Implications of Device Diversity for Organic Localization

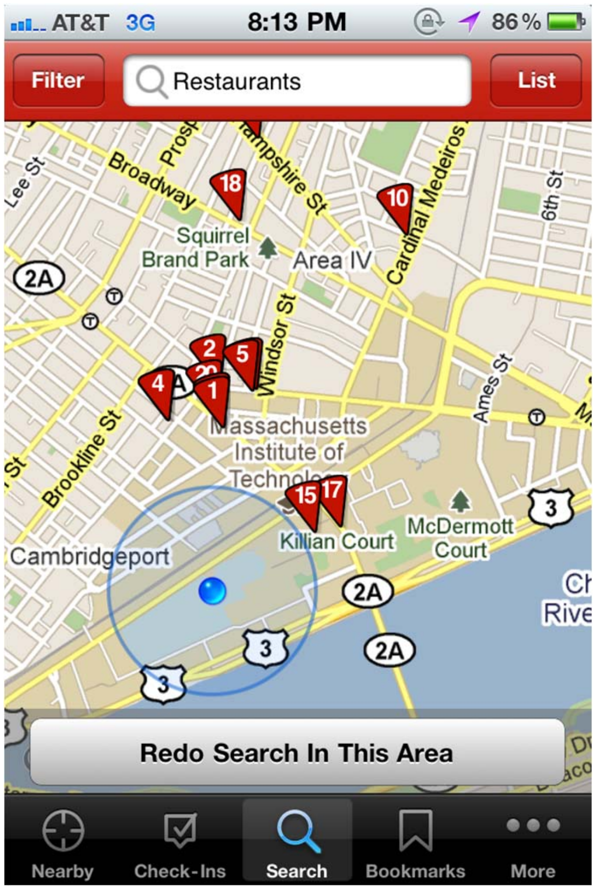
Jun-geun Park[1], Dorothy Curtis[1], Seth Teller[1], Jonathan Ledlie[2]

Computer Science and Artificial Intelligence Laboratory, MIT[1]

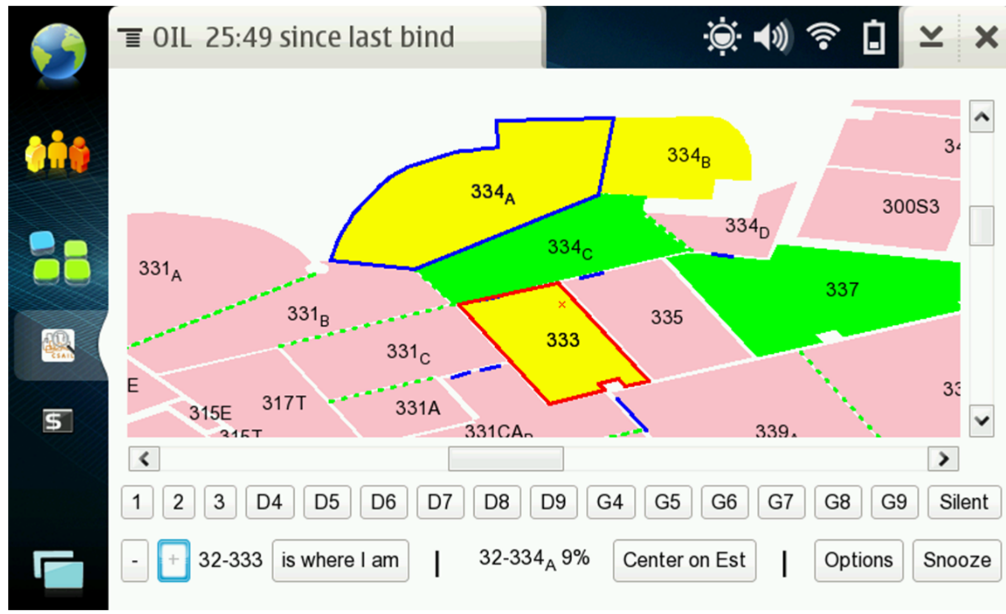
Nokia Research Center Cambridge[2]

IEEE INFOCOM 2011

Motivation: Location Determination for Mobile Applications



yelp (www.yelp.com)



