Algorithms and Systems for the IoT Data Revolution

Demetris Zeinalipour

Associate Professor

Data Management Systems Laboratory (DMSL)

Department of Computer Science

University of Cyprus, Cyprus

https://www.cs.ucy.ac.cy/~dzeina/

dzeina@ucy.ac.cy

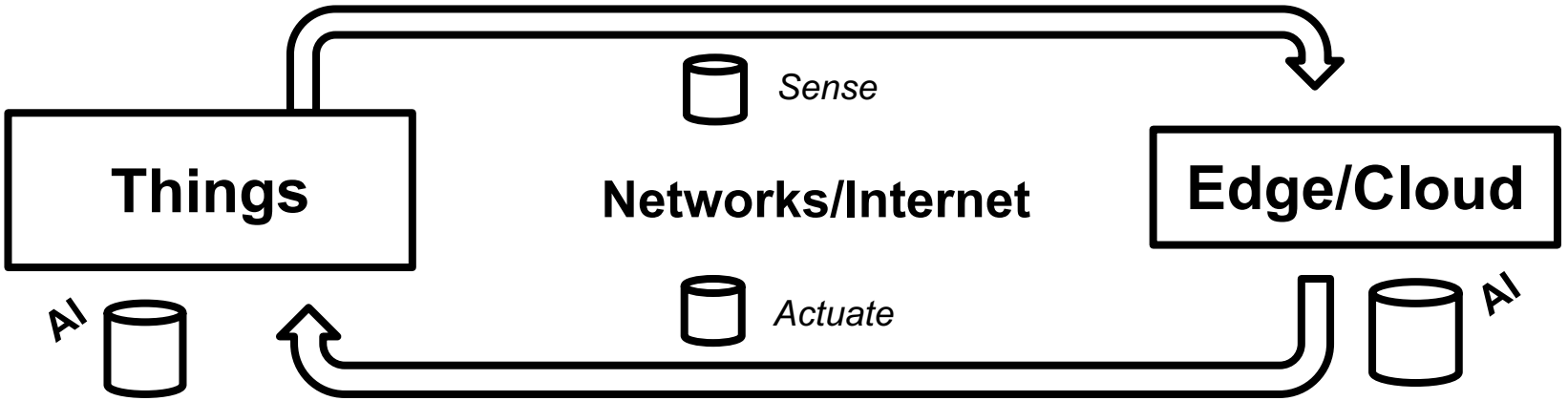
University of Cyprus, Nicosia, Cyprus, February 16th, 2023

DMSL @UCY

University of Cyprus

Internet-of-Things (IoT)

- The **Internet of things** (**IoT**) describes **physical objects** (or groups of such objects) with sensors, processing ability, software and other technologies that connect and **exchange data** with other devices and systems over the Internet or other communications networks - **Wikipedia**



IoT Data Revolution

IoT Data will explode even further from what we already witness today due to: IoT **hardware**, high-bandwidth/low-latency **networks**, cloud/fog **computing.**

Transportation / Smart Cars

Precision Agriculture / Agritech

eHealth

Social (Sophia) Humanoids

Manufacturing / Logistics

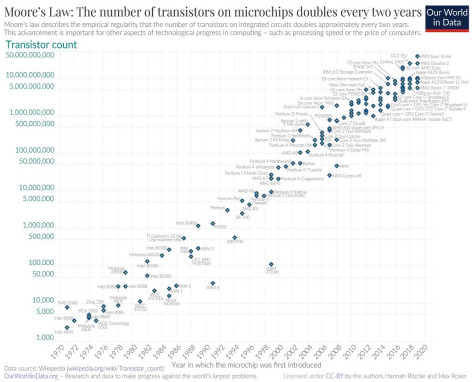
*"The most profound technologies are those that **disappear**. They weave themselves into the fabric of **everyday life** until they are **indistinguishable** from it."*

-Mark Weiser.

The Computer for the 21st Century. Scientific American , September, 1991.

IoT Data Revolution: Devices

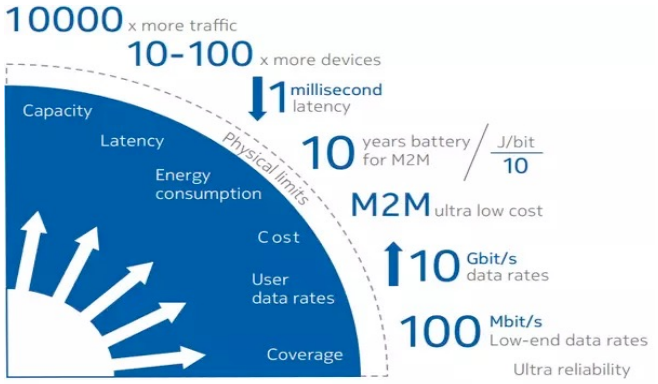
- **Human Population:** 8 Billion (2020), 8.6B (2030), 9.8B (2050), 11.2B (2100) - UN DESA
- **IoT Devices:** 9.7B (2020) to 500B (2030) – cisco.com
- **Smartphones:** 5.9B (2020), 7.6B (2027) – statista.com
  - Nokia CEO, Pekka Lundmark, claims that by 2030, a lot of people will **put down their smartphones** (in replacement of wearable/6G)
- ***Transistors:*** *Moore's Law still going strong!*
  - *number of transistors in integrated circuit (IC) doubles about every two years*
  - *50B/chip (2020)*– Wikipedia
  - *300B/chip (2030)*– Dutch AMSL
  - *CPU => GPU => TPU (AI) => QPU (Quantum)*
    - *2023: 1000 qubits, 2030: 1M qubits - IBM*
    - *Rose's Law: Qubits to double every year*



IoT Data Revolution: Mobility

5G (2020) is pushing forward by popular demand (VR/AR, smart cities, smart factories, …).

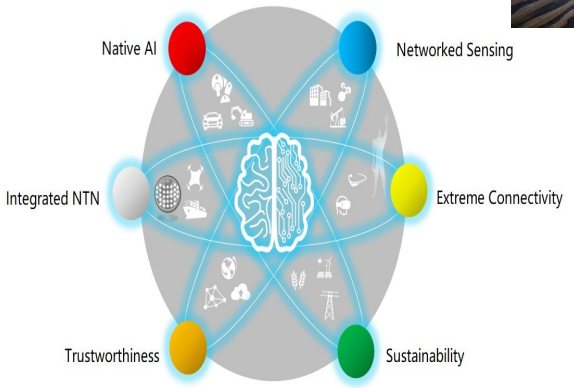
- Capacity (5G: 1M devices/km^2 | 6G: 10M devices/km^2)
- Latency (5G:~4 ms | 6G: sub-ms latency)
- Data Rates (5G: ~1-10Gbps | 6G: 1Tbps)
- Energy Consumption & Cost (ultra low)



5G: https://goo.gl/Pwu4ug

6G (2030) to go one step further:

- Marcus Weldon of Nokia Bell Labs, says that 6G will be a "***sixth*** *sense experience for humans and machines*" where **biology** meets **AI**.
  - 5 Human Senses: eyesight, hearing, taste, touch and smell.



6G: https://www.huawei.com/

Quote: *This is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning -* ***Winston Churchill***

IoT Data Revolution: Data/AI

- **2020:** humanity generated just a **few zettabytes** of data every year – *techtarget*
- **2025:** humanity to generate **175 zettabytes** of data per year. - *IDC*
- **2030:** humanity to create **yottabytes (YB)** of data every year. - *Huawei*
  - IoT will be a main driver of data generation.

gigabyte	GB
terabyte	TB
petabyte	PB
exabyte	EB
zettabyte	ZB
yottabyte	YB

x10^9

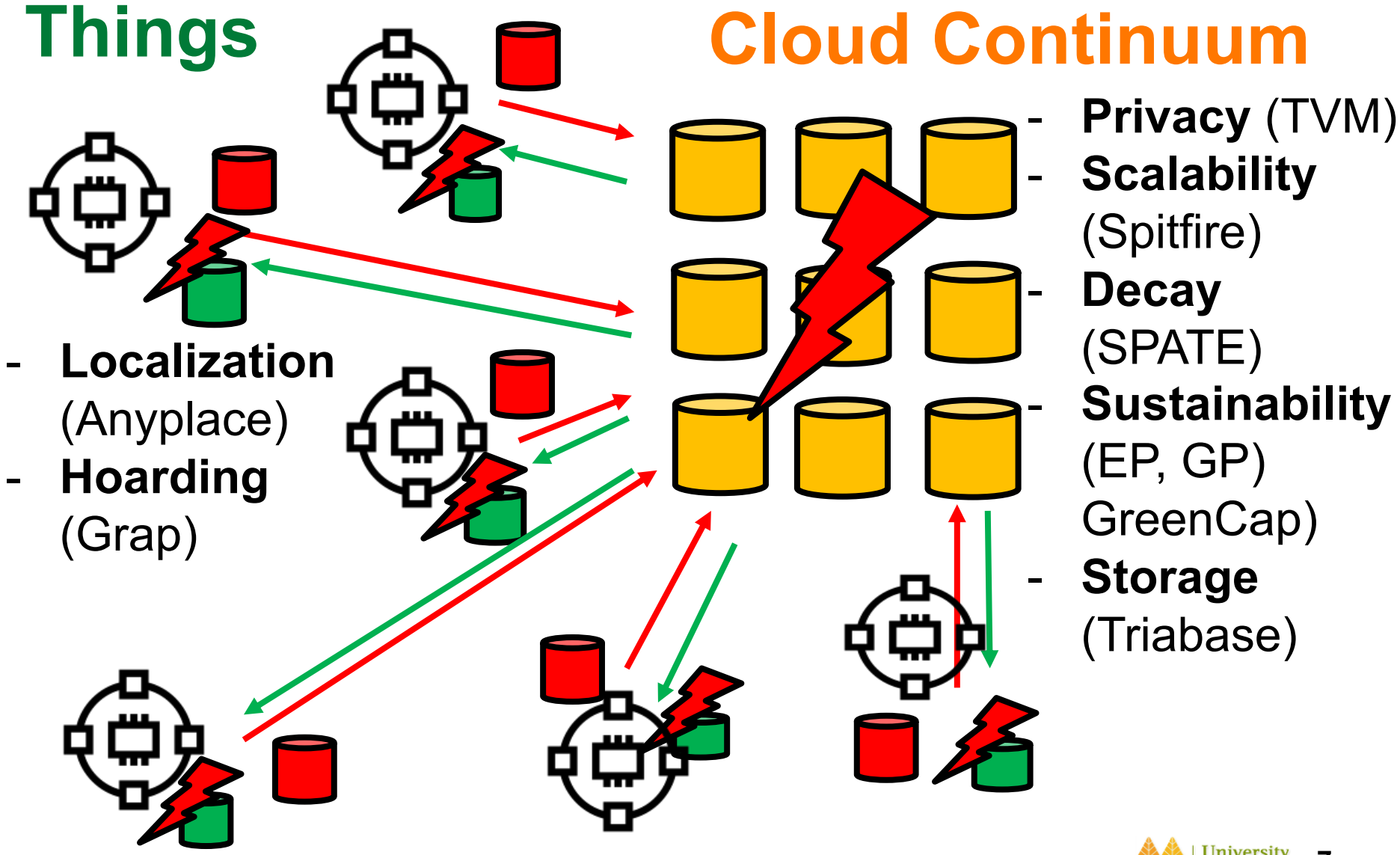
- If current (2020) AI (e.g., GPT-3 Language Model) relies on human generated data then what will 2030 AI look like with data from domain-specific IoT?

Word count

Dataset	Quantity (tokens)	Weight in training mix
Common Crawl (filtered)	410 billion	60%
WebText2	19 billion	22%
Books1	12 billion	8%
Books2	55 billion	8%
Wikipedia	3 billion	3%

Language Models are Few-Shot Learners, OpenAI, https://arxiv.org/abs/2005.14165, 2020.

Challenge: IoT Data Operators



About

- **Short Bio:**
  - 08/2005: Ph.D. Univ. of California - Riverside, CA, USA.
  - 09/2005-06/2007 (**2 years)** Visiting Lecturer @ CSUCY
  - 07/2007-12/2008 (**1.5 years)** Lecturer (Open University of Cyprus)
  - 01/2009-today **(14 years)**: Lecturer/Assistant Prof/Associate Prof. (University of Cyprus)
- **Research Interests:**

  Data Management *in* ***Computer Systems*** *and* ***Networks****:*
  - *Mobile, Sensor and Spatio-Temporal Data Management;*
  - *Big Data Management in Parallel and Distributed Architectures*
  - *Network, Blockchain and Telco Data Management;*
  - *Crowd, Web 2.0 and Indoor Data Management;*
  - *Data Privacy Management; Data Management for Sustainability.*
- ***Approach to Research:***
  - *Problem Formulation, Techniques and Algorithms (Conf. and Workshops)*
  - *Systems and Services (Journals)*
  - *Open-source and Community Development to have wider societal impact*

Research Timeline

01/2009: Lecturer

10/2012: Assistant Prof. / DMSL Founded

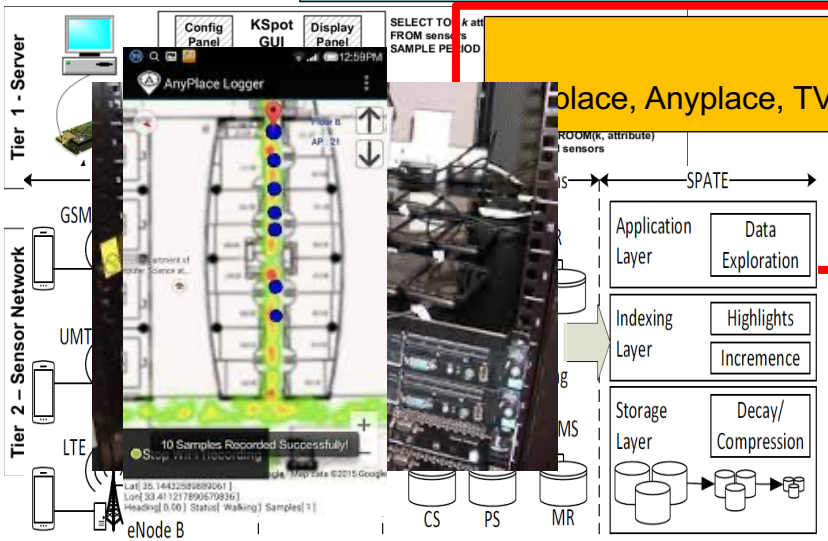
04/2018: Associate Prof.

