

A run a day won't keep the hacker away:

Inference Attacks on Endpoint Privacy Zones in Fitness Tracking Social Networks

Karel Dhondt, Victor Le Pochat,

Alexios Voulimeneas, Wouter Joosen, Stijn Volckaert

Running is enjoying a boom because of the coronavirus pandemic

By Allen Kim, CNN

Updated 0953 GMT (1753 HKT) April 25, 2020



REUTERS

SPORT

SEPTEMBER 23, 2020 / 1:03 AM / UPDATED 10 MONTHS AGO

Exclusive: Brits on bikes as fitness app data shows pandemic boom

By Kate Kelland

2 MIN READ



Bloomberg

The Pandemic Bike Boom Hits in Some Unexpected American Cities

Los Angeles and Houston are hardly cycling capitals. But both saw surges in biking after Covid-19 began, according to new data from the fitness app Strava.

By [Laura Bliss](#)

September 23, 2020, 3:00 PM GMT+2



Fitness apps grew by nearly 50% during the first half of 2020, study finds

15 Sep 2020

Carmen Ang

Reporter, Visual Capitalist

Fitness app Strava lights up staff at military bases

⌚ 29 January 2018



Garmin is slowly coming back online after a massive ransomware hack

By Oliver Effron, [CNN Business](#)

Updated 1937 GMT (0337 HKT) July 27, 2020

Fitness app Polar revealed not only where U.S. military personnel worked, but where they lived

By [Rebecca Tan](#)

July 18, 2018 at 10:00 a.m. UTC

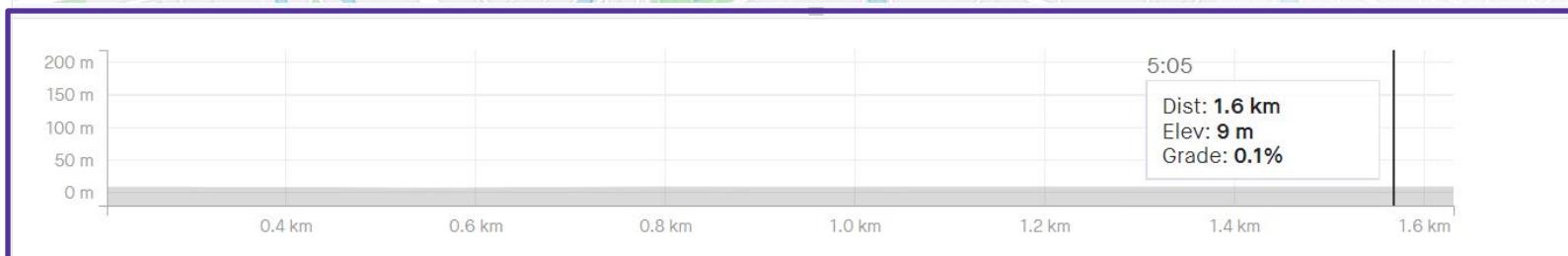
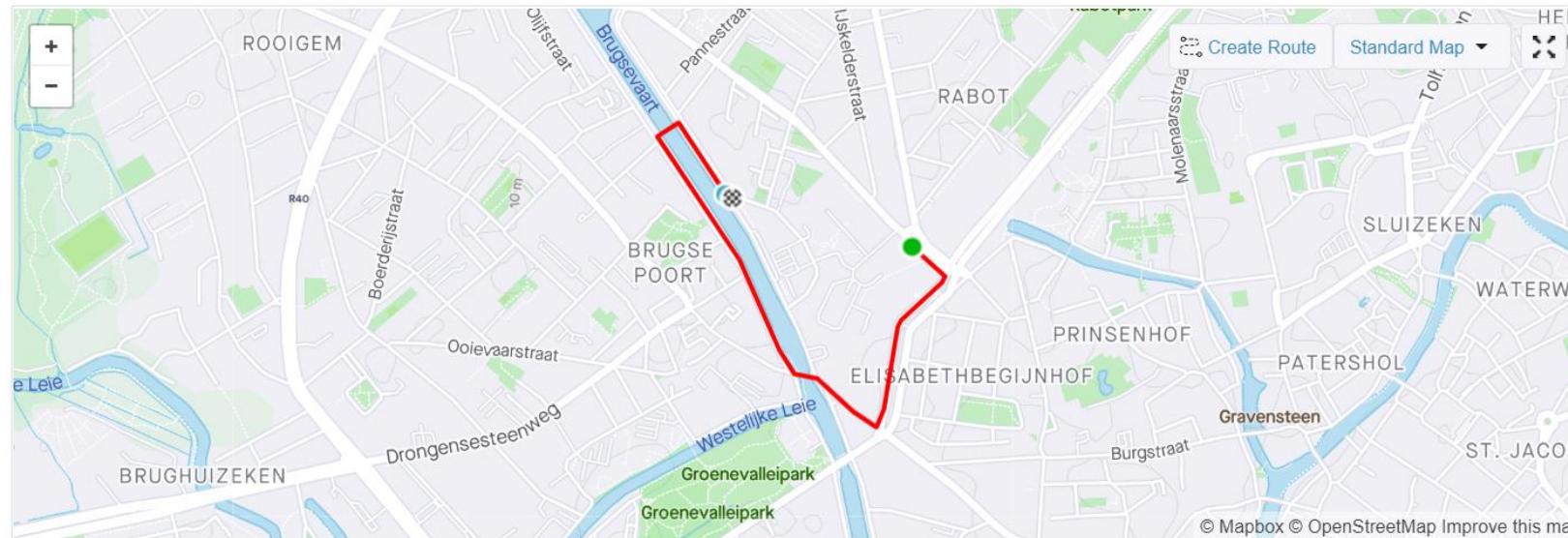
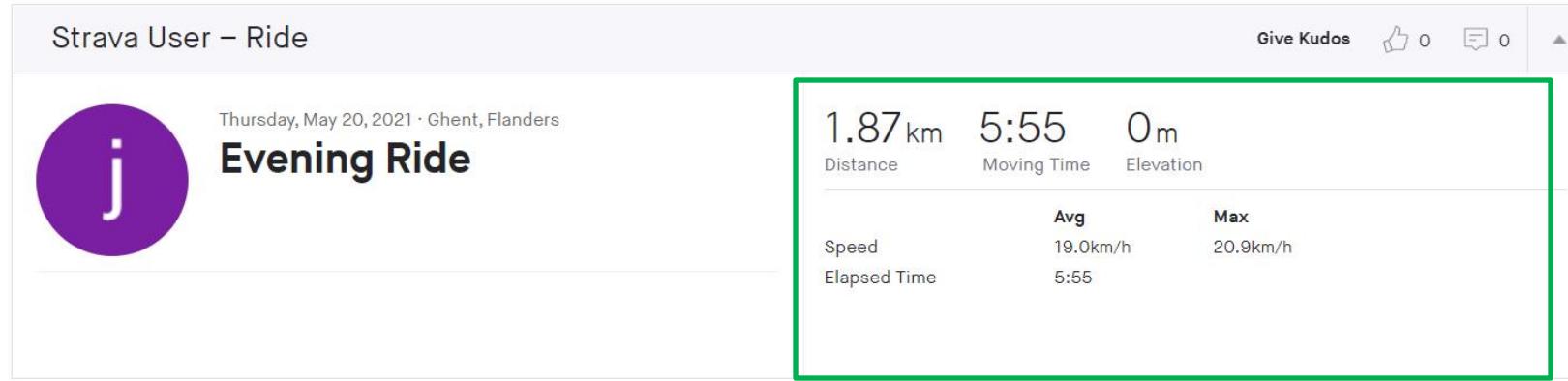


