Technische Universität Berlin

Performance Evaluation of Indoor Positioning Systems

Vlado Handziski (@vlahan)

Telecommunication Networks Group, TU Berlin

GeoIT Talks, Berlin, 03.12.2018

Indoor Positioning

- Almost every application needs context awareness
- User / device position is core context element
- Outdoor: mostly homogenous QoS requirements, domination of GNSS
- Indoor: diverse applications and QoS requirements, no dominant technology
- **Why?**

Microsoft Indoor Localization Competition

- Evaluate and compare technologies from academia and industry in the same, unfamiliar space

2014: Berlin

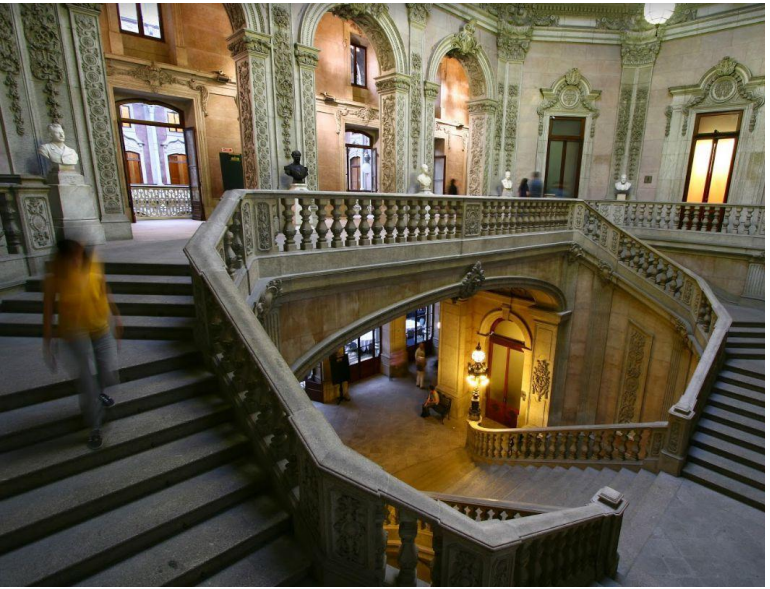
2015: Seattle

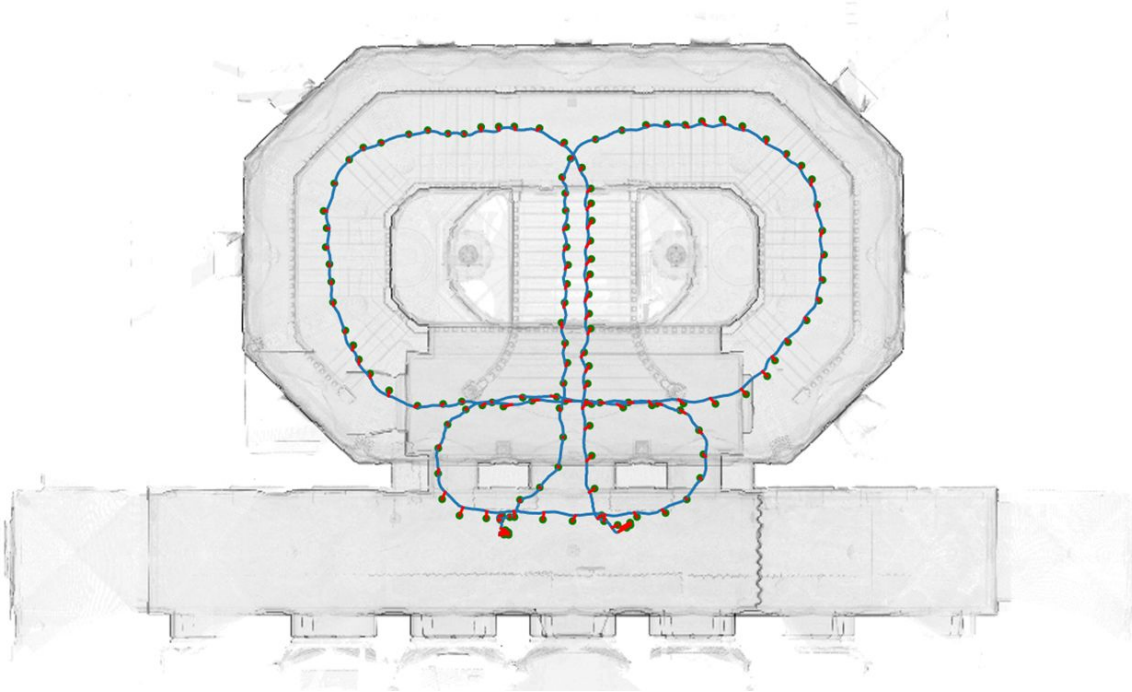
2016: Vienna

2017: Pittsburgh

[Following 5 slides courtesy of Dimitrios Lymberopoulos (Microsoft Research)]

