Location Privacy in Pervasive Computing

Paper by
Alastair R. Beresford and Frank Stajano

Presented by
Nicholas Farnan
11-4-09

The Problem

- Location privacy: the ability to prevent other parties from learning one's current or past location
- As networked computing devices have become increasingly pervasive in modern life, the threat to location privacy has been increasing as well

Why is this a problem?

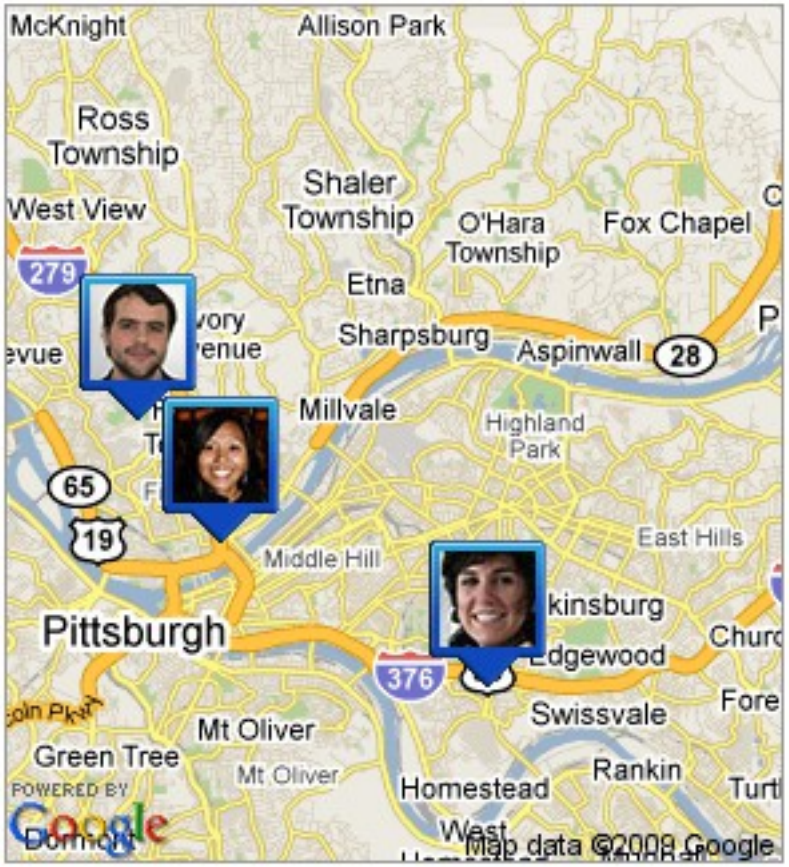
- Did you go to an anti-war rally on Tuesday?
- Were you at Bob Jackson's a week before to plan it?
- Did you walk into an abortion clinic?
- Did you see an AIDS counselor?
- Which church do you attend? Which mosque? Which gay bars?
- Have you been checking into a motel at lunchtimes?
- Why was your secretary with you?
- Did you skip lunch to pitch a new invention to a VC? Which one?

I have nothing to hide!

- Your employer doesn't need to know things about whether, when, and where you went to church.
- Your co-workers don't need to know how late you work or where you shop.
- Your sister's ex-boyfriend doesn't need know how often she spends the night at her new boyfriend's apartment.
- Your corporate competitors don't need to know who your salespeople are talking to.

