MiLA4U: User Driven Adaptation Approach for Microservice-based IoT Applications

Based on joint work with Martina De Sanctis
And

Joint collaboration with Dr. Martina De Sanctis, GSSI, Italy and Prof. Henry Muccini, Univaq, Italy

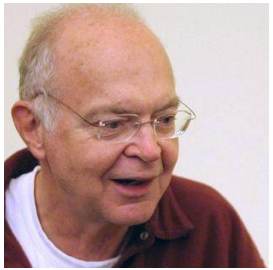
Karthik Vaidhyanathan
Assistant Professor, SERC, IIIT Hyderabad

About Me

Assistant Professor, SERC
IIIT Hyderabad, India
https://karthikvaidhyanathan.com

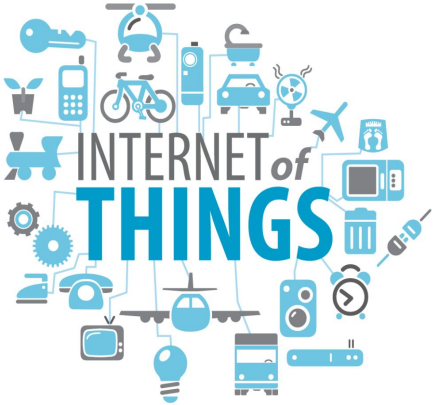
- B.Tech CSE, Amrita University, India
- Dual Master degree (M.Tech, Amrita University and MSc. University of L'Aquila, Italy)
- Industry: Product Lead, Knowledge Lens, India consultant ML architect at Founding Minds, India
- Ph.D in Computer Science, GSSI, Italy
- Google cloud credits for research, AWS Cloud Student Ambassador
- Postdoc, The VASARI Project, Univaq
- Reviewing activities (IEEE TSC, JSS, IST, CAIN@ICSE 2022, SE4RAI@ICSE 2022,.....)
- 2021, 2022 Co-Chair, SAML Workshop@ECSA

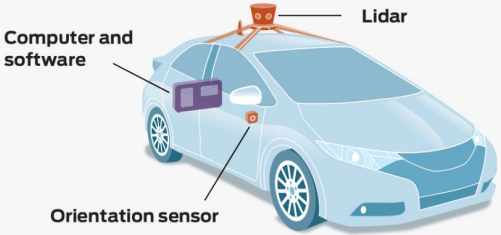
“In fact what I would like to see is thousands of computer scientists let loose to do whatever they want. That's what really advances the field”



Donald Knuth
Computer Scientist,
Turing Award winner

Intelligent and Complex Software Systems





TECH \ AMAZON \

Amazon Web Services says overwhelmed network devices triggered outage

Amazon says it plans to improve its response to outages

By Emma Roth | Dec 11, 2021, 3:34pm EST

Meta services are back ONLINE! Facebook, Instagram and WhatsApp were down for more than one hour worldwide that impacted thousands of users

- **Facebook, Instagram and WhatsApp crashed around 11:30am ET on Friday**
- **The outage hit users worldwide who cited issues with websites and apps**
- **However, users also said they are unable to post on Facebook and Instagram**

By STACY LIBERATORE FOR DAILYMAIL.COM
PUBLISHED: 17:34 GMT, 19 November 2021 | **UPDATED:** 18:28 GMT, 19 November 2021

