Growing an Organic Indoor Location System

Jun-geun Park
MIT CSAIL

Joint work with: Ben Charrow (MIT), Dorothy Curtis (MIT), Jonathan Battat (MIT), Einat Minkov (Nokia, Univ. of Haifa), Jamey Hicks (Nokia), Seth Teller (MIT), Jonathan Ledlie (Nokia)
Applications of Location Information

Location information

Location-based recommendation

Friend-finding

Geotagging
WiFi Localization: Survey Phase

- Survey environment to build WiFi fingerprint database
WiFi Localization: Survey Phase

- Survey environment to build WiFi fingerprint database

| Signal strength (dBm) | 0xa3b | 0x5fe | 0xbc4 | 0x6d2 |
|-----------------------|-------|-------|-------|-------|
| 333                   | -55   | -82   | -39   | -85   |
| 334                   | -30   | -65   | -63   | -45   |
| 335                   | -60   | -55   | -50   | -73   |
| 337                   | -72   | -31   | -73   | N/A   |

Expert surveyor
WiFi Localization: Positioning Phase

- Survey environment to build WiFi fingerprint database

|      | 0xa3b | 0x5fe | 0xbc4 | 0x6d2 |
|------|-------|-------|-------|-------|
| 333  | -55   | -82   | -39   | -85   |
| 334  | -30   | -65   | -63   | -45   |
| 335  | -60   | -55   | -50   | -73   |
| 337  | -72   | -31   | -73   | N/A   |

(-31, -66, -60, -40) dBm

“Where am I?”
WiFi Localization: Positioning Phase

- Survey environment to build WiFi fingerprint database

(-31, -66, -60, -40) dBm
“Room 334”

| Signal strength (dBm) | 0xa3b | 0x5fe | 0xbc4 | 0x6d2 |
|-----------------------|-------|-------|-------|-------|
| Room 333              | -55   | -82   | -39   | -85   |
| Room 334              | -30   | -65   | -63   | -45   |
| Room 335              | -60   | -55   | -50   | -73   |
| Room 337              | -72   | -31   | -73   | N/A   |
Organic Indoor Localization: Motivation

- Who makes the location fingerprints?
  - Survey requires skilled technicians.
  - Survey is expensive and labor-intensive.
  - “I don’t want strangers in my room.”
  - Surveyed data may become outdated.
Organic Indoor Localization: Motivation

- Who makes the location fingerprints?
  - Survey requires skilled technicians.
  - Survey is expensive and labor-intensive.
  - “I don’t want strangers in my room.”
  - Surveyed data may become outdated.

- Our approach
  - Have users collect survey data
  - System facilitates sharing on-line.

- User-generated, or organic localization system
Organic Indoor Localization

I’m in Room 334

Signal strength (dBm)

| 0xa3b | 0x5fe | 0xbc4 | 0x6d2 |
|-------|-------|-------|-------|
|       |       |       |       |
| 334   | -30   | -65   | -63   |
|       | -45   |       |       |

0xa3b
0x5fe
0xbc4
0x6d2
Organic Indoor Localization

I’m in Room 337

Signal strength (dBm)

|       | 0xa3b | 0x5fe | 0xbc4 | 0x6d2 |
|-------|-------|-------|-------|-------|
| 334   | -30   | -65   | -63   | -45   |
| 337   | -72   | -31   | -73   | N/A   |

I’m in Room 337
Organic Indoor Localization

~5 meter avg. distance error

Wikipedia-like Pareto principle in user-contribution

See e.g.: Teller et. al., *Organic Indoor Location Discovery*, Griswold et. al., *ActiveCampus*, Boliger et. al., *Redpin*, Barry et. al., *Long-duration study of user-trained 802.11 localization*
Organic Indoor Location System

Survey-based Location Systems

Survey → Location System → End-user

Organic Location Systems

Location System → End-user
Growing an Organic Indoor Location System

Location System

End-user

Facilitating organic growth of location database

Weeding out erroneous user-inputs
Growing an Organic Indoor Location System

Facilitating organic growth of location database
Conveying Spatial Uncertainty to Users

- At early stage of organic localization, some locations have no fingerprint data
If a user is in a location without fingerprint, localization algorithm will pick one of nearby locations with fingerprint.
Conveying Spatial Uncertainty to Users

- If a user is in a location without fingerprint, localization algorithm will pick one of nearby locations with fingerprint.
Voronoi Diagrams

Voronoi site

Voronoi region
Voronoi Diagrams for Conveying Spatial Uncertainty to Users
Voronoi Diagrams for Conveying Spatial Uncertainty to Users

- Can derive *spatial uncertainty* metrics:
  - Number of rooms / geometric size of the region
Voronoi Diagrams for Conveying Spatial Uncertainty to Users

- Can derive *spatial uncertainty* metrics:
  - Number of rooms / geometric size of the region
- Users get graphical feedback on system’s uncertainty arising from organic growth of location database
GUI Implementation