2018 Evaluation Area -
Two Floors of Palácio da Bolsa, Porto



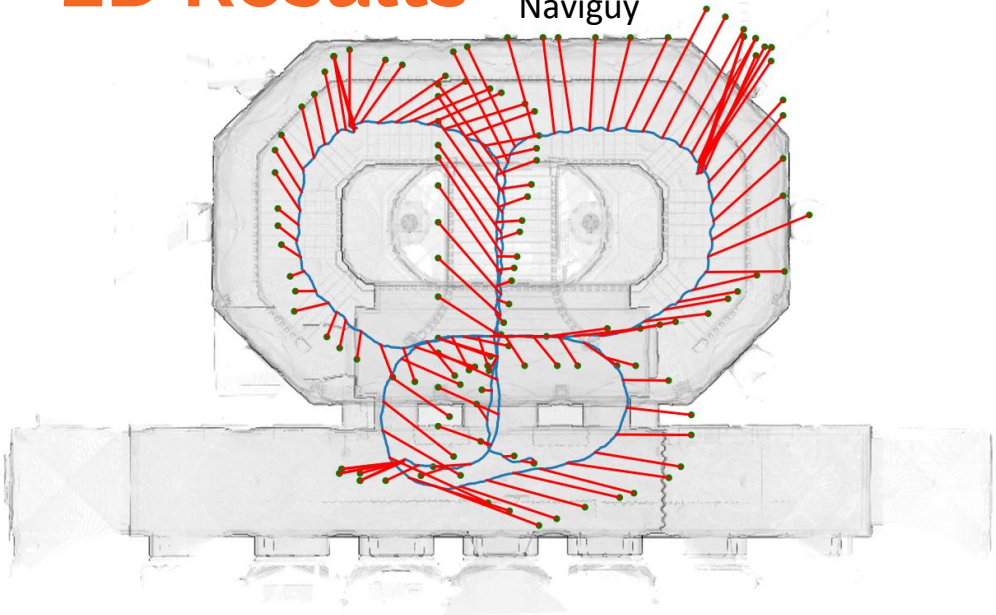


Evaluation Categories

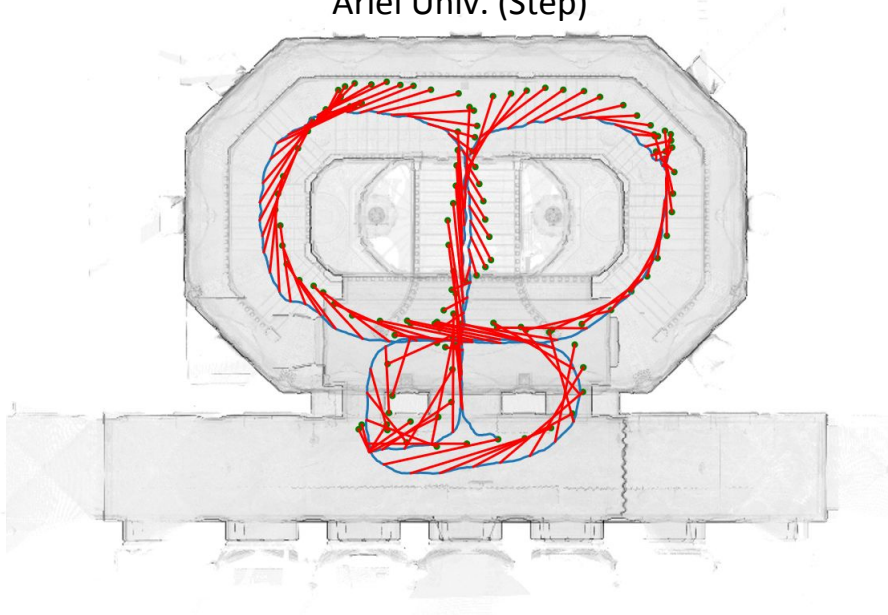
- 2D Category
  - Report (X,Y) locations
  - Disallow deployment of any infrastructure (WiFi and/or IMU based)
- 3D Category
  - Report (X,Y,Z) locations
  - Allow custom hardware deployment (UWB, Ultrasound etc.)
  - Each team can deploy up to 10 anchor devices