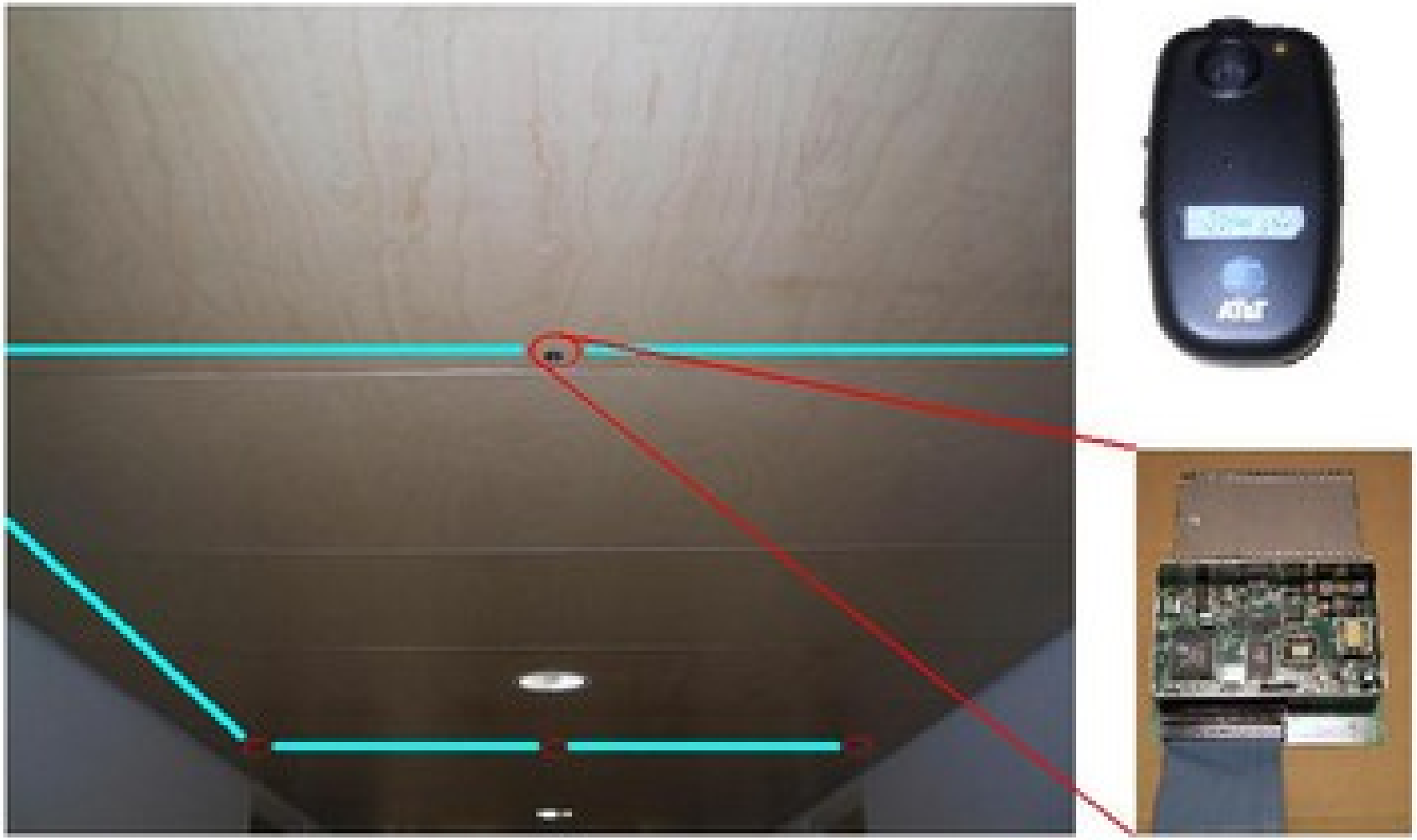
Location-Aware Systems

Google latitude

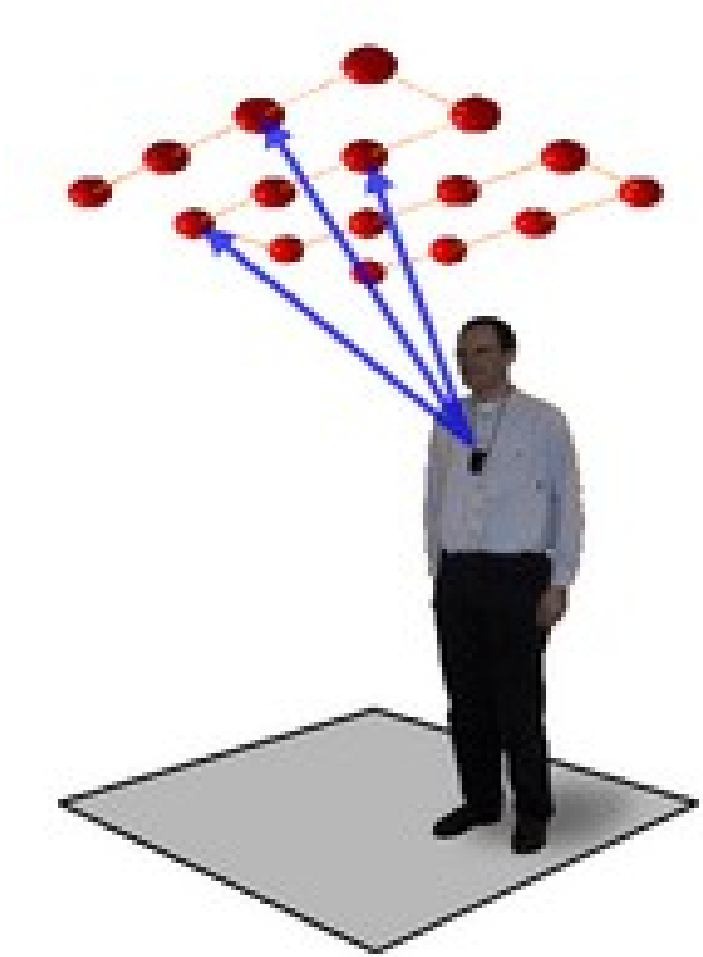


- Allows users to track their friends
- User's locations are updated from smartphones

Active Bat