Voronoi region

Location estimate
Spatial-Uncertainty-Based User Prompting

- Prompt user for location input if spatial uncertainty is too high (large Voronoi region)
  - Many nearby rooms have no fingerprint data

- Other methods for acquiring user input
  - Prompting when localization estimate is unstable
  - Voluntary user contribution

- Users can postpone or turn off prompting
Voronoi Evaluation: Setup

- Compared Voronoi-based user prompting to other basic methods
  - Quantitative analysis by simulation

- Real-world user testing
  - Qualitative analysis by interviewing users
Voronoi Evaluation (1)

The graph shows the coverage or number of binds over time for different scenarios:

- **Voronoï (active)**: The solid green line represents the active Voronoi model. The coverage decreases steadily with time.
- **Voronoï (passive)**: The dotted blue line indicates the passive Voronoi model. It also shows a steady decrease in coverage but at a different rate compared to the active model.
- **Inverse coverage**: The dashed cyan line represents the inverse coverage, which decreases from 0.0 to 0.99 as time progresses.
- **Periodic**: The red dashed line indicates the periodic scenario. The coverage decreases at a consistent rate throughout the time period.

The y-axis represents the coverage or number of binds, while the x-axis represents time in rounds, ranging from 0 to 300.
Responses from top contributors:

“Prompts were the main reason that I made so many binds.”

“Voronoi regions were useful for quickly locating the room that I was in as well as assessing how well the tablet knew my current location.”

“Prompting mechanism had no effect on my behavior.”
Growing an Organic Indoor Location System

- Facilitating organic growth of location database
- Weeding out erroneous user-inputs
Erroneous User Input Filtering: Problem Statement

- Erroneous user inputs result in localization error
Erroneous User Input Filtering: Problem Statement

- Erroneous user inputs result in localization error
Erroneous User Input Filtering

- Common approaches for outlier detection...
  - Density estimation
  - Clustering + majority vote
- ... are not suitable for organic location systems. Why?
  - Organic systems have no data at start
Erroneous User Input Filtering

- Common approaches for outlier detection...
  - Density estimation
  - Clustering + majority vote
- ... are not suitable for organic location systems. Why?
  - Organic systems have no data at start

- Our idea: instead of checking validity directly, check for consistency
  - WiFi scans from nearby locations tend to be similar
  - Given a set of scans from a single location, choose the most consistent subset w.r.t. physically adjacent locations
Erroneous User Input Filtering

- **Step 1: Hierarchical clustering**
Erroneous User Input Filtering

- Step 1: Hierarchical clustering

```
Room A

Cluster B

Two clusters of user inputs
```

AP 1
AP 2
AP 3
AP 4
AP 5

Cluster A

Cluster B
Erroneous User Input Filtering

- **Step 2: Pick the most consistent cluster**

Nearby Room B

Room A

Cluster B

Farther

Cluster A

Closer

Nearby Room C
Erroneous User Input Filtering: Result

- Filtering performance improves with additional data

- If 20~30% of user inputs are erroneous, filtering improves the number of spot-on localization estimates by up to 9%

- Refer to our paper for details
Conclusion & Future Work

**Conclusion**

- Organic localization eliminates survey effort while achieving comparable accuracy
- Organic localization can be improved by adequate methods to facilitate organic process
- Voronoi-diagram-based method for conveying uncertainty and user-prompting
- Clustering-based method for discarding erroneous user inputs

**Future work**

- Adapts to environmental changes (e.g. AP upgrades)
- Handle device diversity
- Combine with “organic” mobile applications
Thank you.

Questions?
Physical Distance vs. Signal Distance

Normalized signal-space Euclidean distance

\[ d_s(b^s, b^t) = \left[ \frac{1}{M} \sum_{i=1}^{k} (b^s_i - b^t_i)^2 \right]^{1/2} \]
Erroneous User Input Filtering: Result

- Filtering performance improves with additional data
- Filtering improves accuracy of location estimates
Clustering Threshold Tuning

- $H_0$: User inputs are from the same location
- $H_1$: User inputs are from different locations
- Select $H_0$ if: $P(H_0|d) > P(H_1|d)$.
User Deployment Statistics

- **9-day user deployment**

| Map Spaces                  | 1,373 |
|-----------------------------|-------|
| Contributing Users          | 19    |
| Bind Intervals (from users) | 604   |
| Scans (from devices)        | 1,142,812 |
| Bound Scans                 | 108,418 (9.4%) |
| Spaces with Bound Scans     | 116 (8.4%) |

- Previous user deployment for 20-days showed similar characteristics
User Deployment Result

- Accuracy over time
  - Pre-installed tablets

- Amount of user input over time
User Deployment Result

- Distribution of per-user contribution
User Deployment Result: Coverage

- Day 1

- Day 9
System Architecture

Clients (Nokia N810 Tablets)

RF scanner
- unbound scans

Cache
- binds
- coverage
- estimates

Localizer

AP

Web server
- relevant signatures

OIL server
- aggregated scans/binds

Database

Comms