2D Results

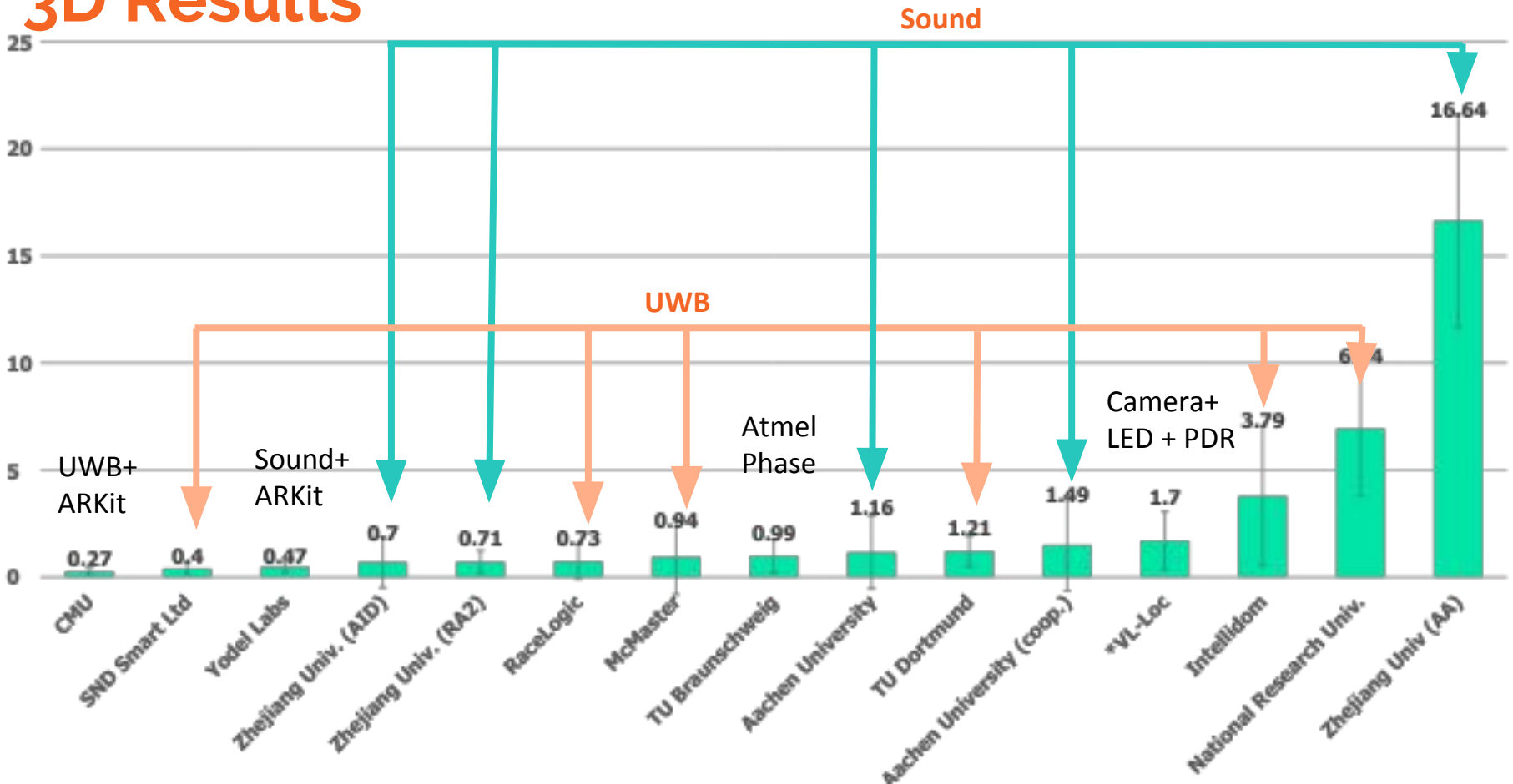
Naviguy

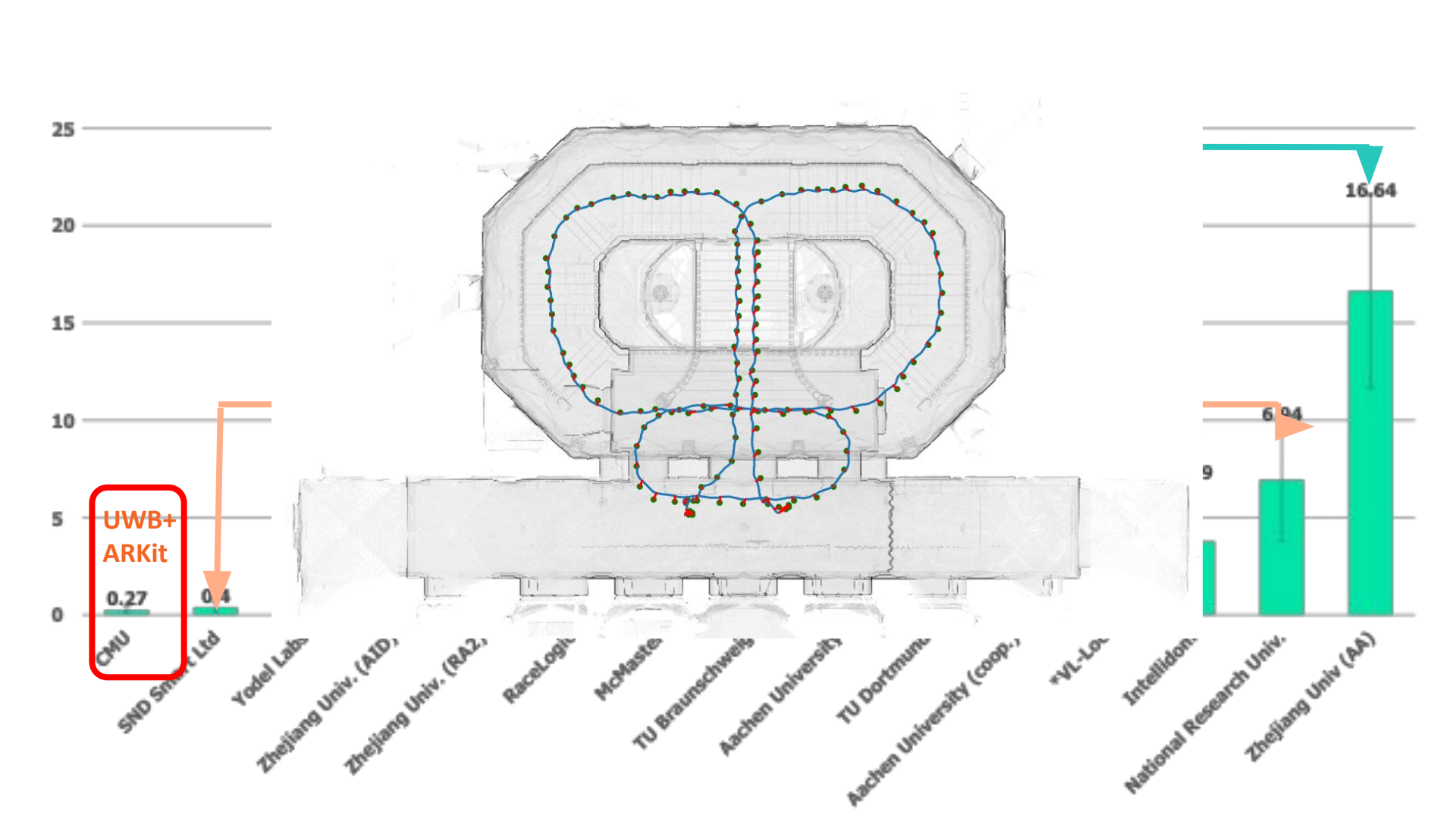


Ariel Univ. (Step)



3D Results


Sound
UWB
25
20
15
10
5
0
UWB+
ARKit
Sound+
ARKit
Atmel
Phase
Camera+
LED + PDR
0.27
0.4
0.47
0.7
0.71
0.73
0.94
0.99
1.16
1.21
1.49
1.7
3.79
6.64
16.64
CMU
SND Smart Ltd
Yodel Labs
Zhejiang Univ. (AID)
Zhejiang Univ. (RA2)
RaceLogic
McMaster
TU Braunschweig
Aachen University
TU Dortmund
Aachen University (coop.)
*VL-Loc
Intelidom
National Research Univ.
Zhejiang Univ (AA)


25
20
15
10
5
0
UWB+
ARKit
0.27
0.4
16.64
6.94
9
CMU
SND Smart Ltd
Yodel Labs
Zhejiang Univ. (AID)
Zhejiang Univ. (RA2)
RaceLogic
McMaster
TU Braunschweig
Aachen University
TU Dortmund
Aachen University (coop.)
*VL-Lo
Intellidon
National Research Univ.
Zhejiang Univ (AA)

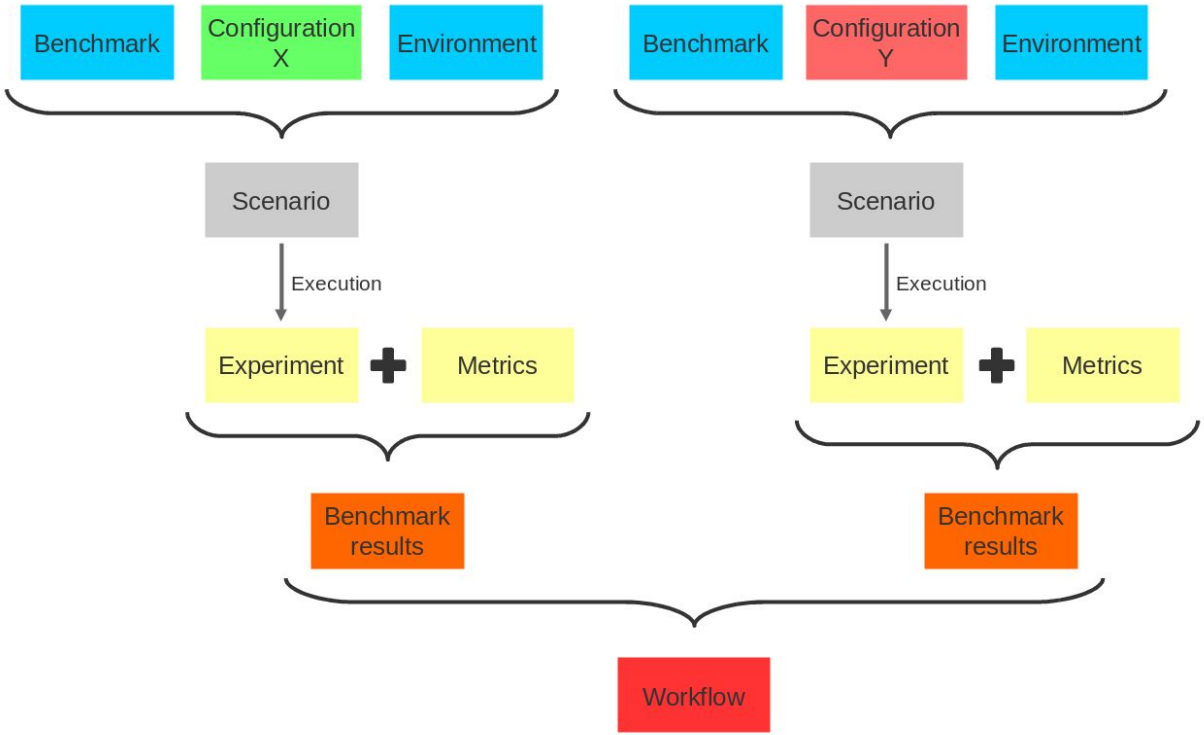
The Problem

- Failure to attain expected performance levels when migrating from development/experimental to target deployment conditions
- **Not missing right “technologies”, but “methodology” for evaluation and predicting performance in target environments!**
- Prevailing praxis of evaluation under custom, non-comparable and non-repeatable conditions

