Analysis of Privacy Protections in Fitness Tracking Social Networks -or- You can run, but can you hide?

Wajih Ul Hassan*, Saad Hussain*, Adam Bates

ILLINOIS

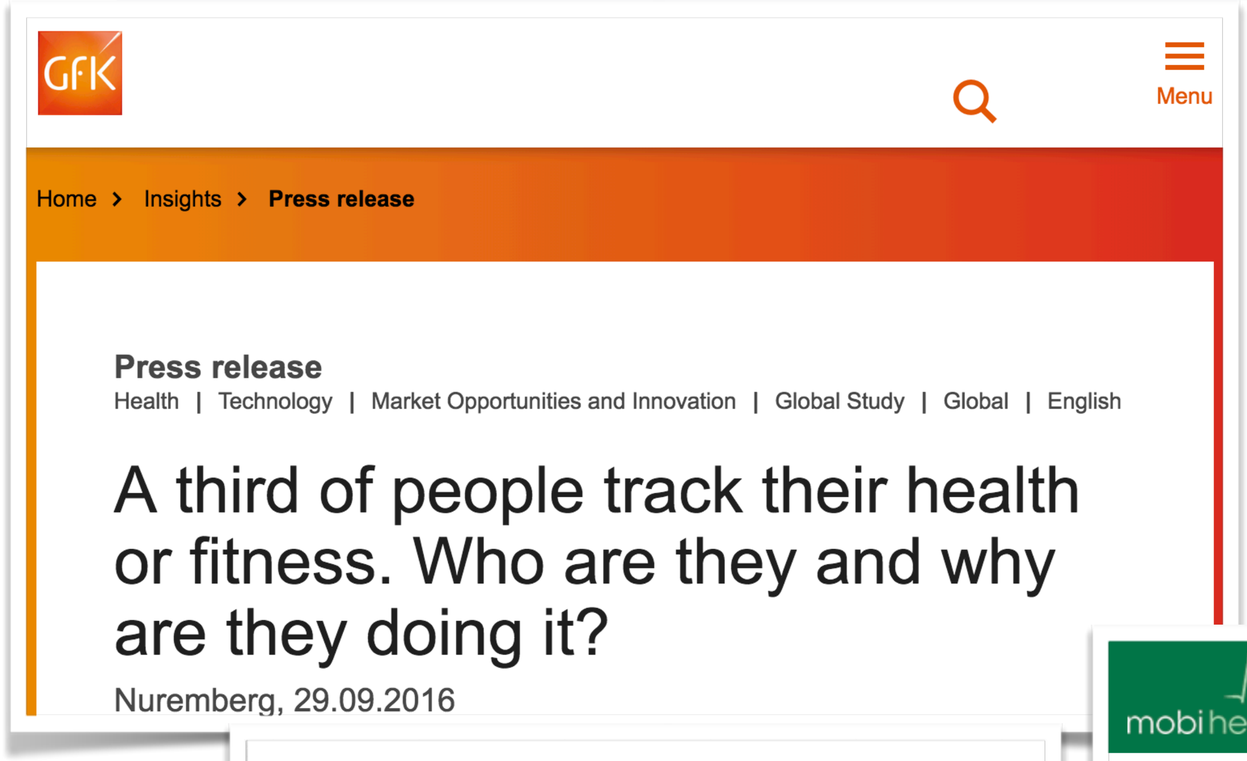
27th USENIX Security Symposium

16 August 2018

*Joint first authors

Self-Tracking Market

Self-Tracking Market


GfK

Menu

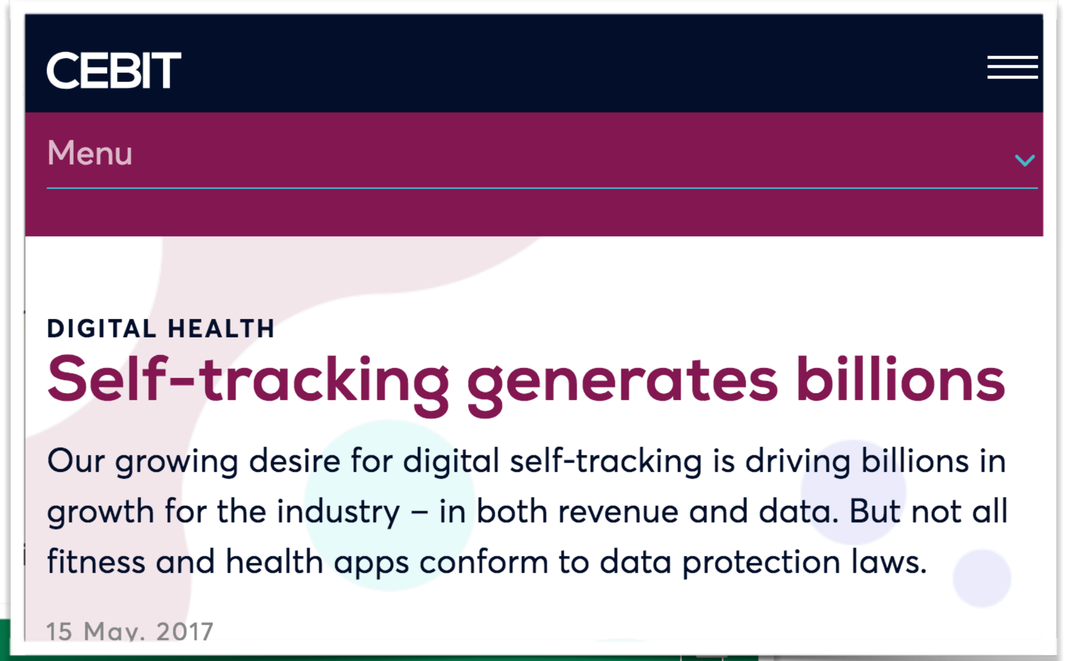
Home › Insights › **Press release**

Press release

Health | Technology | Market Opportunities and Innovation | Global Study | Global | English

A third of people track their health or fitness. Who are they and why are they doing it?

Nuremberg, 29.09.2016


CEBIT

Menu

DIGITAL HEALTH

Self-tracking generates billions

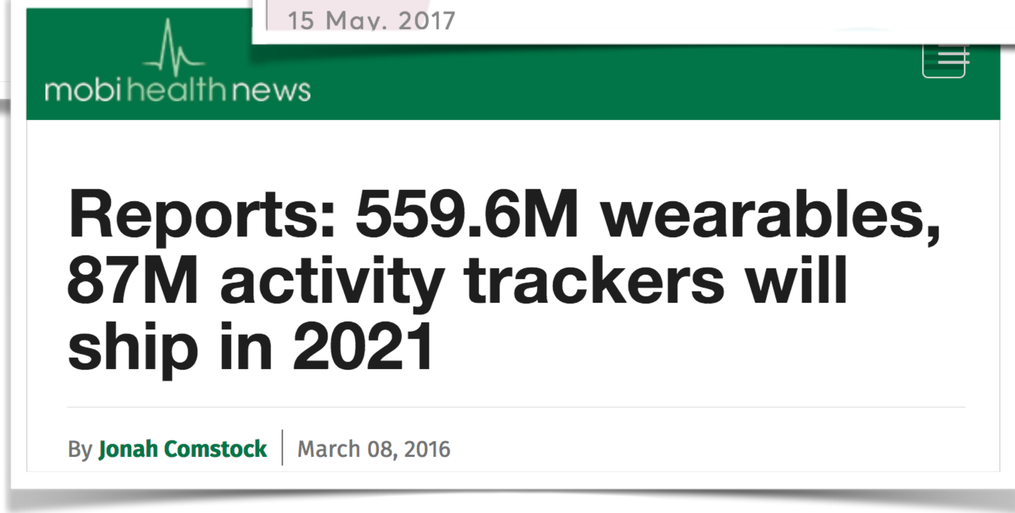
Our growing desire for digital self-tracking is driving billions in growth for the industry – in both revenue and data. But not all fitness and health apps conform to data protection laws.

15 May, 2017


P&S MARKET RESEARCH™
GLOBAL PARTNER IN RESEARCH

Wearable Fitness Trackers Market to Reach $48.2 Billion by 2023: P&S Market Research

