaricho
Tawaramachi
Hikarigaoka
Nerima-kasugacho
Toshimaen
Nerima
Shin-egota
Ochiai-minami-nagasaki
Nakai
Higashi-nakano
Tochomae
Nishi-shinjuku-gochome
Shinjuku-nishiguchi
Yoyogi
Kokuritsu-kyogijo
Sendagaya
Shinanomachi
Yotsuya-sanchome
Ushigome-yanagicho
Ushigome-kagurazaka
Wakamatsu-kawada
Edogawabashi
Gokokuji
Higashi-ikebukuro
Shin-otsuka
Myogadani
Naka-meguro
Ebisu
Hiro-o
Kamiyacho
Toranomon
Toranomon Hills
Daimon
Akabanebashi
Sengakuji
Takanawadai
Gotanda
Togoshi
Nakanobu
Magome
Nishi-magome
Shiodome
Tsukijishijo
Kachidoki
Higashi-ginza
Takebashi
Nijubashimae (Marunouchi)
Sakuradamon
Kodemmacho
Yoyogi-uehara
Yoyogi-koen
Ichinoe
Mizue
Shinozaki
Motoyawatara
Funabori
Higashi-ojima
Ojima
Nishi-ojima
Hamamatsucho
Tamachi
Shinagawa
Takanawa Gateway
Tennozu Isle
Haneda Airport Terminal 1
Haneda Airport Terminal 1･2
Haneda Airport Terminal 2
Haneda Airport Terminal 3
Narita Airport Terminal 1
Narita Airport Terminal 2･3
Chiba
Maihama
Disney Resort Line
Saitama Stadium Line
Tokyo Sakura Tram (Toden Arakawa Line)
Tobu Tojo Line
Seibu Ikebukuro Line
Seibu Shinjuku Line
Seibu Yurakucho Line
Chuo Line
Keio Line
Keio Inokashira Line
Odakyu Line
Tokyu Den-en-toshi Line
Tokyu Oimachi Line
Tokyu Toyoko Line
Tokyu Meguro Line
Tokyu Ikegami Line
Keihin-tohoku Line
Keikyu Line
Tokyo Monorail
Rinkai Line
Yurikamome
Keiyo Line
Sobu Line
Joban Line
Keisei Main Line
Keisei Oshiage Line
Keisei Kanamachi Line
Hokuso Line / Narita Sky Access Line
TOBU SKYTREE Line
Tobu Kameido Line
Toyo Rapid Line
Shibayama Railway Line
Narita Line
Tsukuba Express
Thru to Wako-shi
Thru to Ogawamachi
Thru to Hanno
Thru to Mitaka
Thru to Hashimoto, Takaosanguchi
Thru to Karakida, Isehara
Thru to Chuo-rinkan
Thru to Motomachi-Chukagai
Thru to Hiyoshi
Thru to Misakiguchi
Thru to Urawa-misono
Thru to Minami-kurihashi
Thru to Kuki, Minami-kurihashi
Thru to Toride
Thru to Toyo-katsutadai
*Transfer at ground level
BUREAU OF TRANSPORTATION TOKYO METROPOLITAN GOVERNMENT Tokyo Metro Co., Ltd. ©2020.06

Map

Identifiers

Access cards or credit cards provide a unique user/traveller identifier

Travellers tap the card to enter and exit the system

The distance between entry and exit can determine the fare due

- Designing a privacy-preserving system for charging fares based on route is a homework exercise

Train Tracking

Many subway systems use tracking to monitor usage

Traveller identifiers are logged on entry and exit

- Cash payments or entry-only tracking (as in NYC, right) provide less information