EVARILOS Project

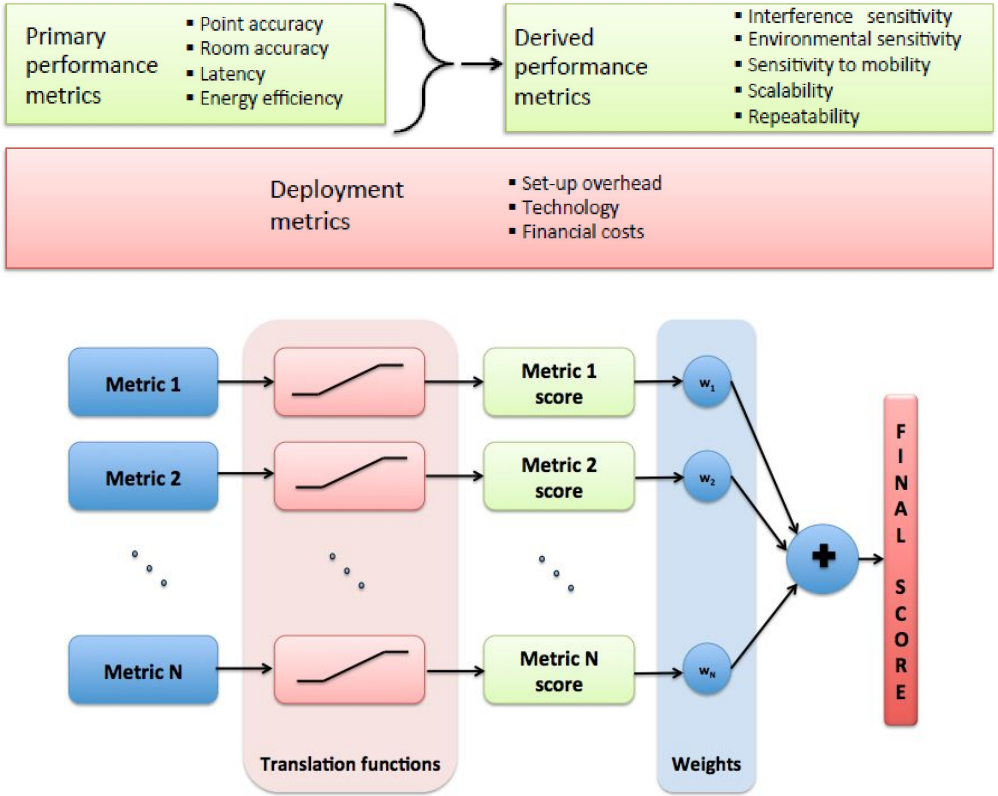
- Partners
  - TU Berlin (coordinator)
  - iMinds, SICS, ADVANTIC, Televic Healthcare
- **Develop a general methodology for benchmarking RF-based indoor localization systems!**
- Validate methodology in healthcare and mining

Standardized Benchmarking Process



EVARILOS Benchmarking Handbook (www.evarilos.eu)

Metrics and Scores



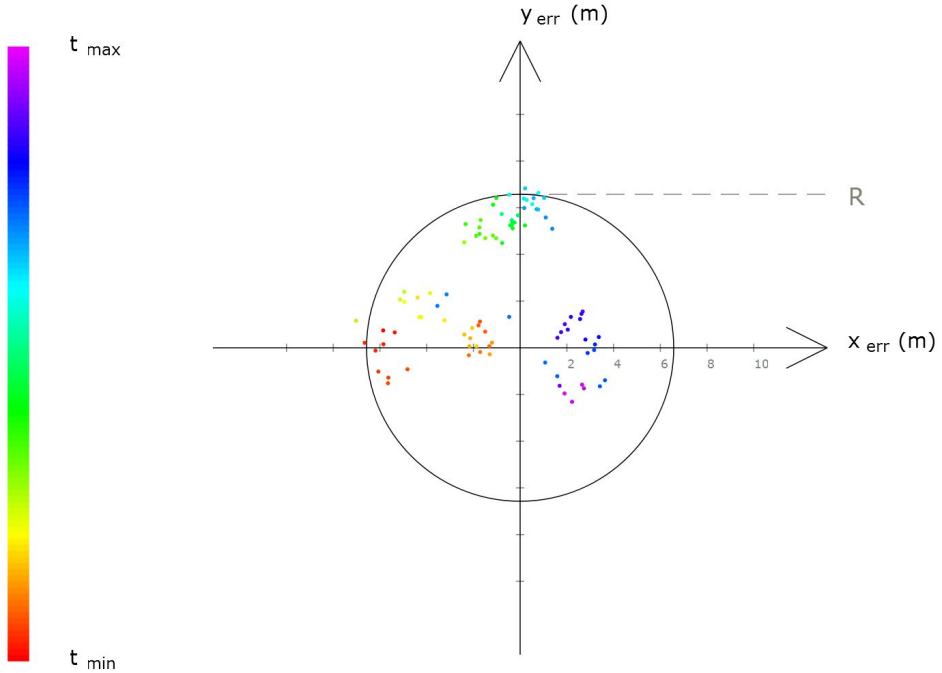
ISO/IEC 18305

- International standard for testing Localization and Tracking Systems
- Initiated by National Institute for Standards and Technology (NIST) in the US
- TKN members of the editorial committee
- Contributed elements of the EVARILOS benchmarking methodology and metrics to the standard
- Ongoing cooperation on remote evaluation of localization systems using online repository of raw sensor data

ISO/IEC 18305 Metrics

- Floor detection probability
- Zone detection probability
- Circular Error 95% (CE95) and Circular Error Probable (CEP)
- Vertical Error 95% (VE95) and Vertical Error Probable (VEP)
- Spherical Error 95% (SE95) and Spherical Error Probable (SEP)
- Means, covariance matrix, variances, RMS
- Coverage
- Relative accuracy
- Latency
- Set-up time
- Location-specific accuracy
- Availability

ISO/IEC 18305 Visualisation of T&E Results

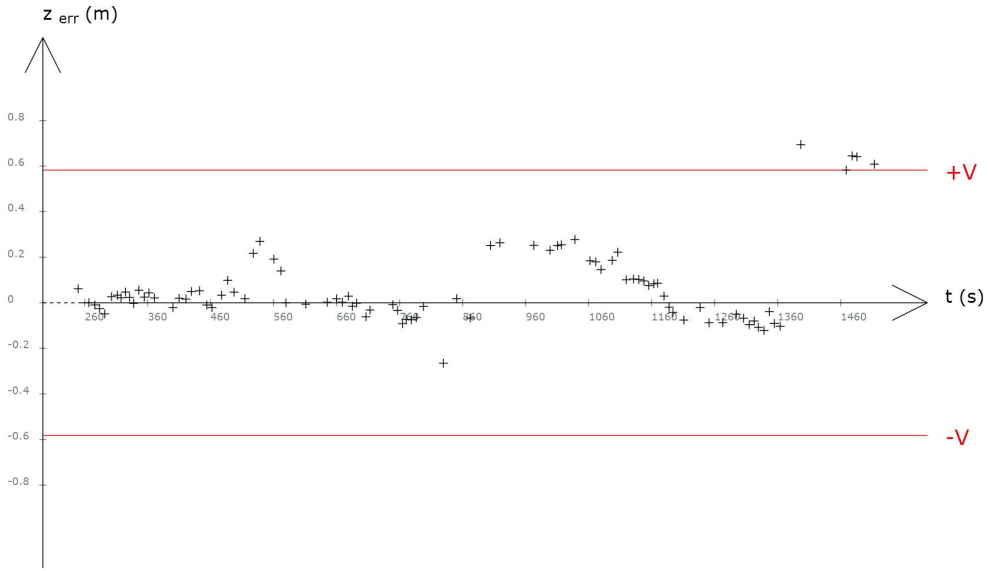


CE95 = R = 6.573 m

Mean = 4.287 m

σ = 1.717 m

ISO/IEC 18305 Visualisation of T&E Results



VE95 = V = 0.582 m

Mean = 0.113 m

σ = 0.148 m

PerfLoc

- Provide an extensive annotated data set to R&D community to facilitate development of smartphone indoor localization apps
- To create a level playing field for performance evaluation of such apps
- More information at perfloc.nist.gov