March 28, 2018 05:30 ET | **Source:** P&S Market Research


mobihealthnews

Reports: 559.6M wearables, 87M activity trackers will ship in 2021

By **Jonah Comstock** | March 08, 2016

Fitness Tracking Ecosystem

Taken from Strava page on google playstore

Fitness Tracking Ecosystem

Taken from Strava page on google playstore

Fitness Tracking Ecosystem

STRAVA™

runtastic

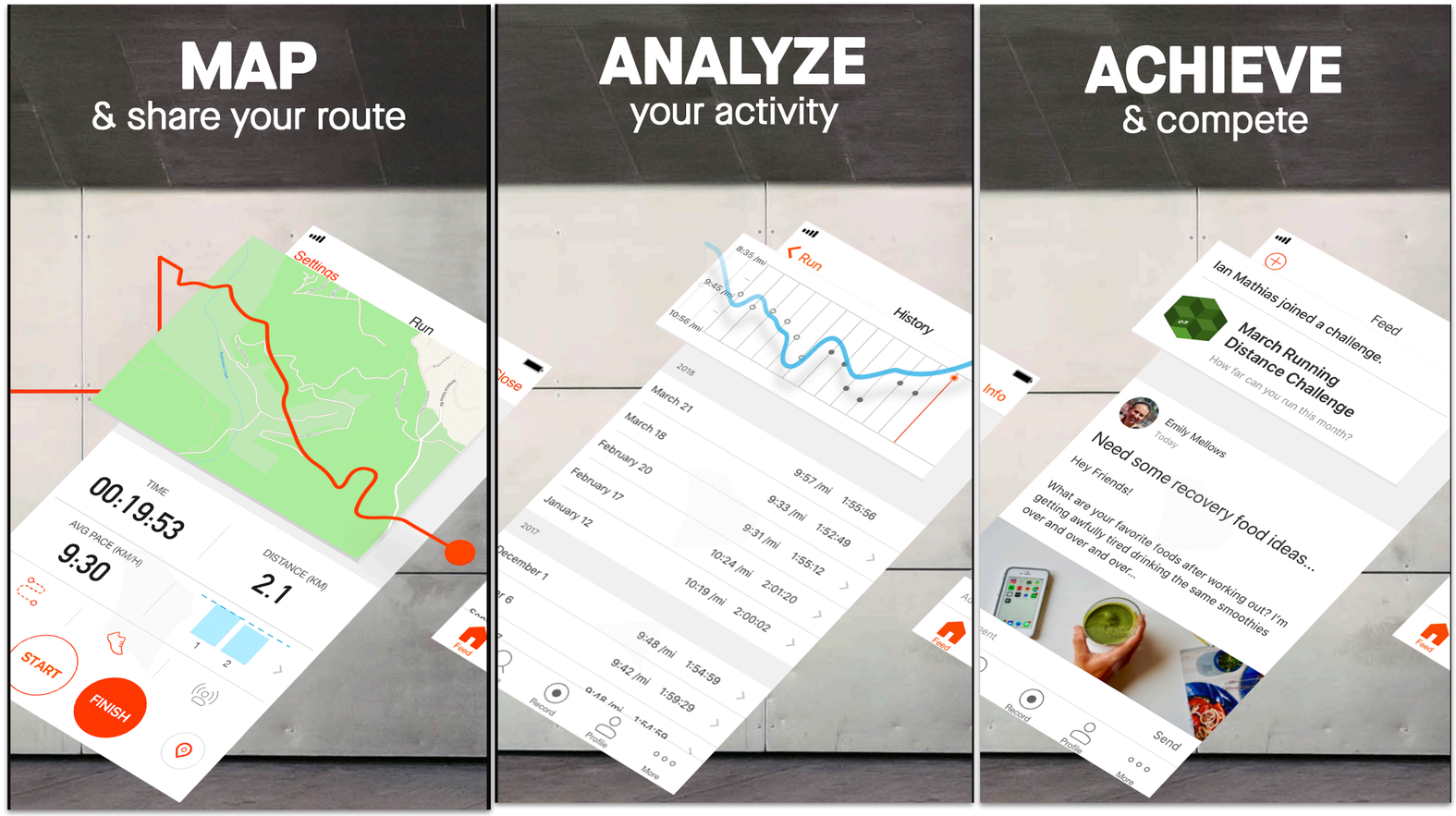
Map My Tracks

runkeeper

endomondo

GARMIN

mapmyrun



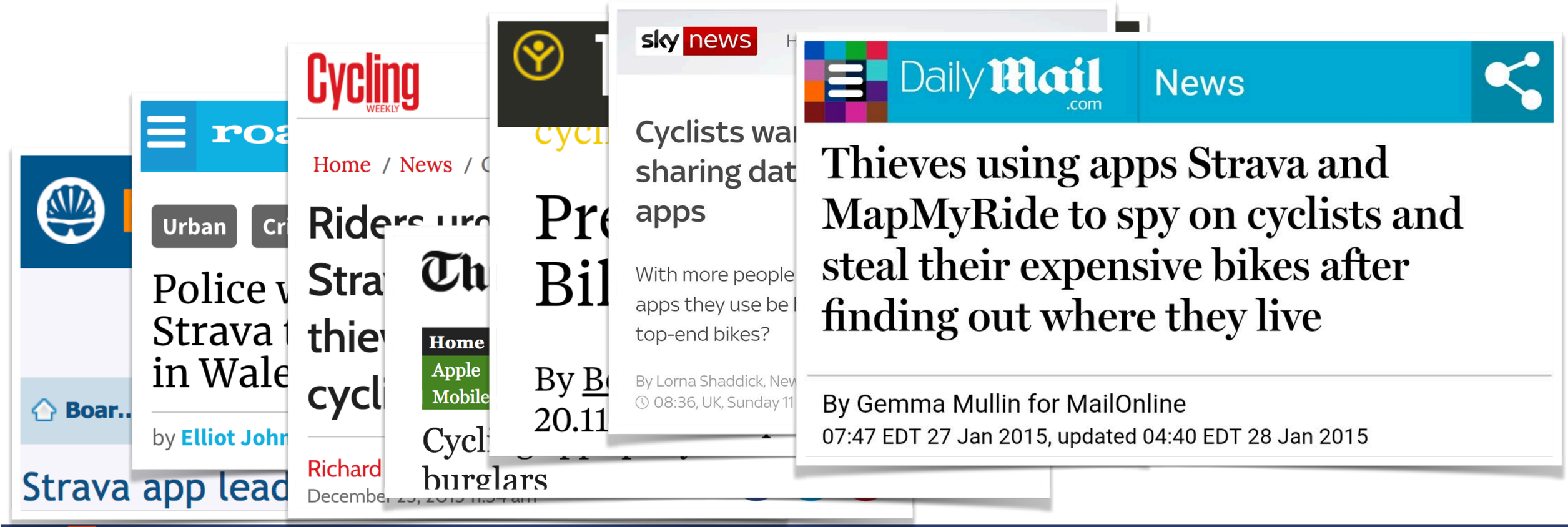
Taken from Strava page on google playstore

Privacy Concerns

- People usually run from sensitive locations (e.g., home)
- Broadcasting routes to public has privacy concerns

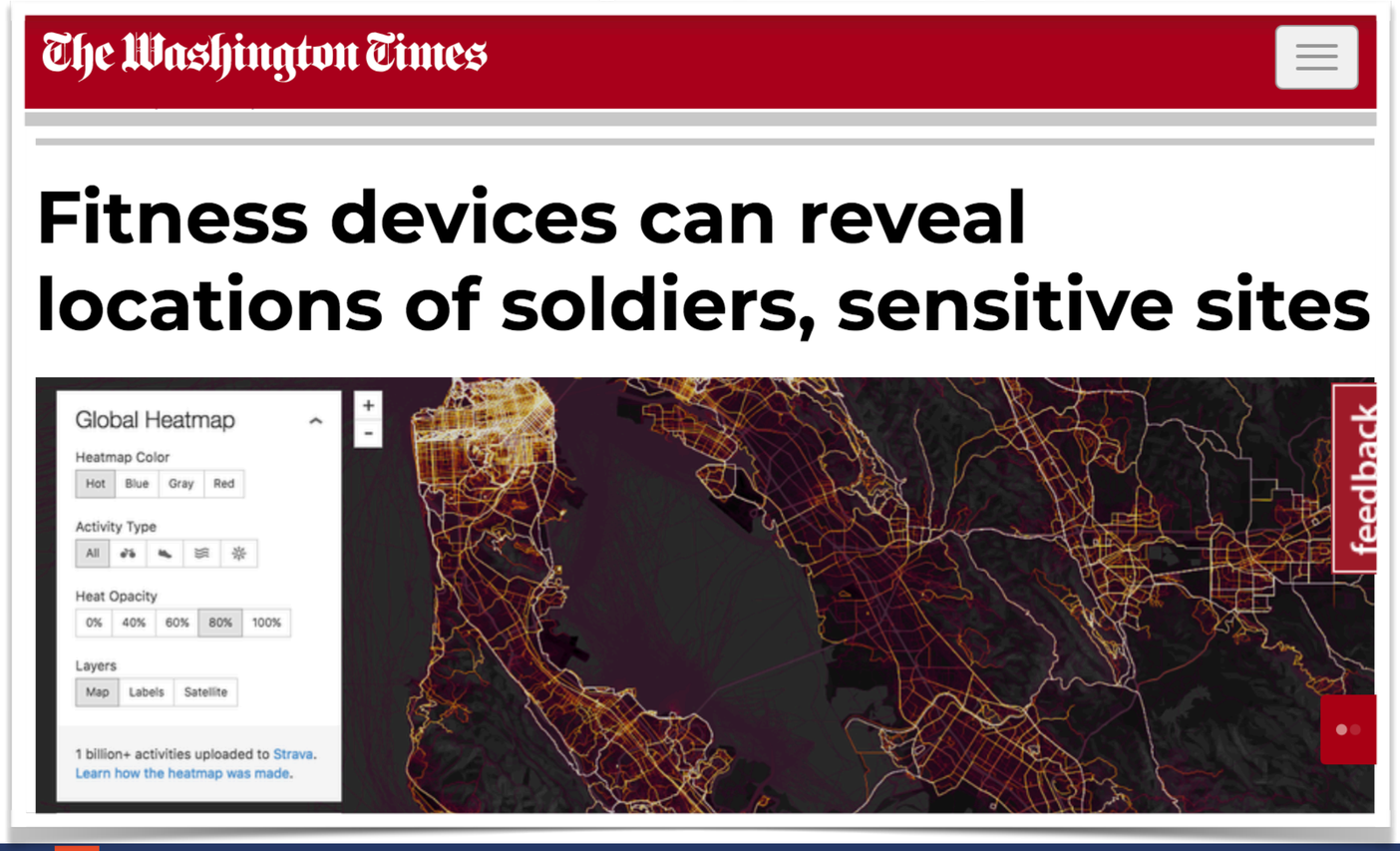
Privacy Concerns

- People usually run from sensitive locations (e.g., home)
- Broadcasting routes to public has privacy concerns



Privacy Concerns

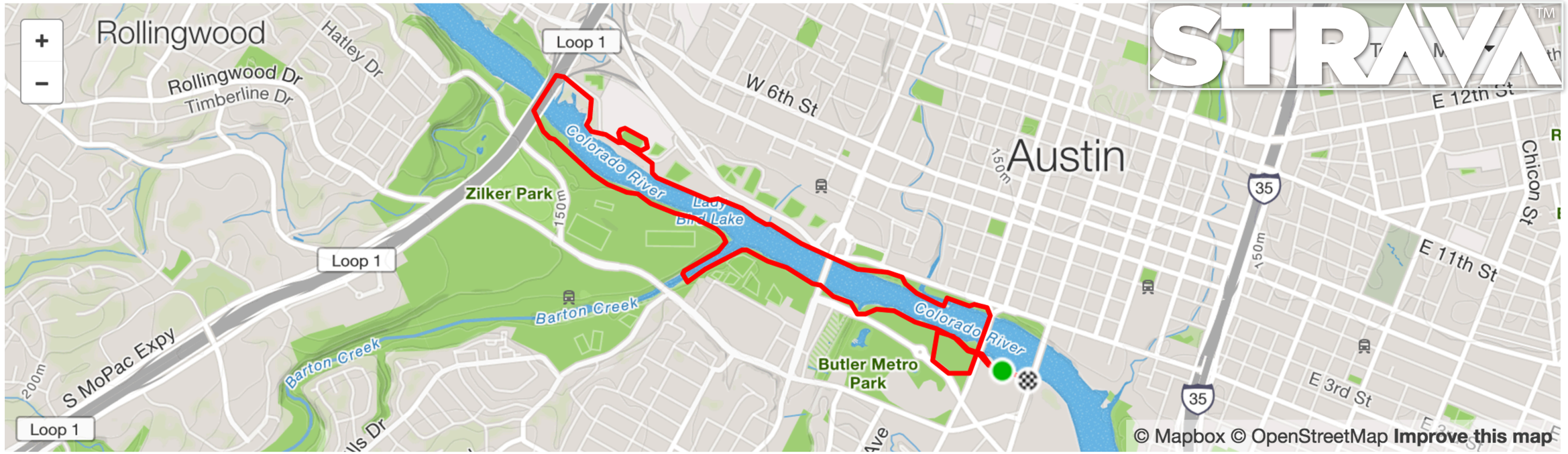
- People usually run from sensitive locations (e.g., home)
- Broadcasting routes to public has privacy concerns



A military base in Helmand Province, Afghanistan with route taken by joggers

Strava Run Activity

5.0 mi	43:37	8:40 /mi	817
Distance	Moving Time	Avg Pace	Calories



USENIX Security 2016

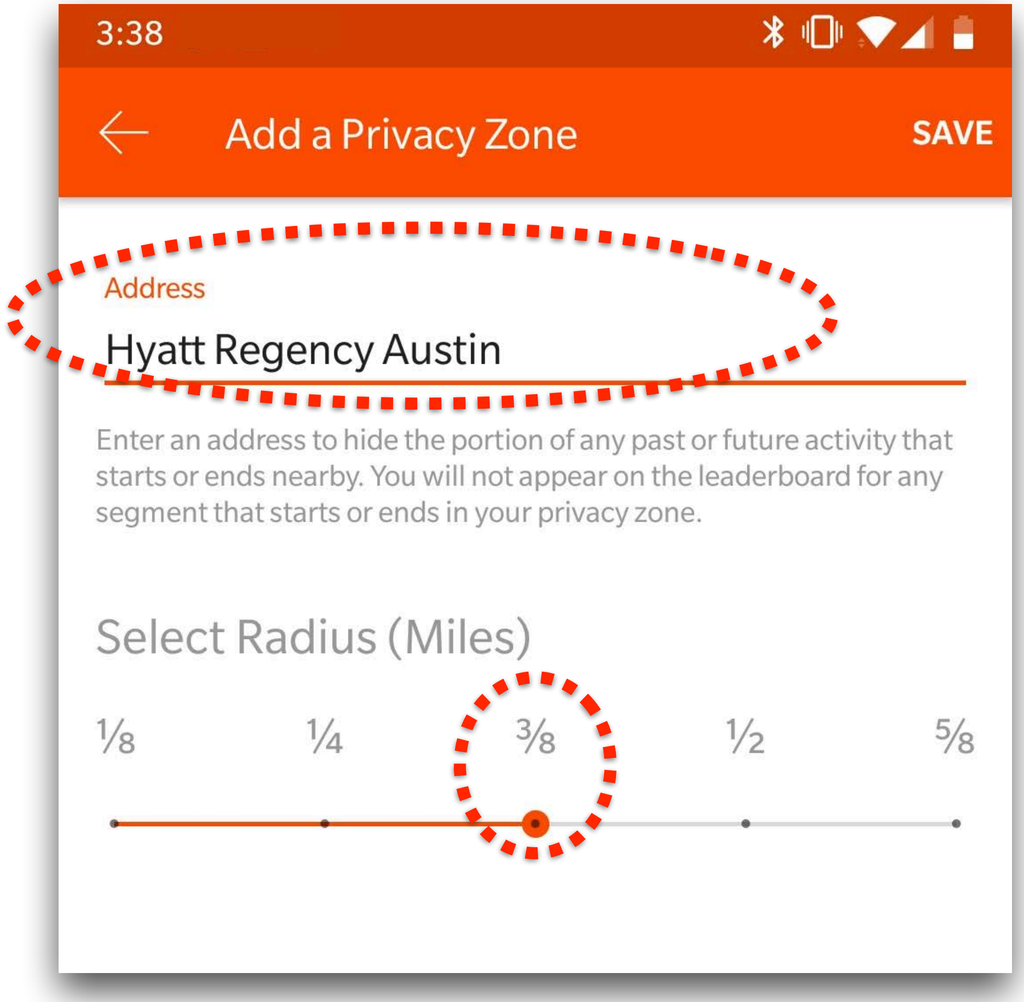
- Before applying Endpoint Privacy Zone

Endpoint Privacy Zone (EPZ)