Now

2009 **2012** **2015** **2018** **2021** **2023**

Sensor Data Management
(MINT, KSpot, MicroPulse, ETC, MHS, SenseSwarm, FlashSort)

Smartphone / Crowd Data Management
(Smartlab, Smarttrace, Spitfire, Rayzit, SmartP2P, CLODA)



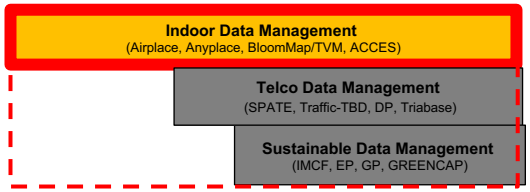
Indoor Data Management
(Indoor Atlas, Anyplace, TVM, Preloc/GRAP, ACCES, FMS, A4IoT, Surface, MDF, SMAS)

Telco/Blockchain Data Management
(SPATE, TBD-DP, Triastore, Triabase)

Route Planning & COVID-19
(EPUI, Healthdist, ASTRO)@PITT

Sustainable DM
(IMCF, EP, GP, GREENCAP)

Presentation Roadmap



- Introduction
- **Indoor Data Management**
  - **ACM TOIT'18, ACM TSAS'21, ACM TSAS'22, ACM DEBS'22, ICCAS'22, IEEE TKDE'15, IEEE JIoT'23**
- Telco Data Management
  - IEEE ICDE'19, Geoinformatica'19, IEEE MDM'18, IEEE ICDE'17
- Sustainable Data Management
  - ACM TIOT'22, IEEE IC'22, IEEE ICDE'21 EDBT'21

Indoor Location

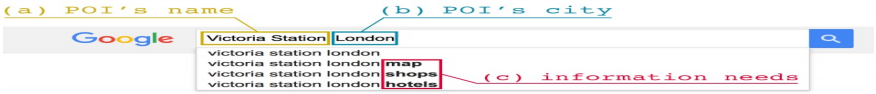
- **People** spend **80-90%** of their time indoors – USA Environmental Protection Agency 2011.
  - This is the place where most human activity, commerce, transactions, etc happen!
- >**85% of data** and **70% of voice** traffic originates from within buildings – Nokia 2012.

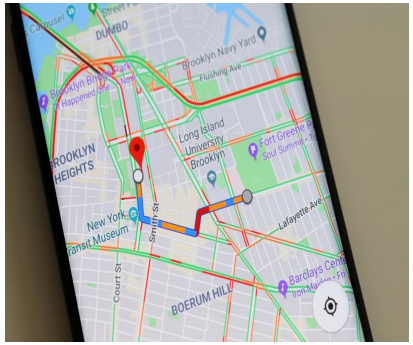




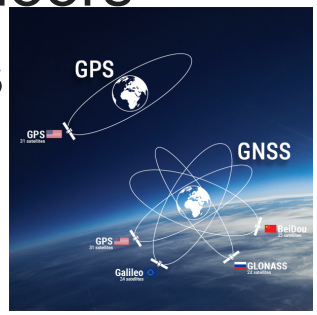

Location

- Humans have a Spatial Intent in their information needs
- 72% of **mobile information** needs are triggered by: **activity**, **location**, **time**, or **conversation**.
  - Google Maps Statistics 2021





- 77% of smartphone users regularly use navigation applications
- GNSS - Global Navigation Satellite System (BeiDou-2/China, Galileo/EU, GPS/US) low availability indoors
  - Blockage or attenuation of the satellite signals
  - High start-up time,
  - Power Demanding (receive signals).
  - Used as secondary option on Smartphones ☹



Indoor Localization

- **Smart Devices** are becoming enablers for modern ***Internet-based Indoor Navigation (IIN)*** services founded on their **measurements.**
- **Technologies:**

Existing Infrastructure

– Wi-Fi APs, Cellular Towers

New Infrastructure

– **Beacons** (BLE Beacons, RFID Active & Passive Beacons)

– **Sound** (Microphone)**, Light** (Light Sensor)

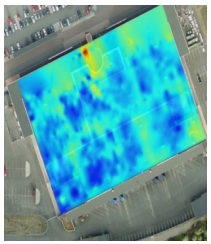
– **Stationary antennas, UWB, SLAM/LIDAR**

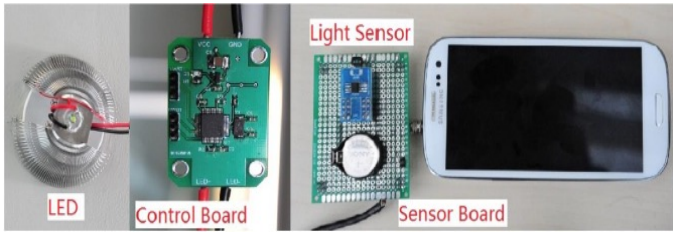
Zero Infrastructure

– **IMU Data** (Gyro, Acceler., Digital Comp)

– **Magnetic Field Sensors**

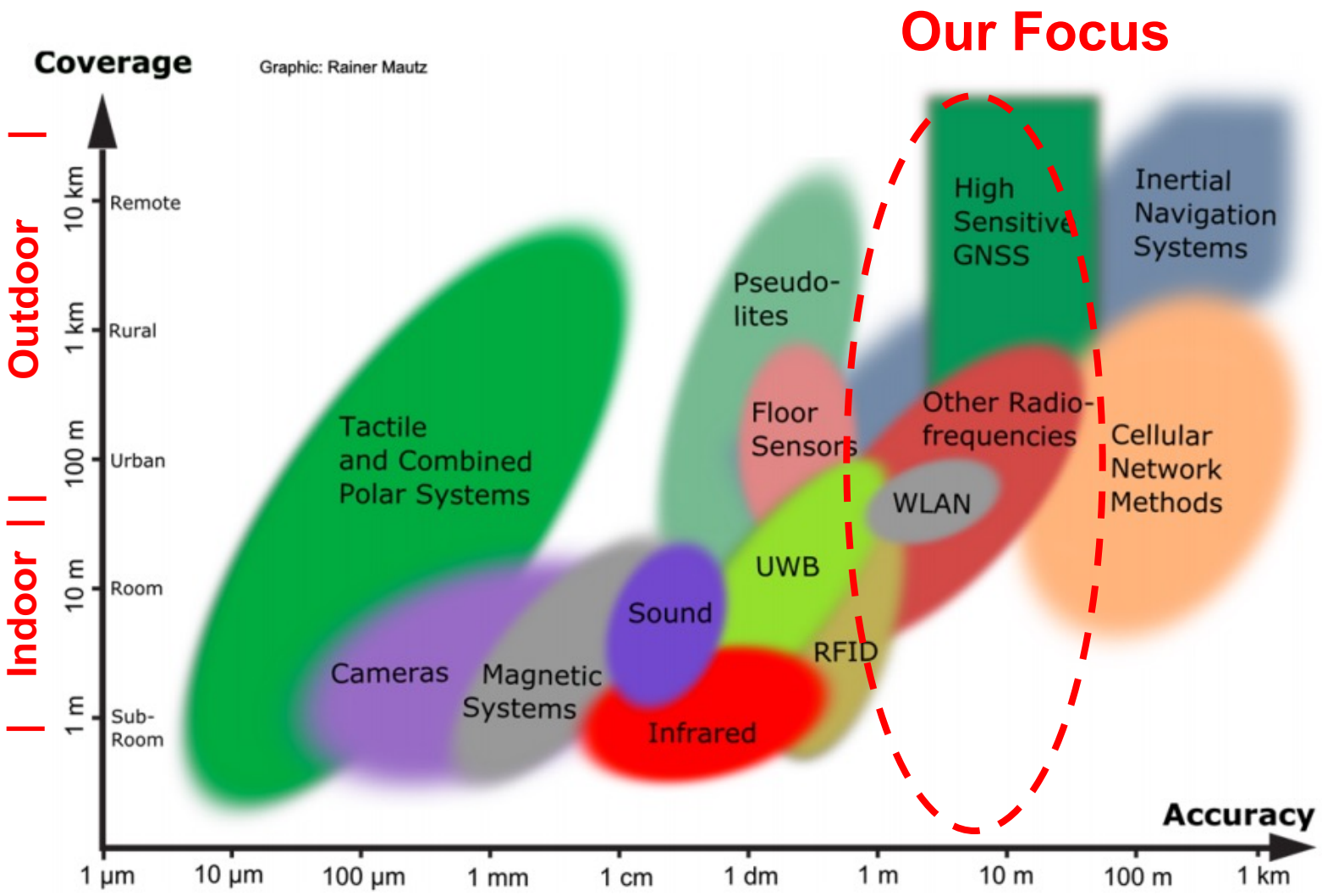
– **Computer Vision (CV) Systems**





"Internet-Based Indoor Navigation Services", Demetrios Zeinalipour-Yazti, Christos Laoudias, Kyriakos Georgiou, Georgios Chatzimilioudis, **IEEE Internet Computing (IC'17),** vol. 21, no. 4, pp. 54-63, July 2017, doi:10.1109/MIC.2017.2911420, IEEE Computer Society, 2017.

Indoor Localization



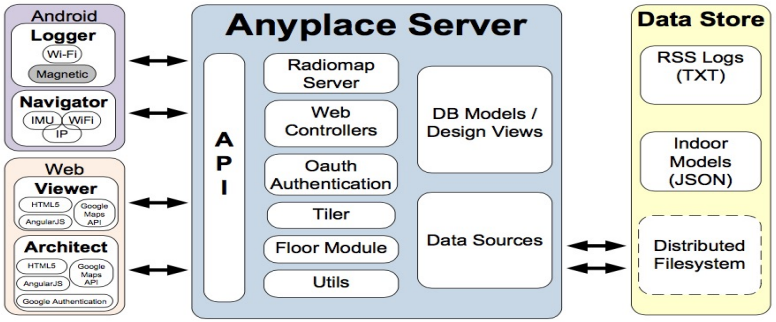
Rainer Mautz, ETH Zurich, 2011

Anyplace

Localization

A complete open-source **Internet-based Indoor Navigation (IIN)** Service

- **predominant IoT open-source** Indoor Localization Service - MIT License.
- **Modular Architecture:** Web, Android, Windows, iOS, JSON API.
- Multiple awards for accuracy & utility.

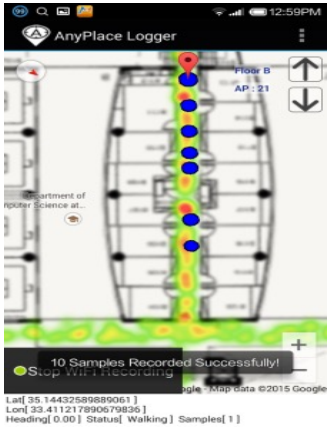


"The Anatomy of the Anyplace Indoor Navigation Service", Demetrios Zeinalipour-Yazti and Christos Laoudias, **ACM SIGSPATIAL Special**, Special Issue on Indoor Spatial Awareness II, Editor: Chi-Yin Chow, Vol. 9, No. 2, pp. 3-10, July 2017.

Fingerprinting in Anyplace

Localization

Logger



a) Upload Measurements

IIN Service

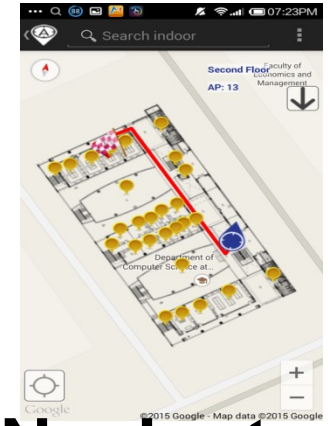
b) Build **RM** (Radiomap)

RadioMap (RM) on s

Location	ap_1	ap_2	…	ap_M
l_1	90	85	…	-1
l_2	-1	-1	…	-1
l_3	-1	40	…	-1
l_4	-1	-1	…	-1
l_5	-1	-1	…	80
l_6	80	-1	…	-1
l_N	…	…	…	…

z

c) Localize



Navigator

Anyplace History

Localization

- **[Airplace]** **Best Demo Award** at IEEE MDM'12, Bangalore, India. (Open Source!)
- **[HybridCywee]** "Indoor Geolocation on Multi-Sensor Smartphones", C.-L. Li, C. Laoudias, G. Larkou, Y.-K. Tsai, D. Zeinalipour-Yazti and C. G. Panayiotou, in **ACM Mobisys'13**, Tapei, Taiwan. Video at: http://youtu.be/DyvQLSuI00I
- **[UcyCywee]** IPSN'14 Indoor Localization Competition (Microsoft Research), Berlin, Germany, April 13-14, 2014. **2nd Position with 1.96m!** http://youtu.be/gQBSRw6qGn4
  - *D. Lymberopoulos, J. Liu, X. Yang, R. R. Choudhury, **..., C.** Laoudias, D. Zeinalipour-Yazti, Y.-K. Tsai, and et. al., "A realistic evaluation and comparison of indoor location technologies: Experiences and lessons learned", **In IEEE/ACM IPSN,** pp. 178-189, Seattle, WA, USA, April 14-16, **2015.***
- **1st** Position at EVARILOS Open Challenge, European Union (TU Berlin, Germany), **2014**.
- **[ACCES]** **Honorable Mention Award** at IEEE MDM'17, S.Korea, Indoor Localization Accuracy Estimation

Cywee / Airplace

Anyplace Wi-Fi

Localization

- Alstom: French Manufacturer of Trains (TGV, Eurostar) that deployed Anyplace in its smart factory in **India**.
  - Anyplace has been compared against Cisco CMX system in an internal study by Alstom

[C81] "The Anyplace 4.0 IoT Localization Architecture", Paschalis Mpeis, Thierry Roussel, Manish Kumar, Constantinos Costa, Christos Laoudias, Denis Capot-Ray, Demetrios Zeinalipour-Yazti, **Proceedings of the 21st IEEE International Conference on Mobile Data Management** (MDM'20), IEEE Computer Society, ISBN:, pp. 218-225, June 30 - July 3, 2020, Versailles, France, DOI: 10.1109/MDM48529.2020.00045, **2020**.

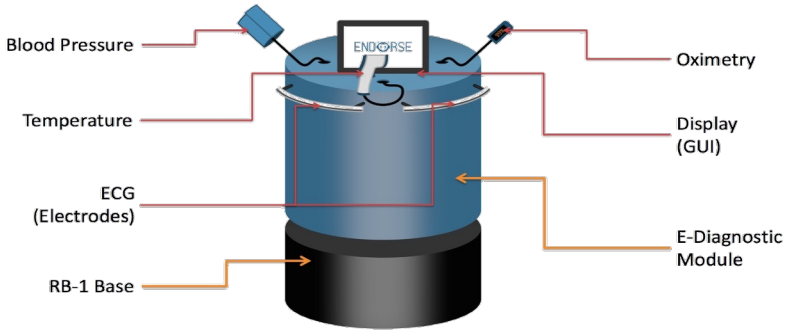
MCSA ENDORSE/RESPECT Projects

- **ENDORSE:** Safe, Efficient and Integrated Indoor Robotic Fleet for Logistic Applications in Healthcare and Commercial Spaces
  - Period: 01/10/2018 - 01/10/2021
  - **Research Topic: Localizing with IoT Integration**

Marie Skłodowska-Curie Actions

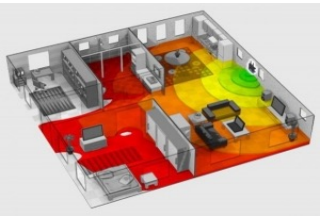
- **RESPECT:** Secure and Privacy-preserving Indoor Robotics for Healthcare Environments
  - Period: 01/05/2021 – 30/04/2024
  - **Research Topic: Privacy Attacks in Localization**

Mobile e-diagnostic

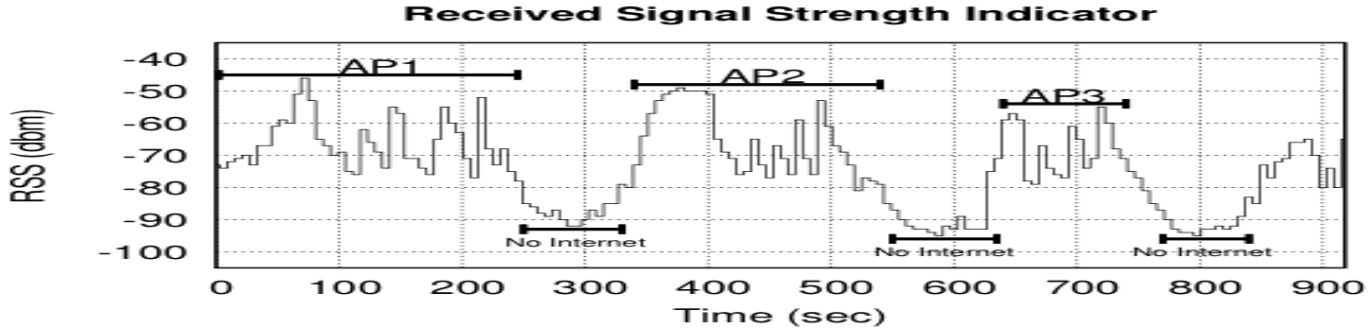


Data Hoarding Challenge

Hoarding



- **Problem:** Wi-Fi coverage might be **irregularly available** inside buildings **due to poor WLAN planning** or due to **budget constraints.**
- A user **walking inside** a **Mall in Cyprus**
  - Whenever the user **enters a store** the RSSI indicator falls **below a connectivity** threshold **-85dBm. (-30dbM to -90dbM)**
  - When **disconnected IIN can't** offer navigation anymore ☹



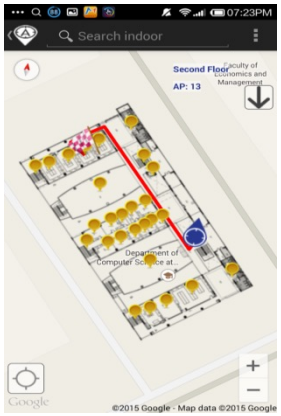
- "IoT Data Prefetching in Indoor Navigation SOAs", Andreas Konstantinidis, Panagiotis Irakleous, Zacharias Georgiou, Demetrios Zeinalipour-Yazti and Panos K. Chrysanthis, ***ACM Transactions on Internet Technology*** (TOIT '18), Vol. 19, Iss. 1, Article 10, pp. 21 pages, **2018**.