Teller et al., “Organic Indoor Location Discovery”, 2008

Wi-Fi Localization

- Objective: Learn a map $f : S \rightarrow P$ from signal space S to physical space P
  - Signal features: signal strength (RSSI)/detection/… for each wireless access point (WAP)
  - Physical location: (x,y) coordinates/location labels
- Employ a learning algorithm with training (calibration) examples $\{(s_i, p_i)\}$ *"fingerprint"*
- Limiting assumptions
  - Need training data for each location
  - Localization samples are drawn from the same distribution as the training samples

Organic Localization*

- Pros
  - Mitigates the need for training data by crowdsourcing
  - The system facilitates sharing fingerprints among users
- Challenge
  - Device diversity due to multiple producers/consumers
  - Different antennas, chipsets, drivers, OS's

* Teller et al, Organic Indoor Location Discovery, MIT-CSAIL-TR, 2008
* Park et al, Growing an Organic Indoor Location System, MobiSys, 2010

Overview

We present an experimental analysis and design considerations for organic localization with heterogeneous devices

- Heterogeneous WiFi Signal Strength Characteristics
- Feature Design of Localization Algorithms for Heterogeneous Devices

Data Collection

- 6 different devices including 5 distinct models
- Data collected at 18 locations in a building
- Data available at http://rvsn.csail.mit.edu/location

Device	WiFi Chipset	OS	Kernel
Clevo D901C laptop	Intel 5300AGN (802.11a/b/g/n)	Linux Ubuntu 10.04	Linux 2.6.32
Asus EEE900A netbook	Atheros AR5001 (802.11b/g)	Linux Ubuntu 10.04	Linux 2.6.32
Lenovo Thinkpad X61 laptop	Intel 4965AGN (802.11a/b/g/n)	Linux Ubuntu 10.04	Linux 2.6.32
Nokia N810 tablet (x2)	Conexant CX3110X (802.11b/g)	Maemo OS2008	Linux 2.6.21
Nokia N95 cellphone	TI OMAP2420 (802.11b/g)	Symbian S60 FP1	EKA2

Algorithm (Signal-Strength-Based)

- Bayes classifier

$$p_{L|O}(l|o) = \frac{p_{O|L}(o|l)\ p_L(l)}{p_O(o)}$$

$$\hat{l} = l_{MAP} = \underset{l \in L}{\operatorname{argmax}} \left[p_{O|L}(o|l) \right]$$

L : Location
O : Signal observation

- Signal-strength feature, independence between APs

$$\hat{l} = \underset{l \in L}{\operatorname{argmax}} \left[\prod_{i \in M} p_{S_i|L}(s_i|l) \right]$$

S_i: Signal strength from AP i

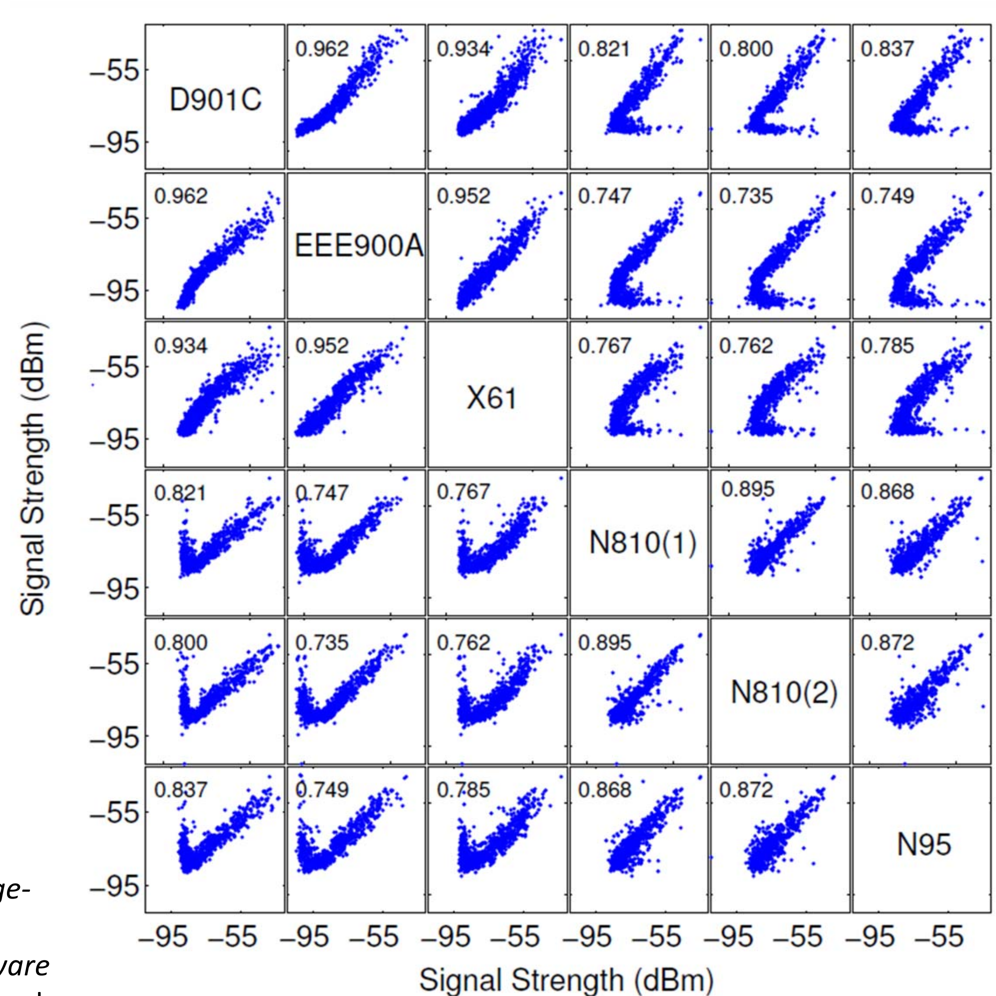
- Training: learning $p_{S_i|L}$ for each WAP i, Location l