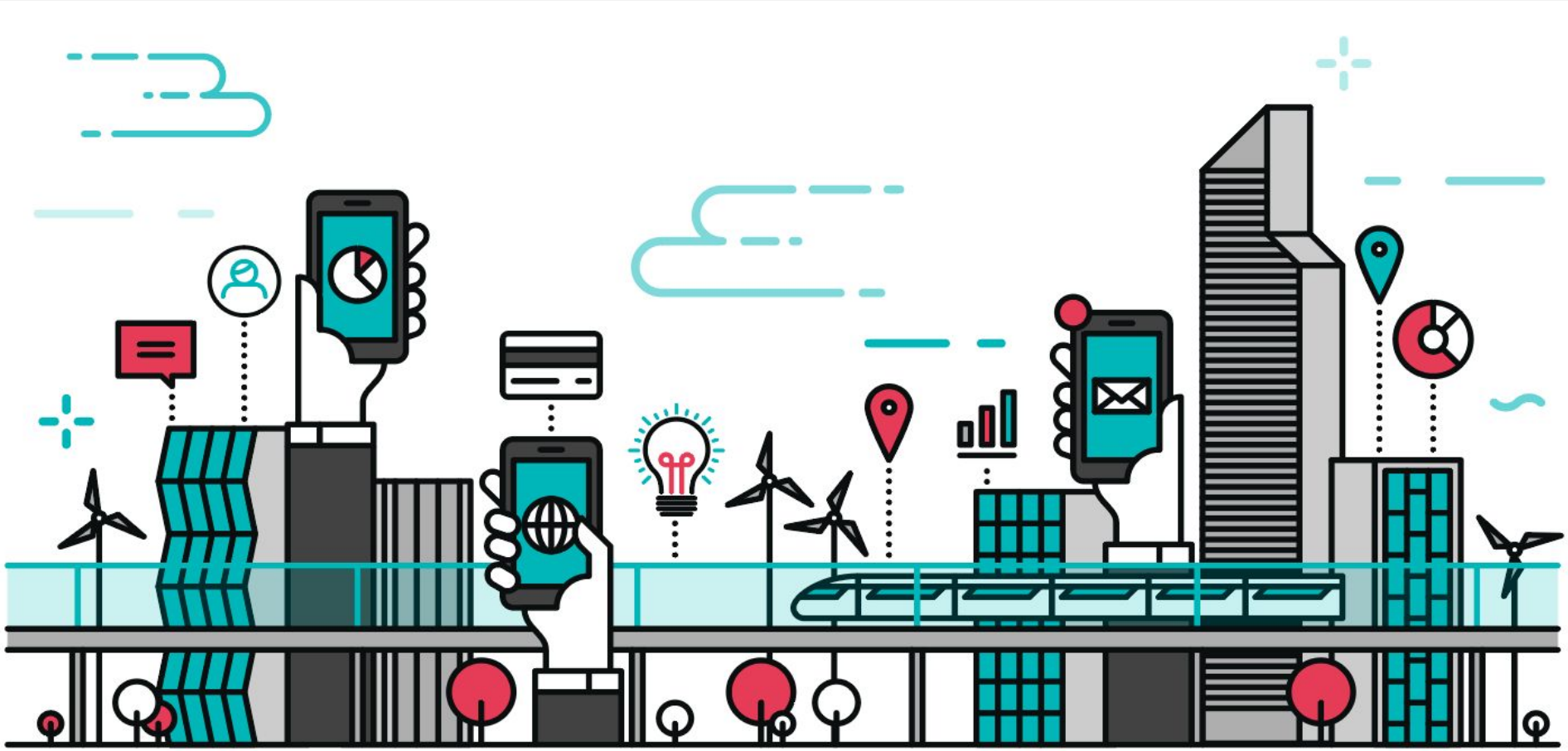
Vox recode

Tesla needs to fix its deadly Autopilot problem

Tesla is facing heat from federal officials following an investigation into a fatal crash involving its Autopilot.

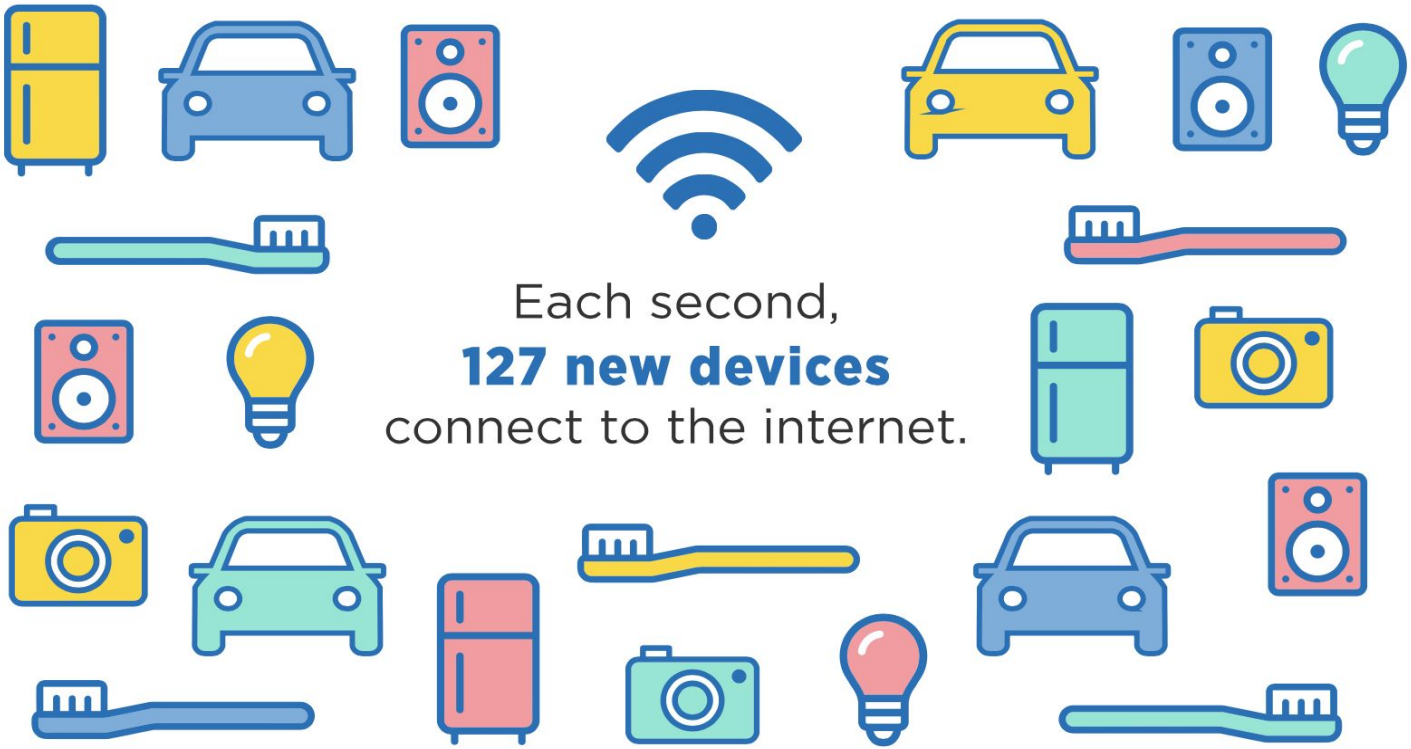
By Rebecca Heilweil | Feb 26, 2020, 1:50pm EST

How to architect these systems to guarantee better QoS (performance, reliability, security, etc.)?