- Strava allows users to hide sensitive endpoints using **privacy zones**
- Endpoint privacy zone is a fixed radius circular area
- It hides the portion of activity that starts or ends in the circular area

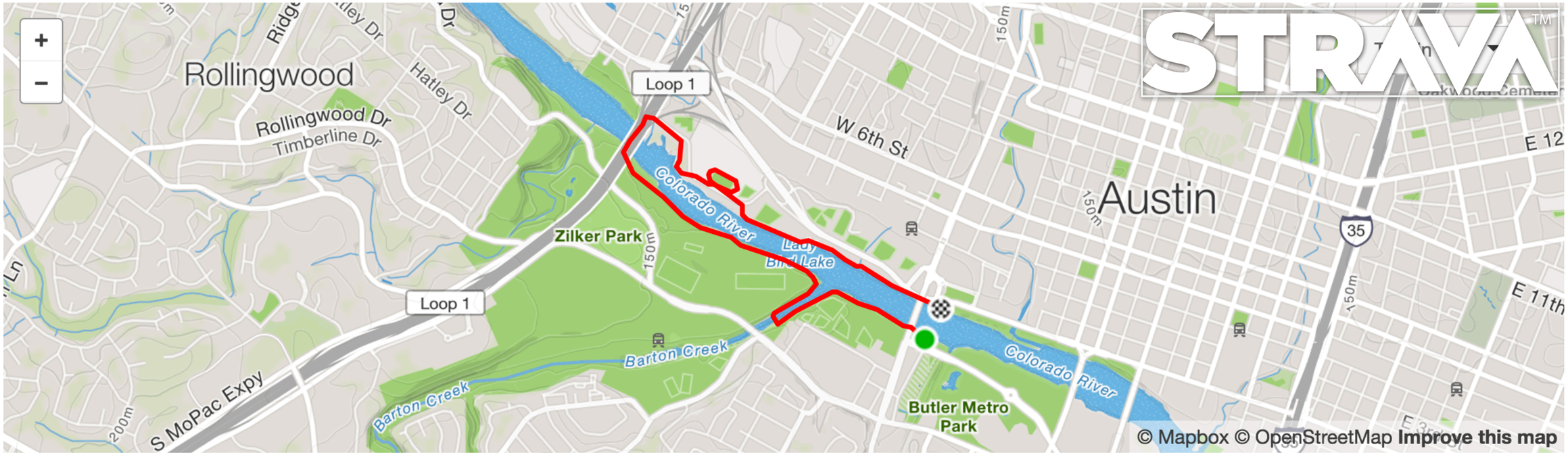
Endpoint Privacy Zone (EPZ)

- Strava allows users to hide sensitive endpoints using **privacy zones**
- Endpoint privacy zone is a fixed radius circular area
- It hides the portion of activity that starts or ends in the circular area



What Other People See

5.0 mi	43:37	8:40 /mi	817
Distance	Moving Time	Avg Pace	Calories

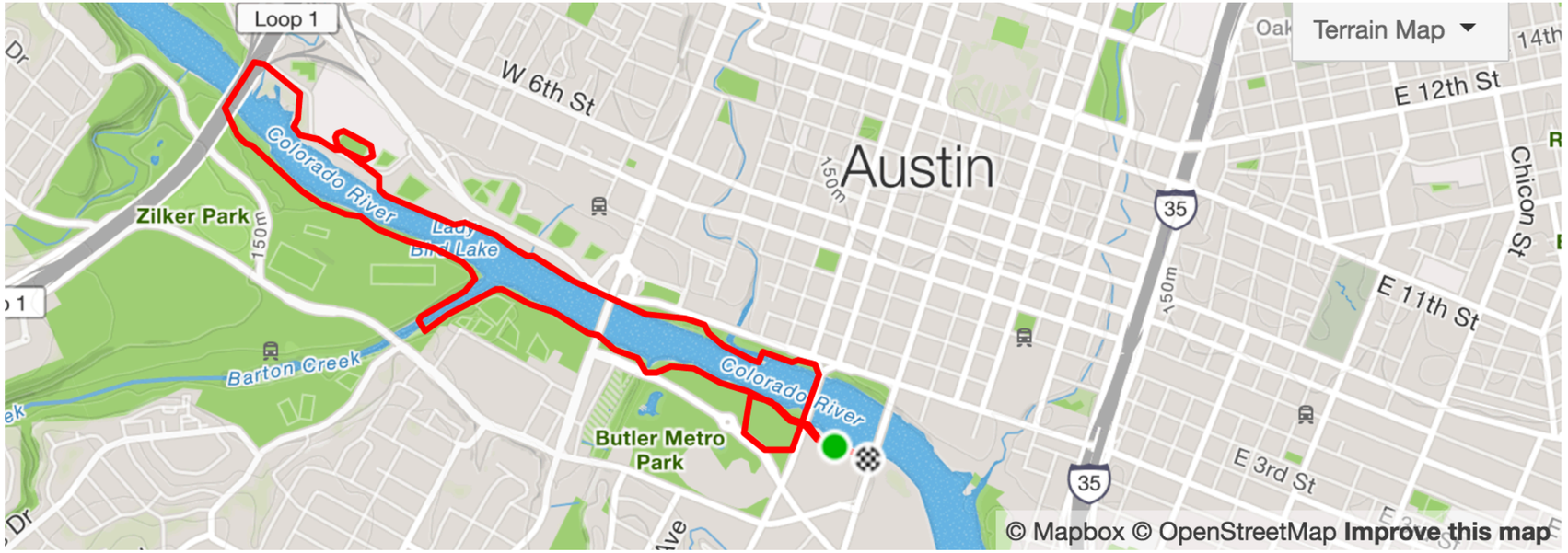


USENIX Security 2016

- After applying Endpoint Privacy Zone

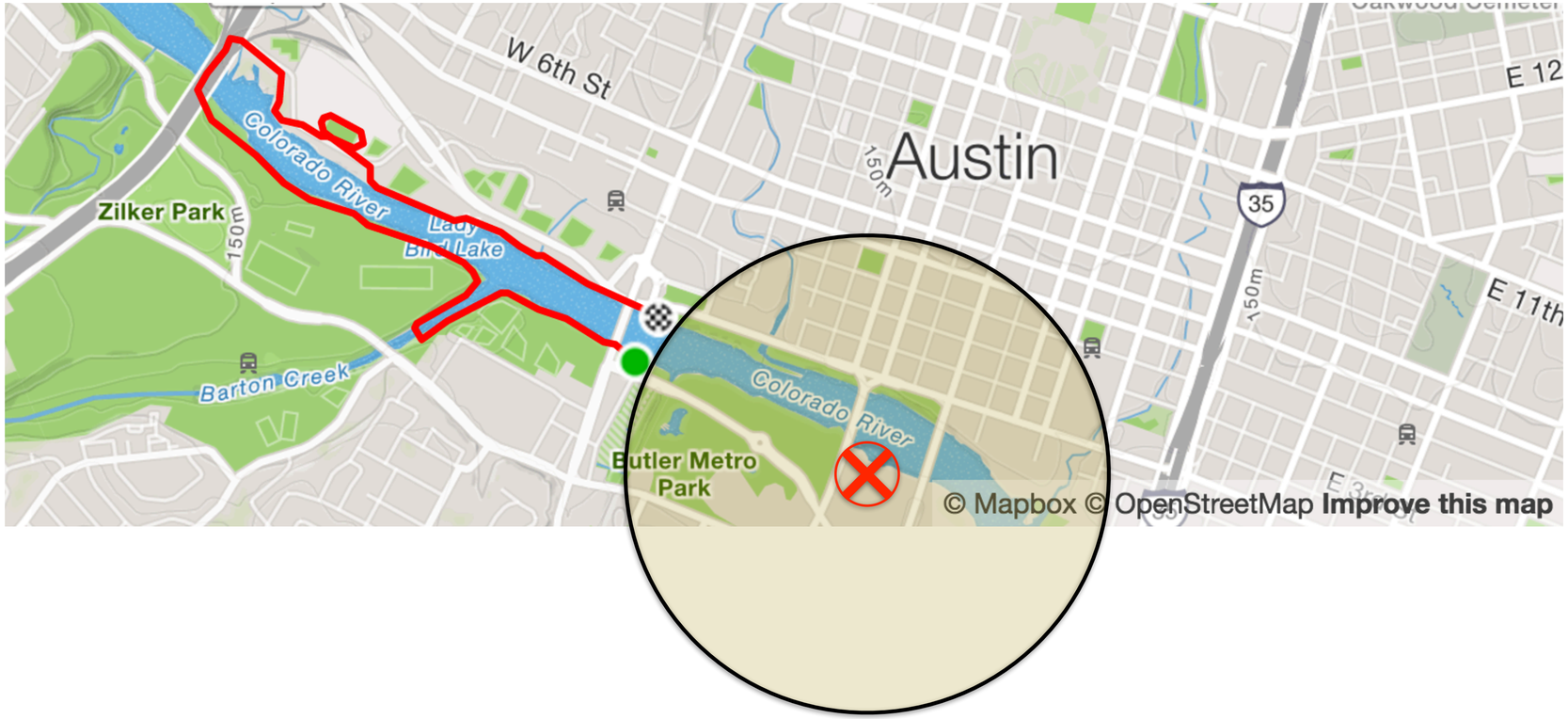
What happened behind the curtain

EndPoint Privacy Zone (EPZ)



EndPoint Privacy Zone (EPZ)