Queries of logs can reveal system utilization and can inform capacity planning

Logs are a Privacy Risk

Logs contain extensive records on the movements of people

Each entry includes a time, a location, and an identifier

...and maybe more

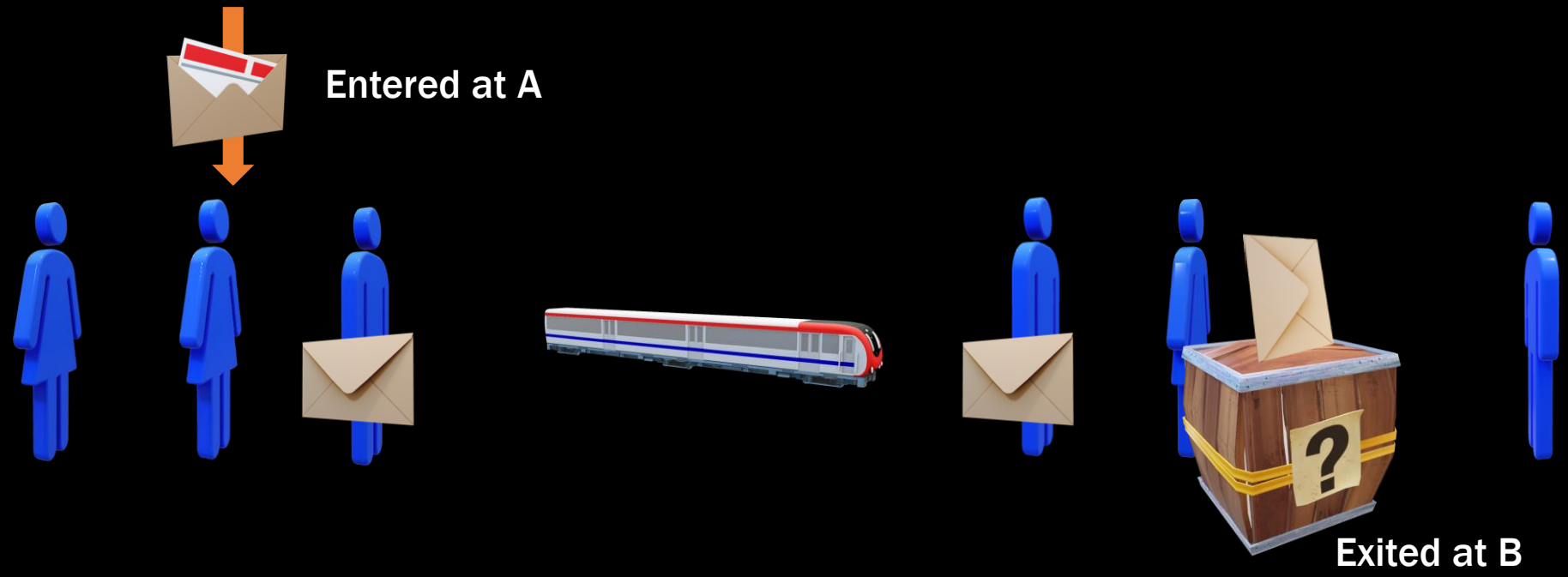
Pseudonymous identifiers provide no meaningful privacy protection

Requirements

- Be able to collected ***aggregated*** information about journeys
- Protect details of individual user journeys

Design

You hand the packets to the attacker to deliver



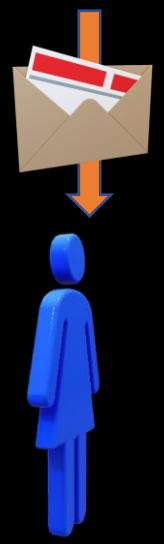
Issue people sealed tokens on entry

...then collect them at the exit

Privacy Mechanisms

- Tokens need to be anonymous (or maybe really low entropy)
  - ...or the token is just another identifier
- Tokens need to be authenticated
  - ...or people can lie
- Opening tokens needs to be delayed
  - ...or the timing reveals who it refers to
  - Random delays and anonymizing proxies might work
  - An aggregation system (PPM WG) can be faster

Token-Based Design Properties



- Tokens are ephemeral
  - They are returned at the exit and only apply for that trip
- Users carry tokens from the entrance to the exit
- The information that a token provides is limited
  - This is generally good for privacy, with some caveats
  - …but this is inherently inflexible
- Aggregation can help some of the worse aspects
  - Delays
  - Unknown anonymity set size


Trains | Advertising

Attribution

Attribution informs just about every aspect of advertising

- Placement
- Creatives
- How much to spend

Attribution measures events that occur

- in different contexts
- to the same person

"How many people saw the ad then came to the show?"

Attribution is More Complex

Entering

.

.

Happens once per trip

Showing ads

...or clicking ads

...or decided not to show an ad

Happens 0..n times

Exiting

.

.