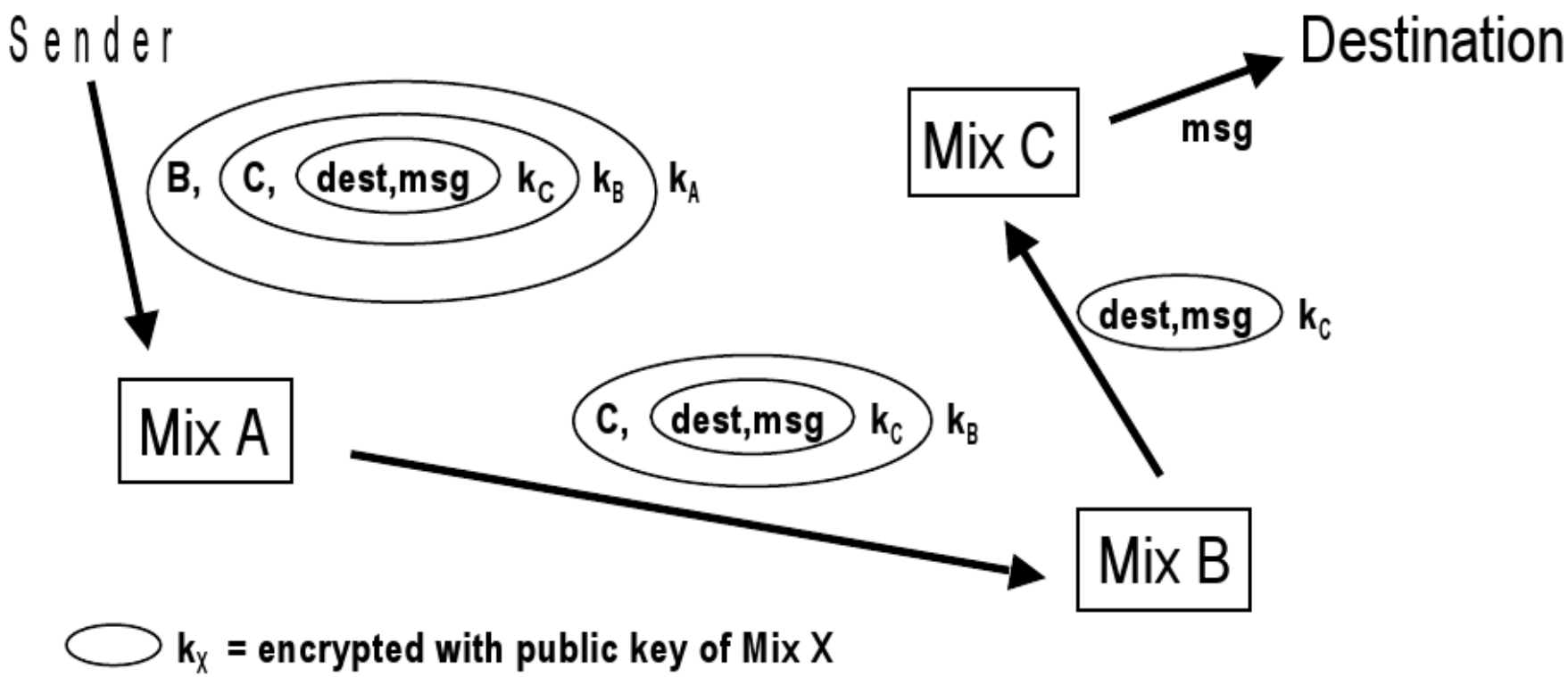


Active Bat



- Bats are polled with some regularity over radio, and emit an ultrasonic chirp
- Ceiling sensors measure time differences in chrips received to determine location