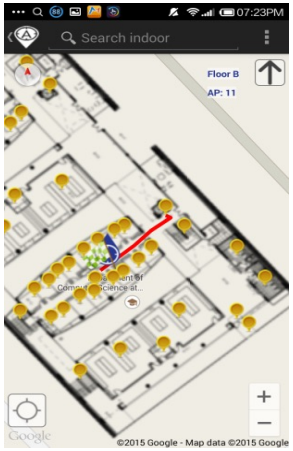
Data Hoarding Challenge

Hoarding

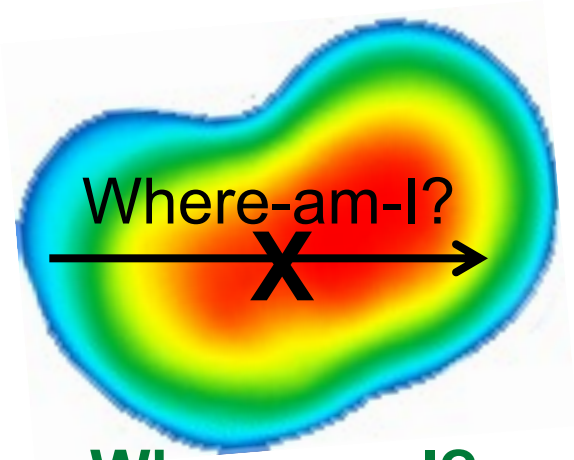
Time

No Navigation





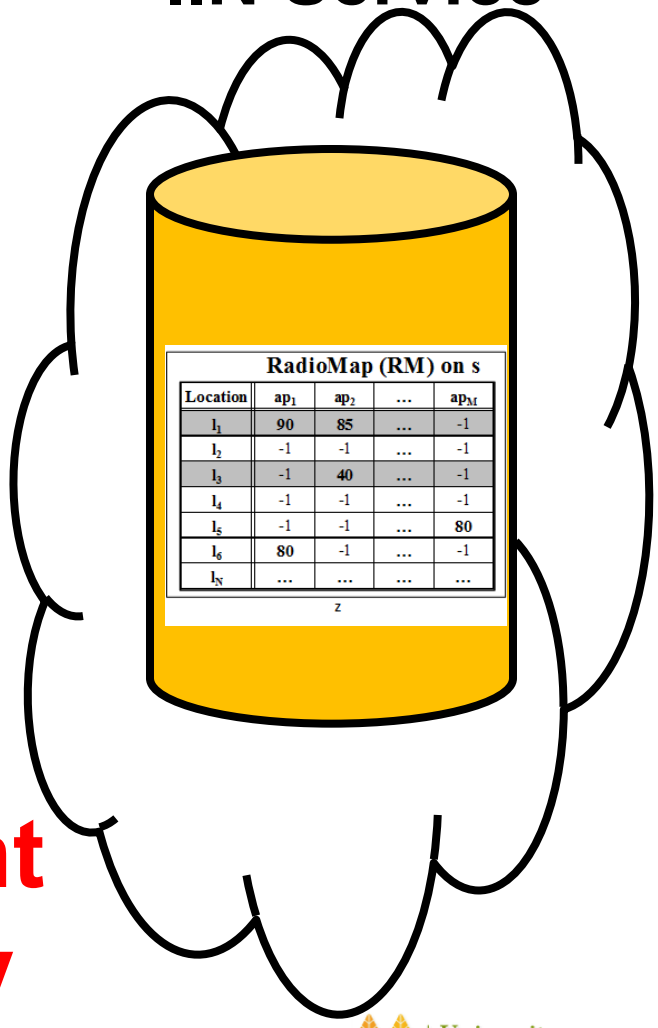
Where-am-I?



Where-am-I?

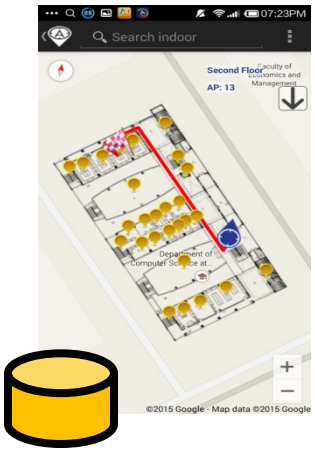
☹ Intermittent Connectivity

IIN Service

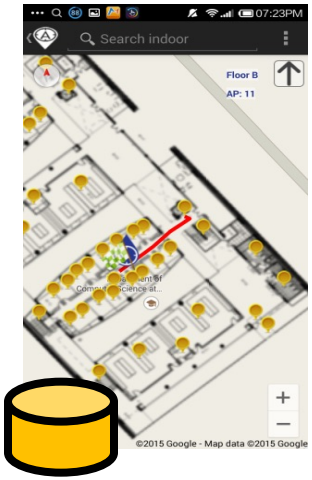


The Preloc Framework

Hoarding: is performed when the connectivity with the server is strong (vs. caching)

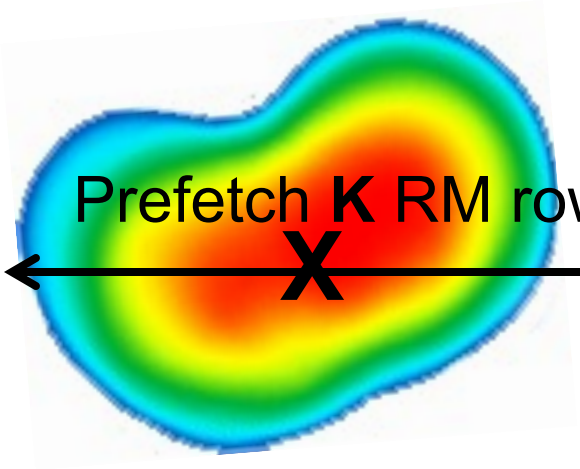


Localize from Cache



Time

Prefetch **K** RM rows



Prefetch **K** RM rows

☺ **Intermittent Connectivity**

IIN Service

RadioMap (RM) on s

Location	ap_1	ap_2	…	ap_M
l_1	90	85	…	-1
l_2	-1	-1	…	-1
l_3	-1	40	…	-1
l_4	-1	-1	…	-1
l_5	-1	-1	…	80
l_6	80	-1	…	-1
l_N	…	…	…	…

z

PreLoc Overview

- Preloc aims to **sequence** the **retrieval** of fingerprint **clusters**, such that the **most important clusters** are downloaded first.
- **Question:** Which **clusters** should a user **download** at a **certain position** if Wi-Fi is **not available** next?
  - PreLoc **prioritizes** the download of **fingerprints** using **historic user traces in a** Dependency Graph.

GraP (Graph Prefetching)

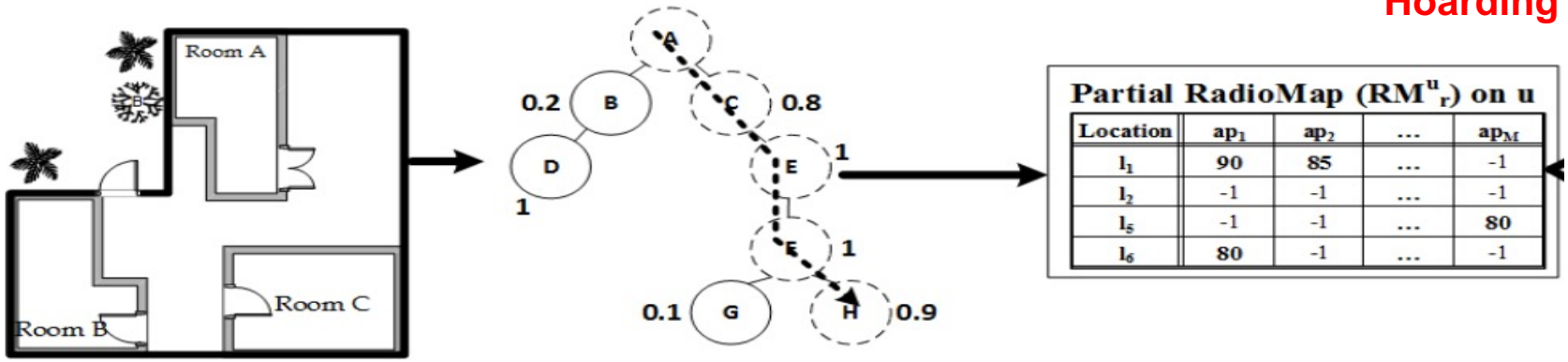
Hoarding

- **Challenges:**
  - **Mobility Traces** inside Buildings are **hard to** obtain.
  - Also goes against our philosophy of **privacy-by-design.**
- **Solution:**
  - We have developed a framework that analyzes building blueprints **to identify hotspots**.
  - These hotspots become **virtual targets** to an **A* search algorithm** we developed.

- "IoT Data Prefetching in Indoor Navigation SOAs", Andreas Konstantinidis, Panagiotis Irakleous, Zacharias Georgiou, Demetrios Zeinalipour-Yazti and Panos K. Chrysanthis, **ACM Transactions on Internet Technology (TOIT'18)**, Vol. 19, Iss. 1, Article 10, pp. 10:1-10:21, 2018.

GraP: Target-less A* Search

Hoarding

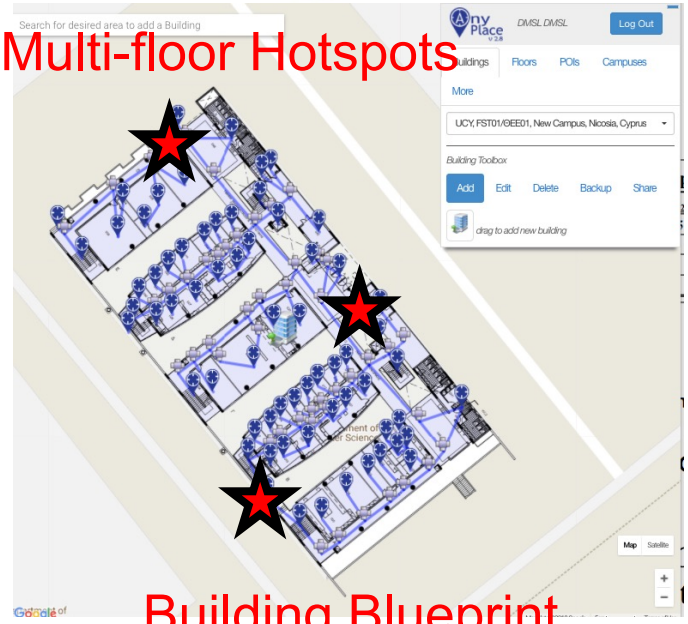


Partial RadioMap (RM^u_r) on u

Location	ap_1	ap_2	...	ap_M
l_1	90	85	...	-1
l_2	-1	-1	...	-1
l_5	-1	-1	...	80
l_6	80	-1	...	-1

B

Multi-floor Hotspots

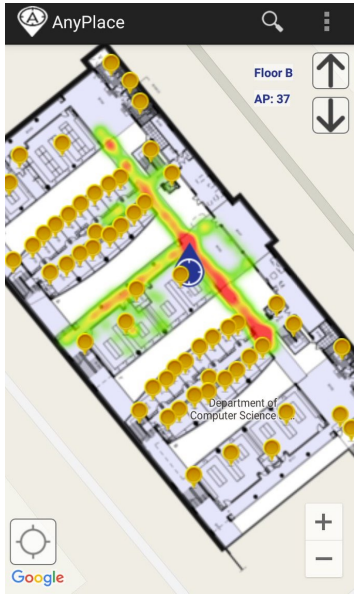


Building Blueprint

DG & GDA

Hotspots become "virtual" targets to a target-less A* prefetching (hoarding) algorithm we've developed.

Prefetching



Location Privacy

-*Towards planet-scale localization on smartphones with a partial radiomap"*, A. Konstantinidis, G. Chatzimilioudis, C. Laoudias, S. Nicolaou and D. Zeinalipour-Yazti. In ACM HotPlanet'12, in conjunction with **ACM MobiSys '12,** ACM, Pages: 9--14, 2012.
- *Privacy-Preserving Indoor Localization on Smartphones*, A. Konstantinidis, G. Chatzimilioudis, D. Zeinalipour-Yazti, P. Mpeis, N. Pelekis, Y. Theodoridis, in **IEEE TKDE'15.**

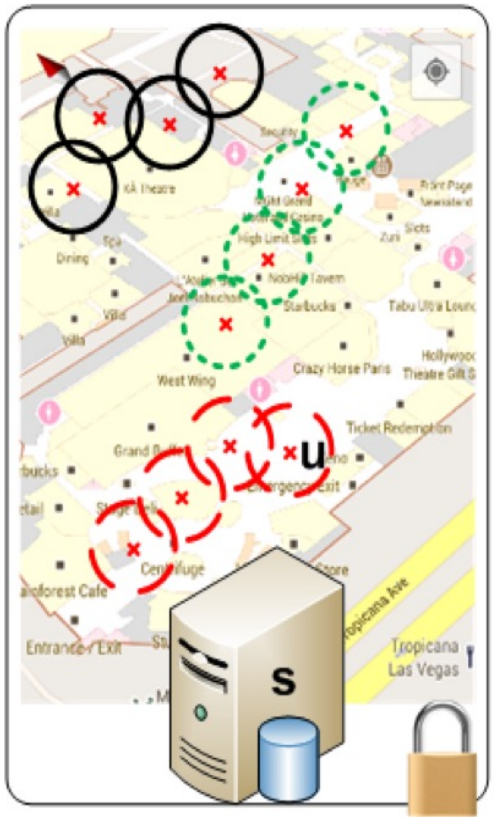
Temporal Vector Map (TVM)

WiFi

WiFi

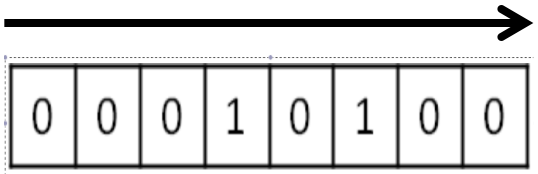
...