Happens once per trip

Purchasing the product

...or just visiting the site

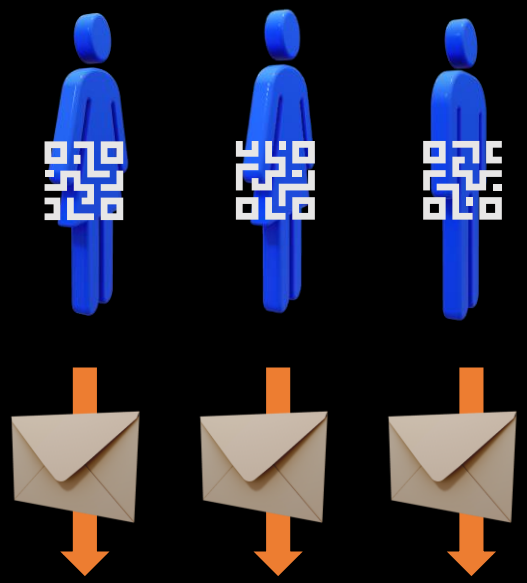
...or any outcome

Happens 0..n times

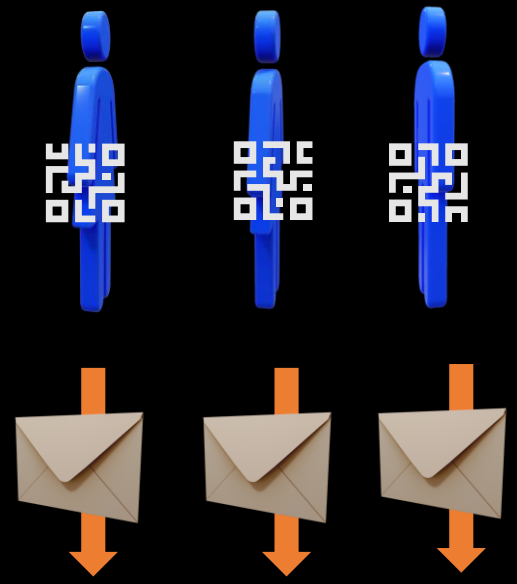
Contextual data is irrelevant

Context is everything

Interoperable Private Attribution



People all have a secret identifier



Sites can ask for an encrypted copy of the identifier

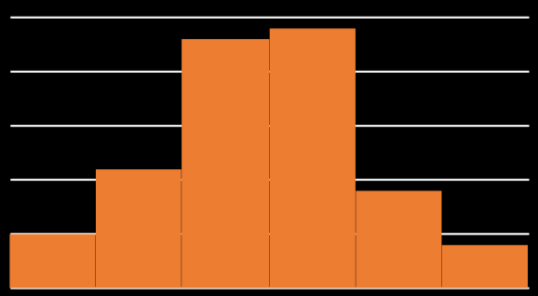
Sites gather identifiers in logs along with contextual info