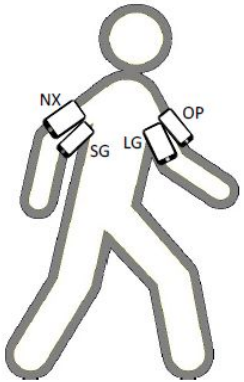
Data Collection Campaign

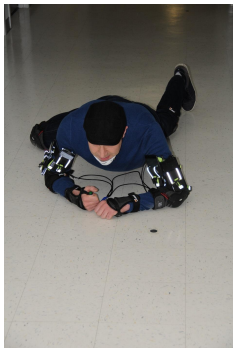
All available sensors on:

- LG G4
- Motorola Nexus 6
- OnePlus 2
- Samsung Galaxy S6

Environments

- Office
- Shop/Warehouse #1
- Shop/Warehouse #2
- Subterranean

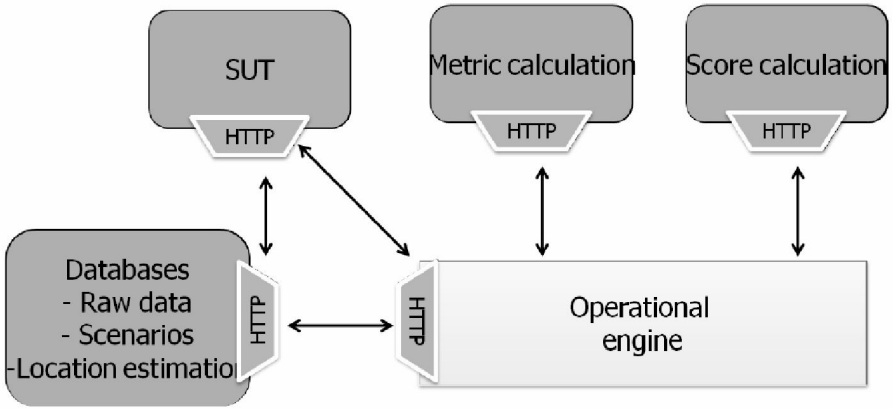




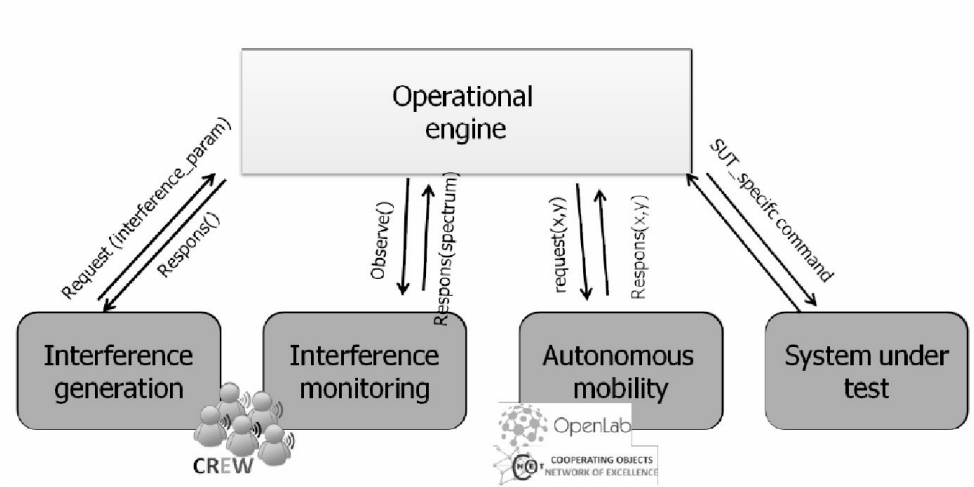
30,000 m^2 of space over 38 test and evaluation scenarios taking roughly 15.6 hours

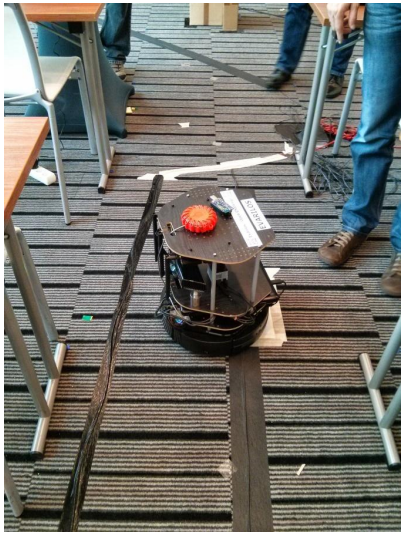
From Methodology to Software Platform

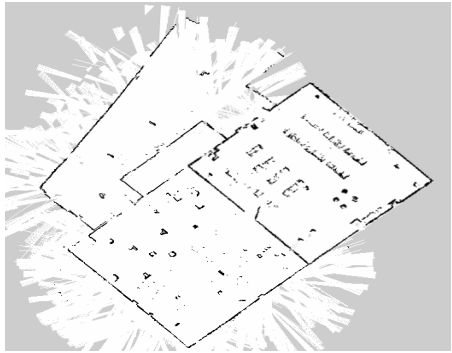
- How to promote use of standardized evaluation methodology?
- Fully automate the benchmarking process
- **Make the “right thing” the path of least resistance**
- **EVARILOS Benchmarking Platform**



Full Automation of T&E Experiments







MILC 2014 Competitor	Manual Avg. Error	EVARILOS Robotic Platform
1. EasyPoint	0.72 m	0.8 m
2. MapUme	1.56 m	1.9 m

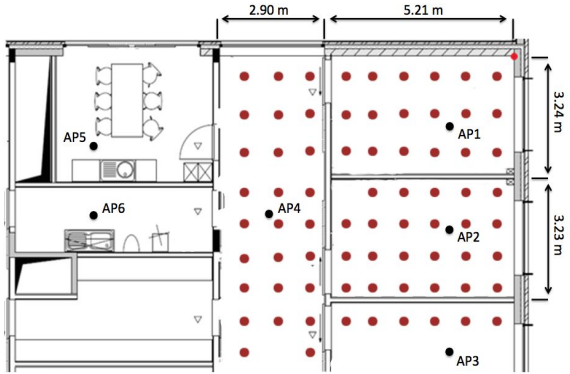
Open Question: Extrapolation

- **To which extent can these evaluation results be extrapolated from one environment to another?**
- How does similarity in environment and algorithm parametrization impact the evaluation results?
- Experiment: two WiFi-based fingerprinting algorithms in four environments with different levels of similarity

