A MEASUREMENT STUDY FOR UNDERSTANDING WIRELESS FORENSIC MONITORING

Yongjie Cai, Ping Ji
John Jay College of Criminal Justice & Graduate Center, City University of New York
Outline

- Background
- Experiment setup
- Observations
- Future work
Increasing number of WiFi and mobile devices

According to wigle.net, 69.1M on 7/31/12, 73.3M wifi reported on 9/20/12
Background – Forensic Challenges

- Lack of traceability through WiFi nets
  - (Open) WiFi – hacker comes and goes
  - Usually no logs kept at APs
- Uncertainty of device identification
  - Mac spoofing, etc.
- Uncertainty of device location
  - Densely populated WiFi in Metropolitan neighborhood – difficulty in localizing device both horizontally and vertically
A Measurement study to

- Explore Metropolitan WiFi nets characteristics
- Explore representative Device Localization Approaches
  - K Nearest Neighbor (KNN): a fingerprinting based approach
  - Log-distance path loss model: a model based approach
Outline

- Background
- Experiment setup
- Observations
- Future work
Equipment Setting

- **Setting I**
  - MacBook Pro laptop with Mac OS X 10.7.4
  - integrated wireless card, BU353 GPS receiver
  - Kismet (Passive Scan)

- **Setting II**
  - MacBook Pro laptop, Backtrack in Virtual Machine running
  - Alfa wireless card, BU353 GPS receiver
  - Kismet

- **Setting III**
  - HTC Stream Tablet, Android,
  - integrated wireless card, integrated GPS receiver
  - **WiFum** (Active Scan)
Testing Path and Duration

- Three-block urban neighborhood of midtown west Manhattan
- Three data sets per test run
Outline

- Background
- Experiment setup
- Observations
- Future work
# Number of AP

<table>
<thead>
<tr>
<th>Tool</th>
<th>Airport + Kismet</th>
<th>Alfa + Kismet</th>
<th>HTC + WiFum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average # of AP/Position</td>
<td>15</td>
<td>11</td>
<td>24</td>
</tr>
<tr>
<td>Average # of AP/Trip</td>
<td>1823</td>
<td>1829</td>
<td>1520</td>
</tr>
<tr>
<td>Average percentage of Encrypted APs</td>
<td>0.706</td>
<td>0.6835</td>
<td>0.8205</td>
</tr>
</tbody>
</table>
Device Manufactory

3807 Access Points and 6615 Mobile Devices from Kismet traces
Average Received Signal Strength (RSS):
External Alfa card > Cards of Mac computer & HTC tablet
- 362,305 packets from Kismet traces
- 50% beacons from APs, 20% CTS & ACK, 11% Probe req/rep
- only approx 3% data
Location Variance

- Choose 10 locations along the path
- At each location gather trace continuously for 2 mins
- Record GPS coordinates from three different experiment settings
- Center Point = (avg latitude, avg longitude)
- Error Distance = |empirical location – center point|
Average Error Distance

- **External GPS (BU353)** provides worse consistency in location estimation.
- **Integrated GPS of HTC tablet** provides relatively stable location info.
- **The HTC Network** approach – combine cell tower & wireless APs – does not provide a significantly better position consistency as we thought.
K Nearest Neighbor (KNN) Approach

- **Hypothesis**
  - The set of APs and their associated signal strengths observed at a position represents a fingerprint that is unique to that position

- **Method:**
  - Build an RF fingerprint database
  - Compare the observed RSS vectors to fingerprints in the database
    - Find K closest neighbors, estimate the new location via their positions
    - Use Euclidean distance as the distance metric between unknown and known locations
Evaluating KNN Approach

- Add 12 (kismet) datasets gathered in one week of April/11
  - Same setting with Setting I

- Data sets:
  - p-airport: the 12 traces from past (April/11)
  - airport: all 16 traces gathered through setting I
    - 12 traces from April/11, 4 traces from December/11
  - alfa+airport: 16 traces from above + 4 Dec. traced gathered through setting II

- AP selection mechanisms
  - ap: use all Aps
  - avg ap: select APs that show larger than Avg RSS at more than one location
  - max ap: only select APs that show Maximum RSS at more than one location
- Mostly, error distance smallest when $K=3$
- Alfa USB wireless card provides significantly worse results than the integrated wireless card
- "Time" plays a role in localization accuracy => change of network parameter?
Outline

- Background
- Experiment setup
- Observations
- Future work
Future Work

- More Measurement!!
  - Establish “ground truth”
  - Static trace gathering: monitoring points close to static APs
  - More thorough investigation on localization algorithms
    - Compare and contrast the performance of existing ones, propose new one (?)
    - Combine traces from both static AP and moving monitoring points
    - Vertical localization
Thanks!

Q&A