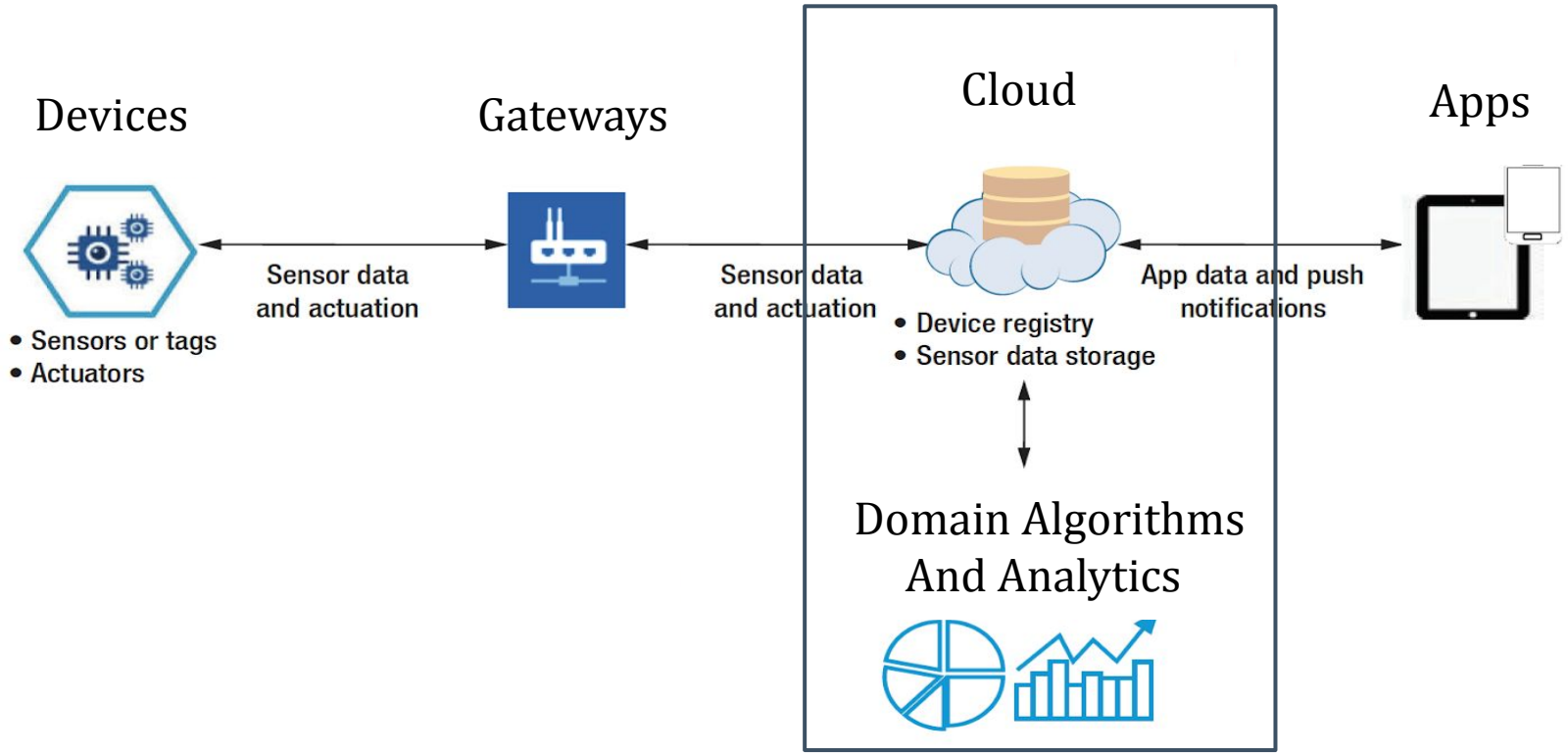
Source: https://ec.europa.eu/digital-single-market/en/policies/next-generation-internet.

The world of Internet of Things



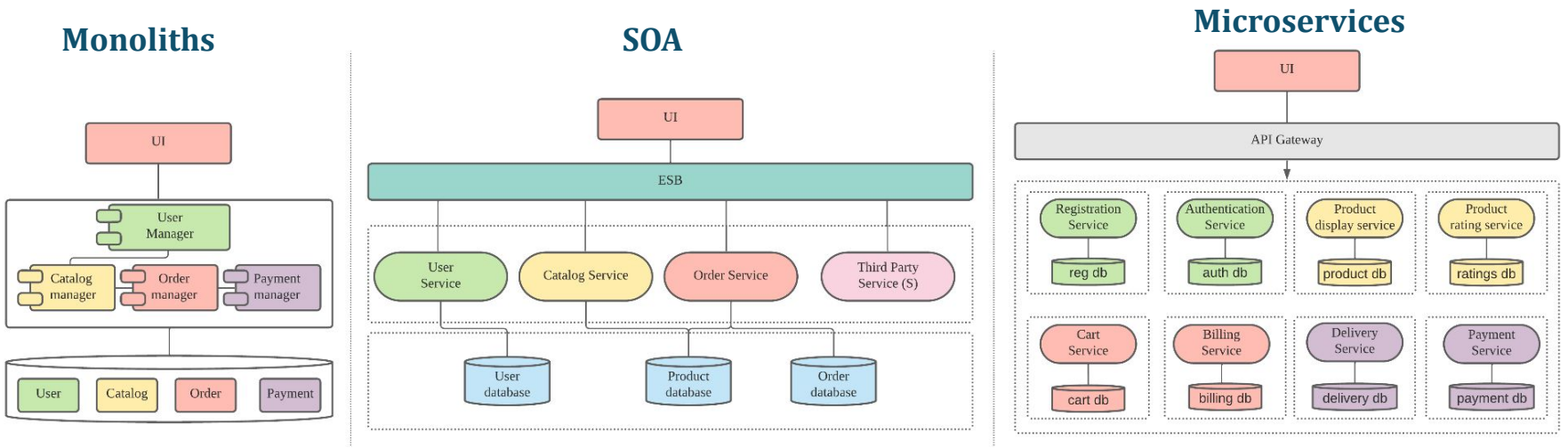
Source: https://www.cpacanada.ca/en/news/world/2019-02-13-internet-of-things-infographic

IoT Architecture



A. Taivalsaari and T. Mikkonen, "**A roadmap to the programmable world: software challenges in the iot era,**" IEEE Software, vol. 34, no. 1, pp. 72–80, 2017.