WiFi



User u

Bloom Filter (u's APs)



Set Membership Queries

IIN Service

K=3 Positions

Visual Analytics / Accuracy Estimation

Visual Analytics Accuracy Estimation / Differential Fingerprints

- **[C72]** "FMS: Managing Crowdsourced Indoor Signals with the Fingerprint Management Studio", Marileni Angelidou, Constantinos Costa, Artyom Nikitin and **Demetrios Zeinalipour-Yazti**, **Proceedings of the 19th IEEE International Conference on Mobile Data Management** (MDM'18), IEEE Computer Society, ISBN: 978-1-5386-4133-0, pp. 288--289, June 25 - June 28, 2018, AAU, Aalborg, Denmark, DOI: 10.1109/MDM.2018.00054, **2018**. **[Best Demo Award!]**
- **[J26]** "Indoor Quality-of-Position Visual Assessment using Crowdsourced Fingerprint Maps", Christos Laoudias, Artyom Nikitin, Panagiotis Karras, Moustafa Youssef, **Demetrios Zeinalipour-Yazti**, ***ACM Transactions on Spatial Algorithms and Systems*** (TSAS'21), Association for Computing Machinery, Vol. 7, Iss. 2, New York, NY, USA, DOI: 10.1145/3433026, **2021**.
- **[J30]** "Cramér–Rao Lower Bound Analysis of Differential Signal Strength Fingerprinting for Crowdsourced IoT Localization", Jiseon Moon, Christos Laoudias, Ran Guan, Sunwoo Kim, **Demetrios Zeinalipour-Yazti** and Christos G. Panayiotou, ***IEEE Internet of Things Journal*** (IoTJ'23), IEEE Computer Society, pp. 13 pages, Los Alamitos, CA, USA, DOI: , **2023**.

Anyplace Computer Vision

CV Localization

- LASH FIRE is an international EU-funded research project aiming to significantly reduce the risk of fires on board ro-ro ships.
  - As part of this objective, we developed a **"zero" infrastructure localization method for smartphones of first responders.**

LASH FIRE

[C92] "Zero Infrastructure Geolocation of Nearby First Responders on Ro-Ro Vessels", Paschalis Mpeis, Jaime Bleye Vicario and Demetrios Zeinalipour-Yazti, **Proceedings of the International Conference on Computer Applications in Shipbuilding** (ICCAS'22), The Royal Institution of Naval Architects (RINA, est. 1860), pp. 249--263, Yokohama, Japan, September 13-15, 2022, DOI: 978-1-911649-35-9, **2022**.

[C91] "SMAS: A Smart Alert System for Localization and First Response to Fires on Ro-Ro Vessels", Paschalis Mpeis, Athina Hadjichristodoulou, Jaime Bleye Vicario and Demetrios Zeinalipour-Yazti, **Proceedings of 16th ACM International Conference on Distributed and Event-based Systems** (DEBS'22), Association for Computing Machinery, pp. 4, 27th June – 30th June 2022, Copenhagen, Denmark, DOI: https://doi.org/10.1145/3524860.3543282, **2022**.

Smart Alert System (SMAS)

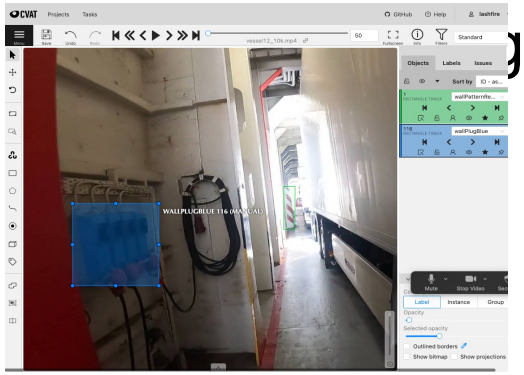
- A **mobile app** for **First Responders** implementing **"Zero Infrastructure" + Offline** localization system.
  - ➔ requires no localization infrastructure
  - ➔ requires no network
  - ➔ requires no video transfer
  - ➔ only on-device computation
- If network is available, then also supports chat and alert.
  - Edge server on-premise (vessel)

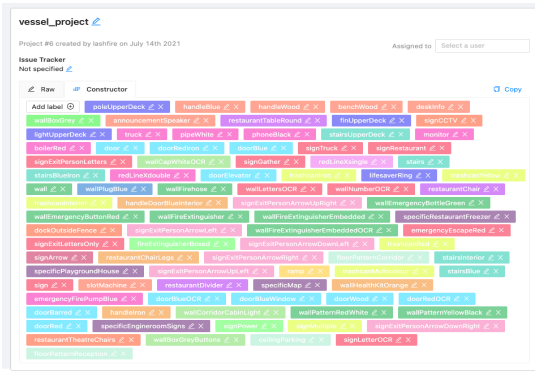
CV Localization

ML Training

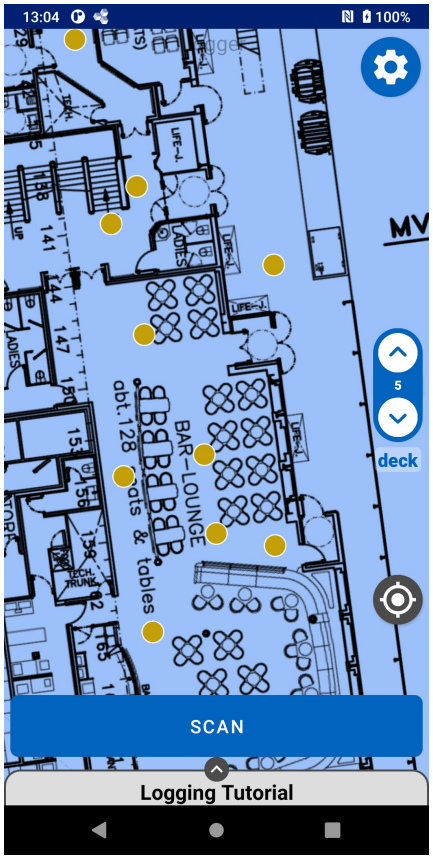
CV Localization

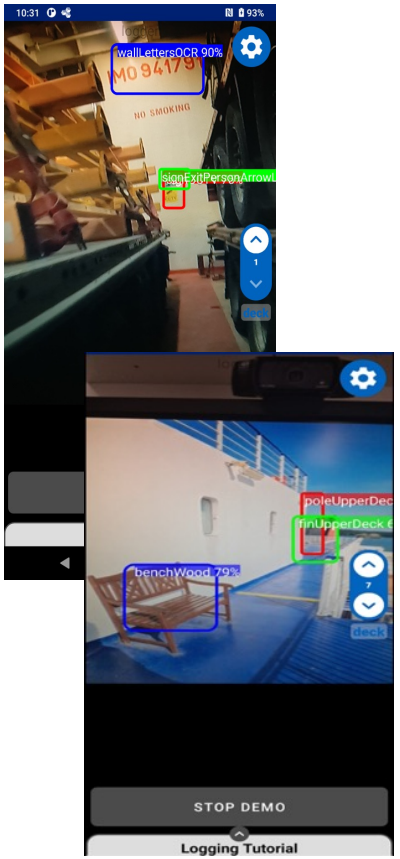
- **Purpose:** Train a Machine Learning model to recognize objects using **Computer Vision** and **Deep Learning.**
- Computer Vision Annotation Tool (CVAT)
  - a free, open source, web-based image and video annotation tool by Intel
  - HP DL380 Gen10 with 80 logical processors and a powerful NVIDIA V100 card.





CV Logging





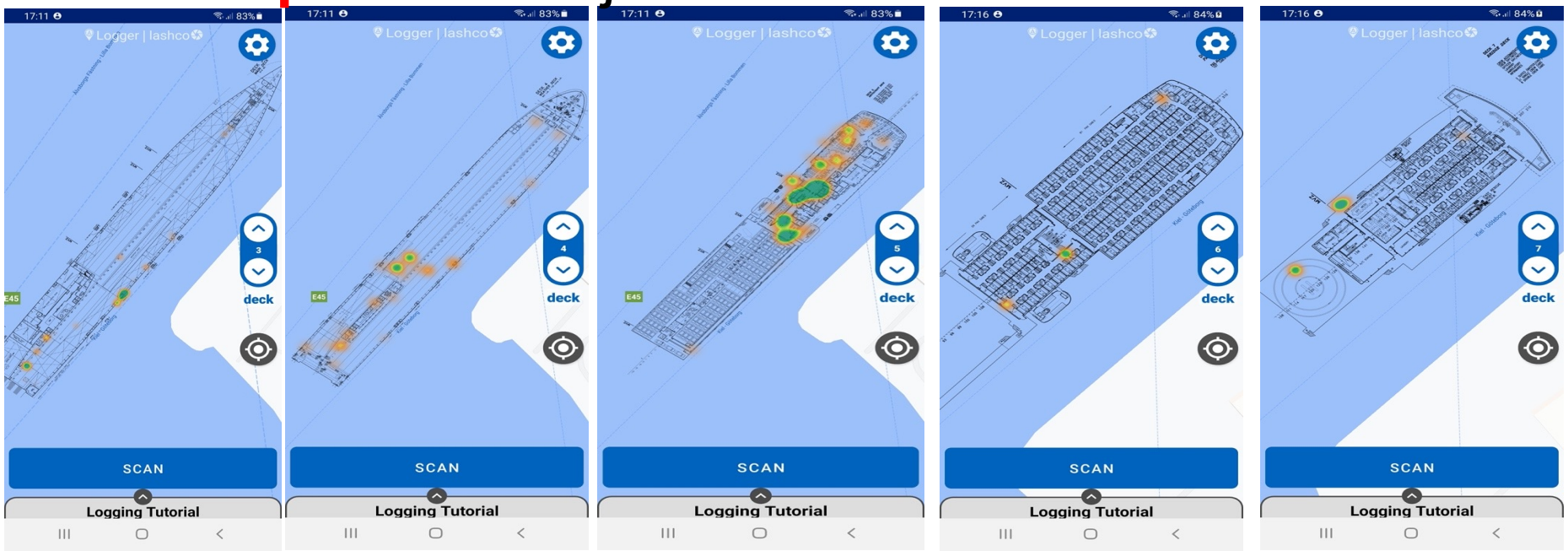
CV Localization

Google's ML kit on-device **Optical Character Recognition (OCR)** library (available in multiple languages) can refine an object recognition. For example, for the last figure it reads **"STB PILOT DOOR BUNKER TTION"**

CV Logging

heatmaps show object collection

CV Localization



Estimations
(based on onboard study)

Logging 7 decks: 16-20 hours

Objects 7 decks: 11K objects

Vessel Characteristics

- **Length overall (LOA):** 186.42
- **Width:** 25.6 meters.
- **Gross Tonnage:** 26904
- **Draught:** 6.4m

Surface Algorithm

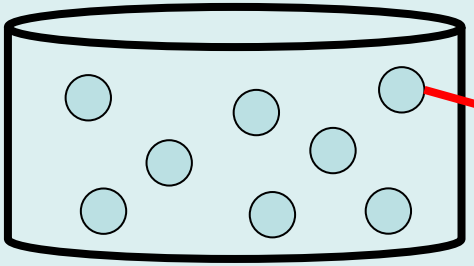
- **Surface:** Our data-driven localization algorithm using Computer Vision (CV).

CV Localization

- **Problem:** How to rank objects in a way that the correct location is estimated?

Query: Where am I?
I see the below objects …

{drencher, door}



Fingerprint Database (FDB)

Record:
[***x,y,deck,***{drencher, charger, door}]

- Additional Challenges:
  - How to filter out the **proximity** of similar fingerprints?
  - How **important objects** can be ranked higher?
  - How to support **tracking** (Continuous Localization)?
  - How to reduce the **quadratic lookup cost**?

Surface Algorithm

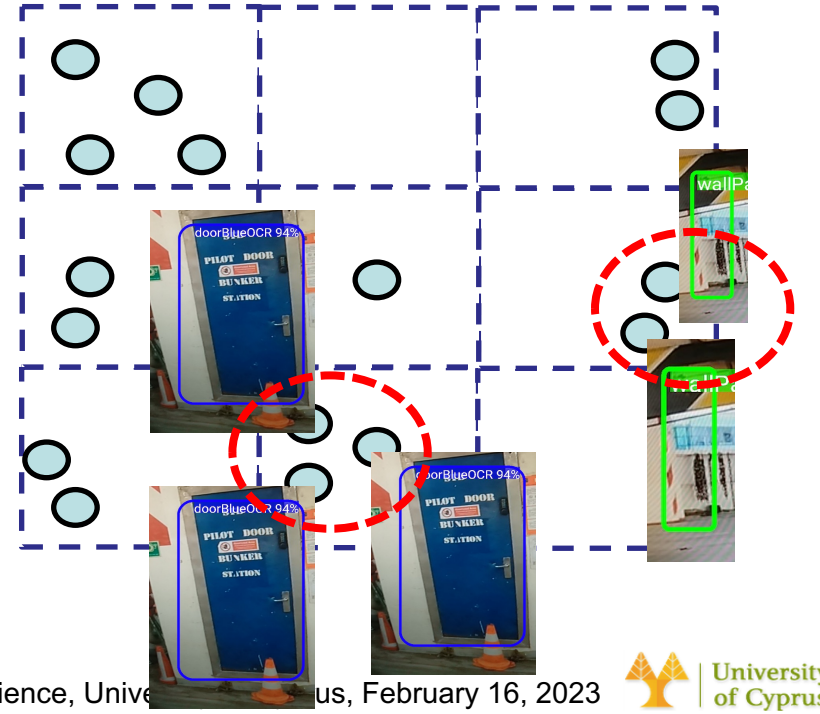
CV Localization

- Surface (Data Management / SQL) Concepts
  - A) Multiset Subtraction
  - B) Spatial Partitioning of Fingerprints
  - C) Implemented in SQLite (Android and backend) – I/O tuned
  - D) Bounding Box Filtering of Fingerprints

```
SELECT F.fid, F.X, F.Y, F.deck,
       ABS(x - "$prevX") as xDiff, ABS(y - "$prevY") as yDiff, ABS(deck - "$prevDeck") as deckDiff,
       (SELECT IFNULL(COUNT(*),0) FROM
(-- Ranking Criterion A: Multi-set Subtraction as
SELECT ROW_NUMBER() OVER (PARTITION BY FLT.oid) AS RowNum, FLT.oid
FROM FINGERPRINT_LOCALIZE_TEMP FLT WHERE FLT.uid="$uid"
EXCEPT
SELECT ROW_NUMBER() OVER (PARTITION BY FO.oid) AS RowNum, FO.oid
FROM FINGERPRINT_OBJECT FO WHERE FO.fid=F.fid
) ) AS dissimilarity, (SELECT IFNULL(AVG(weight),1) FROM
(-- Ranking Criterion B: Multi-set Subtraction on Global Partitioned Frequency Counting with Spatial Partitioning of Fingerprints
SELECT ROW_NUMBER() OVER (PARTITION BY FLT.oid) AS RowNum, FLT.oid, OF.weight as weight
FROM FINGERPRINT_LOCALIZE_TEMP FLT, OBJECT_FREQUENCY OF
WHERE FLT.oid = OF.oid and FLT.uid="$uid"
EXCEPT
SELECT ROW_NUMBER() OVER (PARTITION BY FO.oid) AS RowNum, FO.oid, OF.weight as weight
FROM FINGERPRINT_OBJECT FO, OBJECT_FREQUENCY OF
WHERE FO.oid = OF.oid and FO.fid=F.fid
)
) AS weight
FROM FINGERPRINT F
-- Bounding Rectangle Filtering of Fingerprints
WHERE F.modelid="$modelid" and F.build="$build" and
      (x between "$prevX" - "$smas_db_location_bound_meters" and "$prevX" + "$smas_db_location_bound_meters") and
      (y between "$prevY" - "$smas_db_location_bound_meters" and "$prevY" + "$smas_db_location_bound_meters")
      and (deck between "$prevDeck" - 1 and "$prevDeck" + 1)
-- keep only x,y,deck rounded by 10m bounding box
GROUP BY ROUND(F.X,"$smas_db_location_bound_rounding"), ROUND(F.Y,"$smas_db_location_bound_rounding"), F.deck
-- include results that have at least 1 common object
HAVING dissimilarity < (SELECT COUNT(*) from FINGERPRINT_LOCALIZE_TEMP FLT WHERE FLT.uid="$uid")
-- rank by dissimilarity, then OBJECT_FREQUENCY weight than by distance from prior location
ORDER BY dissimilarity, weight, ABS(x - "$prevX") + ABS(y - "$prevY") + ABS(deck - "$prevDeck") ASC
-- return highest ranked result only
LIMIT 1;
```