Strava removes automatic flybys after safety concerns

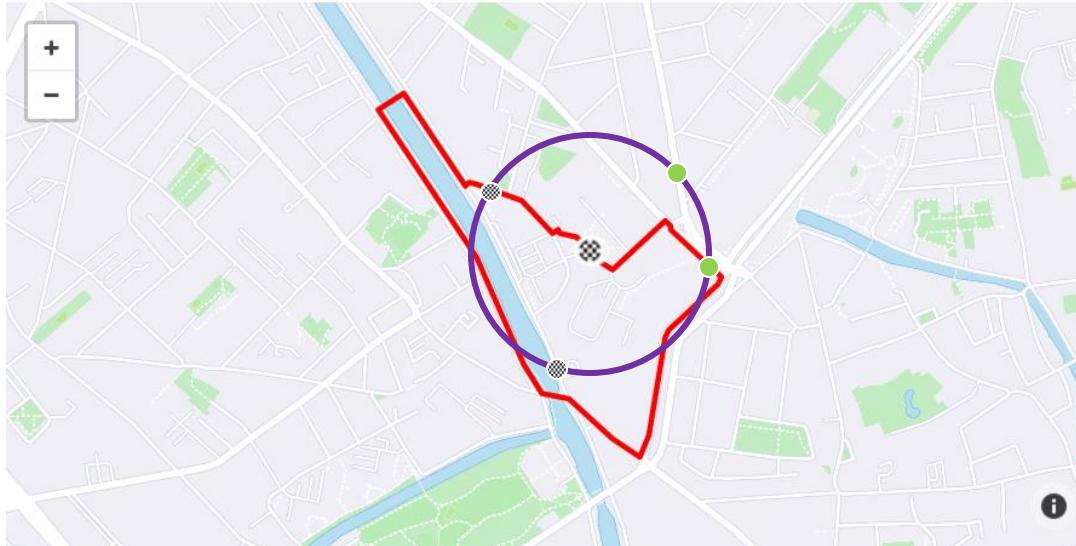
The ride-tracking app has now made the comparison feature opt-in

BY [ALEX BALLINGER](#) OCTOBER 15, 2020

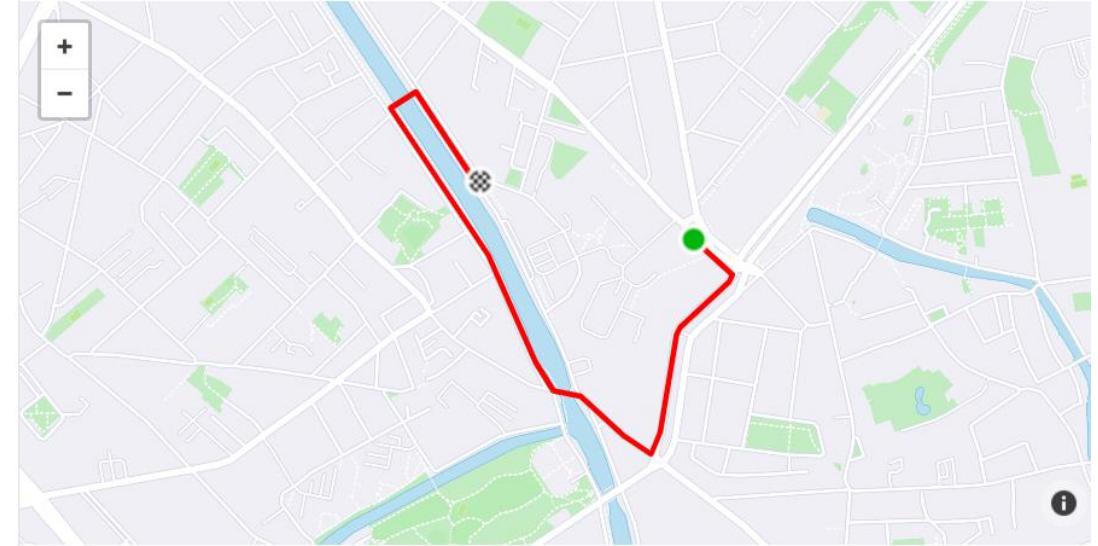
Fitness Tracking Social Networks: Activities