Sites cannot correlate user activity without help

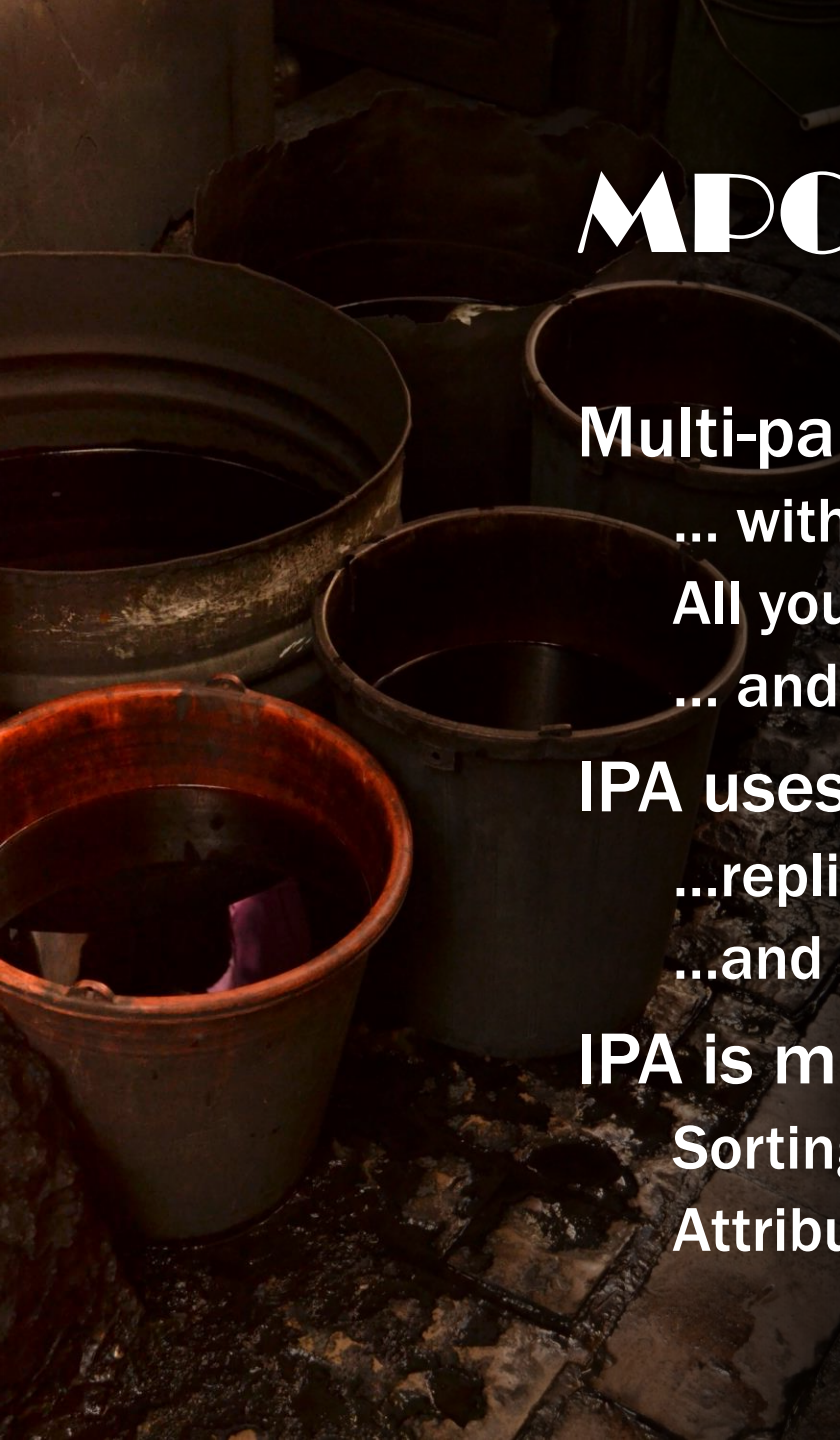
IPA: Attribution in MPC

A site gathers events from multiple sites and uses contextual data to formulate a query

MPC decrypts identifiers and performs attribution



The result is aggregated results



MPC in IPA

- Multi-party computation can perform any computation
  - … without revealing individual inputs
  - All you need is additions and multiplications
  - … and money: complex computations can be *very* expensive
- IPA uses a three-party, honest-majority MPC
  - …replicated secret sharing provides performance
  - …and almost information theoretic security guarantees
- IPA is mostly generic MPC
  - Sorting groups inputs by the (hidden) identifier
  - Attribution is computed over adjacent inputs