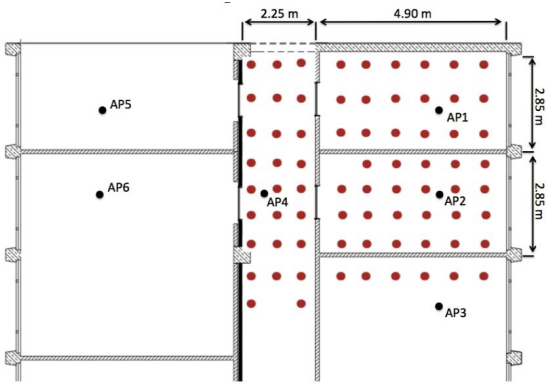
[F. Lemic et al, „Toward Extrapolation of WiFi RSSI-based Fingerprinting Performance Across Environments“, ACM HotMobile 2016]

Hospital vs. Mock-up

- Real-life hospital environment:

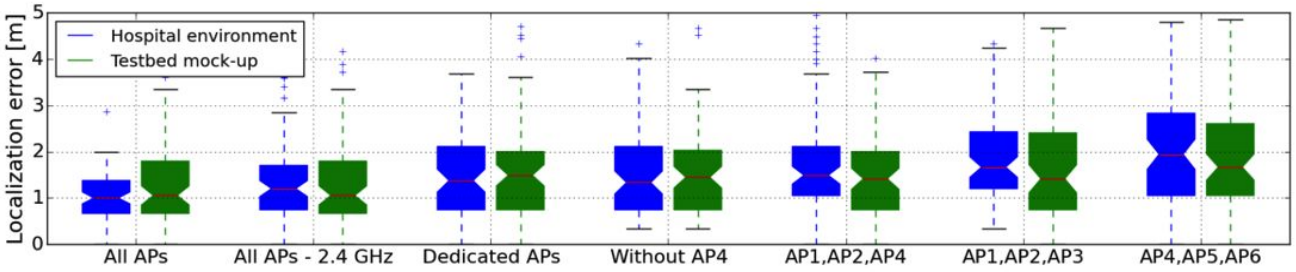


- TWIST testbed mock-up:

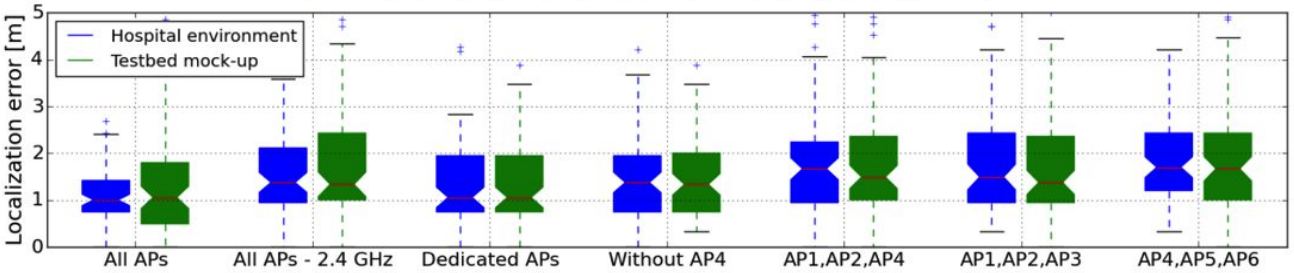


- Environments are similar in:
  - Location and configuration of anchor points
  - Location of evaluation points
  - Outline of the environments

Evaluation Results Hospital and Mock-Up



(a) Euclidean distance of averaged RSSIs



(b) Pompieu-Hausdorff distance of RSSI quantiles

Predicting Performance Results

- Extrapolation of the performance of WiFi fingerprinting from one environment to another is indeed possible
- **We need a formal measure of the similarity!**
- Importance of both environmental attributes and algorithm parametrization attributes
- ML-approaches show biggest promise
- **We need a lot more publicly available data!**

Thank You

Comments and Questions:

e-mail: vlado.handziski@tu-berlin.de

twitter:@vlahan

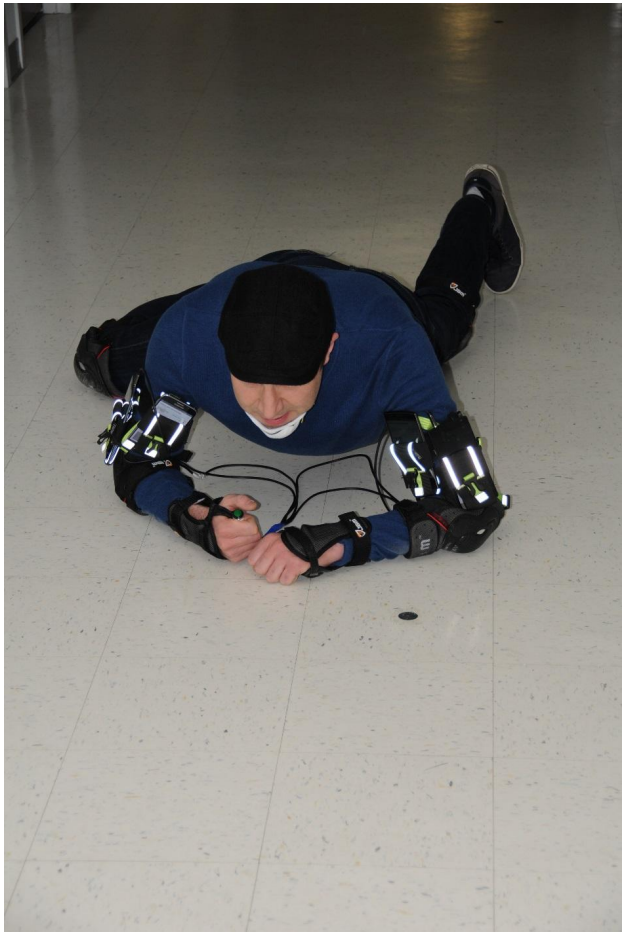
More on T&E Scenarios

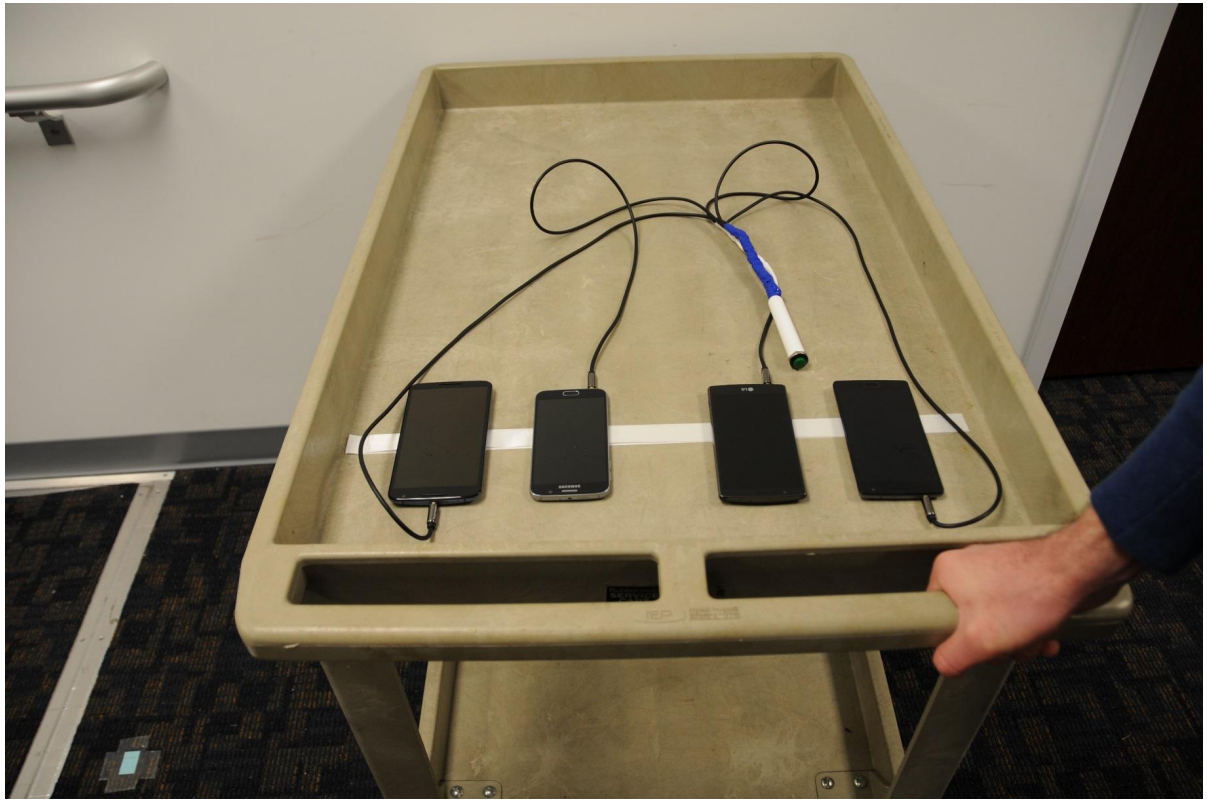
Data collected over scenarios using different types of mobility:

- Walking
- Walking with pauses
- Running
- Crawling
- Walking backwards
- Sidestepping
- Use of stairs
- Use of an elevator

Types of Data Collected (I)

- Any sensor available in each Android phone
- RF signal strength data (Wi-Fi and cellular), and GPS fixes when available.



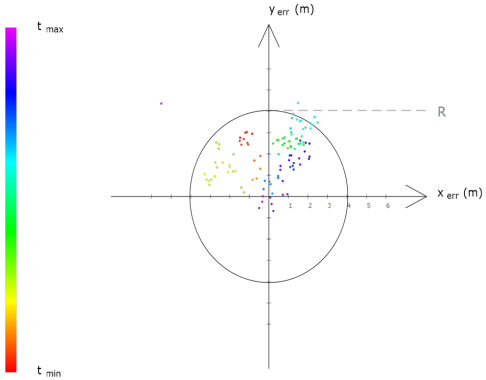


Training vs. Test Data

- 900+ test points (dots) deployed in the buildings and their locations professionally surveyed.
- All sensor values, RF signal strength data, and GPS data was timestamped.
- Also timestamped instances of going over dots.
- Training data provides ground truth locations at certain dots / time instances. Test data does not.

T&E of Developed Apps

- Web-based
- App developers will upload location estimates for certain time instances in each test data trace.
- Performance evaluation based on ISO/IEC 18305 international standard.
- Performance metrics will mostly deal with



Spatial distribution of horizontal error with color-coding to show evolution with test time

Prize Competition

- Contingent upon approval
- Web portal will publish a list of the apps evaluated in the order of *average* performance.
- The top three apps 6-12 months after the launch of the performance evaluation of capability will be awarded cash

Questions?

Contact Nader Moayeri

moayeri@nist.gov

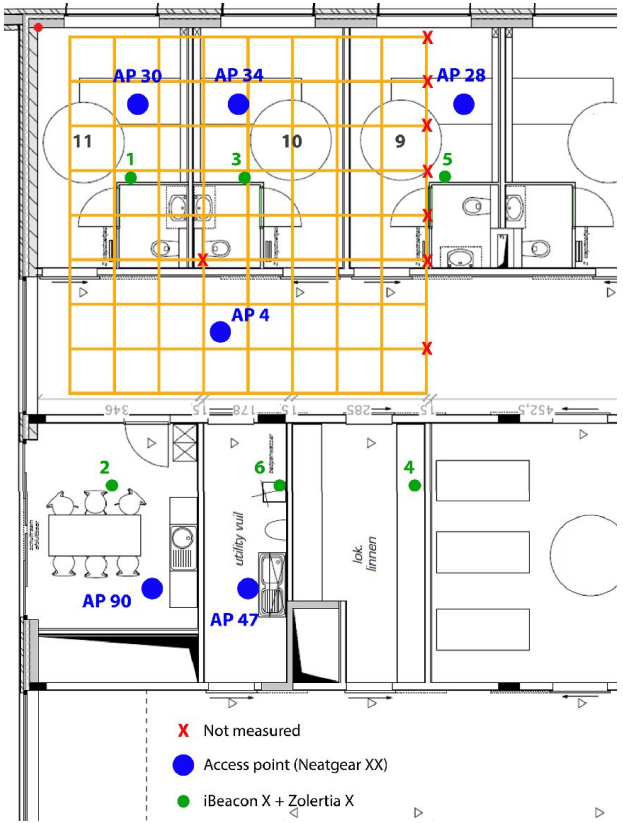
+1 301-975-3767

EVARILOS Benchmarking in St. Jozefs Kliniek Izegem

[Van Haute et al., “Performance Analysis of Multiple Localization Solutions in a Healthcare Environment”, Int. J. Health Geographics]

Deployment and Evaluation Plan

- Three widely available technologies:
  - WiFi (Netgear)
  - BLE (iBeacon)
  - ZigBee (Zolteria)
- Placed following typical existing deployment rules
- Diversity of rooms: patient rooms, corridor, functional areas
- Dense 2-D evaluation grid (8x8)