Endpoint Privacy Zones



View of owner of activity



View of user that doesn't own activity

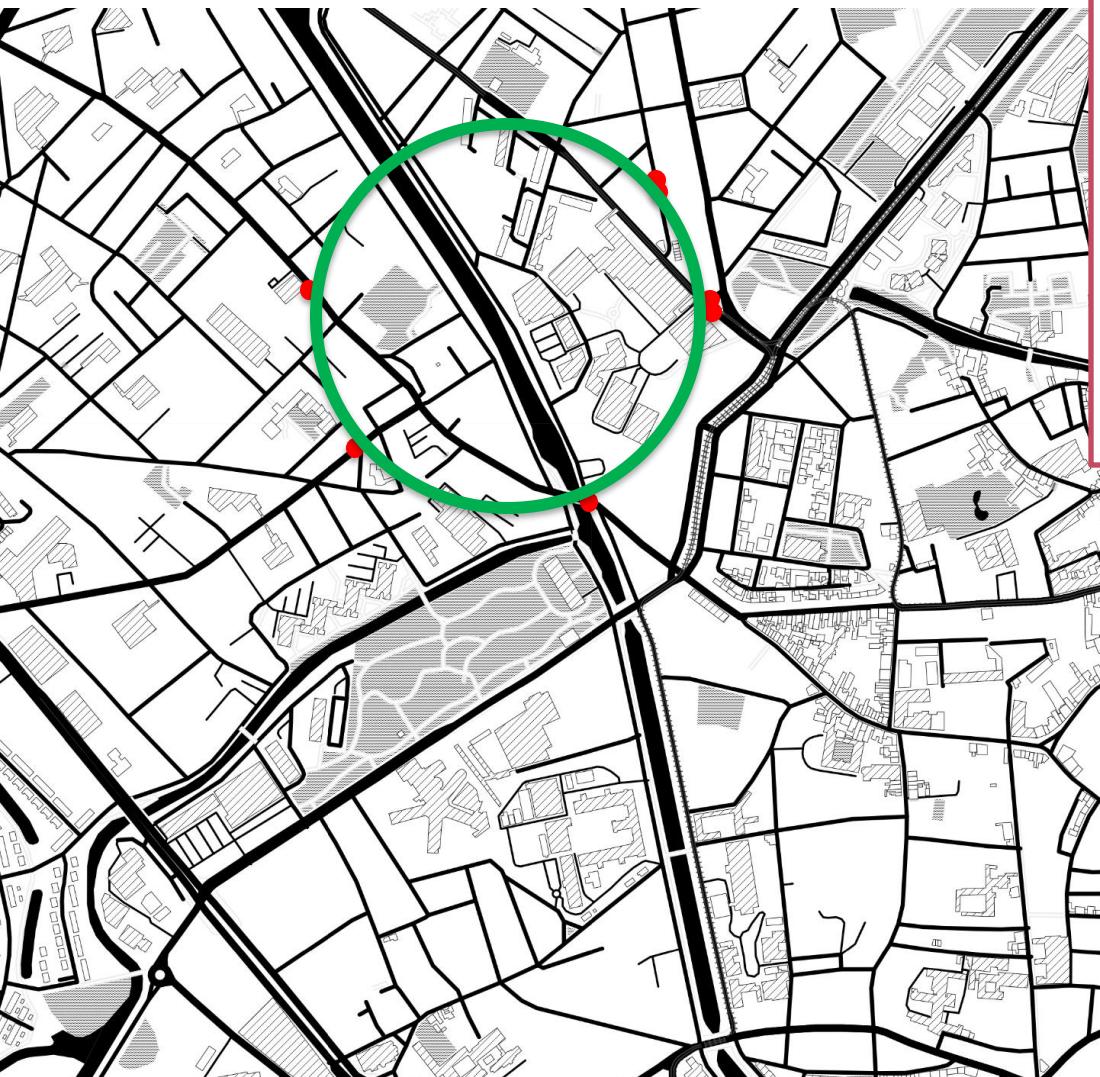
[1] Hassan et al. Analysis of Privacy Protections in Fitness Tracking Social Networks -or- You can run, but can you hide? In USENIX (2018)

[2] GRUTESER et al. Anonymous usage of location-based services through spatial and temporal cloaking. In Proceedings of the 1st international conference on Mobile systems, applications and services (2003)

Attack

- › Threat model
 - capabilities of *regular* user
 - only based on *public* (meta)data
- › Two subproblems:
 1. Discovering EPZs
 2. Finding protected location inside EPZ

Attack: Discovering EPZs



Adaptation of K-Means

repeat

 assign each endpoint to closest fitted circle of cluster

 Isq fit new circle for cluster

until convergence criterium is met



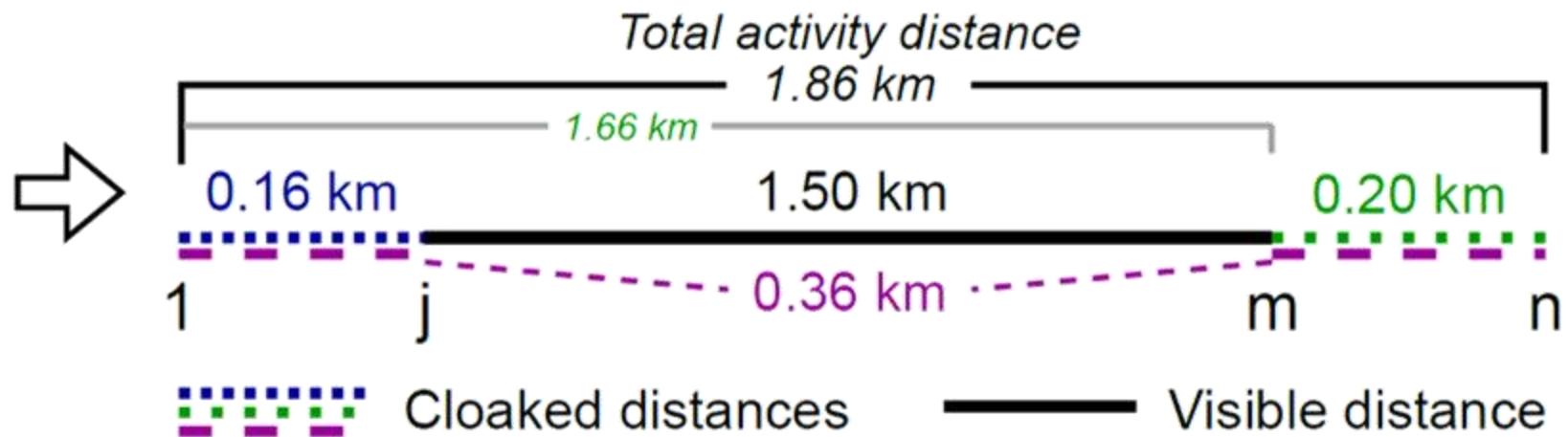
Attack: Protected Location Inside EPZ

› Two scenarios:

1. Inner Distance
2. Total Distance

Activity metadata

total_distance: 1.86,
visible_distances:
[0.16, 0.18,
...,
1.65, 1.66]



Available distances:

Inner distance scenario: 0.16 km + 1.50 km + 0.20 km = 1.86 km
Total distance scenario: 0.36 km + 1.50 km = 1.86 km

Inner Distance Scenario

Strava User – Ride

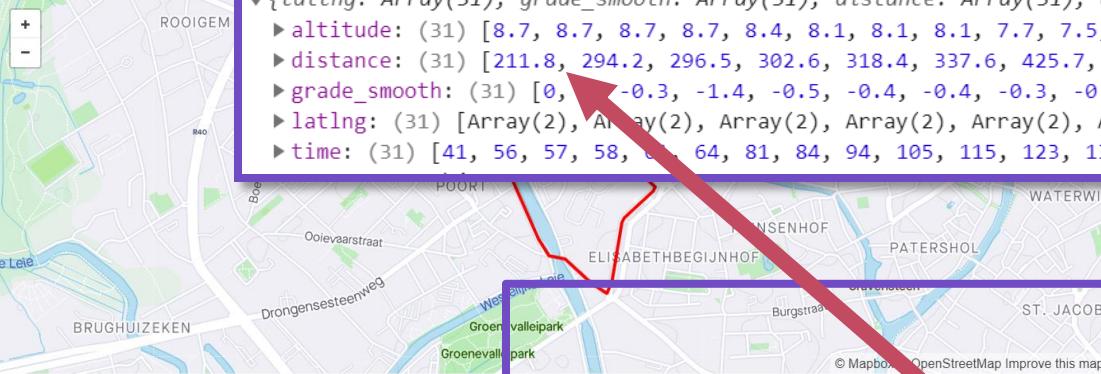
Thursday, May 20, 2021 - Ghent, Flanders
Evening Ride

1.87 km 5:55 0m
Distance Moving Time Elevation

Speed Avg Max
Elapsed Time 19.0km/h 20.9km/h

pageView._streams.streamData.data

```
▼ {latlng: Array(31), grade_smooth: Array(31), distance: Array(31), altitude: Array(31), time: Array(31)} ⓘ
  ► altitude: (31) [8.7, 8.7, 8.7, 8.7, 8.4, 8.1, 8.1, 8.1, 7.7, 7.5, 8, 8.2, 8.3, 8.5, 9, 9.2, 9, 9.1, 9.2, 9.3, 9.1, 9.2, 9.2, 9.3, 9.3, 9.3, 9.4]
  ► distance: (31) [211.8, 294.2, 296.5, 302.6, 318.4, 337.6, 425.7, 440.5, 496.2, 551.9, 607.3, 645.8, 699.2, 737.4, 780.4, 797.7, 844.3, 851.1, 901.8, 982.4, 1063,
  ► grade_smooth: (31) [0, -0.3, -1.4, -0.5, -0.4, -0.4, -0.3, -0.1, 0, 0.3, 0.5, 0.6, 0.7, 0.5, 0.4, -0.3, -0.3, 0, 0, 0.2, 0.2, 0, 0, 0.1, 0.1, 0.1, 0, -0
  ► latlng: (31) [Array(2), Array(2), Array(2),
  ► time: (31) [41, 56, 57, 58, 60, 64, 81, 84, 94, 105, 115, 123, 133, 140, 148, 151, 160, 162, 171, 186, 202, 208, 226, 244, 263, 272, 280, 292, 305, 307, 309]
```



ROOIGEM B40 Oudestraat POORT Drongensesteenweg ELISABETHBEGIJNHOF PATERSHOL ST. JACOBS BRUGHUIZEN Waterwijk Leie Groenvalleipark Groenvalleipark © Mapbox OpenStreetMap Improve this map

Segments

Name	Time	Speed	Power	VAM	HR
Groendreef tussen Noordstraat en Beukelaarstraat	1:32	16.7km/h	—	—	—

Learn more about segments

Inner distance scenario
Distance covered inside EPZ leaked

Total Distance Scenario

- › distance covered inside EPZ = total distance – track distance



Attack

- › Two scenarios:
 1. Inner Distance
 2. Total Distance

	Total Distance Attack	Inner Distance Attack
Strava	✓	✓
Garmin Connect	✓	
Komoot	✓	
Map My tracks	✓	✓
Map My Run	✓	
Ride With GPS	✓	✓