Cross-Device Positioning with Signal Strength Features

- Training data from device A
  - $p_{S_i|L}$ was estimated with data from device A
- Localization on device B
  - $p_{S_i|L}$ is different for data from device B
  - Prediction performance degrades

Linear Transformation for Calibration?

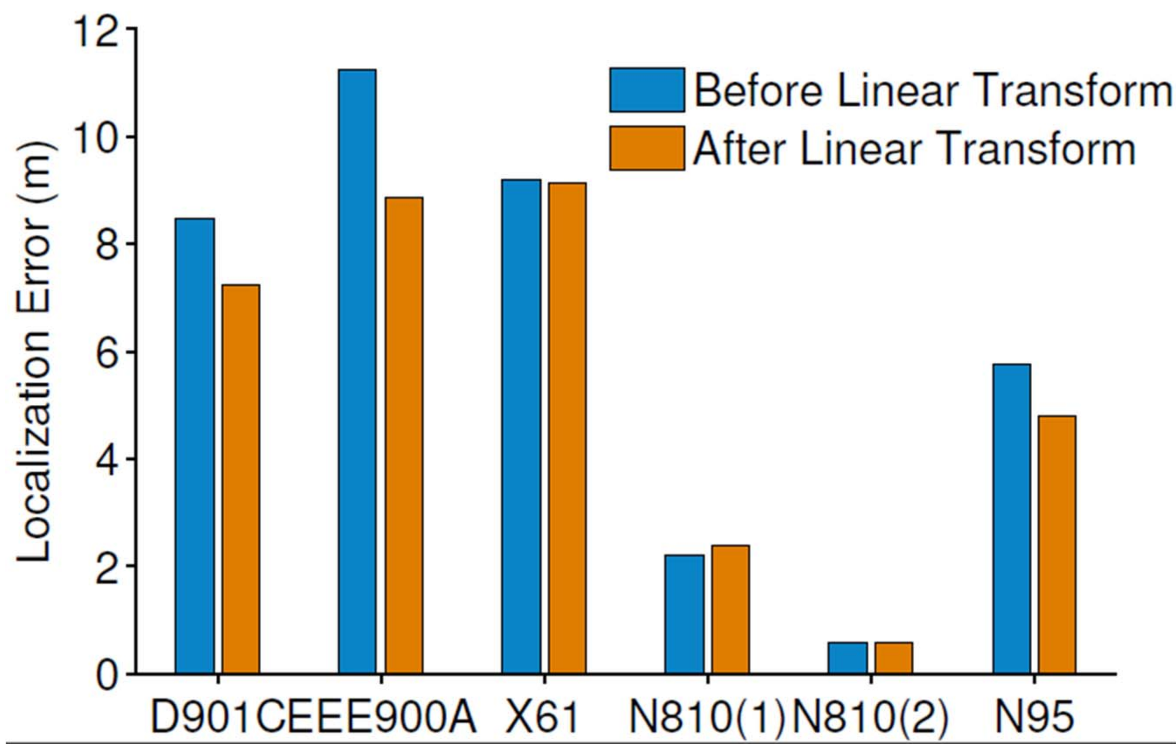
- Scatterplot matrix of signal strengths from different devices
- High linear correlations exist
- Previous work* suggested linear transformation would be sufficient