SA: Over the years



1980- 2000 (monoliths and distributed)

2000-2010 (Internet connected)

2010-2020 (internet native)

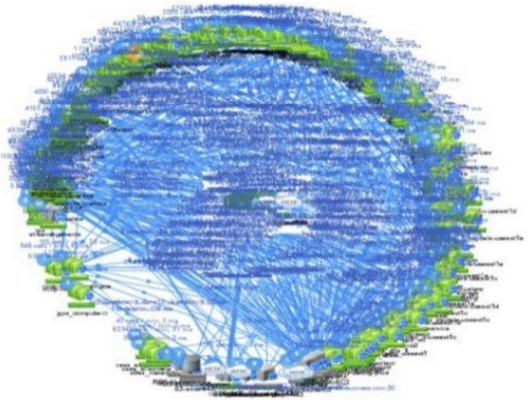
now

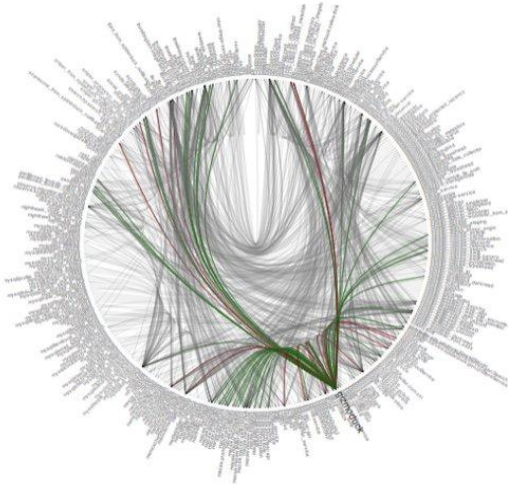
Serverless has come up, age of intelligent connected systems!!

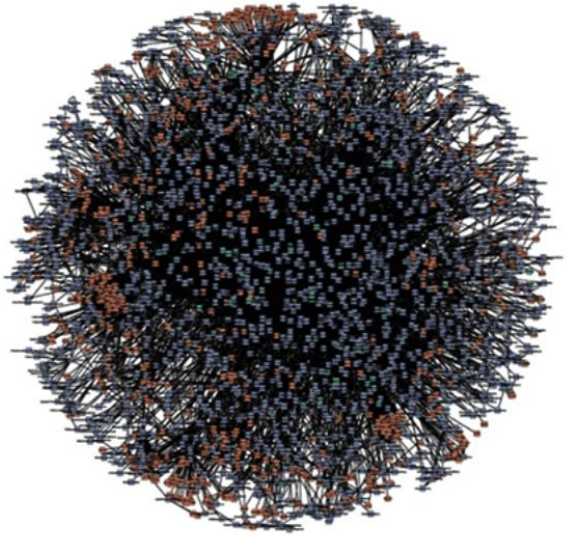
Woods, E., 2016. **Software architecture in a changing world**. *IEEE Software*, *33*(6), pp.94-97.

The world of Microservices

"It is an approach to developing a single application as a suite of small services, each running in its own process and communicating with lightweight mechanisms, often an HTTP resource API"
- Martin Fowler







NETFLIX

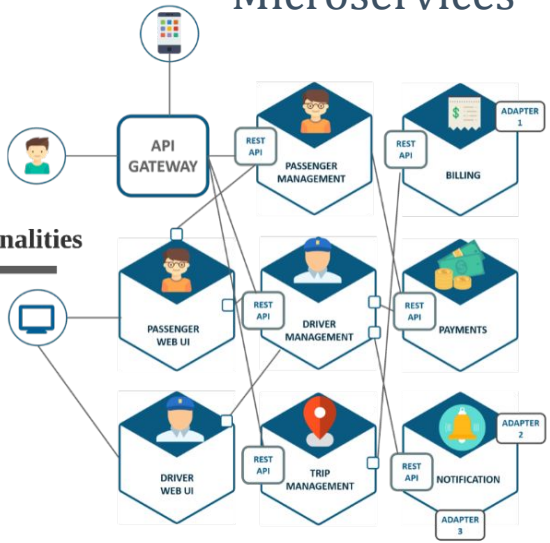
amazon.com

source: tinyurl.com/3kswbtak (Twitter, Google images)

Research Context

Users

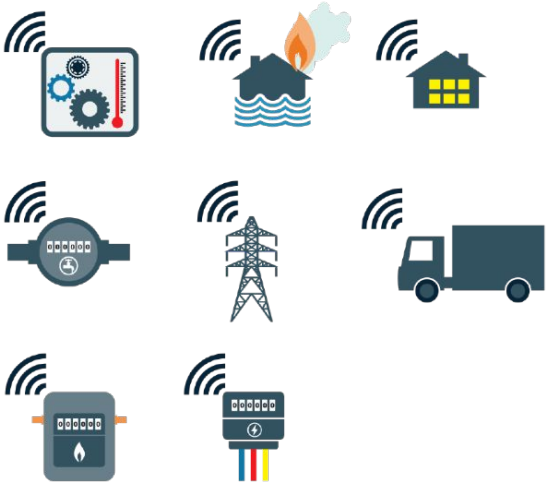
Microservices



Functionalities

Data

IoT Devices



- Uses of web/mobile apps for visualization
- Requires flexibility and better experience

- Implements the functionalities
- Subjected to resource constraints

- Captures and provides data
- Subjected to resource and environment constraints

Source: Google images

Microservices can be tricky!

Source: google images

Two Dimensional Open Challenges

Social challenges

- *changing user's preferences and needs:*
- the user typically adapts to how the software is designed, thus affecting the user's engagement.