Differential Privacy

IPA uses (ε, δ)-differential privacy to hide individual contributions

Sites get a query budget that renews each epoch/week

Privacy loss is bounded by time and number of sites involved

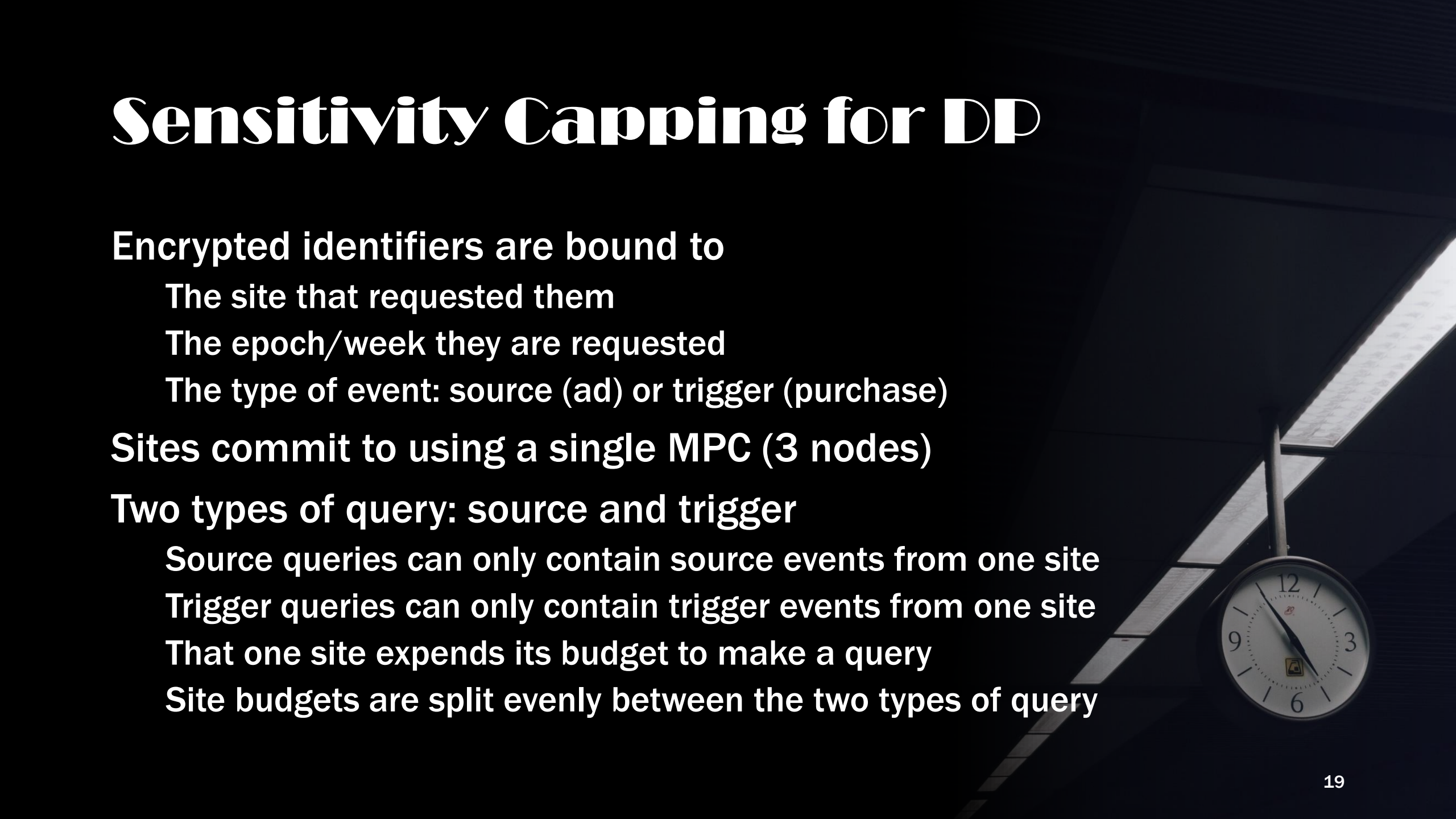
- Each site has their own budget
- Budgets are renewed weekly
- Goal is to limit privacy loss ***rate***

Each query of the MPC uses up budget

- Sites trade off noise with the number of queries
- Values for ε and δ not decided

Sensitivity Capping for DP

- Encrypted identifiers are bound to
  - The site that requested them
  - The epoch/week they are requested
  - The type of event: source (ad) or trigger (purchase)
- Sites commit to using a single MPC (3 nodes)
- Two types of query: source and trigger
  - Source queries can only contain source events from one site
  - Trigger queries can only contain trigger events from one site
  - That one site expends its budget to make a query
  - Site budgets are split evenly between the two types of query



IPA: Advantages and Challenges

- IPA offers more flexibility for advertisers than alternatives
  - Contextual information can be selected at query time
  - Less need for special fraud prevention mechanisms
- Flexibility might hurt accountability
  - DP provides bounds on privacy loss, but no one understands DP
  - The content of queries cannot be easily inspected and understood
- MPC performance is a challenge
  - Current implementation has plausible costs at small scale
  - Scaling to meet needs of large advertising businesses is hard

Status

- IPA is still active research
  - Feasibility largely established
  - Finer details of algorithms still being worked out
  - Meta and partners are running trials
- Ongoing work in the PATCG and PATWG in the W3C
  - Other proposals are also being considered
  - Protocols will likely go to IETF PPM WG


PAT
CG
https://patcg.github.io/