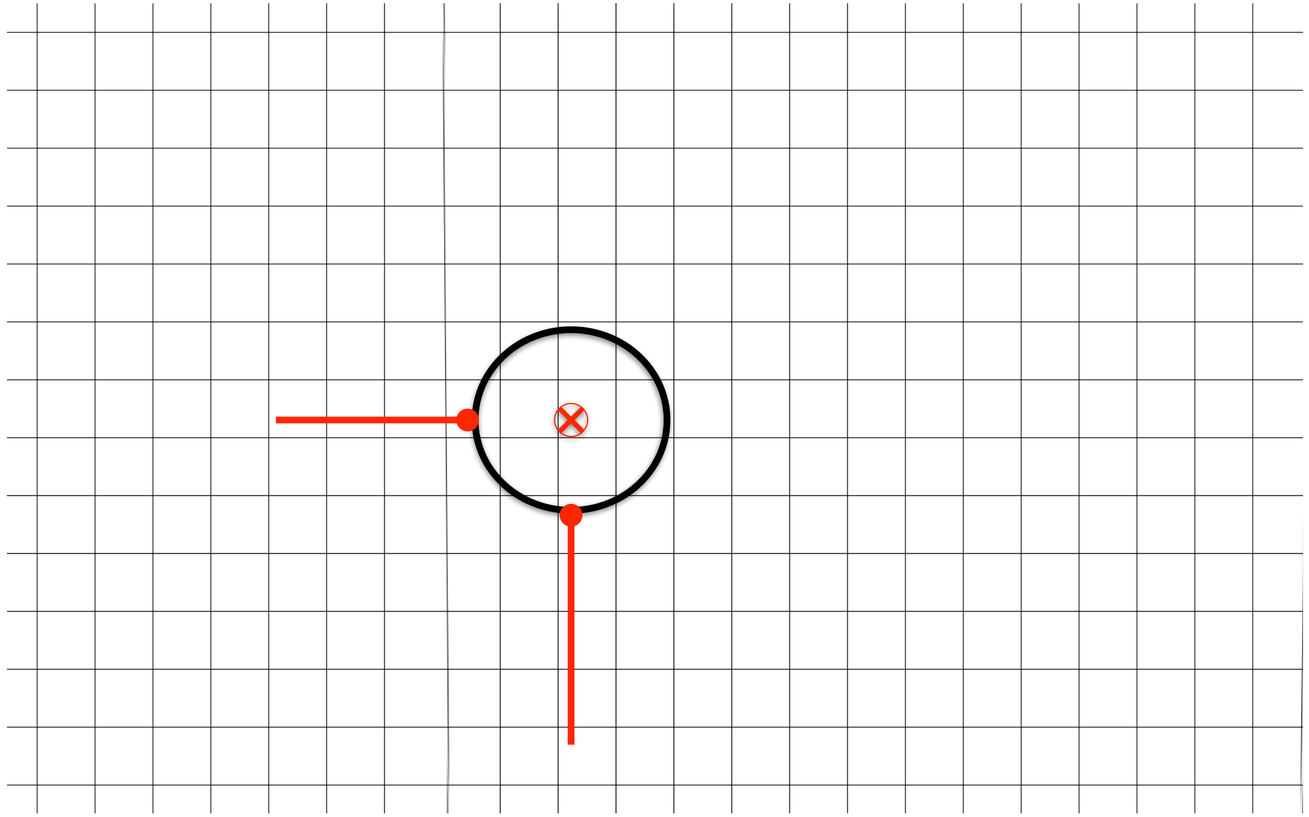
- Generate a EPZ at user specified center
- Center is the sensitive location
- Remove all the activity GPS points lying inside circle



Guessing EPZ Center

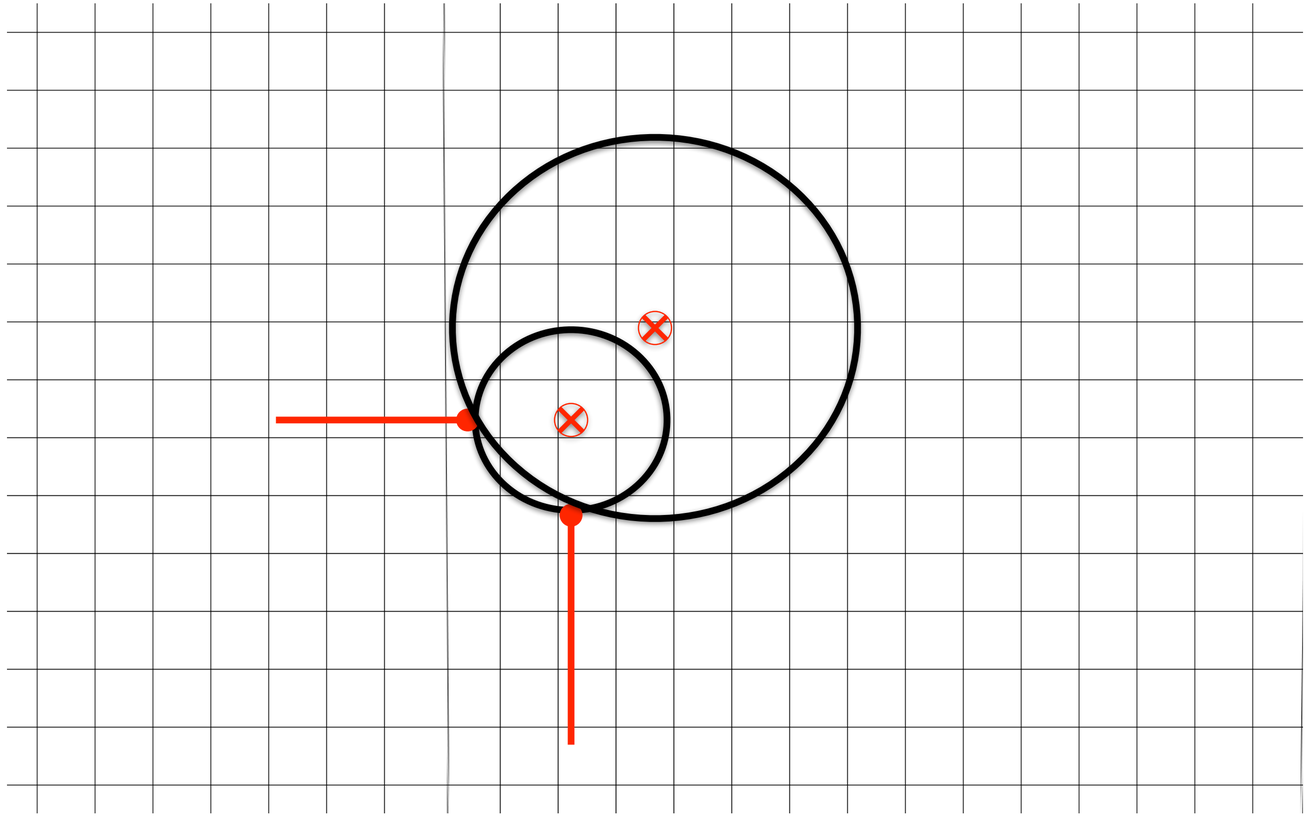
Guessing EPZ Center

Given two points generate circles that passes through those points



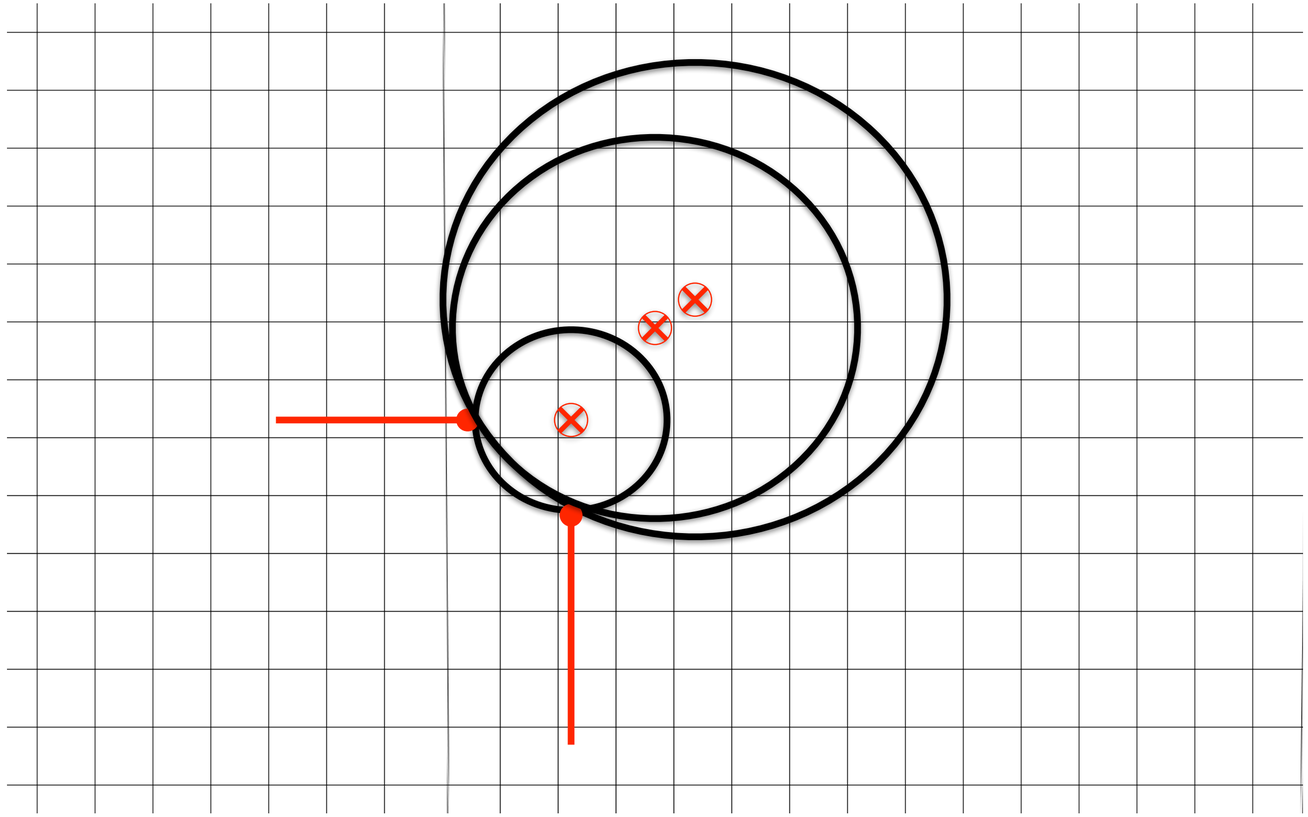
Guessing EPZ Center

Given two points generate circles that passes through those points



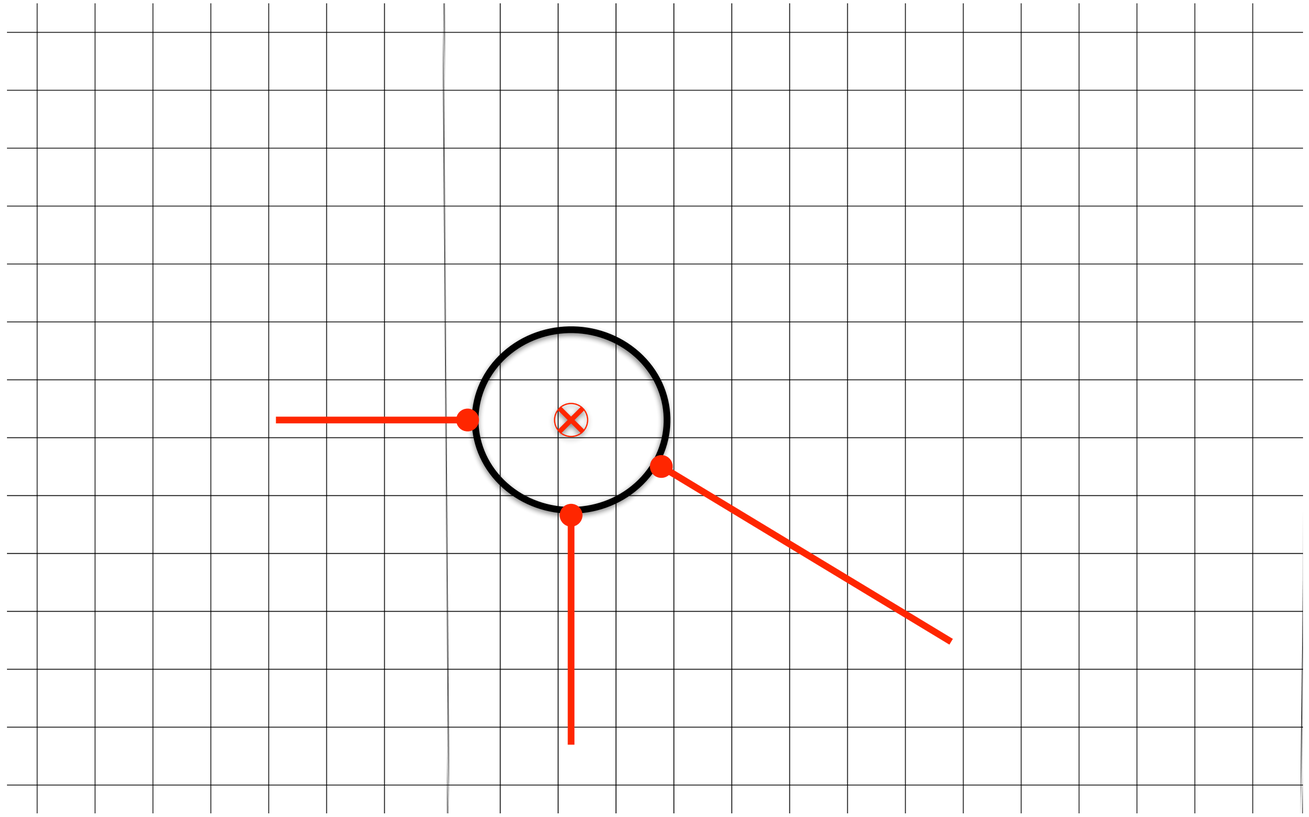
Guessing EPZ Center

Given two points generate circles that passes through those points



Guessing EPZ Center

Given two points generate circles that passes through those points



With three points you will find right EPZ

Is it that Simple?