Technical challenges

- arise *when microservice-based solutions are applied to IoT systems:*
- due to uncertainties faced by IoT devices and those of microservices themselves (e.g., battery level, VMs/containers resource constraints).

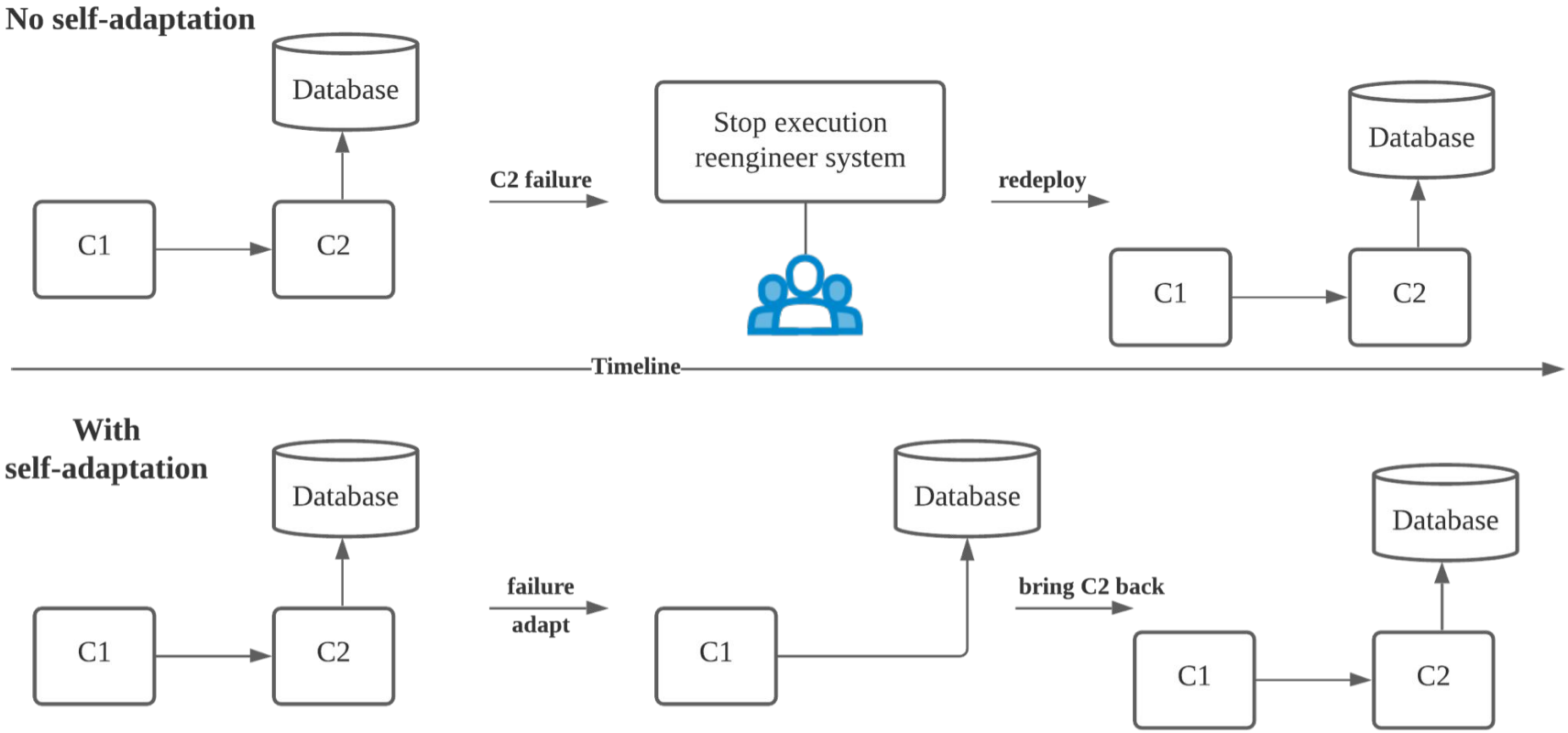
Self-adaptation to the rescue!!