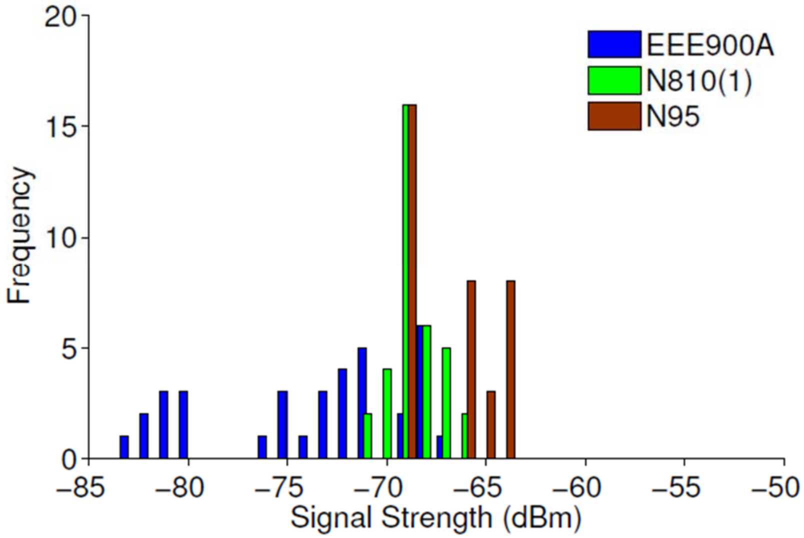
* Haeberlen et al., *Practical Robust Localization over Large-Scale 802.11 Wireless Networks*, MobiCom, 2004,
Tsui et al., *Unsupervised Learning for Solving RSS Hardware Variance Problem in WiFi Localization*, Mobile Networks and Applications, 2009

Linear Transformation for Calibration?

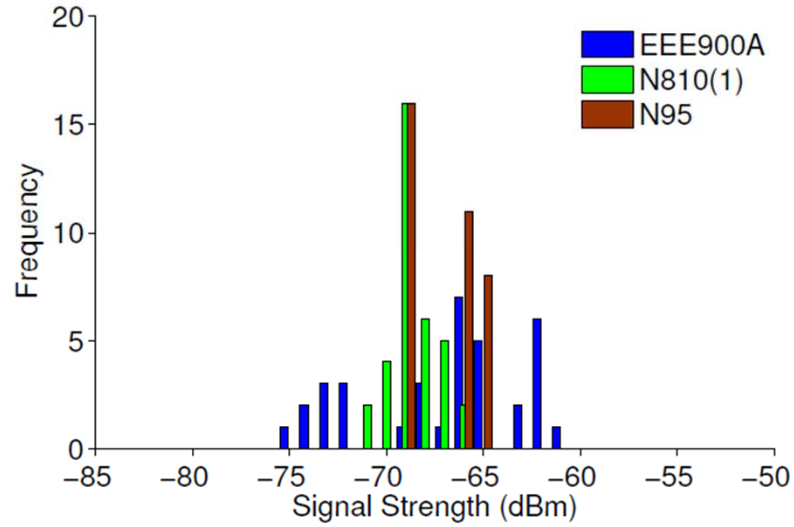
- Linear transformation alone does not solve the problem
- When N810(2) is used for training:



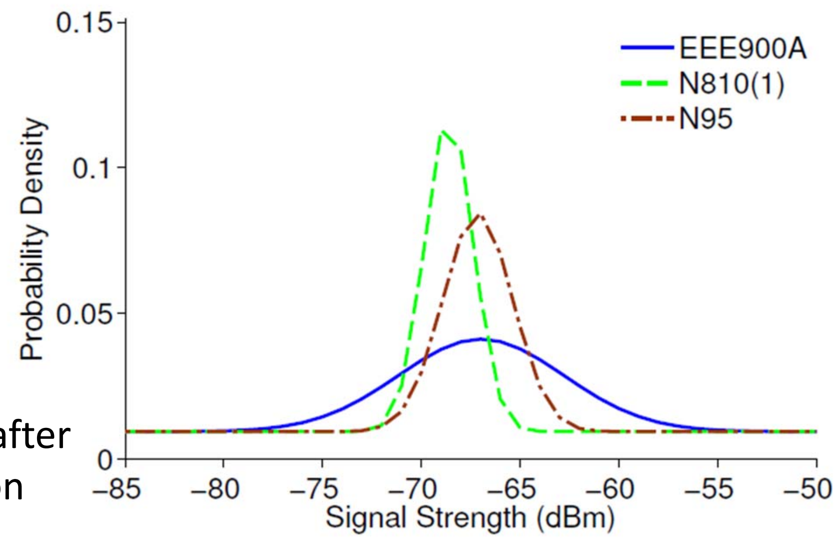
Linear Transformation for Calibration?



(a) Histograms before linear transformation



(b) After linear transformation



(c) Gaussian fitting after linear transformation

"Smoothing" by Kernel Density Estimation

- We need to take into account the variation in individual samples from different devices
- Kernel density estimator

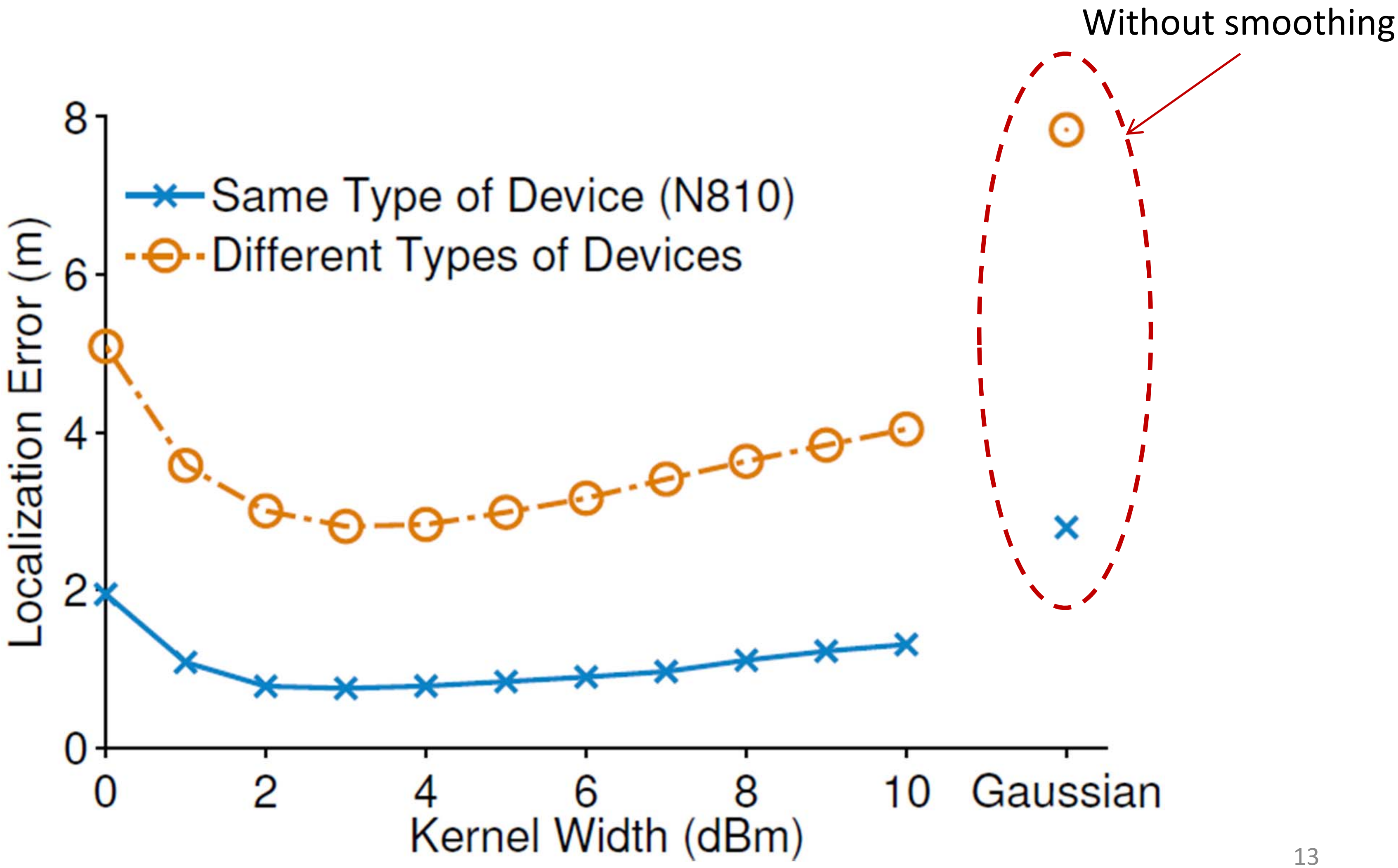
$$\hat{p}_X^k(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