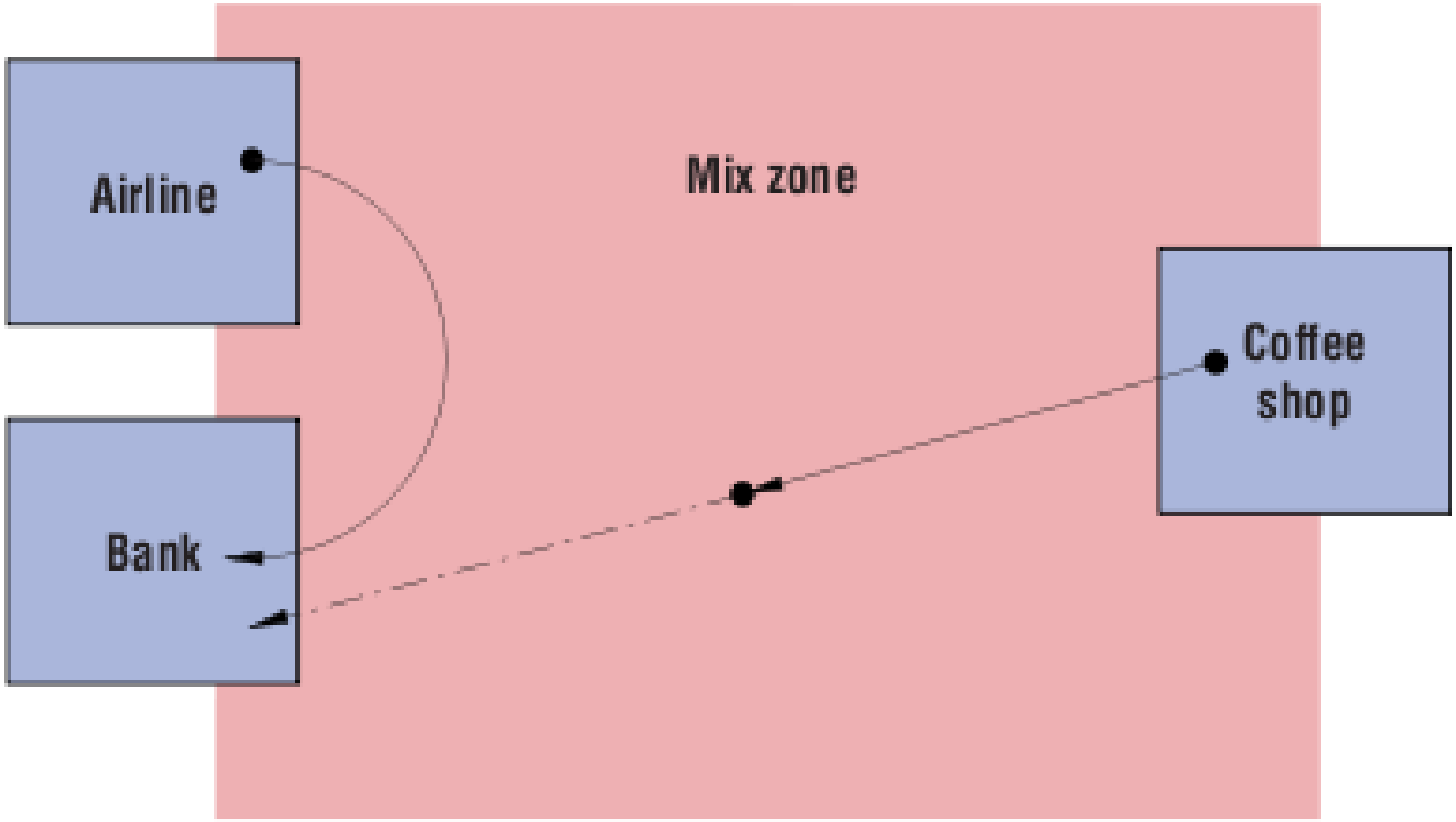
Mix Network


Sender
Destination
Mix C
msg
B, C, dest,msg k_C k_B k_A
dest,msg k_C
Mix A
C, dest,msg k_C k_B
Mix B
k_X = encrypted with public key of Mix X