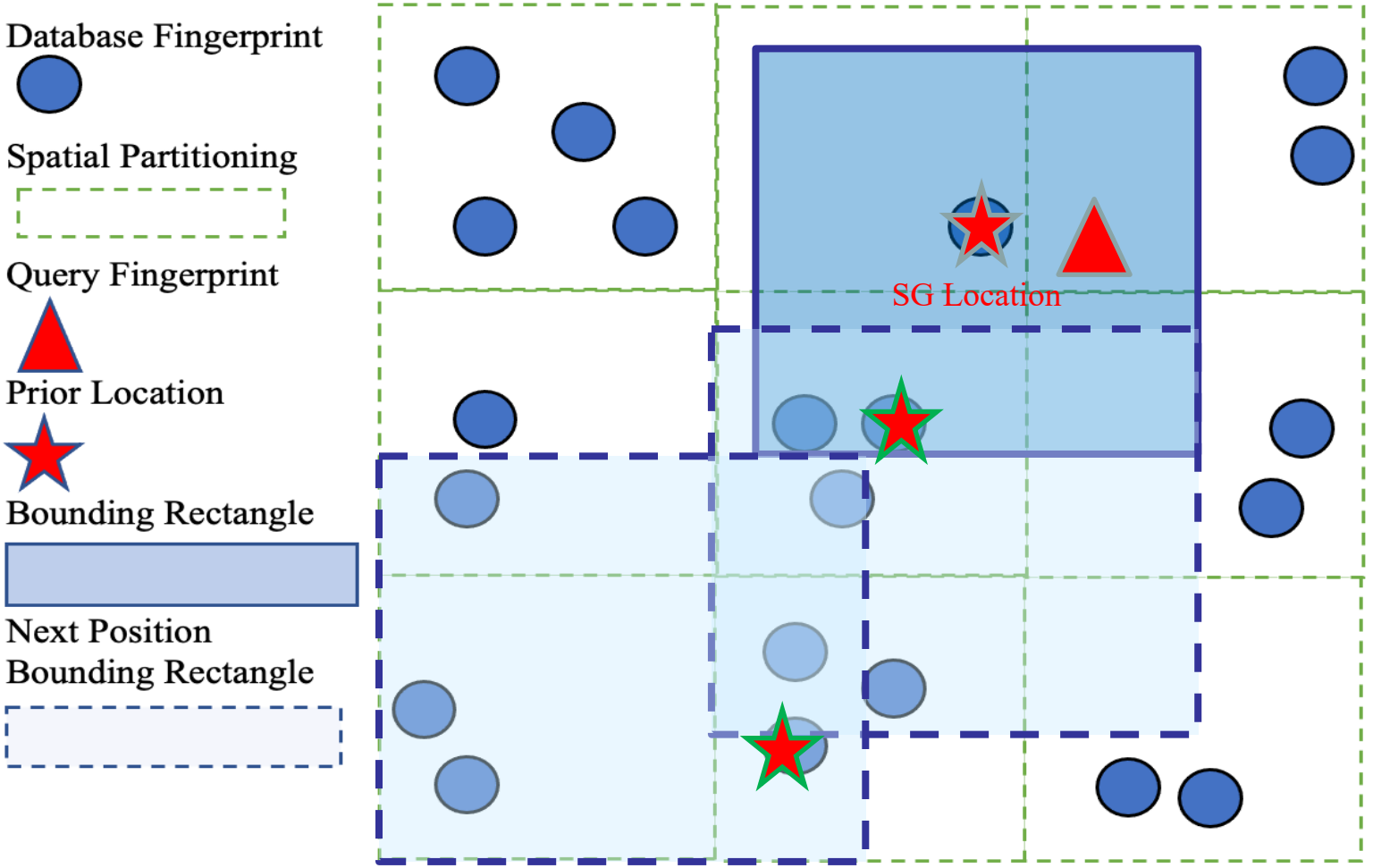
Dissimilarity is 1

(signTruck not found)



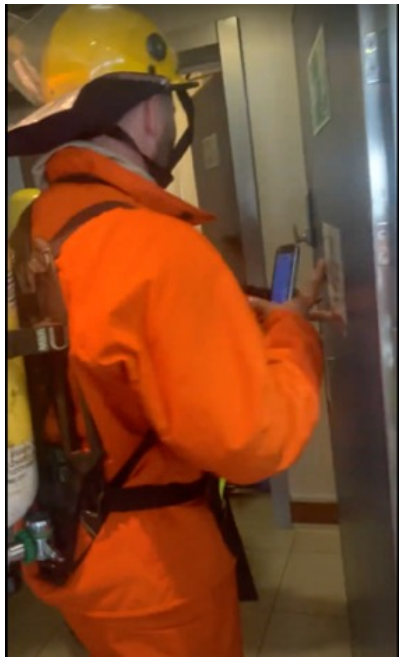
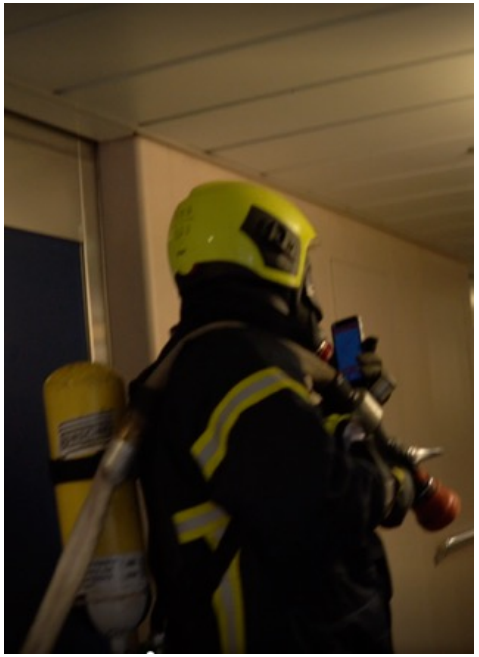
Surface Algorithm

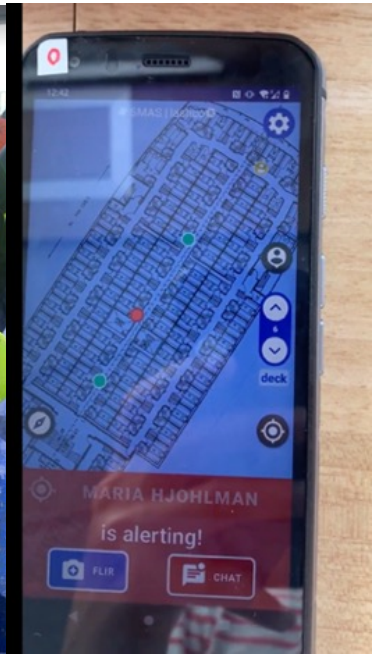
- Bounding Box Filtering of Fingerprints

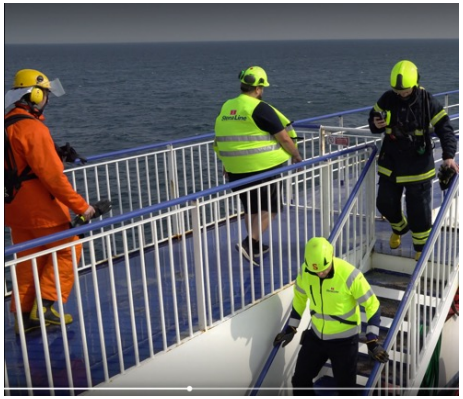
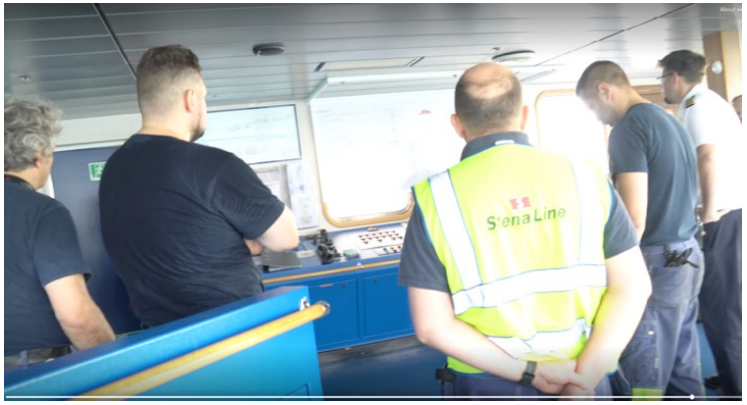


CV Onboard Study

CV Localization




future

Motivation

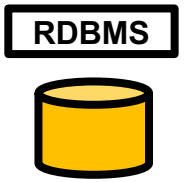
Decaying

- The expansion of mobile networks and IoT have contributed to an explosion of data inside **Telecommunication Companies (Telcos)**

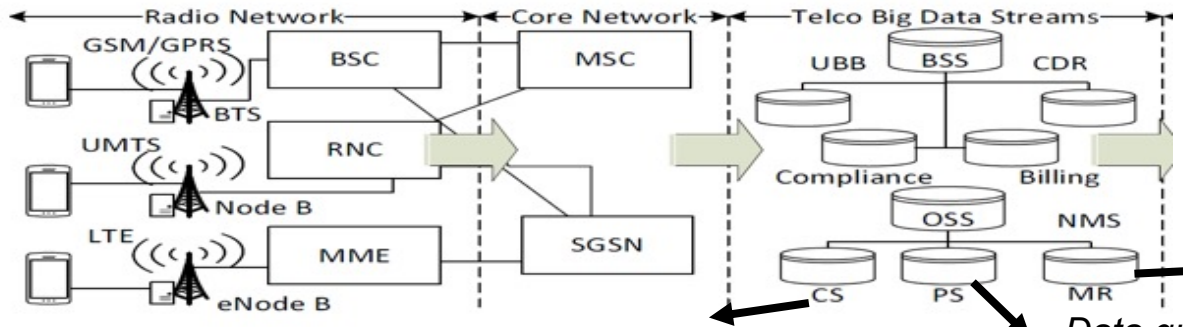
- **[Conference]** "Efficient Exploration of Telco Big Data with Compression and Decaying", Constantinos Costa, Georgios Chatzimilioudis, Demetrios Zeinalipour-Yazti, Mohamed F. Mokbel, **Proceedings of the IEEE 33rd International Conference on Data Engineering** (ICDE'17), IEEE Computer Society, pp. 1332-1343, April 19-22, 2017, San Diego, CA, USA, DOI: 10.1109/ICDE.2017.175, ISBN: 978-1-5090-6543-1, **2017**.
- [Tutorial] "Telco Big Data Research and Open Problems", Constantinos Costa and Demetrios Zeinalipour-Yazti, Proceedings of the **35th IEEE International Conference on Data Engineering (ICDE'19),** IEEE Computer Society, 8-12 April 2019, Macau SAR, China, **2019.**

Telco Big Data (TBD)

Decaying

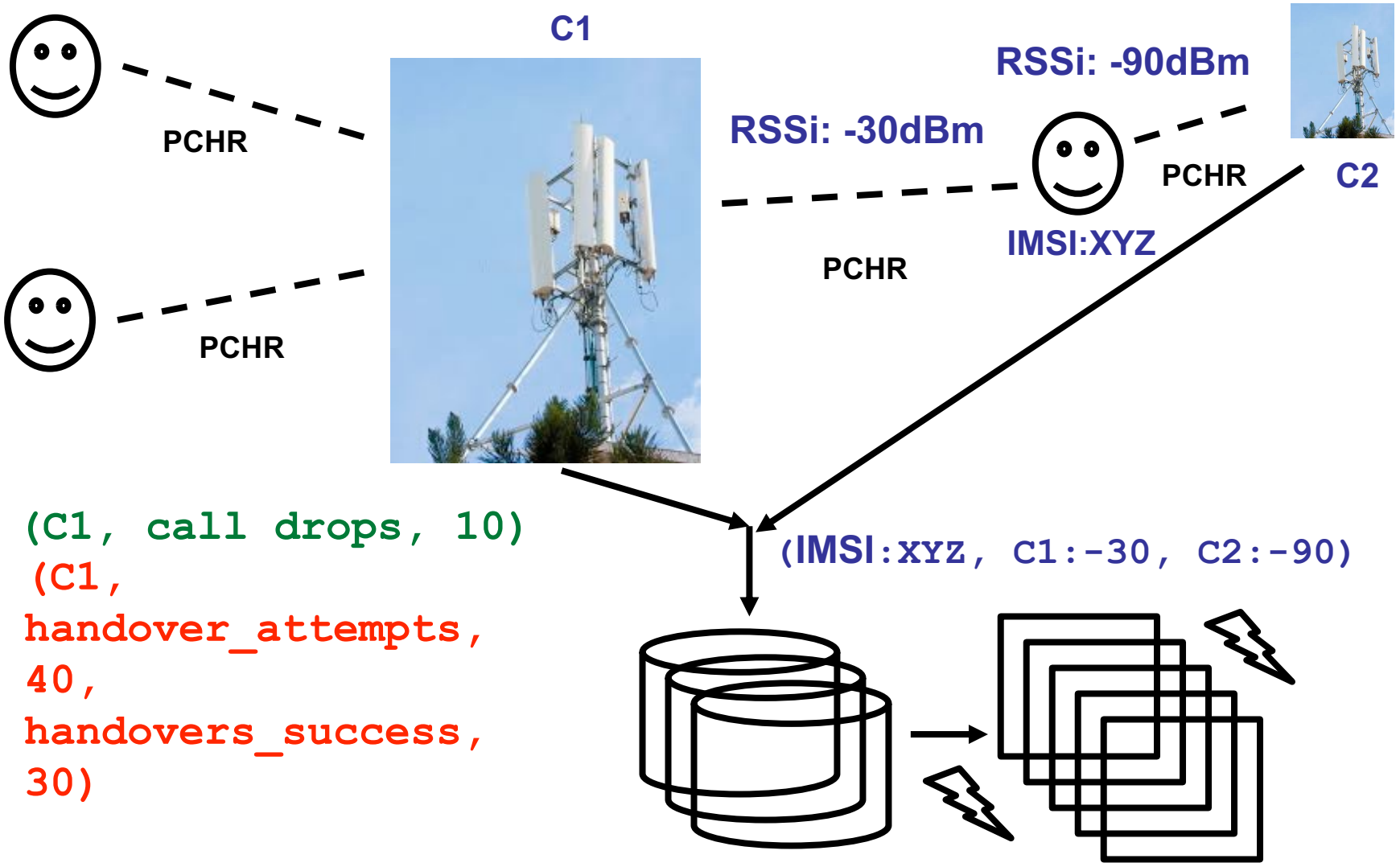


- **Telco Data:** Traditional source for OLAP Data Warehouses and Analytics.
  - e.g., Accounting, Billing, Session data.
  - **Problem:** Inadequate data resolution to address biggest challenges:
    - e.g., 5G network optimization, user-experience assessment (churn prediction), road network traffic mapping.
- **Telco Big Data (TBD):** Velocity data from cell towers.
  - e.g., signal strength, call drops, bandwidth measurements.
  - **Size:** 5TBs/day for 10M clients (i.e., 2PB/year).