Is it that Simple?
NO

Confounding Factors:

1. *Multiple endpoints*
    - Multiple EPZs can be predicted but only **one** is correct
2. *GPS sampling errors*
    - Endpoints of routes may not lie exactly on EPZ
    - Uncertainty when attempting to infer a protected location.

You can run, but can you hide?

You can run, but can you hide?

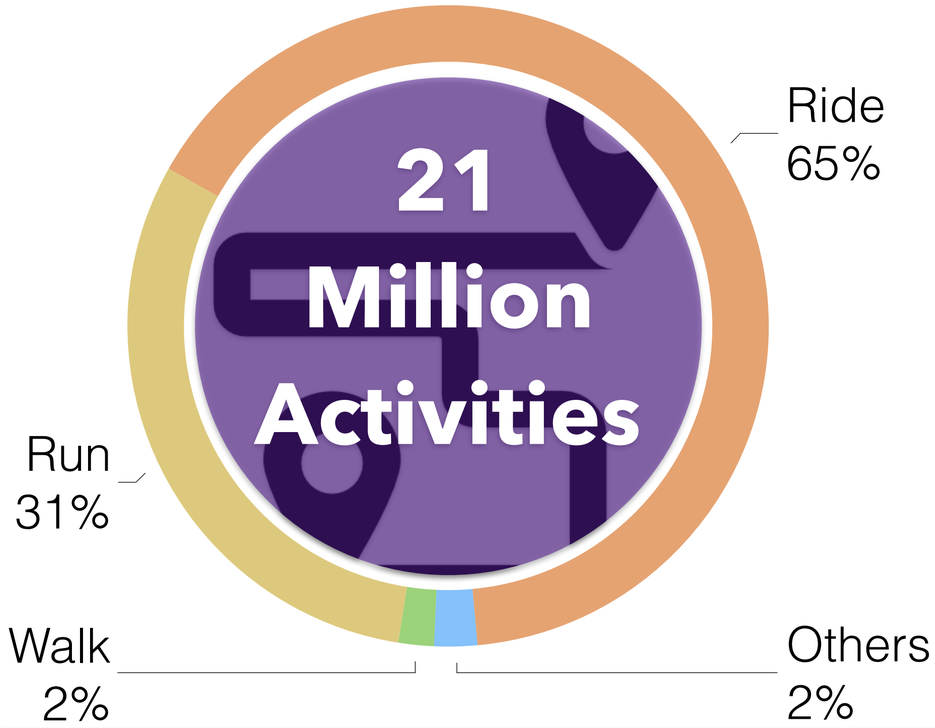
1. Collected activity dataset from strava.com
2. Designed robust algorithm to search EPZ circles
   a. Validated our algorithm on unprotected activities
   b. Ran an attack on protected activities
3. Designed mitigation strategies

Dataset

- Scraped May 2016 activity dataset from strava.com
- Collected GPS coordinates, activity type, duration, distance, athlete ID etc.

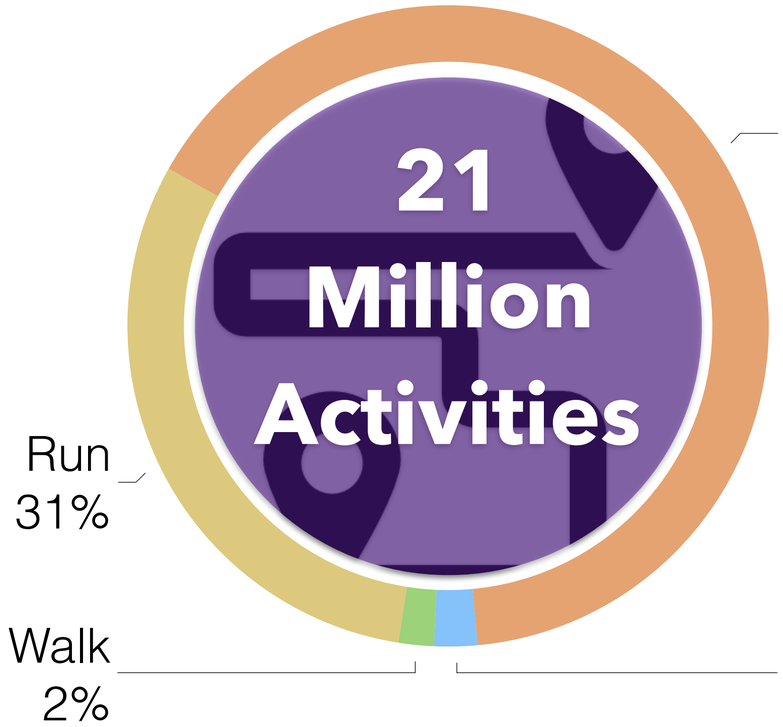
Dataset

- Scraped May 2016 activity dataset from strava.com
- Collected GPS coordinates, activity type, duration, distance, athlete ID etc.



Dataset

- Scraped May 2016 activity dataset from strava.com
- Collected GPS coordinates, activity type, duration, distance, athlete ID etc.



EPZ Search

- Least Square Circle Fit (LSF) algorithm
- But it does not fit our needs:
  - Considers all the endpoints
    - We don’t know which endpoints associated with EPZ
    - Wildly inaccurate results if one endpoint is wrong
  - Considers all possible radii while only finite are allowed by services
  - Does not cater jitter GPS coordinates
  - Slow

Our EPZ Search Algorithm

1. Generate candidate EPZs as follows:
    - For each pair of activities, perform pairwise inspection of each possible combination of endpoints

Our EPZ Search Algorithm

1. Generate candidate EPZs as follows:

- For each pair of activities, perform pairwise inspection of each possible combination of endpoints
- For each pair of endpoints `(x1,y1)`,`(x2,y2)`, solve the simultaneous equations

$$(xc-x1)^2 + (yc-y1)^2 = r^2$$
$$(xc-x2)^2 + (yc-y2)^2 = r^2$$

where `(xc,yc)` is center and `r` is possible radius

- Each solution of simultaneous equation is one candidate EPZ

Our EPZ Search Algorithm

2. We use consensus procedure among candidate EPZs to give confidence scores to each EPZ

- Confidence score is the number of activity start/end points that independently agree on the location of the EPZ

Our EPZ Search Algorithm

2. We use consensus procedure among candidate EPZs to give confidence scores to each EPZ

   - Confidence score is the number of activity start/end points that independently agree on the location of the EPZ

3. To accounts for jitters in GPS coordinates we use parametrized threshold

   - We considers two EPZs as same if the distance between the centers is less than or equal to the specified threshold

Our EPZ Search Algorithm

4. Finally, rule out candidate EPZs which do not make sense
    - Candidate EPZs intersecting other portion of routes
    - Candidates EPZs with less than certain confidence score

Our EPZ Search Algorithm

4. Finally, rule out candidate EPZs which do not make sense
    - Candidate EPZs intersecting other portion of routes
    - Candidates EPZs with less than certain confidence score

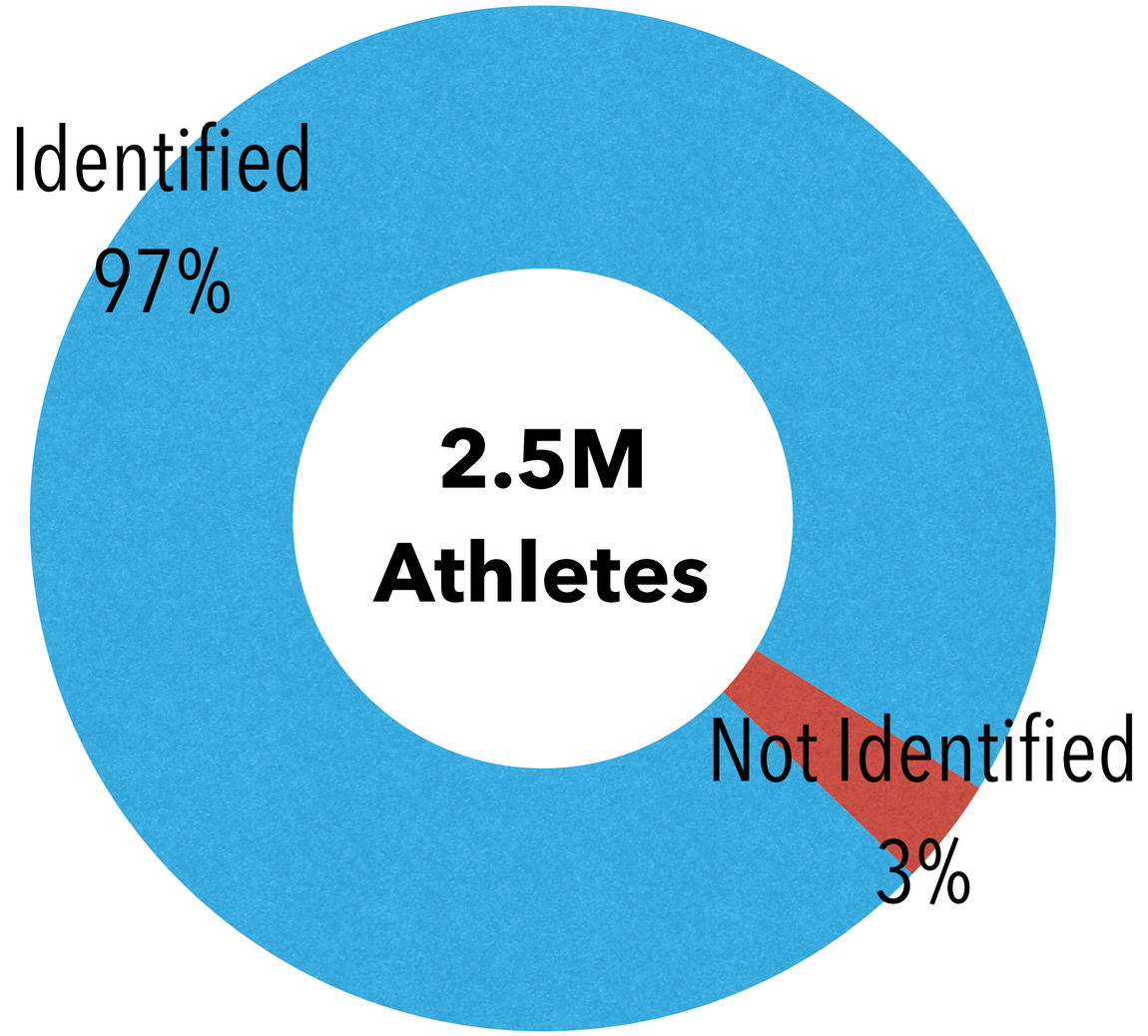
5.In end, our algorithm predicts one EPZ with highest confidence score
    - We use training run to automatically find best threshold for a specific service

Attack Validation

- Test EPZ search algorithm
- Created synthetic EPZ with radius 0.25 miles on unprotected activities so that we know exact center
- Use our algorithm to predict the center