Attack: Finding Protected Locations Inside EPZ

Intuition of attack



Attack: Finding Protected Locations Inside EPZ

Intuition of attack



Attack: Finding Protected Locations Inside EPZ

Preprocessing



Downloaded road graph



Node resolution increased through chaining

Attack: Finding Protected Locations Inside EPZ

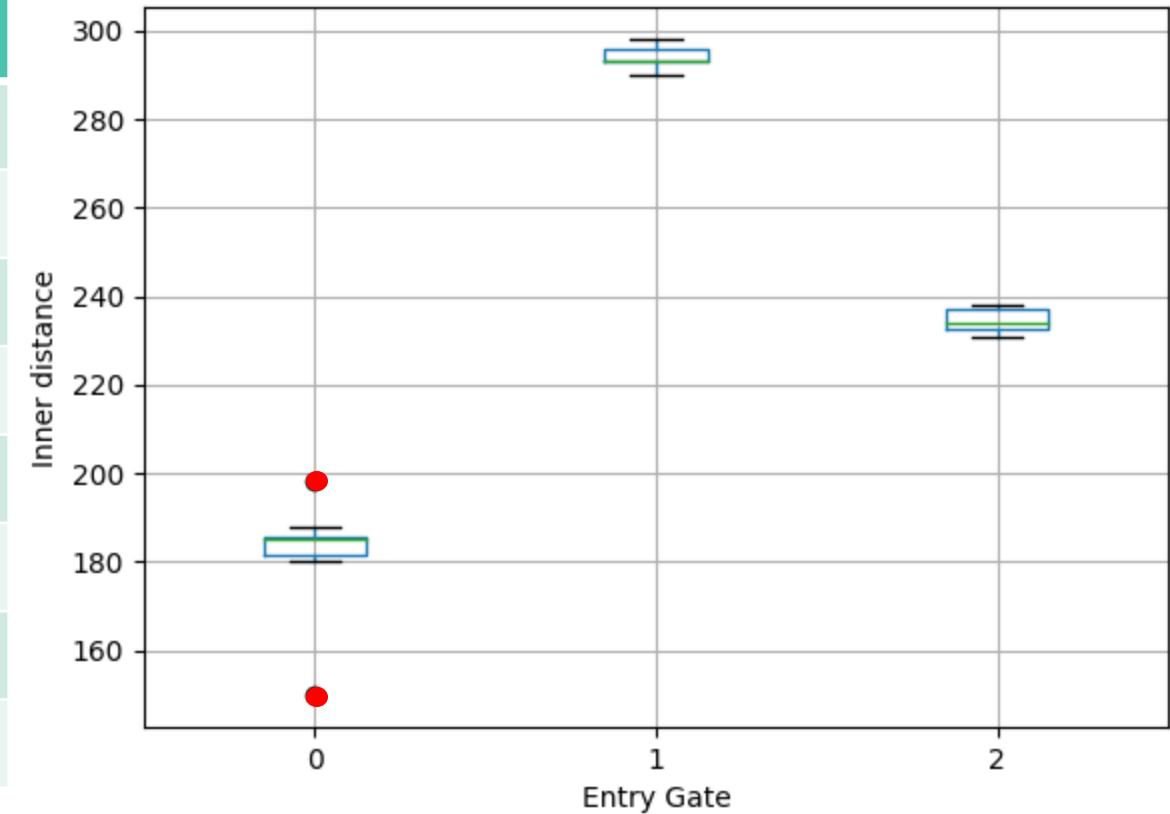
Identifying Entry Gates



Attack: Finding Protected Locations Inside EPZ

Filtering outliers

activity_id	entry_gate	type	inner_distance
1	EG0	START	184.8
1	EG1	END	293.2
2	EG2	START	236.4
2	EG0	END	199.1
3	EG0	START	152.3
3	EG1	END	289.7
...
N	EG0	START	186.9



Attack: Finding Protected Locations Inside EPZ

Predicting Location

- For each node of interpolated road graph:

LAD fit of N observed distances and M theoretical distances

activity_id	entry_gate	type	EPZ_distance
1	EG0	START	184.8
1	EG1	END	293.2
2	EG2	START	236.4
3	EG1	END	289.7
...
N	EG0	START	186.9

Observed Activity Distances

node_id	EG_0	EG_1	EG_2
0	$d_{0,0}$	$d_{0,1}$	$d_{0,2}$
1	$d_{1,0}$	$d_{1,1}$	$d_{1,2}$
2	$d_{2,0}$	$d_{2,1}$	$d_{2,2}$
3	$d_{3,0}$	$d_{3,1}$	$d_{3,2}$
...
M	$d_{M,0}$	$d_{M,1}$	$d_{M,2}$

Theoretical Distances

Attack: Finding Protected Locations Inside EPZ

Predicting Location



Constructing Confidence Intervals

activity_id	entry_gate	type	inner_distance
1	EG0	START	184.8
1	EG1	END	293.2
2	EG2	START	236.4
3	EG1	END	289.7
...
N	EG0	START	186.9

Observed Activities

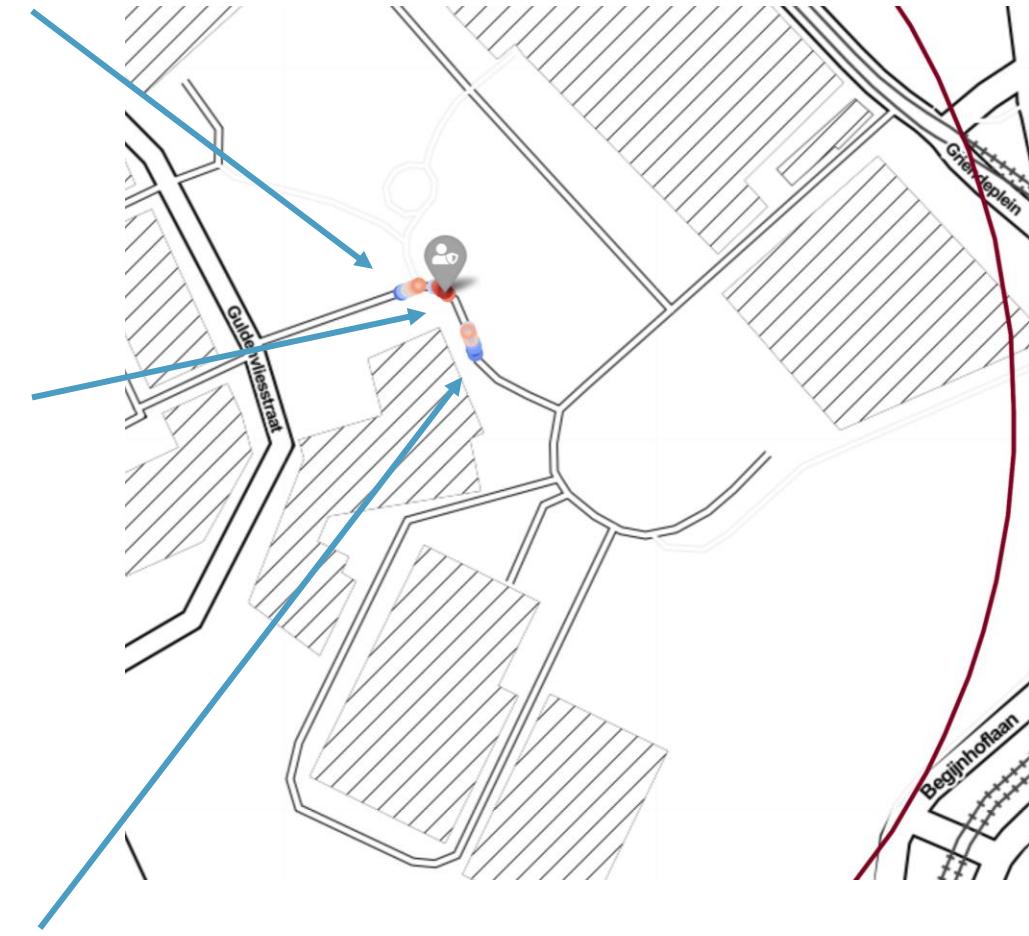
activity_id	entry_gate	type	inner_distance
1	EG0	START	184.8
1	EG1	END	293.2
2	EG2	START	236.4
2	EG2	START	236.4
...
N	EG0	START	186.9

activity_id	entry_gate	type	inner_distance
1	EG0	START	184.8
1	EG1	END	293.2
1	EG1	END	293.2
1	EG0	START	184.8
...
N	EG0	START	186.9