Measurement Reports (can be used to localize users with triangulation ~30-50 meters)

Measurement Reports (MR)*

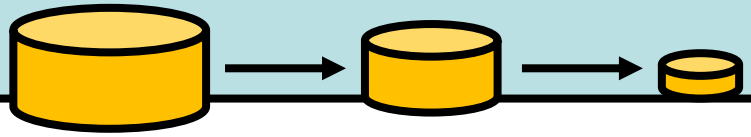


* **Alternative:** Mobile BroadBand (MBB) OR Network Measurement System (NMS) data

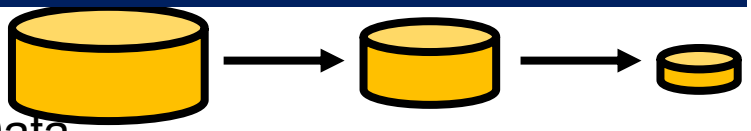
TBD Challenge

Decaying

- The Big data era lead us to a point where organizations collect **more than they can**!
  - Global **volume** of stored data **doubling** every **2 years**.
  - Costs for data storage **decline** only at a rate of **~15% per year**. Datacenter Journal, https://goo.gl/o4MnJp
- TBD is **straining** telco datacenters that can not benefit from economies-of-scale available on public clouds (due to confidentiality/security).
- **Our Approach:** Introduce a complete TBD analytic stack that makes **Compression** and **Decaying** a first class citizen.



How to Reduce Data ?

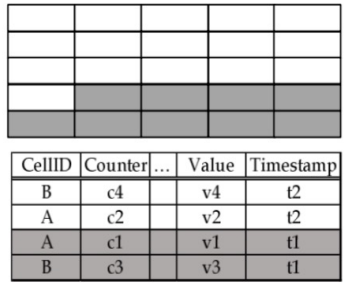


- **Data Sampling**
  - Uniform or Random Data < Original Data
- **Data Aggregation (Data Cubes)**
  - Query Result Data < Original Data
- **Data Reduction (SVD, DFT, DWT)**
  - Principal Components (Patterns) << Original Data
- **Data Synopsis** (e.g. equiwidth/depth histograms, Bloom filters, sketches)
  - "Statistics" about Data (e.g., bit vector) << Original Data
- **Data Compression** (lossless or lossy)
  - Combination of techniques, e.g., GZIP=DEFLATE(LZ77+HUFF)
- **Problem:** None of the above considers how to outdate (decay) data as time elapses.

"Efficient Exploration of Telco Big Data with Compression and Decaying", Constantinos Costa, Georgios Chatzimilioudis, Demetrios Zeinalipour-Yazti, Mohamed F. Mokbel, Proceedings of the **IEEE 33rd International Conference on Data Engineering (ICDE '17),** IEEE Computer Society, pp. 1332-1343, April 19-22, 2017, San Diego, CA, USA, ISBN: 978-1-5090-6543-1, 2017.

Indexing Layer: Decaying

- **Decaying** refers to the *"progressive loss of detail in information as data ages with time until it has completely disappeared."*
  - M. L. Kersten, "Big data space fungus," in CIDR'15
  - M. L. Kersten, L. Sidirourgos "A Database System with Amnesia" in CIDR'17



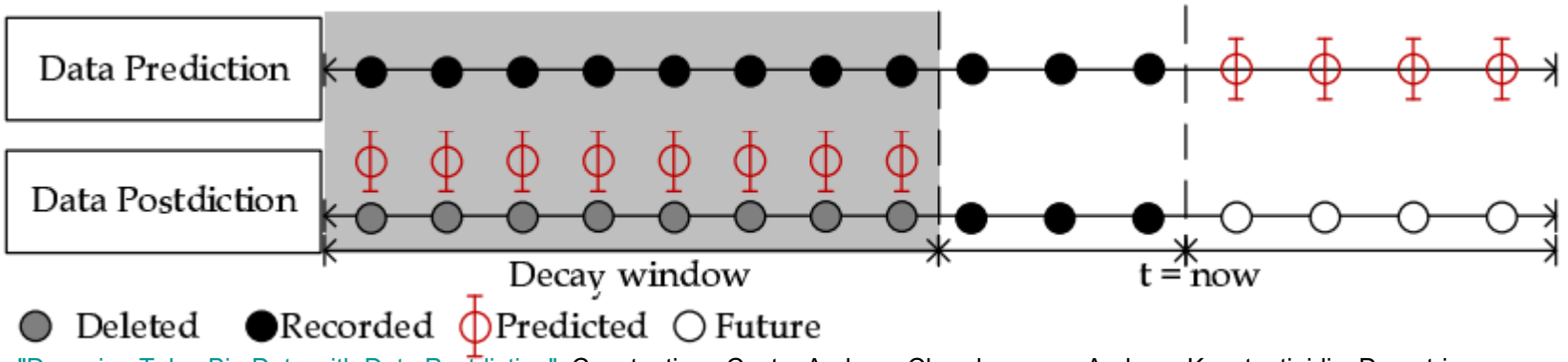
CellID	Counter	...	Value	Timestamp
B	c4		v4	t2
A	c2		v2	t2
A	c1		v1	t1
B	c3		v3	t1

FIFO-Amnesia

- **Benefits:**
  - Retain aggregate data exploration capabilities.
  - Save enormous amounts of storage and I/O.
- Alternative definition referring to **DB Schema Decay** also exists, but is not applicable here.
  - M. Stonebraker, R. Castro, F. Dong Deng, and M. Brodie, "Database decay and what to do about it." 2016. [Online]. BLOG@CACM: https://goo.gl/tJNa9m.

Data Postdiction

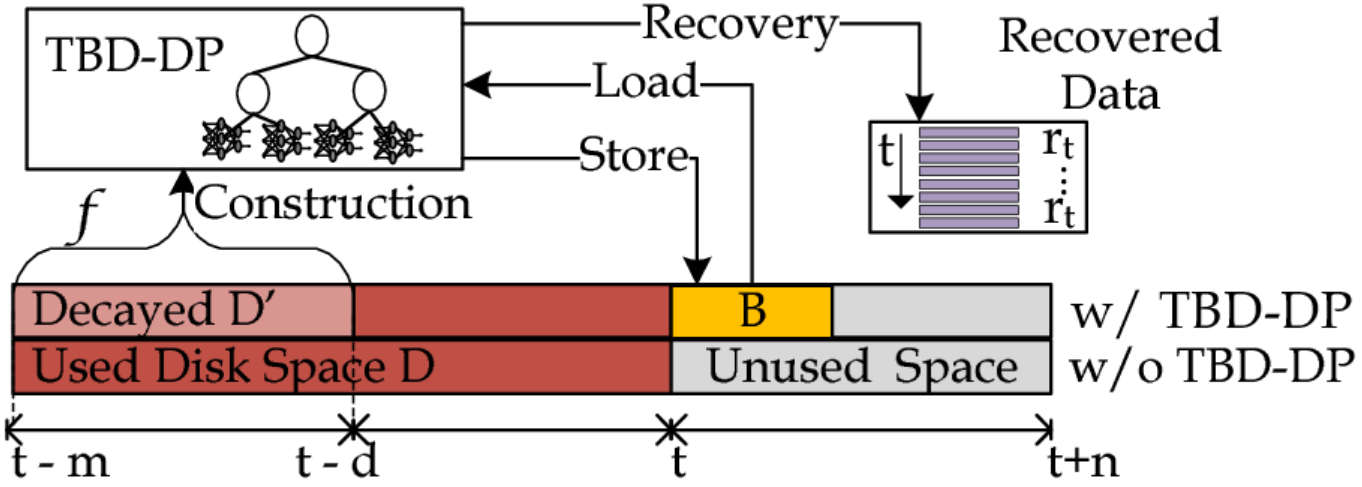
- **Data Postdiction (DP)**: aims to recover the past value of some tuple, which has been **decayed** for efficiency purposes, using a ML model.
  - **Data Prediction**: aims to make a statement about the future value of some tuple using a ML model.
  - **Data Postdiction:** a technique to carry out data decaying (other: FIFO-amnesia, UNIFORM-amnesia, Exponential-amnesia)



- "Decaying Telco Big Data with Data Postdiction", Constantinos Costa, Andreas Charalampous, Andreas Konstantinidis, Demetrios Zeinalipour-Yazti and Mohamed F. Mokbel, 19th IEEE International Conference on Mobile Data Management (MDM'18), ISBN: 978-1-5386-4133-0, pp. 106--115, June 25 - June 28, 2018, AAU, Aalborg, Denmark, 2018. (Best of IEEE MDM'19, Aalborg, Denmark)
- **"Continuous Decaying of Telco Big Data with Data Postdiction"**, Constantinos Costa, Andreas Charalampous, Andreas Konstantinidis, Demetrios Zeinalipour-Yazti, Mohamed F. Mokbel, ***International Journal on Advances of Computer Science for Geographic Information Systems*** (**GeoInformatica'19**), Vol. 23, Iss. 4, pp. 25 pages, DOI: 10.1007/s10707-019-00364-z, **2019**.

TBD-DP Operator

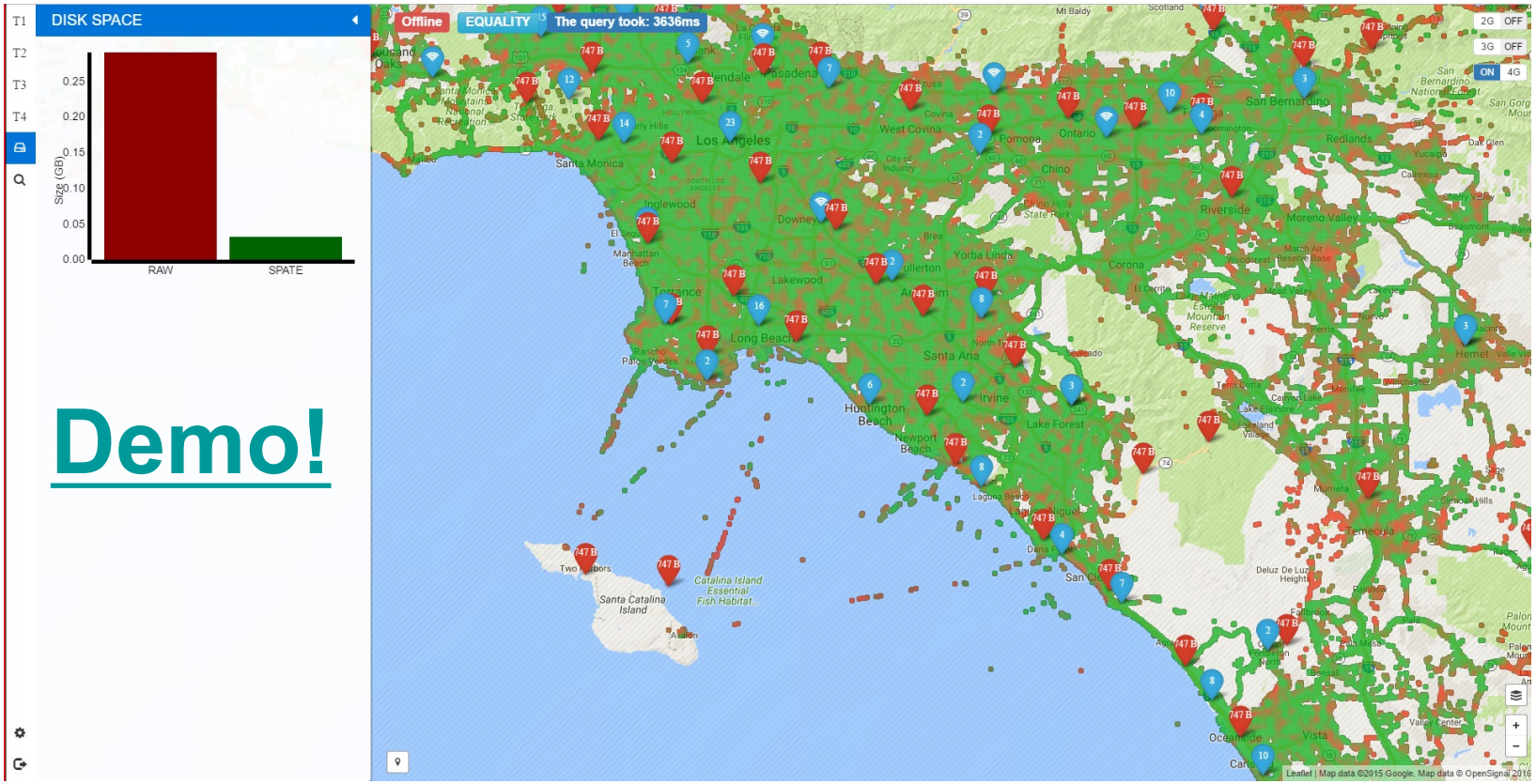
- TBD-DP Operator Overview
  - ***Construction* algorithm:** construct a **DP-tree (B)** for a percentage **(f)** of the historic data **(D), denoted as D'**.
    - Delete **D'** and retain **B** for data recovery**.**
  - ***Recovery* algorithm:** use **{B, D-D'}** to recover **any past** data blocks, upon demand.
    - **D =** full data resolution and **B =** model data resolution



Application Layer: SPATE-UI

Decaying

- The **SPATE UI** over the San Diego area with data from https://www.cellmapper.net/map



Demo!

Experimental Methodology

- To evaluate *TBD-DP operator*, we have implemented a **trace-driven** experimental testbed:
  - Compared Approaches/Algorithms:
    - **RAW:** does not apply any decaying.
    - **COMPRESSION:** the decayed dataset is compressed with the GZIP library (used in *SPATE* @ ICDE'17)
    - **SAMPLING:** retain 50% (i.e., (1-f)%) full resolution + every second item in the rest input stream (i.e., average 75%).
    - **RANDOM:** retain 50% full resolution + uniformly and randomly select one additional record.
    - **TBD-DP:** retain 50% full resolution + Models.
  - Metrics:
    - **Storage Capacity(MB):** the total space to store data and index
    - **Normalized Root Mean Square Error (MB):** the error of the recovered data D′ using the well known NRMSE.