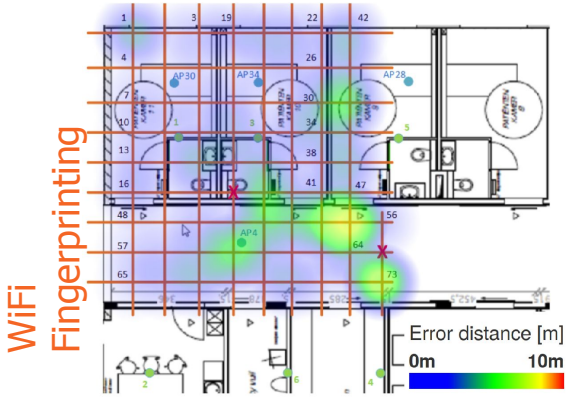


Summary of Evaluation Results

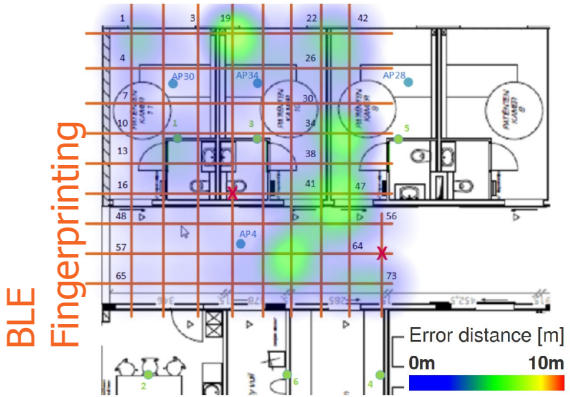
Technology	Point accuracy (m)	Room accuracy (%)	Latency (s)
Fingerprinting			
ZigBee	1.99	88	1.65
Wi-Fi	1.21	96	5.43
BLE	2.13	79	3.06
RSSI MLAT			
ZigBee	4.06	49	0.50
Wi-Fi	3.65	47	3.00
BLE	3.85	61	2.50
Spray RSSI + ToA			
ZigBee	3.89	47	0.50

Spatial Distribution of the Errors

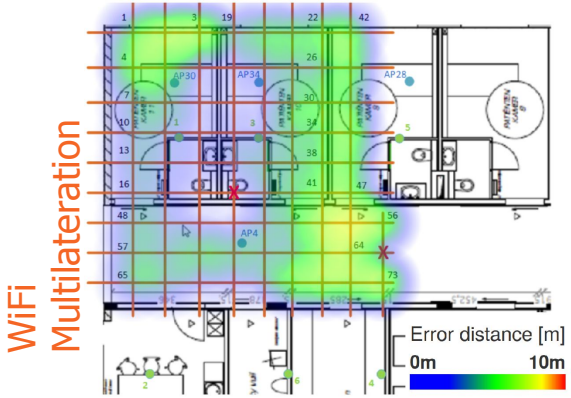
WiFi Fingerprinting



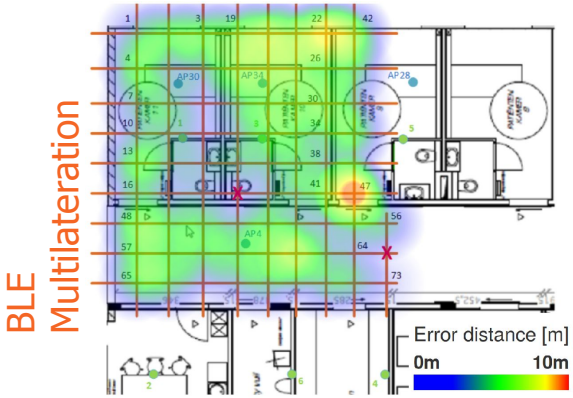
BLE Fingerprinting



WiFi Multilateration



BLE Multilateration



Impact on Standardization

- EVARILOS Methodology aligned with upcoming ISO/IEC 18305 Standard "Test and Evaluation of Localization and Tracking Systems"
  - Taxonomy of localization solutions (also non-RF-based)
  - Wide range of evaluation scenarios and performance metrics
  - Tests designed to trigger known sensor failure modes
- Participation in the Editing Committee for the standard led by NIST
- Ongoing cooperation with NIST on building a web platform for remote evaluation of localization systems using an online repository of raw sensor data

Future Challenges

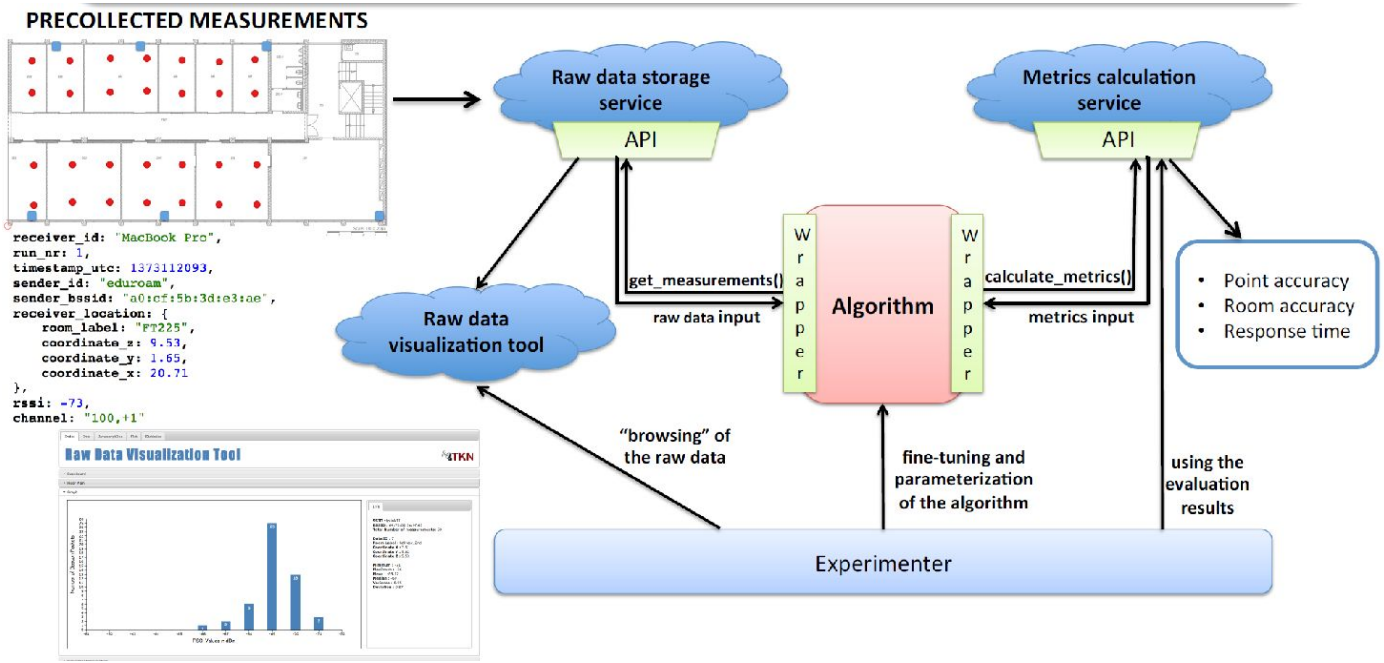
- A lot of work on Testbed Federation Architectures
  - WISEBED, CONET, FED4FIRE, etc.
  - Good APIs and software infrastructure
- But standardized methodology for cross-validation on multiple testbeds is still lacking
- Lack of formal support for extrapolating results from testbed environments to real-life deployments
- Fundamental trade-off between controllability and realism
  - Need for real users in the loop
  - We need to transition to "inverted testbeds"

THANK YOU!

VLADO HANDZISKI
E-MAIL: VLADO.HANDZISKI@TU-BERLIN.DE
TWITTER:@VLAHAN

BACKUP

Virtual Experimental Evaluation of RF-based Indoor Localization Algorithms



EBS: Multi-Workflow Support