...

activity_id	entry_gate	type	inner_distance
1	EG0	START	184.8
2	EG2	START	236.4
2	EG2	START	236.4
3	EG1	END	289.7
...
N-1	EG0	START	185.3

Resamples



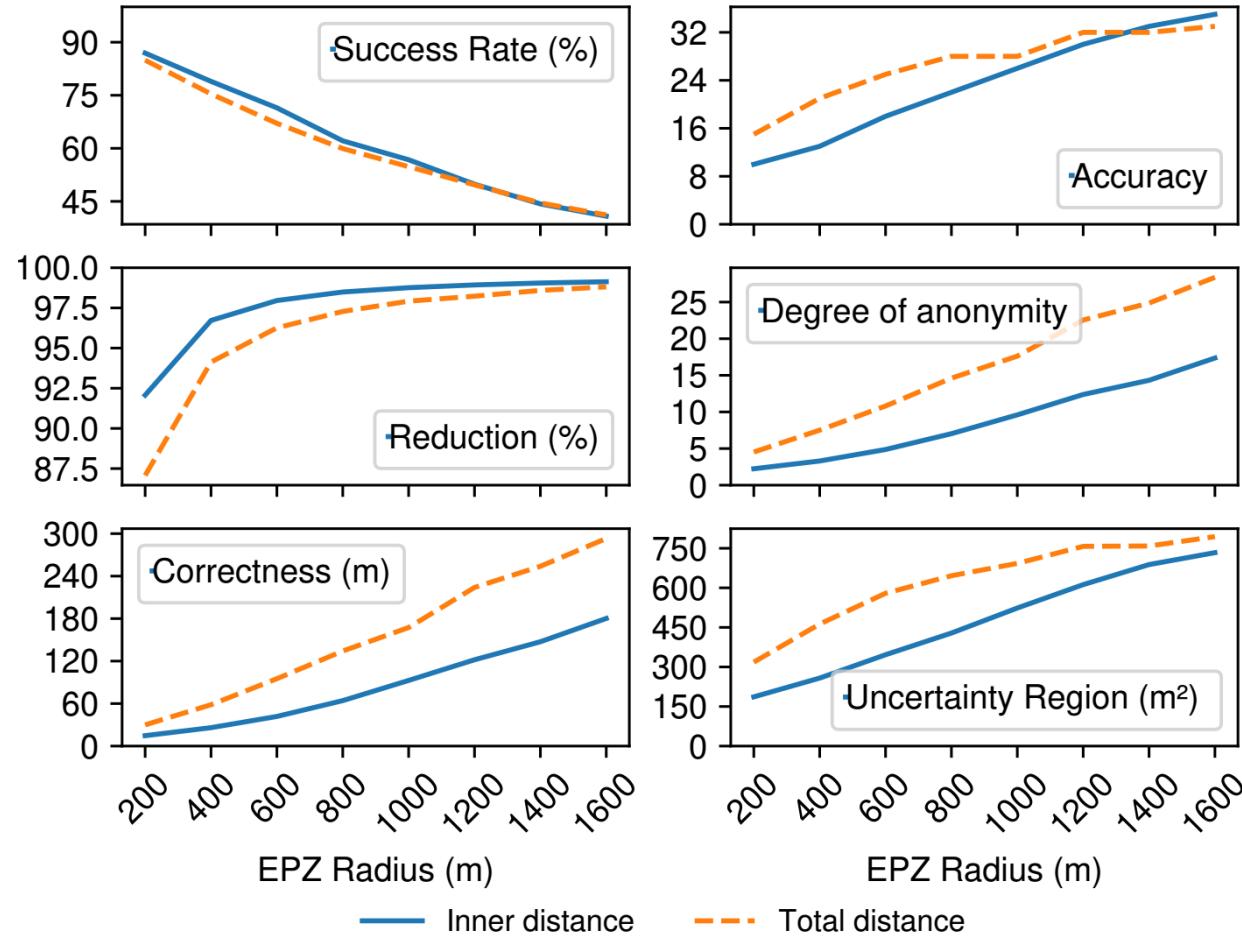
Confidence Interval

Privacy Metrics



- › **Success:** prediction within threshold of GT
- › **Accuracy:** # unique predicted locations
- › **Reduction:** Accuracy / # locations inside EPZ
- › **Correctness:** avg distance between predictions and GT
- › **Uncertainty region:** joint area around predictions