Self-adaptive systems: An Intuition

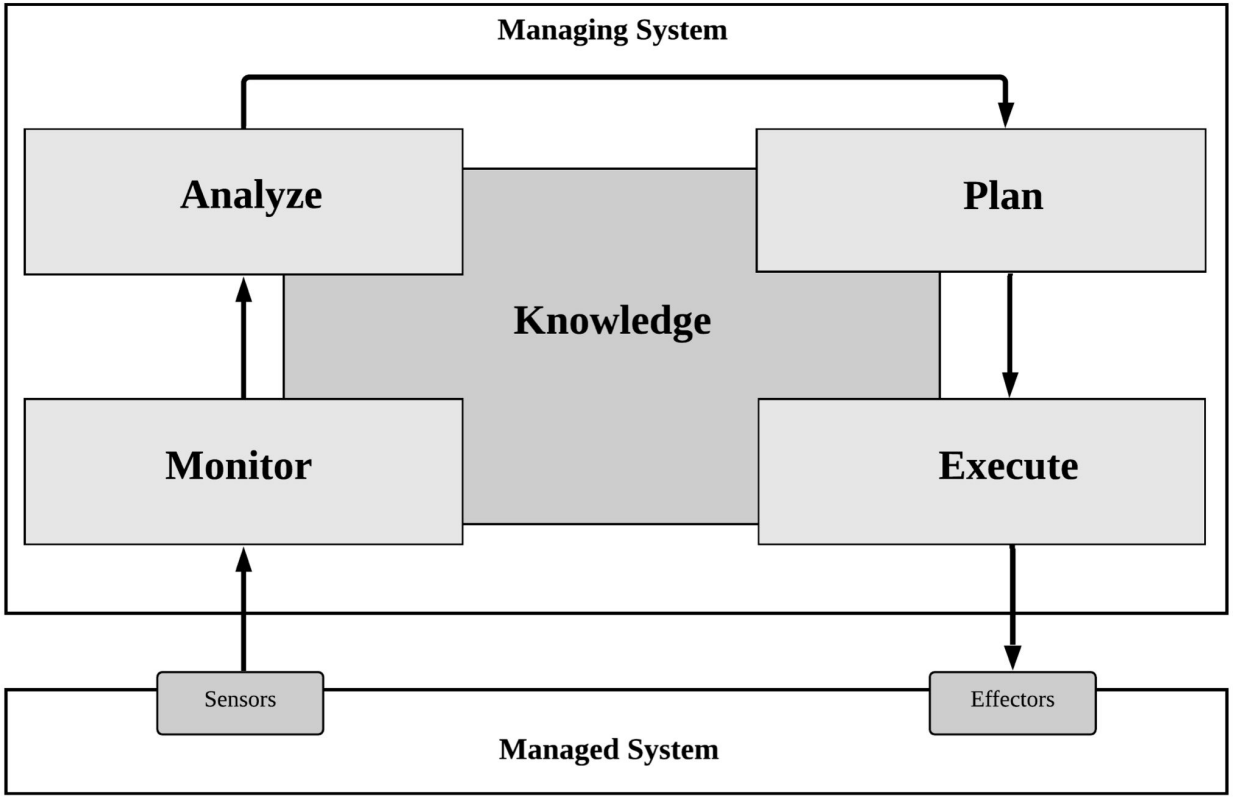
What if he had an umbrella?

What if softwares have the ability to autonomously adapt just like humans?

Self-adaptive systems: An Intuition

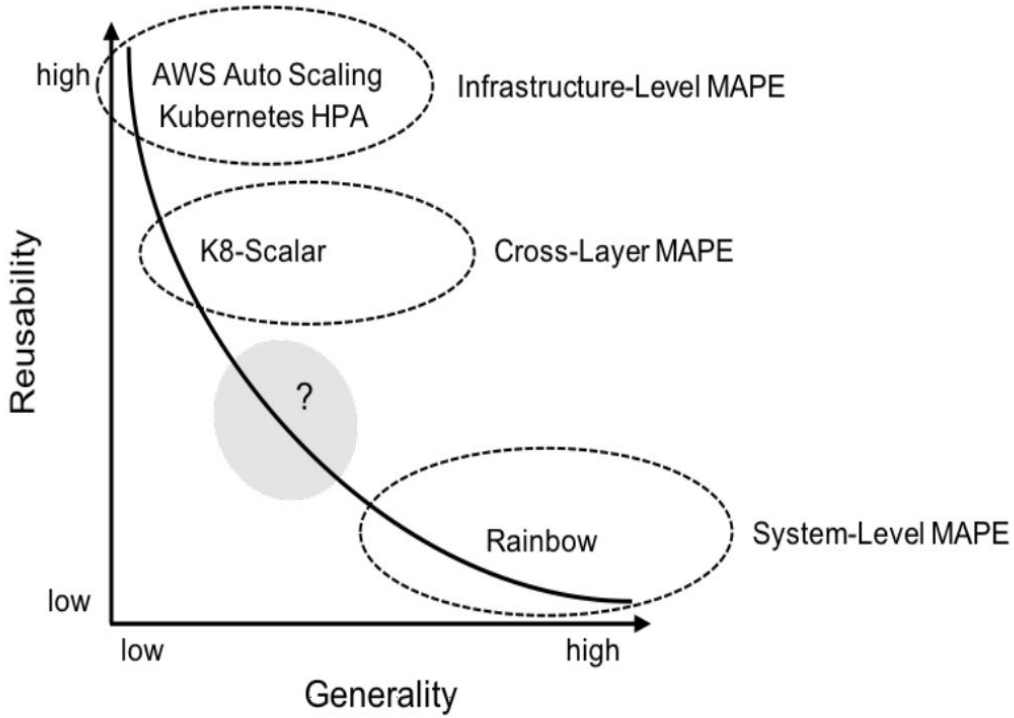


The MAPE-K Framework



Kephart, Jeffrey O., and David M. Chess. **The vision of autonomic computing,** *Computer* 36, no. 1 (2003): 41-50.

Issue with Traditional Approaches

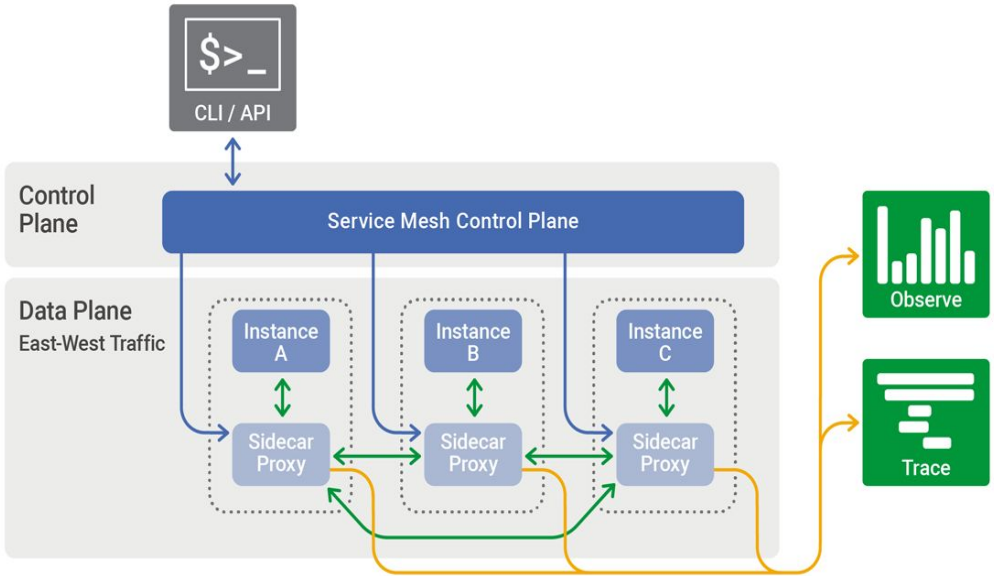


N. C. Mendonça, D. Garlan, B. R. Schmerl, and J. Camara, "**Generality vs. reusability in architecture-based self-adaptation: the case for self-adaptive microservices**," in Proceedings of the 12th European Conference on Software Architecture ECSA 2018