$K(x, x_i)$: kernel function
h : kernel width

- M-estimator on kernel density estimate

$$p_X(x) = \frac{N\hat{p}_X(x) + \Phi\bar{p}_X(x)}{N + \Phi}$$

Kernel Density Estimator Improves Cross-Device Localization



Overview

We present an experimental analysis and design considerations for organic localization with heterogeneous devices

- Heterogeneous WiFi Signal Strength Characteristics
- Feature Design of Localization Algorithms for Heterogeneous Devices

To Use or Not To Use AP Detection Feature?

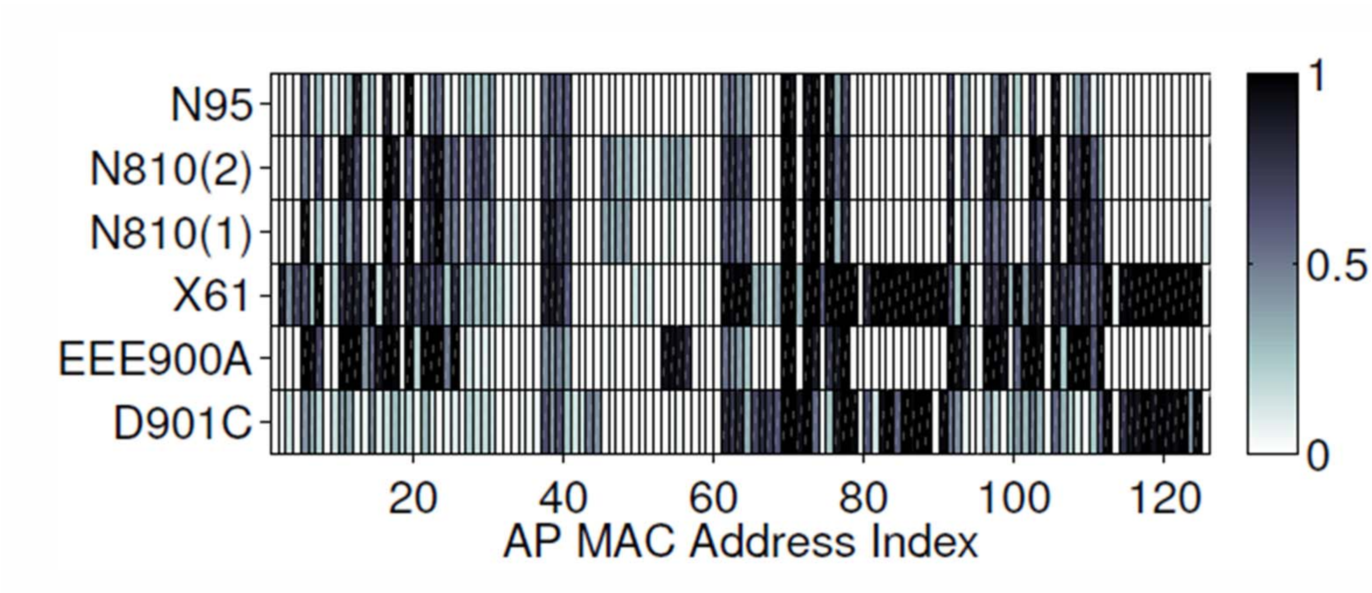
- $J_i = \begin{cases} 1 & if\ AP_i\ detected \\ 0 & otherwise \end{cases}$
- Multivariate Bernoulli model ($J_i, 1 \leq \mathrm{i} \leq \mathrm{k}$, k WAPs)

$$\hat{l} = \operatorname*{argmax}_{l \in L} \left[\prod_{1 \leq i \leq k} \left\{ p_{J_i|L}(1|l) \right\}^{J_i} \left\{ 1 - p_{J_i|L}(1|l) \right\}^{1-J_i} \right]$$