Results



Success: prediction within threshold of GT

Accuracy: # unique predicted locations

Reduction: Accuracy / # locations inside EPZ

Correctness: avg distance between predictions and GT

Uncertainty region: joint area around predictions

Recommendations

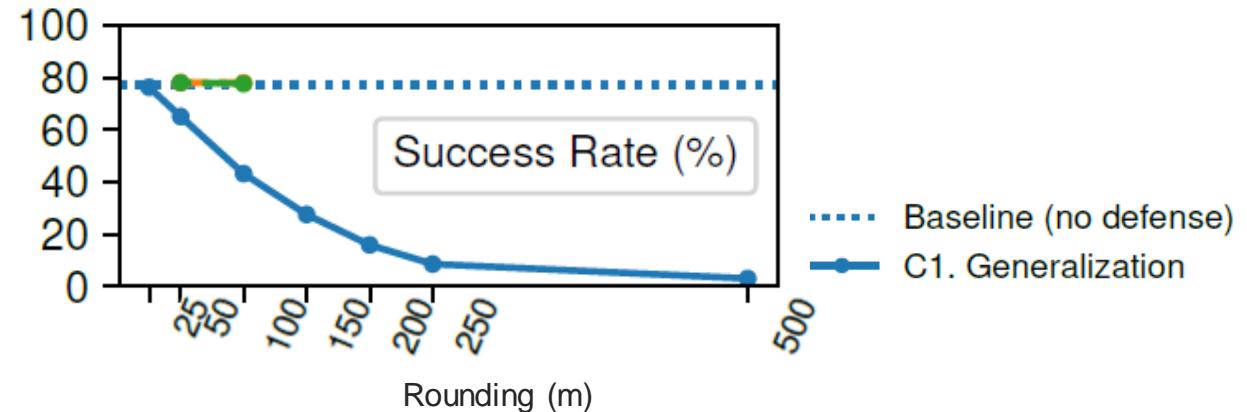
› Data minimization

» *"What you don't have, you can't leak"*

» (On-device) Generalization

» Truncation

- Trade-off with usability: activity gets shorter

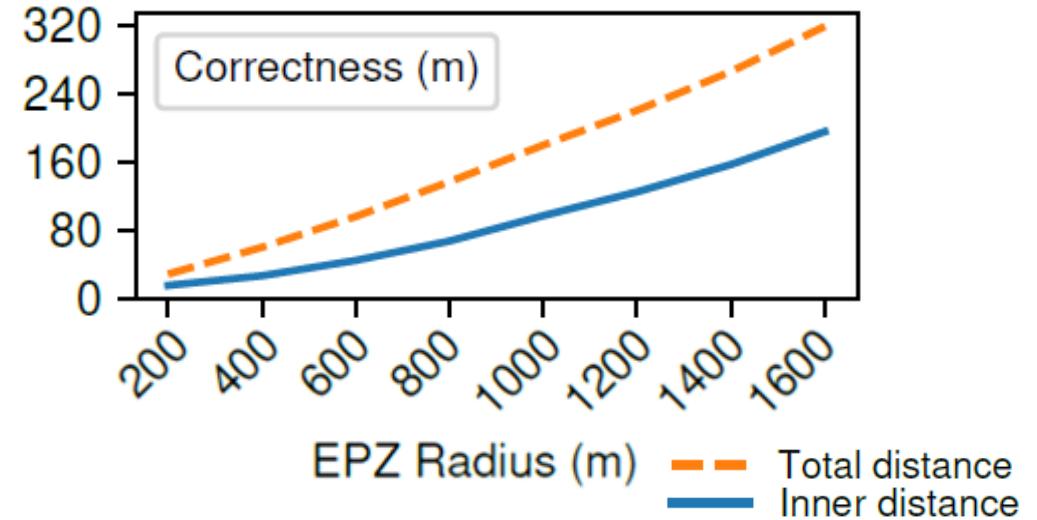


Reflect on data minimization at design time

Recommendations

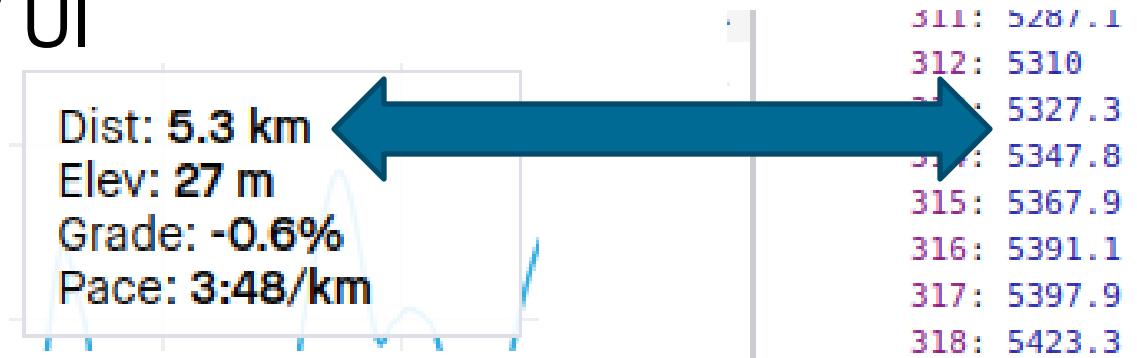
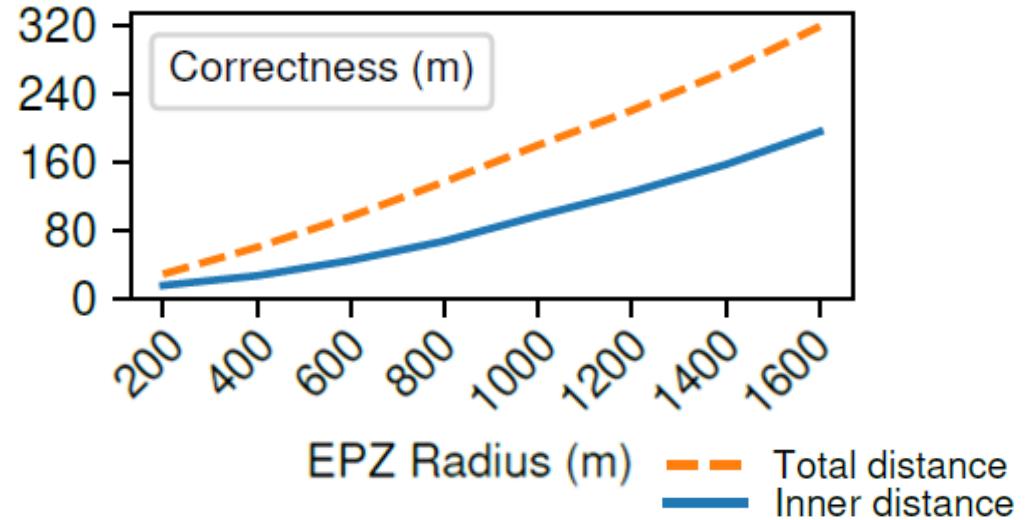
› Data leak prevention