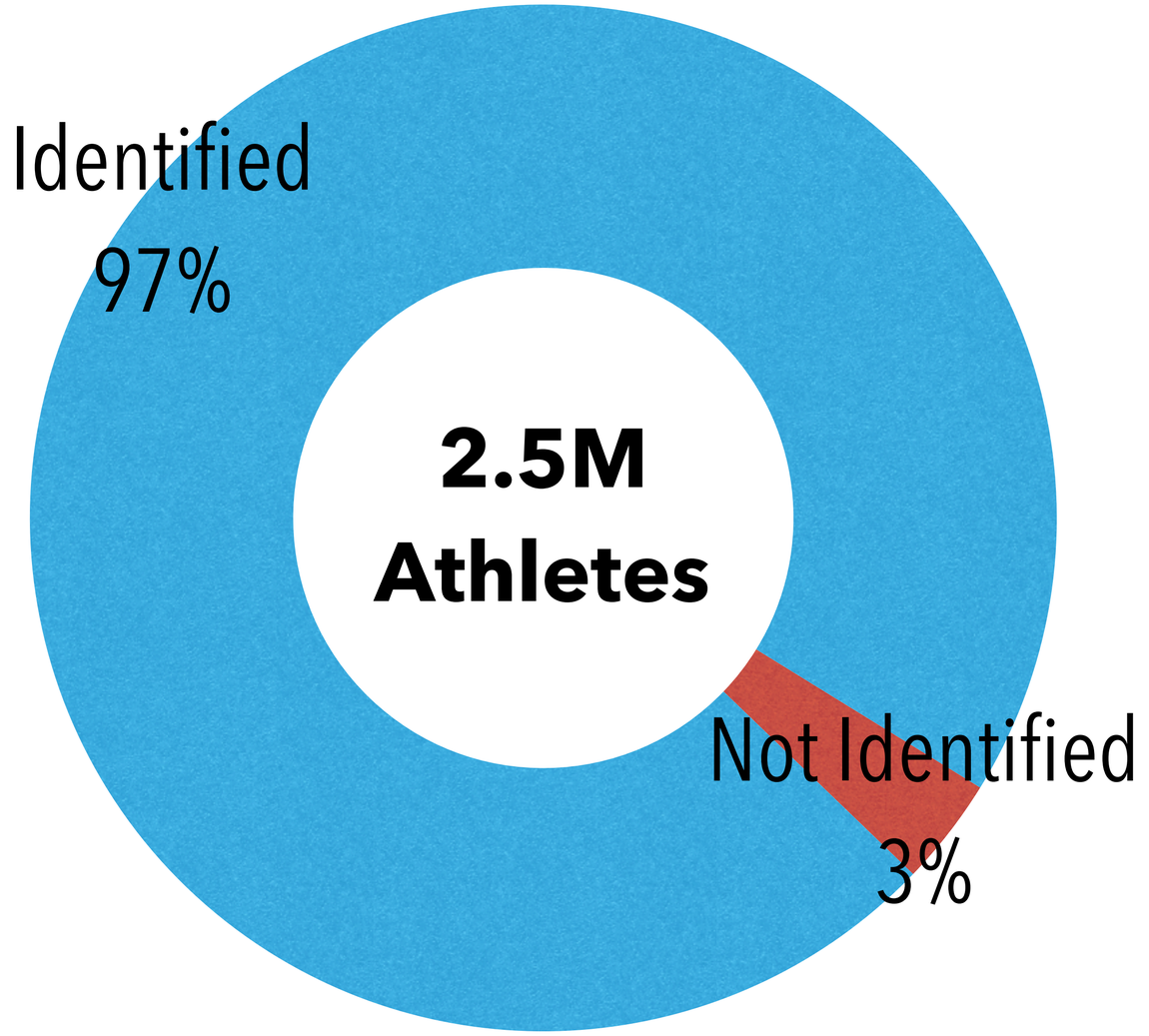
Attack Validation

- Test EPZ search algorithm
- Created synthetic EPZ with radius 0.25 miles on unprotected activities so that we know exact center
- Use our algorithm to predict the center



Attack Validation

- Test EPZ search algorithm
- Created synthetic EPZ with radius 0.25 miles on unprotected activities so that we know exact center
- Use our algorithm to predict the center



- Lack of available observations e.g. athletes has only <2 activities

Attack Evaluation

- After attack validation we identified actual protected location
- Deployed our search algorithm on protected activities

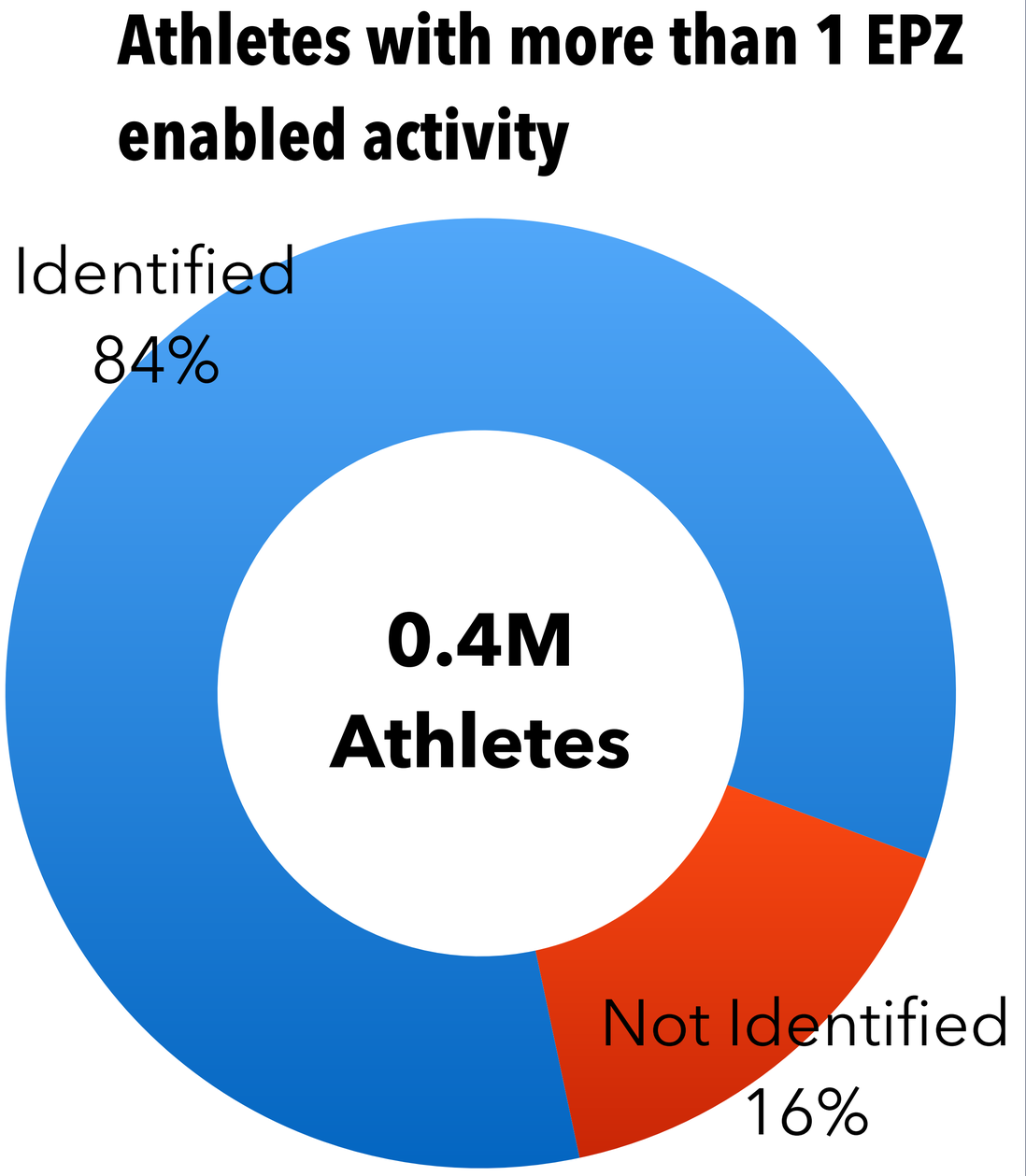
Attack Evaluation

- After attack validation we identified actual protected location
- Deployed our search algorithm on protected activities

Athletes with more than 1 EPZ enabled activity

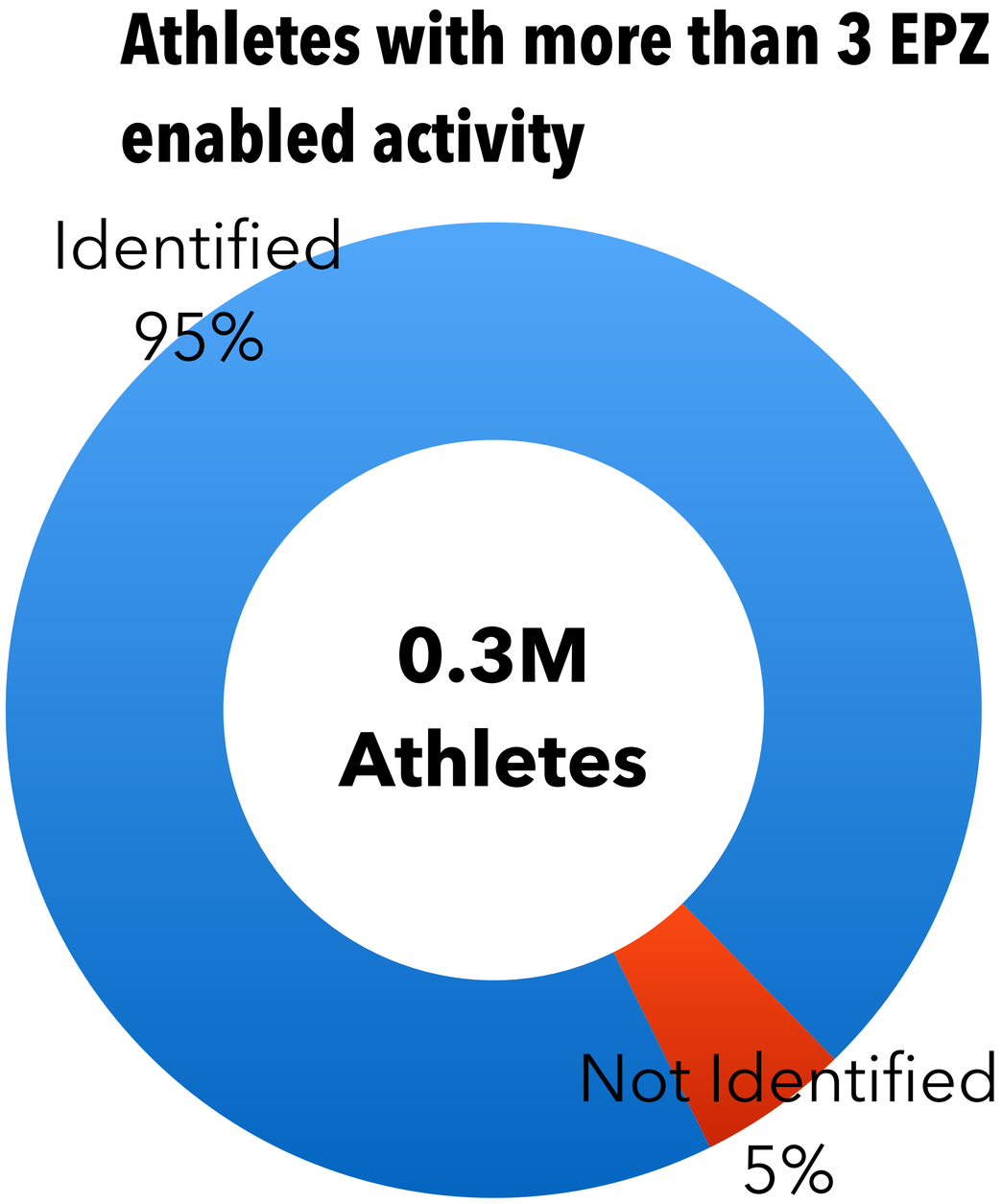
Attack Evaluation

- After attack validation we identified actual protected location
- Deployed our search algorithm on protected activities



Attack Evaluation

- After attack validation we identified actual protected location
- Deployed our search algorithm on protected activities