- Combine with signal strength (hybrid)

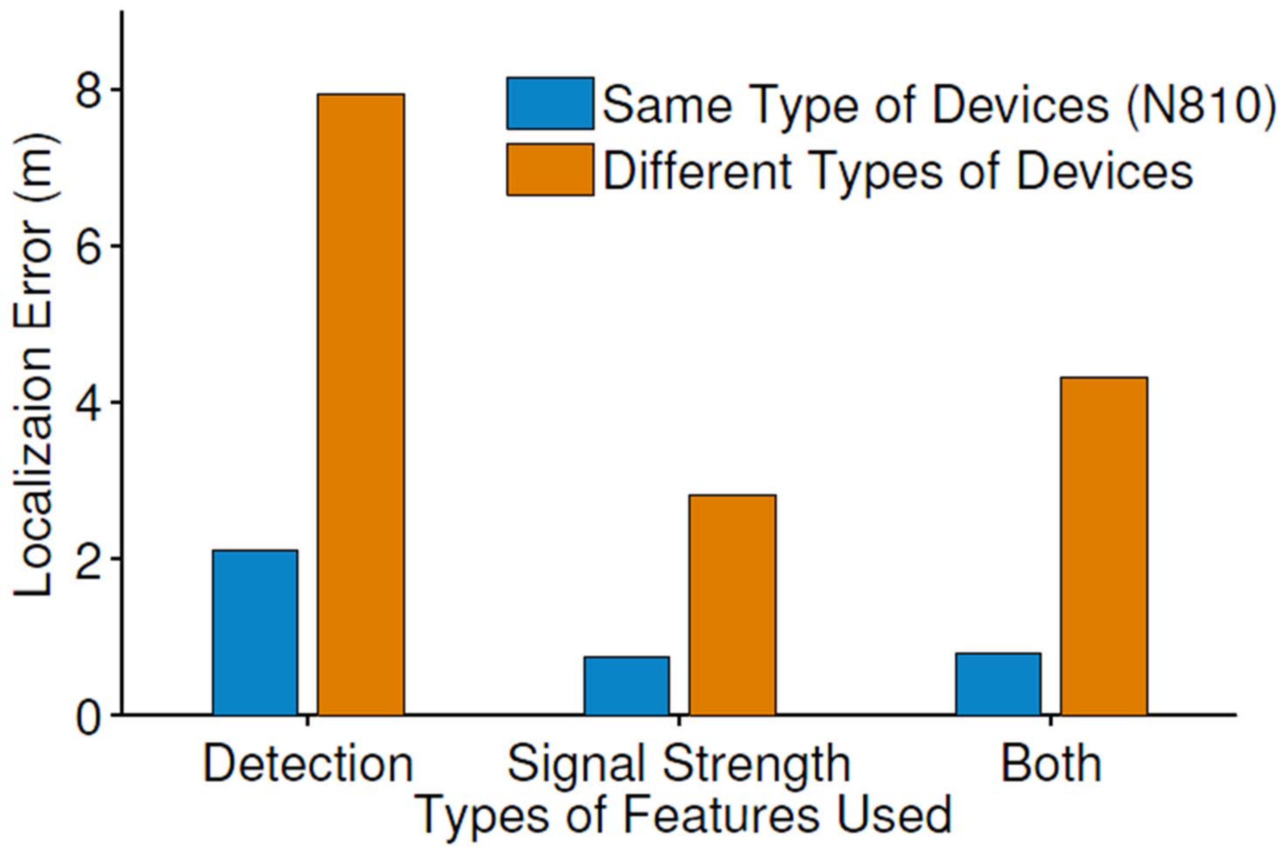
$$\hat{l} = \operatorname*{argmax}_{l \in L} \left[\prod_{1 \leq i \leq k} \left\{ p_{J_i|L}(1|l)\ p_{S_i|J_i,L}(s_i|1,l) \right\}^{J_i} \left\{ 1 - p_{J_i|L}(1|l) \right\}^{\alpha(1-J_i)} \right]$$