Mix Zones

- Zones where no users have registered callbacks (where no location-based activity occurs)
- Application zones are just the opposite
- Everytime a user enters a mix zone, they adopt a new pseudonym for interacting with applications

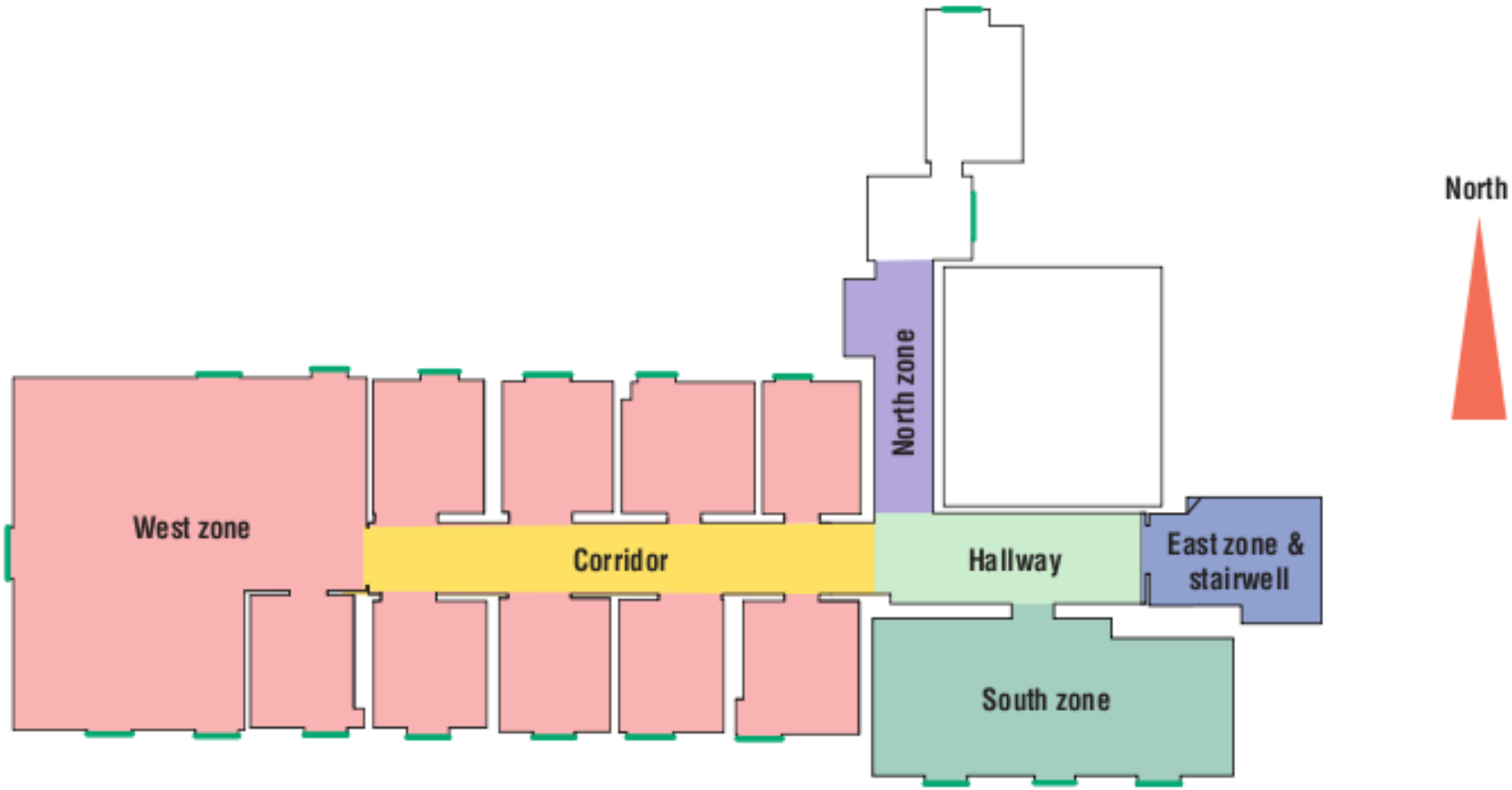
Mix Zones



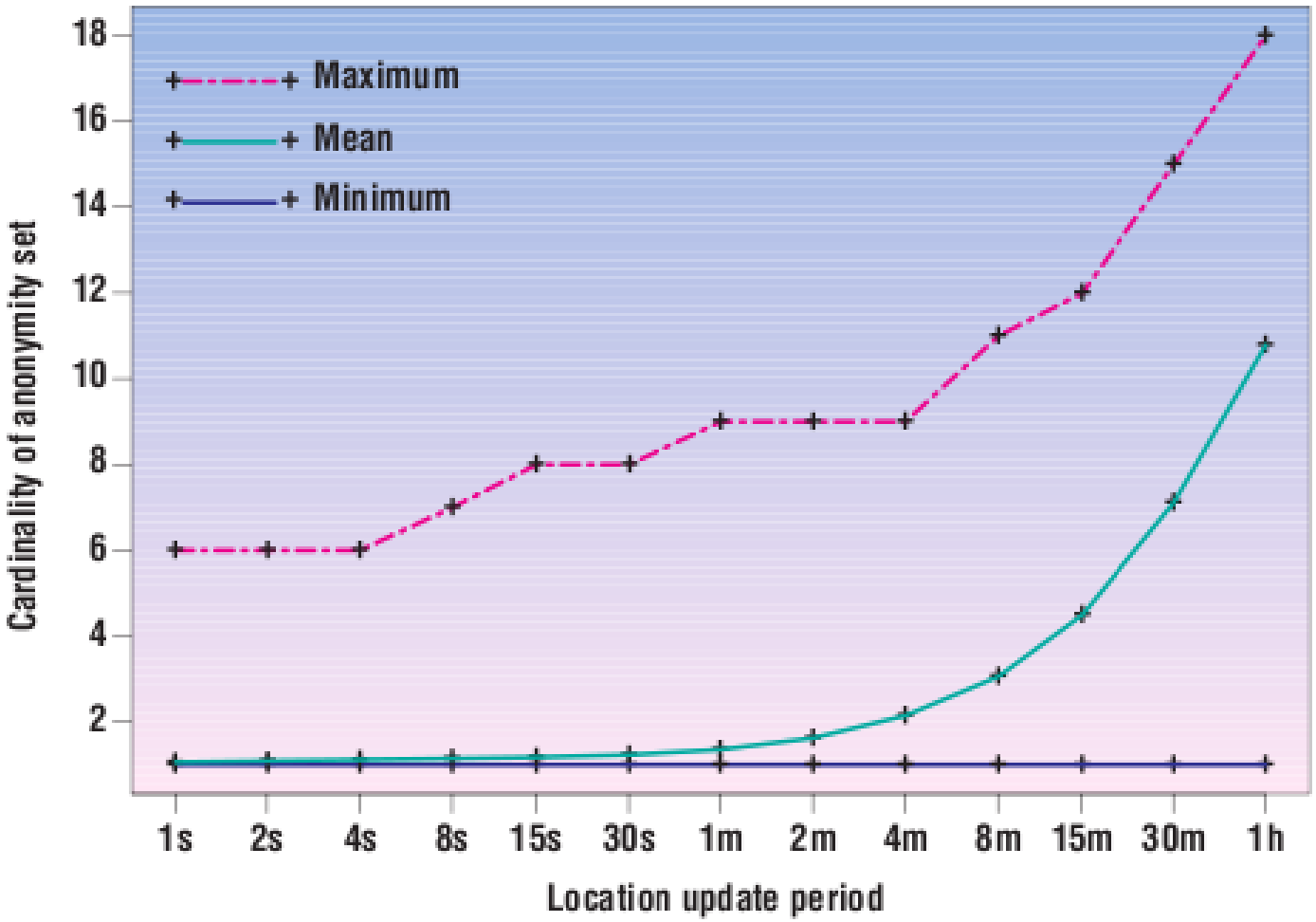
Anonymity Sets

- The *anonymity set* of a given mix zone during a given time period is the set of users in that mix zone during that time period
- Users could be notified of the anonymity set size (average or current) of neighboring mix zones before they begin using location-aware applications in that area