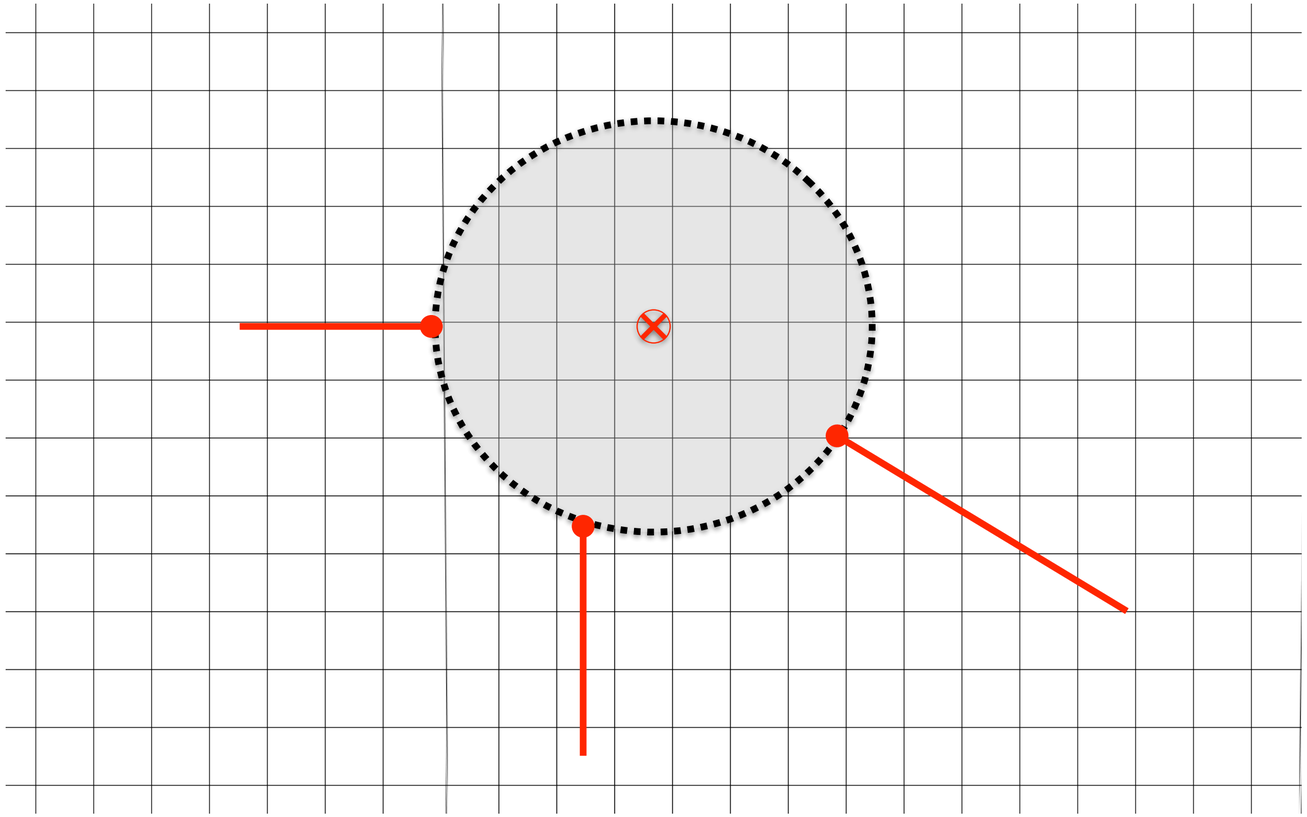
Countermeasures?

Countermeasures?

- Modify Radius Size
- Fuzz EPZ Intersection Points
- Spatial Cloaking

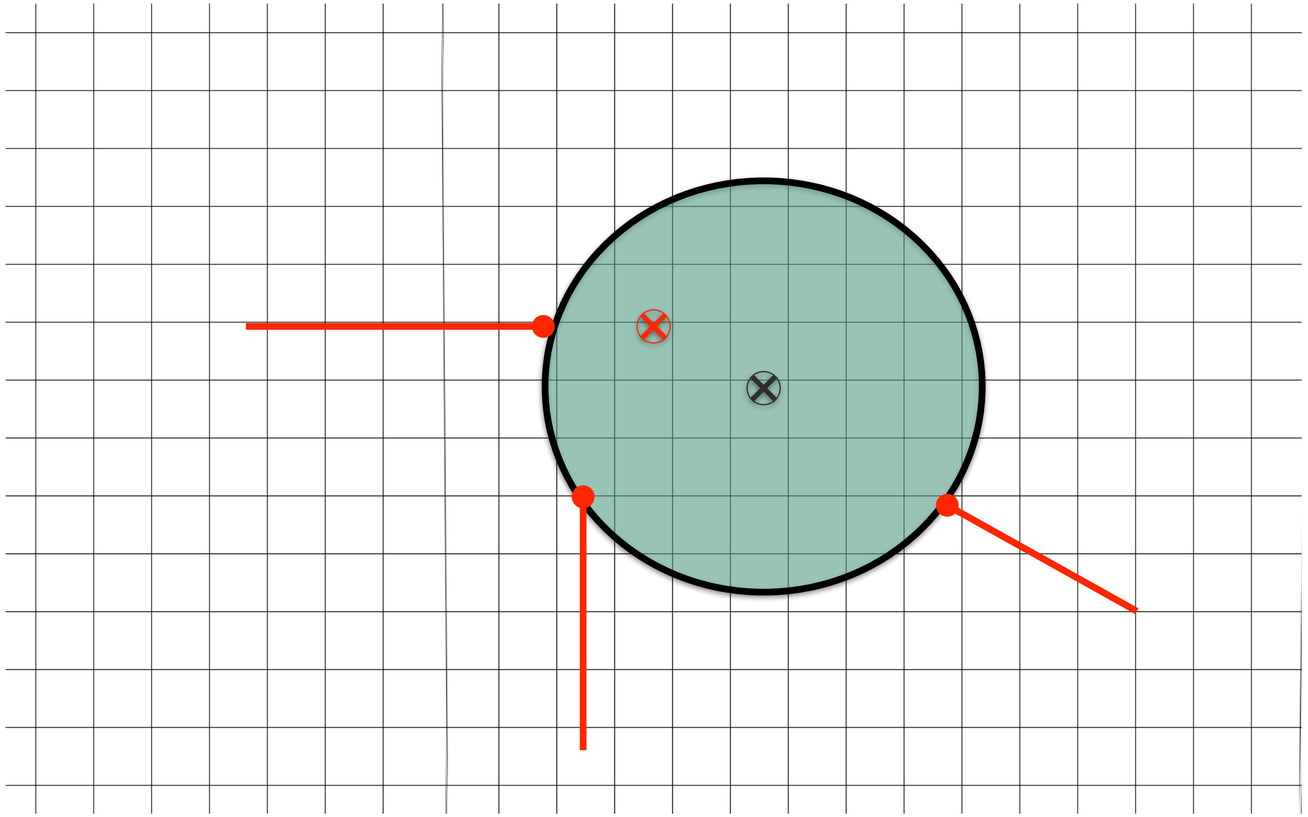
Spatial Cloaking

- Adapt geo-indistinguishability in the context of fitness tracking services



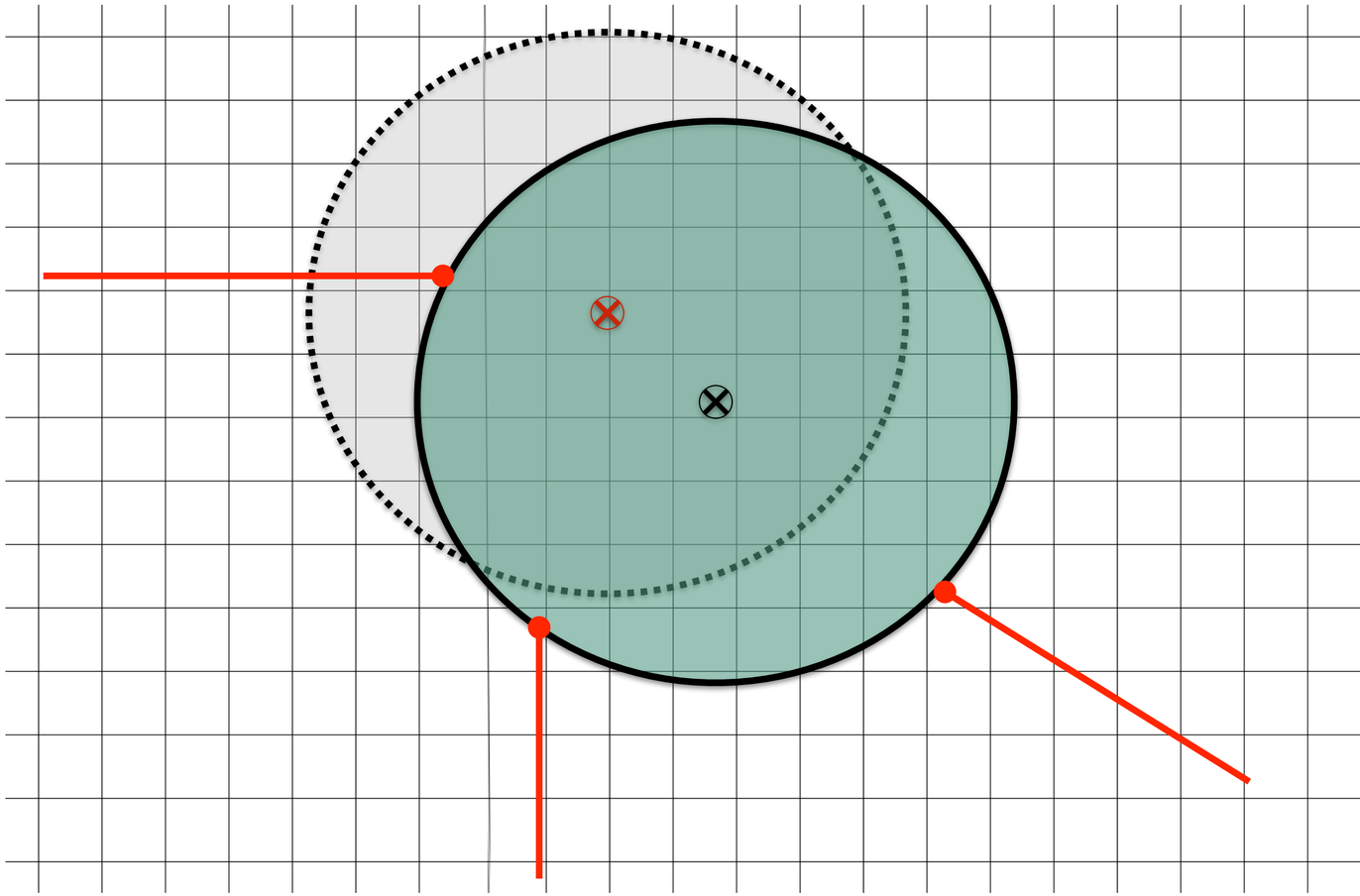
Spatial Cloaking

- Adapt geo-indistinguishability in the context of fitness tracking services
- Shift the center of EPZ
- Sensitive location can be anywhere inside new EPZ