What can be done? - Service Mesh MAPE-K

Configurable low-latency infrastructure layer

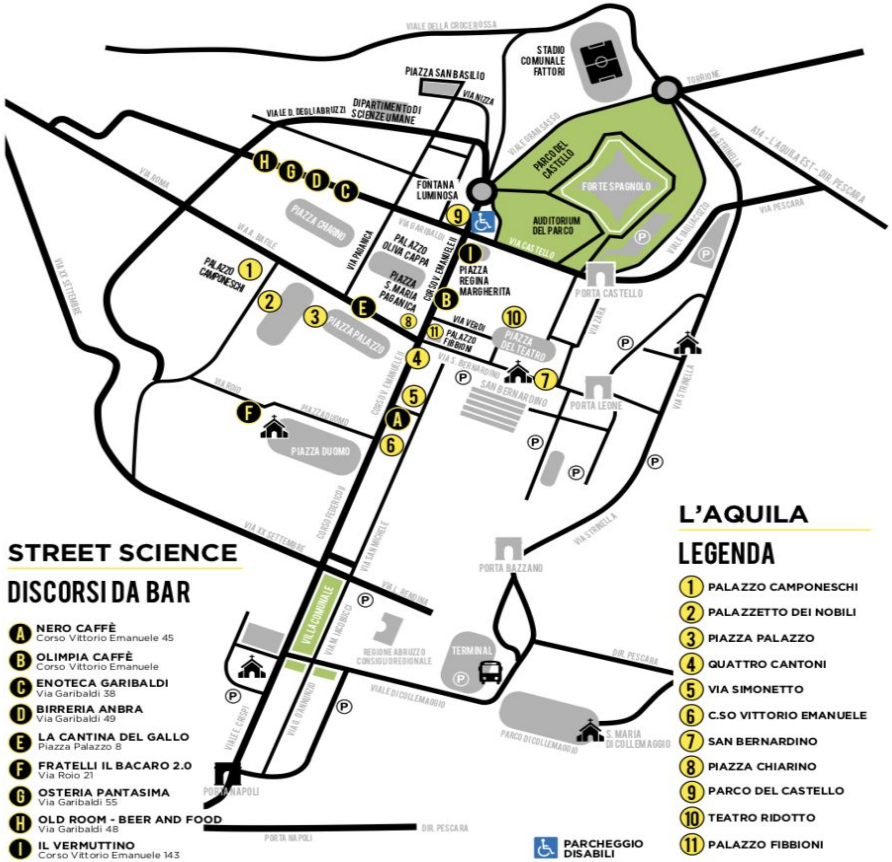
- Features of using service mesh:
  - Load balancing
  - Traffic routing
  - Security
  - ….
- Provides observability!!
- Use the metrics data for performing adaptations



https://www.nginx.com

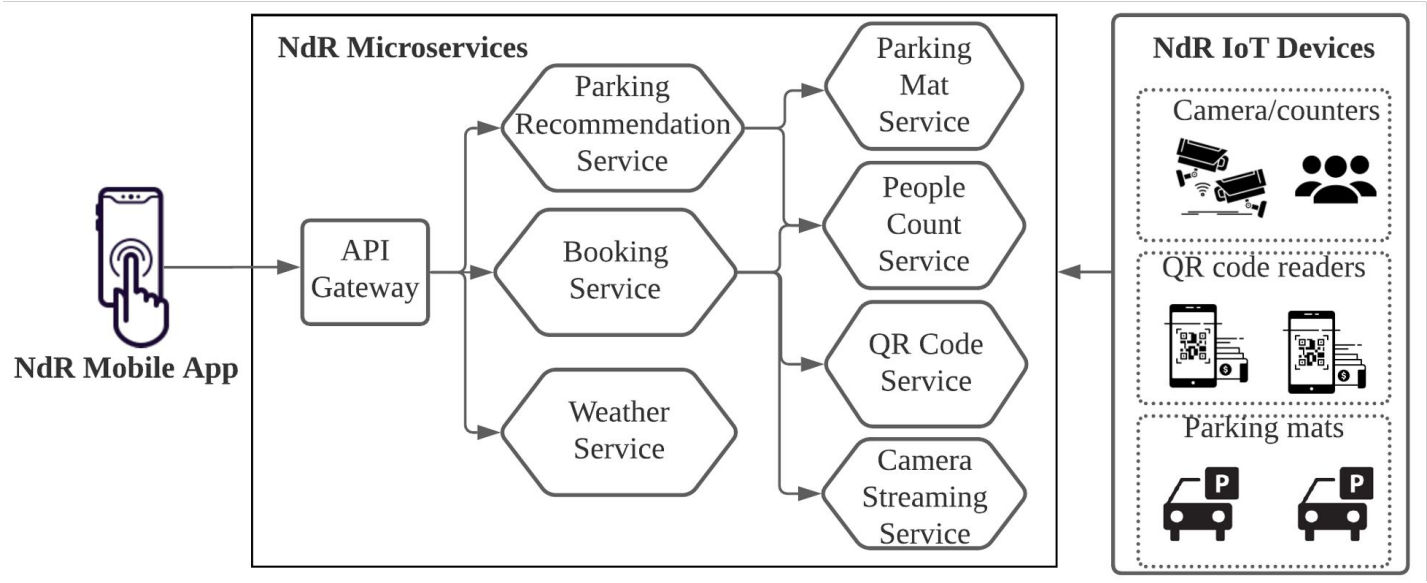
N. C. Mendonça, D. Garlan, B. R. Schmerl, and J. Camara, "**Generality vs. reusability in architecture-based self-adaptation: the case for self-adaptive microservices**," in Proceedings of the 12th European Conference on Software Architecture ECSA 2018