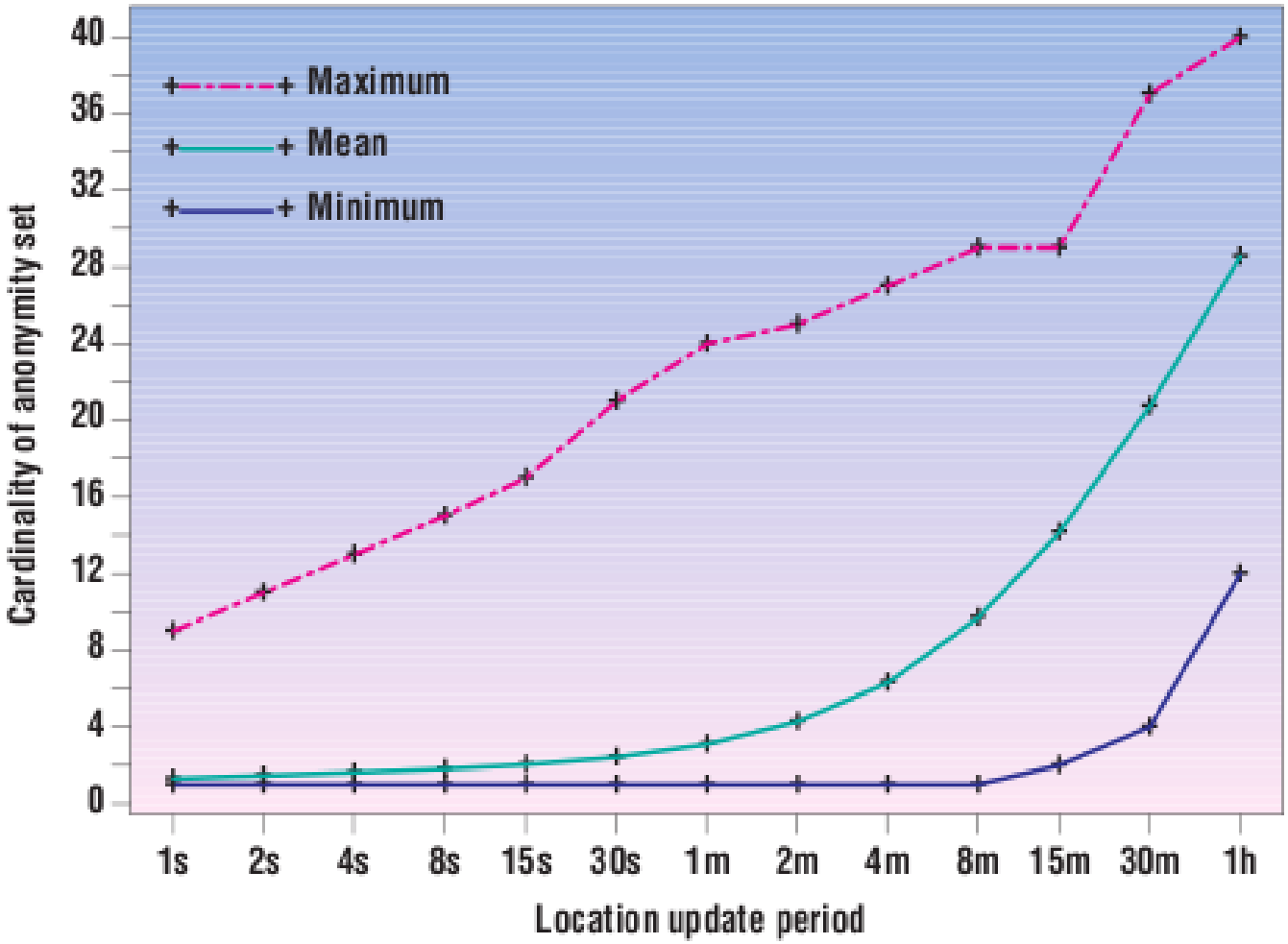
Testbed


North
North zone
West zone
Corridor
Hallway
East zone &
stairwell
South zone

Corridor


Maximum
Mean
Minimum
Cardinality of anonymity set
18
16
14
12
10
8
6
4
2
1s
2s
4s
8s
15s
30s
1m
2m
4m
8m
15m
30m
1h
Location update period