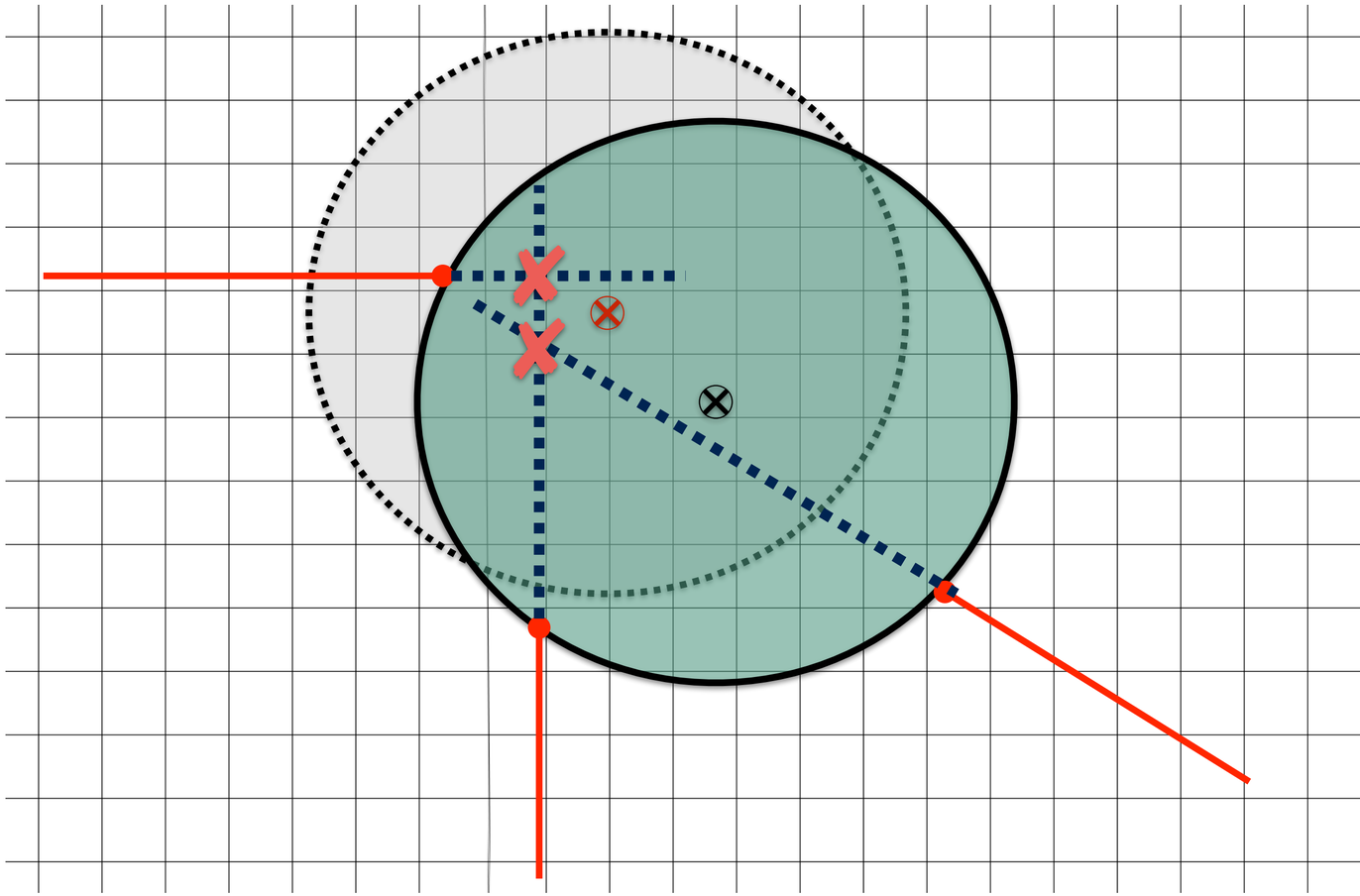
Attack on Spatial Cloaking

- Interpolates the direction of routes as they enter EPZ
- Group all the intersection together that fall within certain threshold and find centroid
- Prediction successful if centroid falls without 50 meters of protected location



Attack on Spatial Cloaking

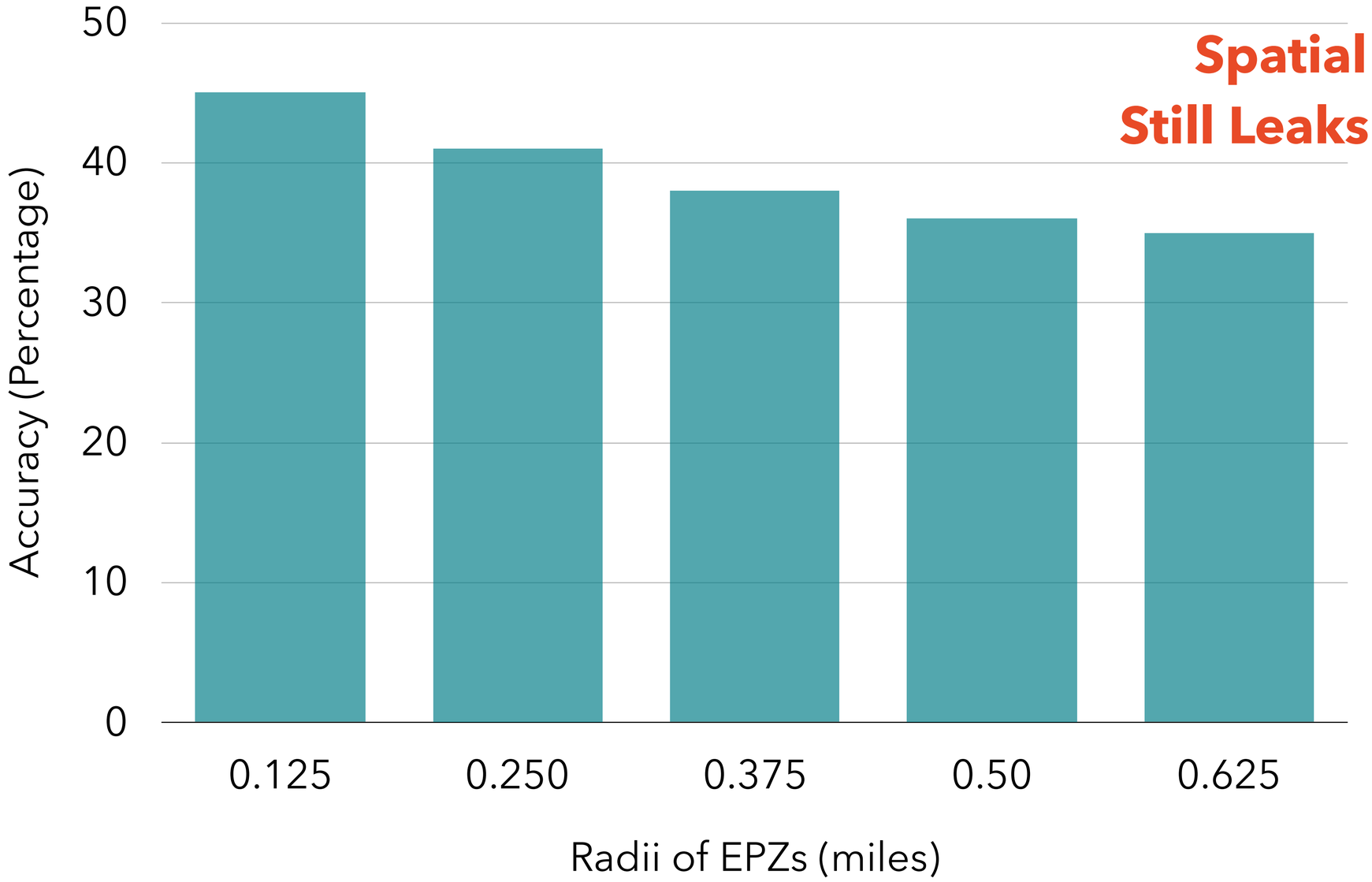
- Interpolates the direction of routes as they enter EPZ
- Group all the intersection together that fall within certain threshold and find centroid
- Prediction successful if centroid falls without 50 meters of protected location



Spatial Cloaking Results

Spatial Clocking
Still Leaks Information

Spatial Cloaking Results



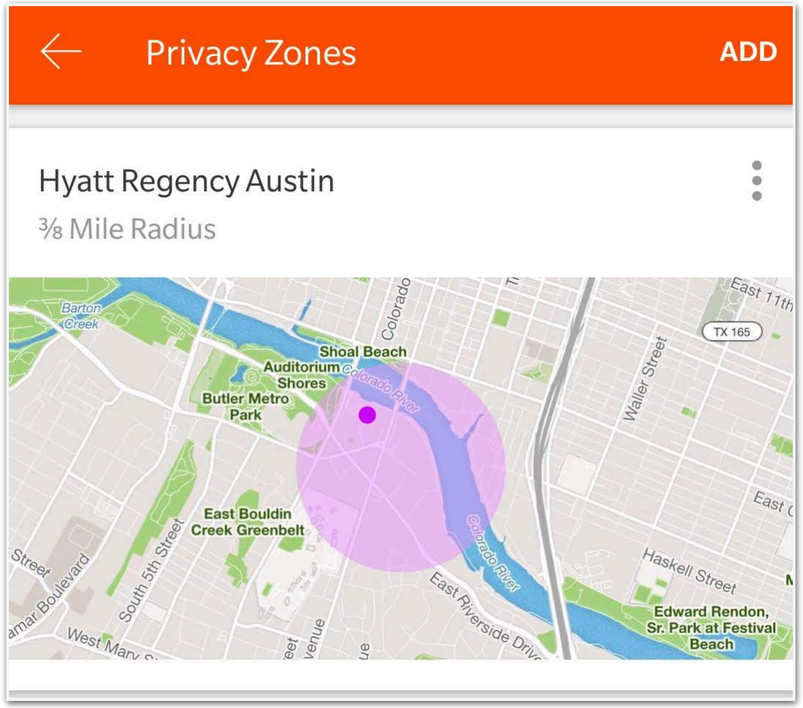
Spatial Clocking Still Leaks Information

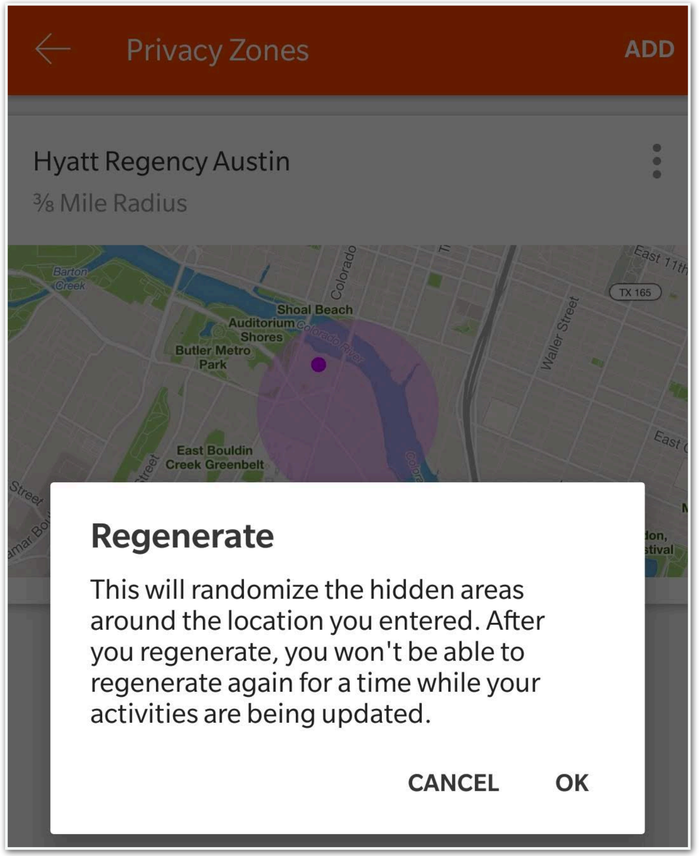
Impact

Our defense was adopted by:

Impact

Our defense was adopted by:





Takeaways

- Demonstrate that EPZ was a bad privacy measure from its infancy
- Intrinsic risk between desire to publish route information while concealing sensitive endpoints.
- Need efficient and usable privacy features in fitness tracking apps

Thanks & Questions

 whassan3@illinois.edu

Backup Slides

Finding EPZ Activity

- Through experimentation we discovered if an activity is using EPZ.
- There is a discrepancy between the advertised distance on the activity page and the final distance traveled according to the GPS samples

Heatmap Incident

THE VERGE

TECH SCIENCE CULTURE MORE

APPS US & WORLD TECH 23

Strava's fitness tracker heat map reveals the location of military bases

Geolocation isn't a new problem for the military

By Andrew Liptak | @AndrewLiptak | Jan 28, 2018, 3:51pm EST

EPZ found in our dataset:

- 1 of 7 athletes at GCHQ
- 1 of 8 athletes at Pine Gap
- 1 of 13 athletes at Kandahar