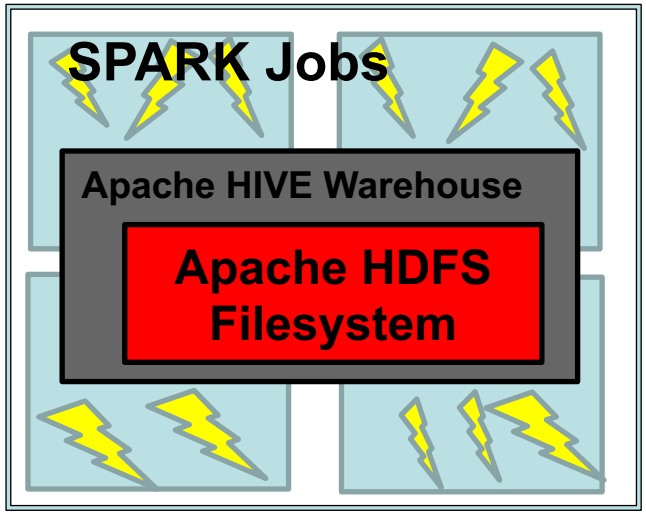
Experimental Datasets

- We constructed 6 realistic KPI Queries based on an **anonymized TBD dataset** [C64]:
  - **Dataset:** 3660 cells (CELL) coming from 2G, 3G and LTE antennas,100M NMS records, 300K users, ~10GB.
  - **KPIs (Key Performance Indicators) Queries**: Calls (CS), Call Drops (CSD), ThroughPut (TP), Handover Attempts (HA), Handovers (HS), Call Setup Attempts (CSA), Call Setups (CS)

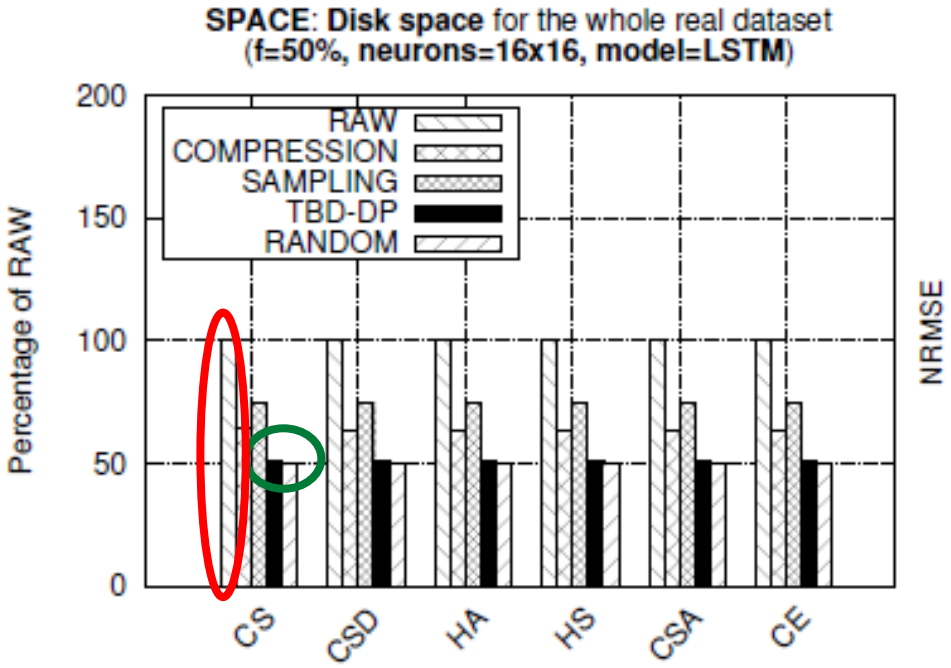
[C64] "**Efficient Exploration of Telco Big Data with Compression and Decaying**", Constantinos Costa, Georgios Chatzimilioudis, Demetrios Zeinalipour-Yazti, Mohamed F. Mokbel, **Proceedings of the IEEE 33rd International Conference on Data Engineering** (**ICDE'17**), IEEE Computer Society, pp. 1332-1343, April 19-22, 2017, San Diego, CA, USA, DOI: 10.1109/ICDE.2017.175, ISBN: 978-1-5090-6543-1, **2017**.

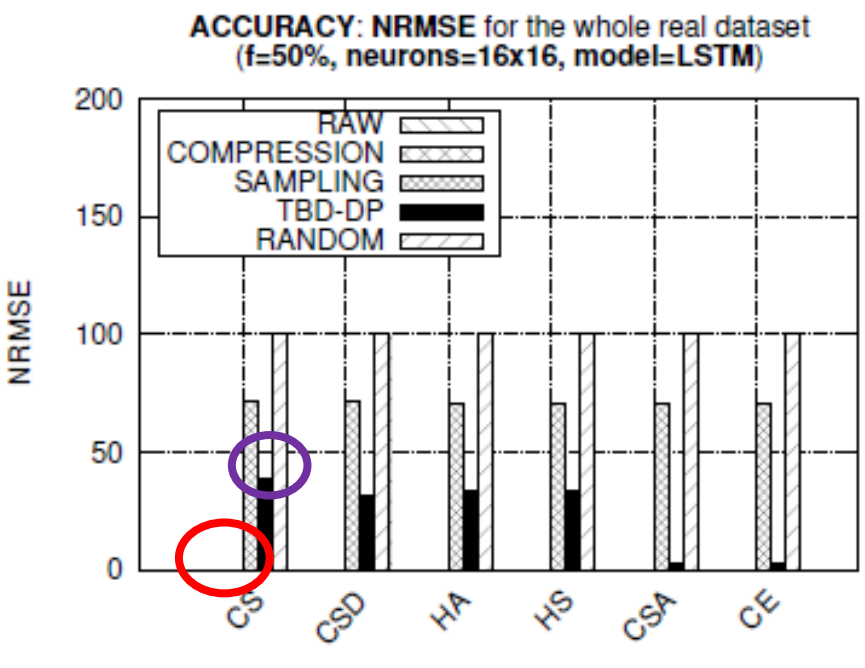
Experimental Testbed

- **TBD Operating System Stack**
  - **Datacenter:** VMWare ESXi 5.0.0 Hosts
  - **VMs:** 4 Ubuntu 14.04 server images, each featuring: 8GB of RAM with 2 virtual CPUs (2.40GHz)
  - **Storage Element:** Slow 7.2K RPM RAID-5 SAS, 6 Gbps disks. Each disk formatted in VMFS 5.54 (1MB block)
- **TBD Framework Stack**
  - Hadoop Distributed File System (HDFS) v2.5.2
  - Apache Hive 2.0 (online querying)
  - Apache Spark 1.6.0 (micro-batching => future Dataflow/Flink on-the-fly process)
  - Python 3.6.3
  - Tensorflow 1.7 (LSTM, RNN, GRU)



Space Capacity

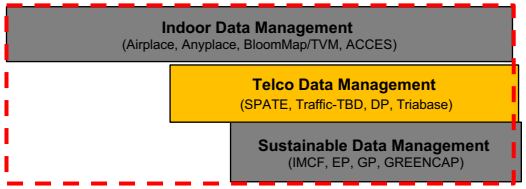




- **Observations:**
  - **TBD-DP** provides **25%, 50%** better **space capacity** than **COMPRESSION** and **RAW**, respectively.
  - **TBD-DP** outperforms the **SAMPLING** approach by **50%** in terms of **NRMSE**, on average, in all datasets.
  - **COMPRESSION** approach provides an optimal **NRMSE = 0** (GZIP is lossless, but requires more space)
  - **TBD-DP** could have been configured with a decay factor **f=100%** rather than **f=50%,** yielding even less space (i.e., decay everything, retain only model approach)

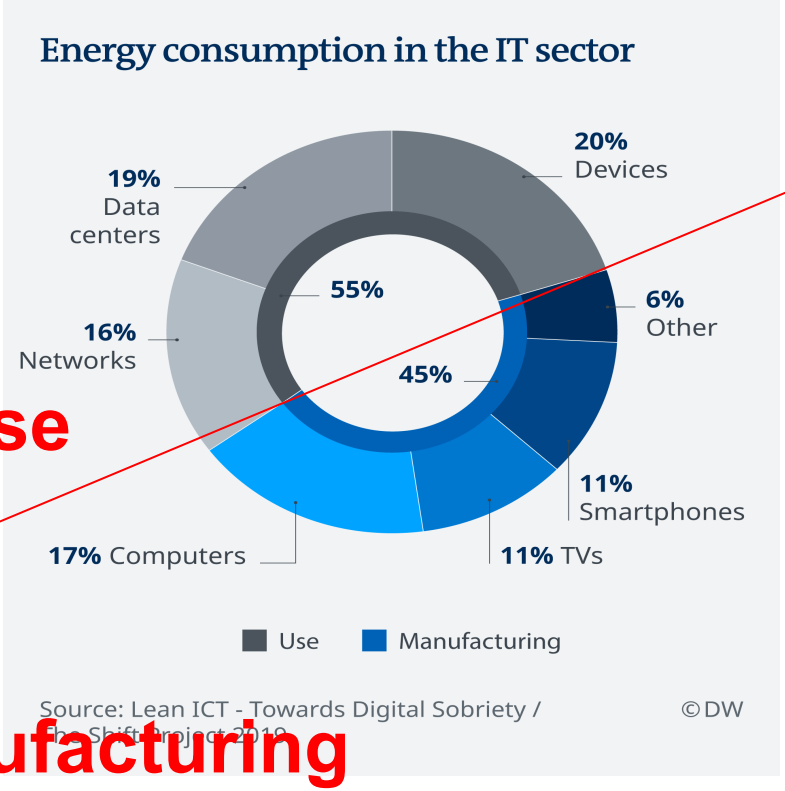
Presentation Roadmap

- Introduction
- Indoor Data Management
  - ACM TOIT'18, ACM TSAS'21, ACM TSAS'22, ACM DEBS'22, ICCAS'22, IEEE TKDE'15, IEEE JIoT'23
- Telco Data Management
  - IEEE ICDE'19, Geoinformatica'19, IEEE MDM'18, IEEE ICDE'17
- **Sustainable Data Management**
  - **ACM TIOT'22, IEEE IC'22, IEEE ICDE'21 EDBT'21**



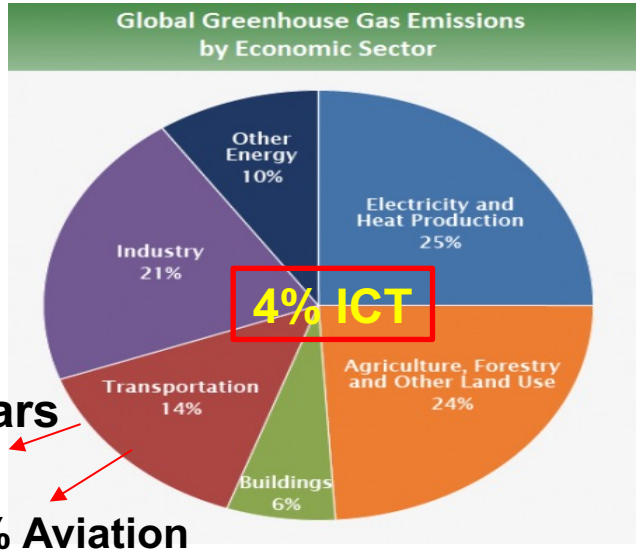
Motivation

- ICT contributes **~2-4%** of world CO2 emissions and will increase to **8%** by 2025!



55% use

45% manufacturing



10.5% ICE Cars

2.5% Aviation

How to make impactful research to reduce this increase from 4% to 8%?

Green Smart Homes

- 32.3% of Australian Homes have Solar - Roymorgan.com

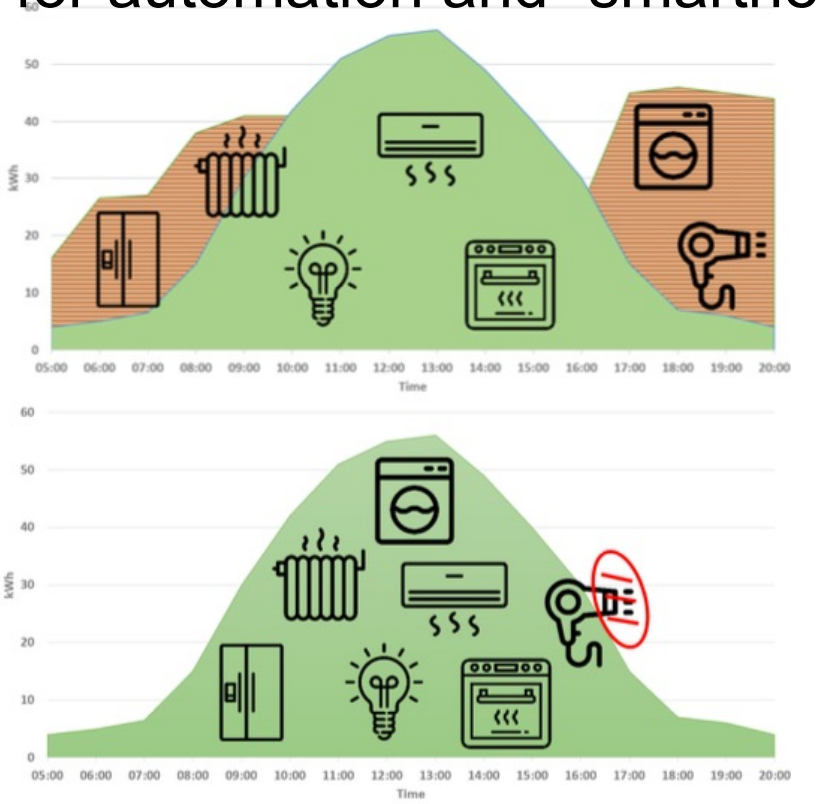
Self-consumption

The Self Consumption Problem

Alignment of Consumption with Production patterns.
Laborious task that calls for automation and “smartness”!