» Avoid inner distance scenario



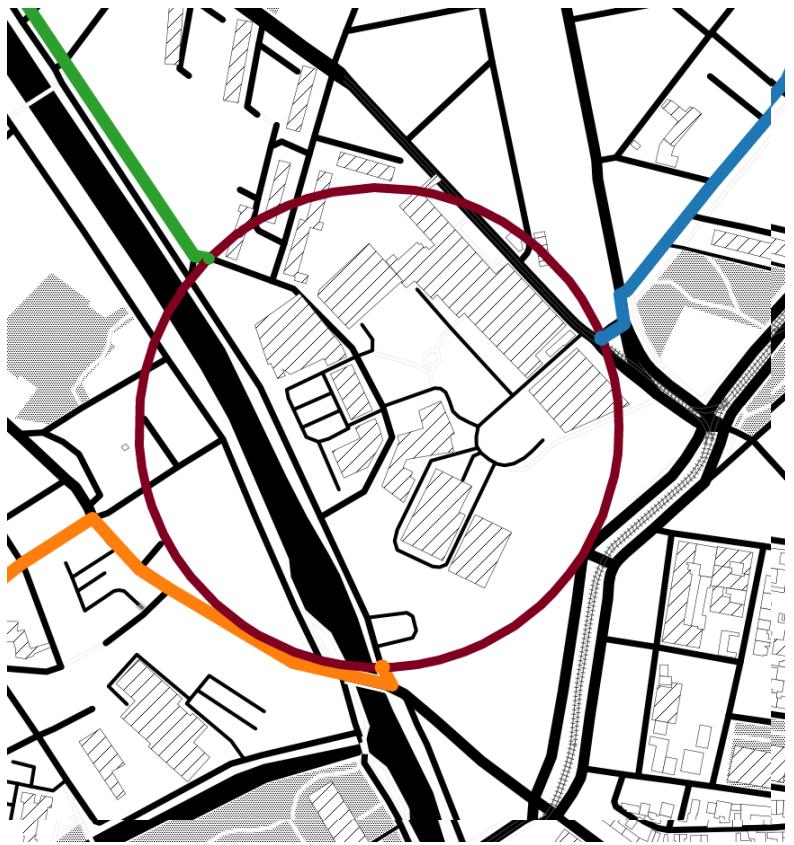
Recommendations

- › Data leak prevention
 - » Avoid inner distance scenario
 - » Fixing API leaks
 - » Matching data precision API / UI



Recommendations

- › Reduce the possibility of inferences



Recommendations

- › Reduce the possibility of inferences
 - » Metadata leaks may enable inferences!
 - » Model and mitigate possible inferences during design
 - »» May require some out-of-the-box thinking



Consider inferences during algorithm design

Recommendations

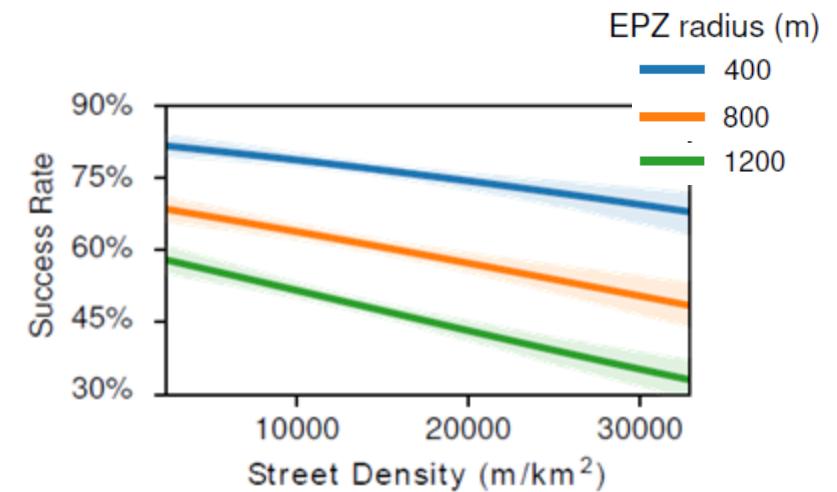
- › Noisy distances?
 - » Random noise distributions average out!
- › Shifting distances?
 - » No influence on total distance scenario!
- › Regenerating EPZs yields more diverse data
- › Smoothing tracks makes regression more accurate



Apparent solutions might not work!

Recommendations

- › Nudge and support users towards privacy-friendly options
 - » Enable privacy zones by default
 - » Suggest EPZ radius based on street density
 - » *Requires effective solutions
that do not violate user privacy perception*

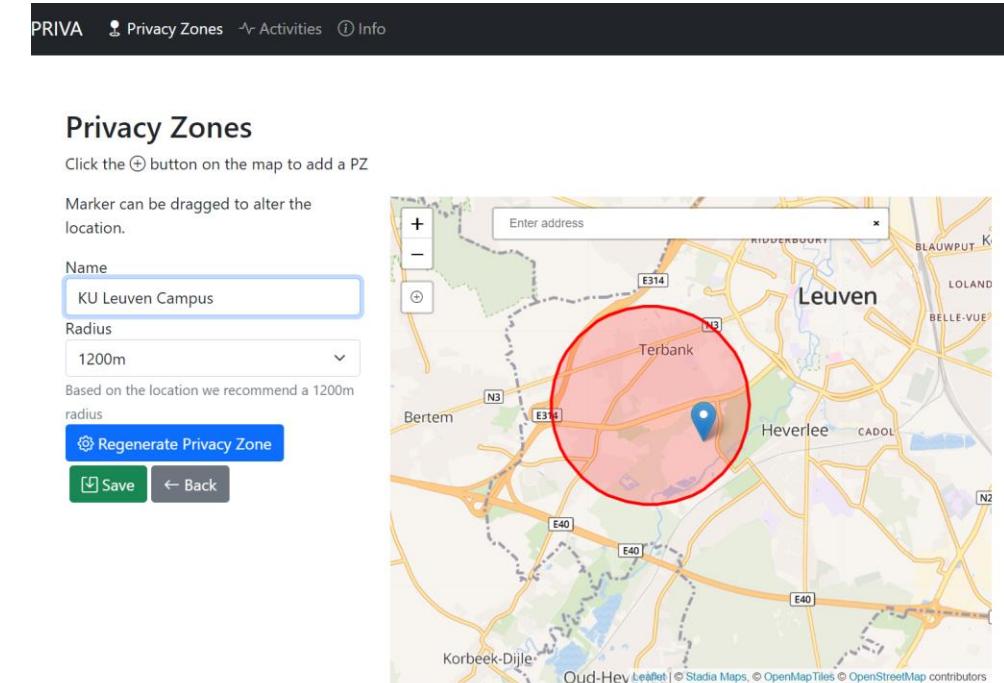


Provide users with clear privacy options

Proof-of-concept Service

› 'Sanitize' sports activities

- » Create privacy zone based on street density
- » Avoiding the "inner distance" scenario
- » Applying generalization
- » Upload sanitized activity to service



<https://priva.distrinet-research.be/>

Disclosure to Networks

- › All affected networks were contacted
- › 3 out of 6 acknowledged our report
- › Strava has engaged in a substantial discussion

Conclusion

- › We develop a novel **inference attack** on privacy zones
- › Intuition: distance metadata + street grid = protected location

Black Hat Sound Bytes

1. Thoroughly test API implementations for leaks 
2. Consider inferences during algorithm design 
3. Provide users with clear privacy options 



Thank you!

karel.dhondt@kuleuven.be

victor.lepochat@kuleuven.be

<https://distrinet.cs.kuleuven.be/>