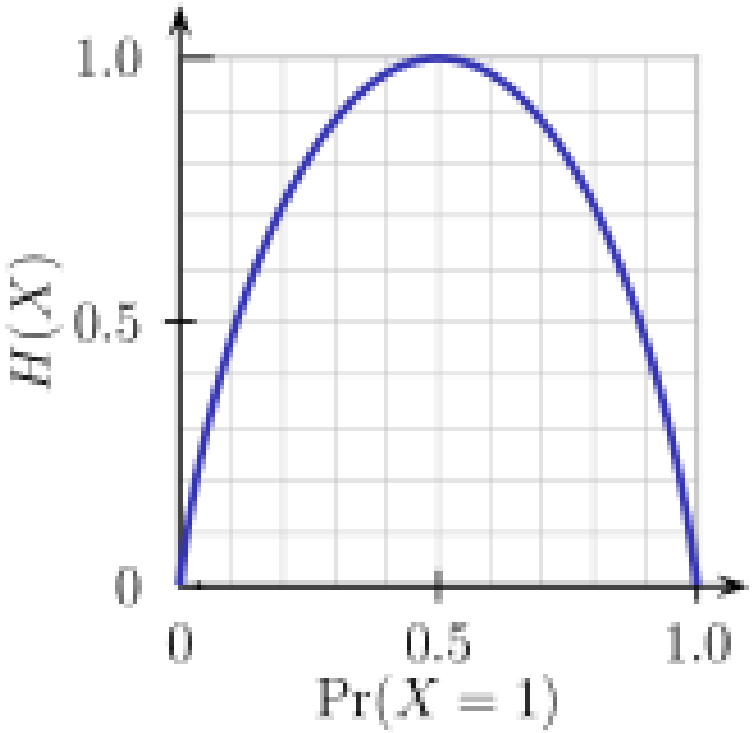
Corridor, Hallway, and Stairwell



User Movement Analysis

- This still doesn't address the problem of an attacker analyzing user movement
- Information theoretic approaches have been used to analyze the anonymity provided by mix networks, so why not use the same idea here?

Entropy



- For a set of *n* equiprobable elements, need lg*n* bits to identify one
- The entropy of file is the theoretical limit to the size of a losslessly compressed version

User Movement Analysis

Preceding zone, *p* \ Subsequent zone, *s*	East	North	South	West	Mix zone
Mix zone	96	47	66	101	814
West	125	13	17	1	43
South	29	55	7	22	30
North	39	6	64	39	66
East	24	47	74	176	162

Entropy Analysis

- Consider colluding applications in the east and west ends of the lab
- Each sees one of its users disappear during a time period t and a new pseudonym appear at the start of $t + 2$ ($t + 1$ assumedly spent in the mix zone)
- Does each have the same user again, or did the users change places?

Entropy Analysis

- *observed* = EEWW v EWWE v WWEE v WEEW
- *uturn* = EEWW v WWEE
- P(EWWE) = P(E, W) * P(W, E), according to historical data presented in the matrix
- P(*uturn* | *observed*) = 0.001
- There is a *99.9%* chance that each has the opposite's previous user, and the pseudonyms can be linked with incredibly high probability

Wait, I Said *Entropy* Analysis

- $H(X) \equiv -\sum_{i=1}^{n} p(x_i) \log_b p(x_i)$

- The entropy of this situation (number of bits needed to describe whether the users switched application spaces or returned to their origins) is 0.012 bits

Future Work

- How are user preferences managed?
  - If you request the price of the same type of coffee, and you're the only one who likes mocha-frappe-latte-chinos, the coffee shop can rather uniquely identify you
- How should a user react to an environment of insufficient anonymity?
- This was a relatively small-scale test, what would happen on, say, a city level?