Detection Frequency Varies Across Devices



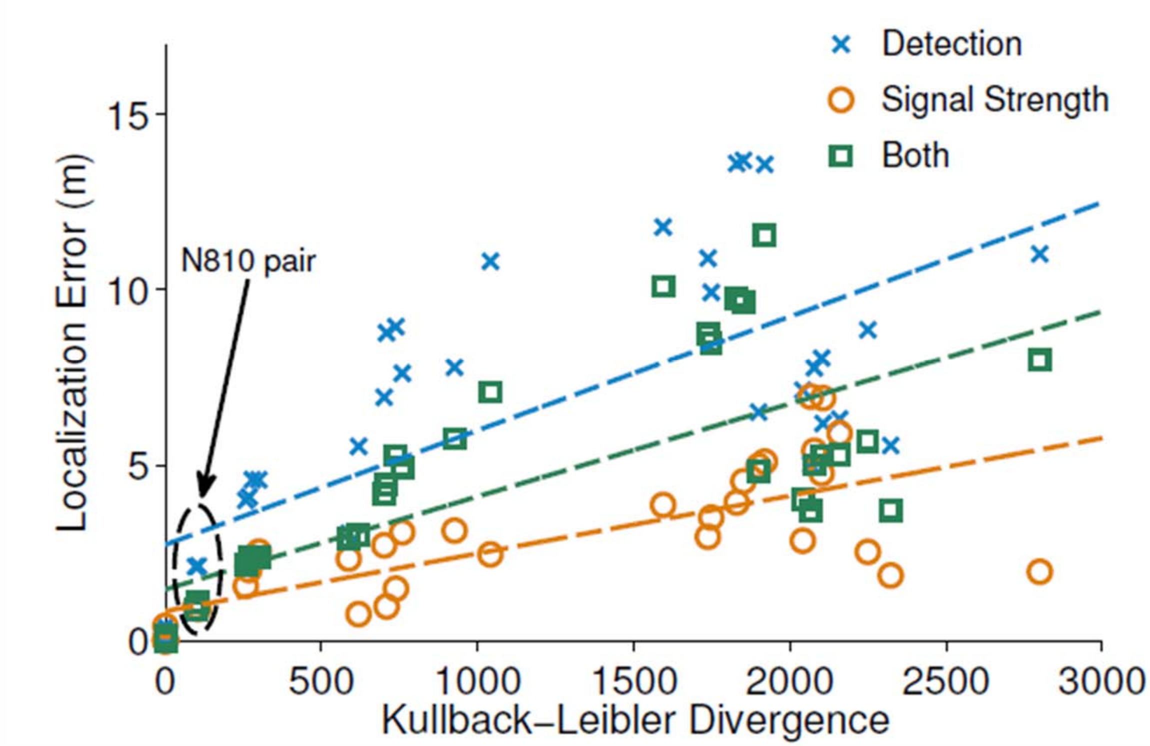
Localization Performance with Different Sets of Features

- Detection feature does not help cross-device localization



Empirical Characterization

- Quantified difference in AP detection by Kullback-Leibler divergence of detection rate probabilities J_i between different devices



Discussion

- Detection frequency feature does not give much extra information
- Detection frequency feature is largely incompatible across devices
- Signal-strength-only construction does not explicitly model detection process

- Other non-Bayesian algorithms (k-NN, SVM, …) are also affected because missing entries need to be filled in

Concluding Remarks

- Localization algorithms for cross-device positioning need to account for different dispersions in signal strengths
  - Smoothing by kernel density estimation is an easy way to improve cross-device localization
- Localization algorithms should use only "transferrable" information across different devices
  - AP detection frequency feature may be harmful because it varies widely across different 802.11 devices.

Concluding Remarks

- Localization algorithms for cross-device positioning need to account for different dispersions in signal strengths
  - Smoothing by kernel density estimation is an easy way to improve cross-device localization
- Localization algorithms should use only "transferrable" information across different devices
  - AP detection frequency feature may be harmful because it varies widely across different 802.11 devices.

- Organic indoor localization
  - Teller et al., *Organic Indoor Location Discovery*, MIT-CSAIL-TR-2008-075, 2008
  - Park et al., *Growing an Organic Indoor Location System*, MobiSys 2010
  - Ledlie et al., *Molé: a Large-Scale, User-Generated Positioning Engine*, IPIN 2011, submitted
  - http://rvsn.csail.mit.edu/location

Detection Frequency Varies Across Devices