Bad Alignment ☹

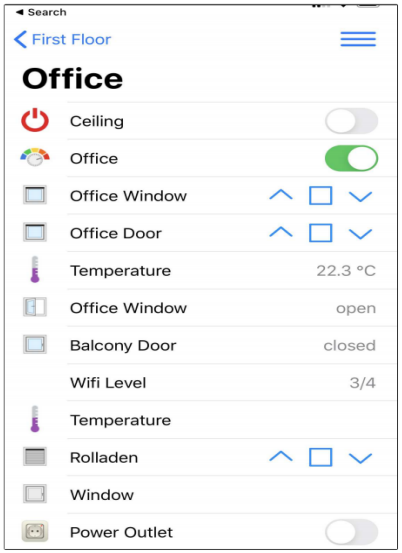
Good Alignment ☺

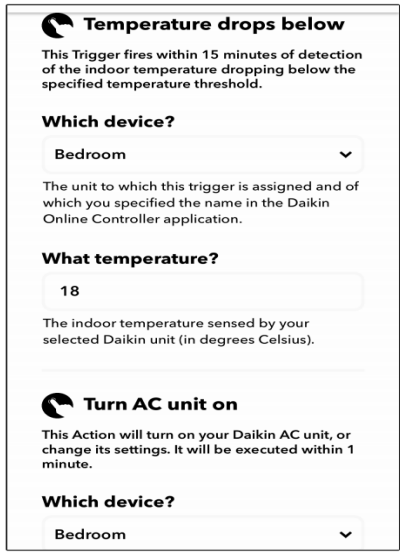


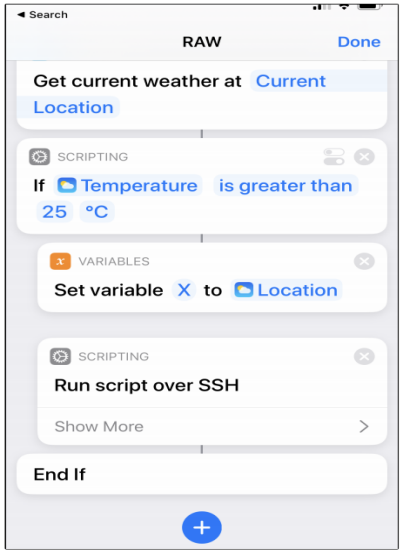
[J28] "Green Planning Systems for Self-Consumption of Renewable Energy", Soteris Constantinou, Andreas Konstantinidis and Demetrios Zeinalipour-Yazti, **IEEE Internet Computing (IC'22),** IEEE Computer Society, pp. 7 pages, Los Alamitos, CA, USA, DOI: 10.1109/MIC.2022.3164581, **2022.**

Rule Automation Workflows

- Span from simple **predicate statements** to **procedural workflows** capturing a **smart actuation pipeline**
  - Not optimized for **self-consumption :-(**
  - Not optimized for a **long-term objective** (e.g., 30% energy reduction - Paris Agreement)



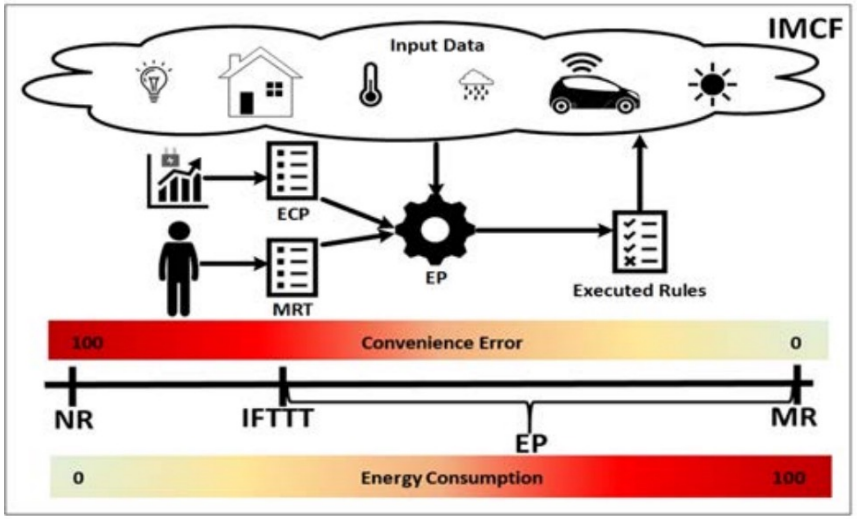




The EP and GP Algorithms

Energy Planner & Green Planner

- Data-driven algorithms for solar self-consumption having a **long-term energy budget** (e.g., 8000 kWh / year)

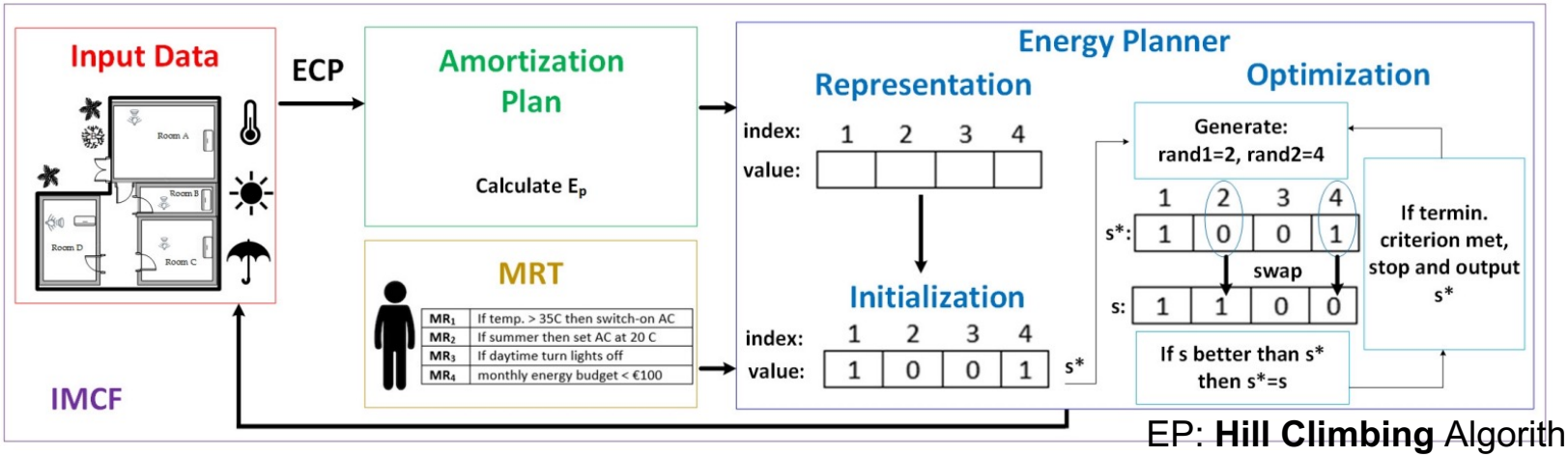


[C85] "The IoT Meta-Control Firewall", Soteris Constantinou, Andreas Konstantinidis, Demetrios Zeinalipour-Yazti, Panos K. Chrysanthis **"37th IEEE International Conference on Data Engineering" (ICDE '21),** IEEE Computer Society, ISBN:, Pages: 2523-2534, April 19 - April 22, 2021, Chania, Crete, **2021.**

[C84] "IMCF: The IoT Meta-Control Firewall for Smart Buildings", Soteris Constantinou, Antonis Vasileiou, Andreas Konstantinidis, Panos K. Chrysanthis, Demetrios Zeinalipour-Yazti **"24th International Conference on Extending Database Technology"** (EDBT '21), OpenProceedings.org, Pages: 658--661, March 23 - March 26, 2021, Nicosia, Cyprus, **2021.**

[J29] "Green Planning of IoT Home Automation Workflows in Smart Buildings", Soteris Constantinou, Andreas Konstantinides, Panos K. Chrysanthis and Demetrios Zeinalipour-Yazti, ***ACM Transactions on Internet of Things*** (**TIOT'22**), ACM, Vol. 3, Iss. 4, pp. 1--30, New York, NY, USA, DOI: 10.1145/3549549, **2022**.

The IMCF framework



EP: **Hill Climbing** Algorithm with restarts until solution is found.

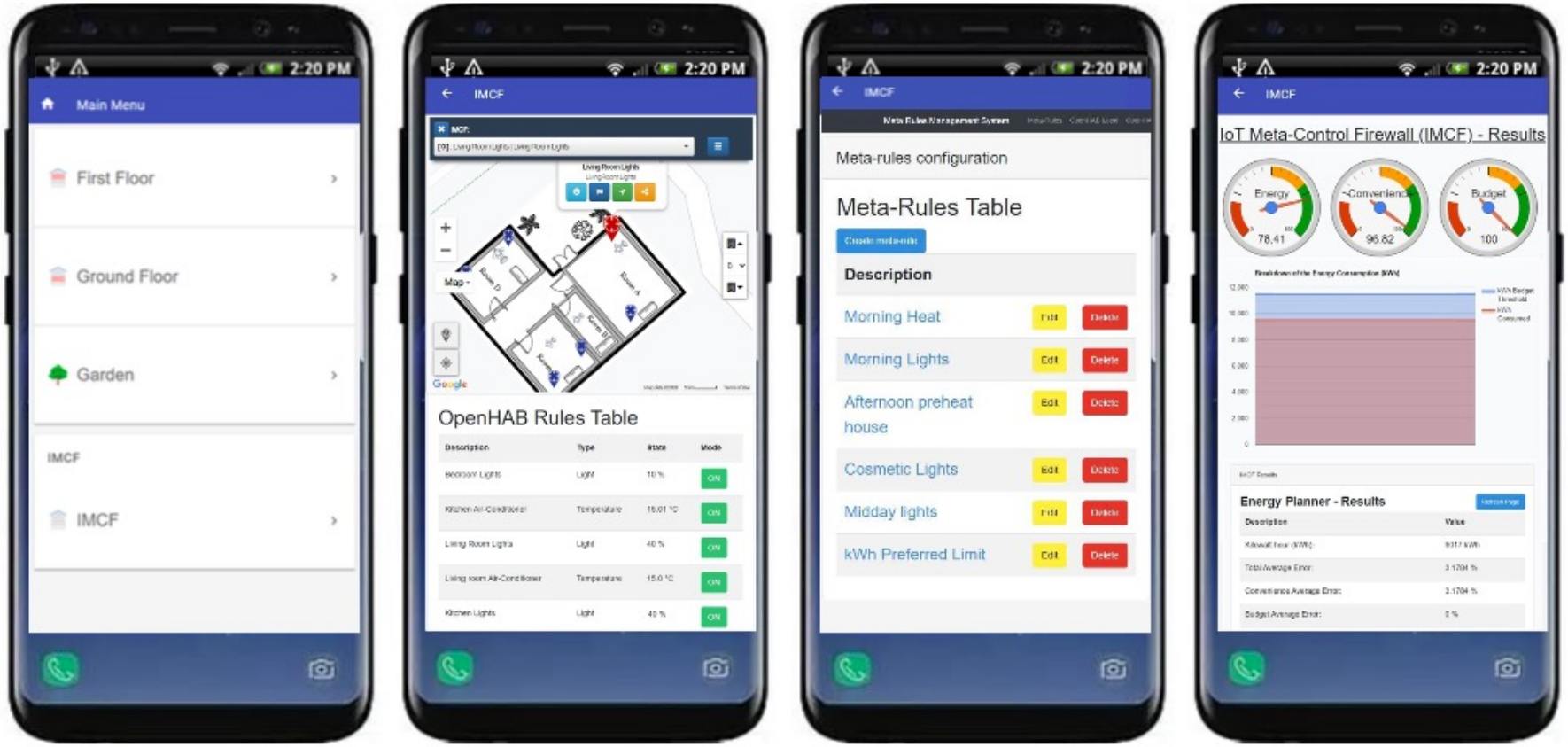
The IMCF algorithm is composed of two subroutines:

(i) the **Amortization Plan (AP):**
- Calculates the maximum energy budget constraint through a preselected amortization formula.

(ii) the **Energy Plan (EP):**
- Executed every few seconds over a time period for generating an energy plan solution for optimizing Convenience Error and satisfying Energy Consumption.

A hill-climbing algorithm has been adopted
- Doesn't require a learning history
- Doesn't require a target function
- Straightforward to be implemented in a resource-constraint setting

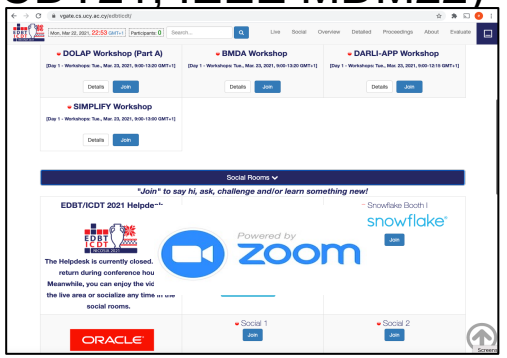
The IMCF Platform



https://imcf.cs.ucy.ac.cy/

Virtual Conferences + COVID Actions

- **E-Conferences:** We implemented a Zoom + Google Sheet Conference Management Platform. https://vgate.cs.ucy.ac.cy/
  - Used it as General Chair in two conferences (EDBTICDT21, IEEE MDM22)
  - Now used in more than five events!



- **COVID Contact "Tracing" and "Avoidance"**

- **[C83]** "COVID-19 Mobile Contact Tracing Apps (MCTA): A Digital Vaccine or a Privacy Demolition?", Demetrios Zeinalipour-Yazti and Christophe Claramunt, **Proceedings of the 21st IEEE International Conference on Mobile Data Management** (MDM'20), IEEE Computer Society, ISBN:, pp. 1--4, June 30 - July 3, 2020, Versailles, France, DOI: 10.1109/MDM48529.2020.00020, **2020**.
  - Panel Discussion
- **[C88]** "A Context, Location and Preference-Aware System for Safe Pedestrian Mobility", Constantinos Costa, Brian Nixon, Sayantani Bhattacharjee, Benjamin Graybill, Demetrios Zeinalipour-Yazti, Walter Schneider, Panos K. Chrysanthis, **The 22nd IEEE International Conference on Mobile Data Management** (MDM'21), IEEE Press, pp. 217-224, June 15 - June 18, 2021, Toronto, Canada, DOI: 10.1109/MDM52706.2021.00042, **2021**. **[Best App Paper Award!]**
  - BLE System for **"Contact Avoidance"** (not CT) at Upitt / Medical School
- **[J27]** "ASTRO: Reducing COVID-19 Exposure through Contact Prediction and Avoidance", Chrysovalantis Anastasiou, Constantinos Costa, Panos K. Chrysanthis, Cyrus Shahabi, and Demetrios Zeinalipour-Yazti, ***ACM Transactions on Spatial Algorithms and Systems*** (TSAS'22), Association for Computing Machinery, Vol. 8, Iss. 2, New York, NY, USA, DOI: 10.1145/3490492, **2022**.
  - Predict Congestion in Corridors in CM-Structure and then avoid congestion using an AI search algorithm

Acknowledgements

Collaborators, Students & Funding **Thank You!**

- Panos K. Chrysanthis (28)
- **Andreas Konstantinidis (27)**
- Dimitrios Gunopulos (22)
- **Constantinos Costa (21)**
- **Georgios Chatzimilioudis (21)**
- **Christos Laoudias (20)**
- **Panayiotis Andreou (20)**
- Vana Kalogeraki (16)
- George Samaras (15)
- Marios D. Dikaiakos (10)
- **Georgios Larkou (10)**
- Christos Panayiotou (7)
- Mohamed F. Mokbel (7)
- **Paschalis Mpeis (6)**
- **Soteris Constantinou (5)**
- Song Lin (5)
- Walid A. Najjar (5)
- Wang-Chien Lee (5)
- **Artyom Nikitin (4)**
- Xiaoyu Ge (4)
- Konstantinos Pelechrinis (4)
- Michail Vlachos (4)
- Chryssis Georgiou (3)
- **Andreas Pamboris (3)**
- Zografoula Vagena (3)
- Maria I. Andreou (3)
- Yannis Theodoridis (3)
- **Silouanos Nicolaou (3)**
- Panagiotis Karras (3)
- **Andreas Charalampous (3)**
- **George Constantinou (3)**
- **Marios Mintzis (3)**
- Cyrus Shahabi (2)
- Jaime Bleye Vicario (2)
- Jannik Strötgen (2)
- **Athina Hadjichristodoulou (2)**
- Dhruv Gupta (2)
- K. Harald Gjermundrød (2)
- Divesh Srivastava (2)
- Klaus Berberich (2)
- **Kyriakos Georgiou (2)**
- **Kyriakos Neocleous (2)**
- **Lambros Petrou (2)**
- **Erodotos Demetriou (2)**
- Vassilis J. Tsotras (2)
- Evaggelia Pitoura (2)
- Theodoros C. Kyprianou (2)
- **Stavroulla Koumou (2)**
- Anirban Banerjee (2)
- **Marios Constantinides (2)**
- Christian S. Jensen (2)
- **Panagiotis Drakatos (2)**
- Christophe Claramunt (2)
- Nick Koudas (2)
- Moustafa Youssef (2)
- George Panayi (2)
- Nikos Pelekis (2)
- **Chrysovalantis Anastasiou (2)**

Algorithms and Systems for the IoT Data Revolution

Demetris Zeinalipour

Thanks! Questions?

Associate Professor
Data Management Systems Laboratory (DMSL)
Department of Computer Science
University of Cyprus, Cyprus

https://www.cs.ucy.ac.cy/~dzeina/
dzeina@ucy.ac.cy

University of Cyprus, Nicosia, Cyprus, February 16th, 2023

DMSL @UCY

University of Cyprus