NdR: A Case Study (European Researcher Night)



https://www.streetscience.it

NdR Case Study



Architecture of NdR Application

NdR Predefined application flow

Check availability -> book venues -> get parking recommendations -> check weather

Constraints in NdR

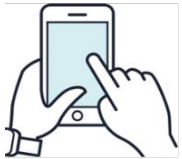
Adaptation concerns of multiple entities

- IoT devices - Resource constraints, failures, ...
- Microservices - Resource management, failures, ...
- Users - Changing needs/requirements, ...

Instances from NdR

- QR code reader has limited battery capacity
- Booking microservice gets sudden surge of requests
- Online booking of venue and preferred transportation

Chain of Multi-level Adaptation



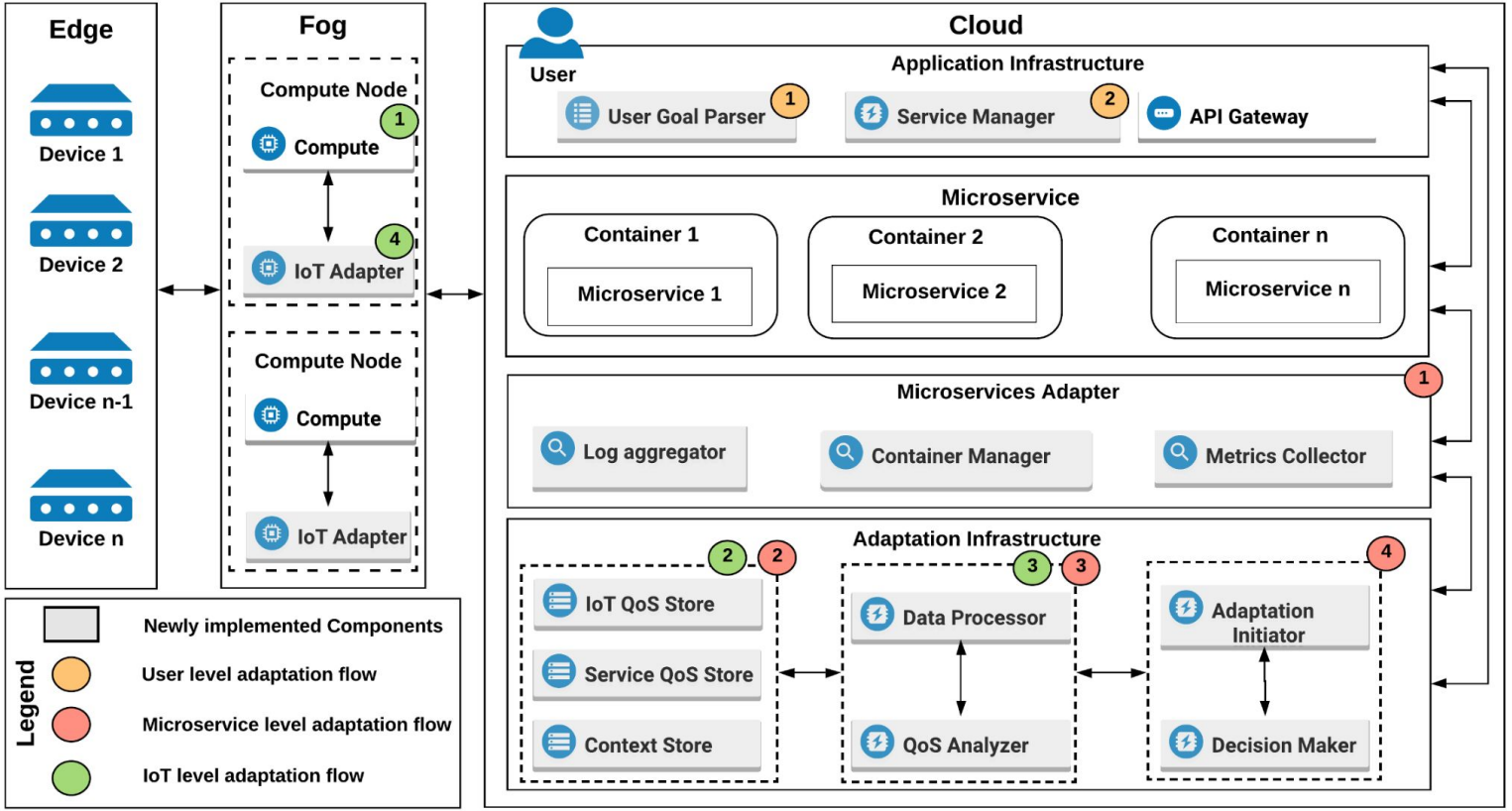
User goal:
«first check the weather conditions and then check for parking lots»

Application level: dynamically combine the weather and parking microservices to accomplish the user's goal.

Microservices level: combine the instances of weather and parking microservices that offer the least response time (to minimize the overall response time perceived by the user).

IoT devices level: reduce the data transfer frequency of IoT devices except parking mats to save more power.

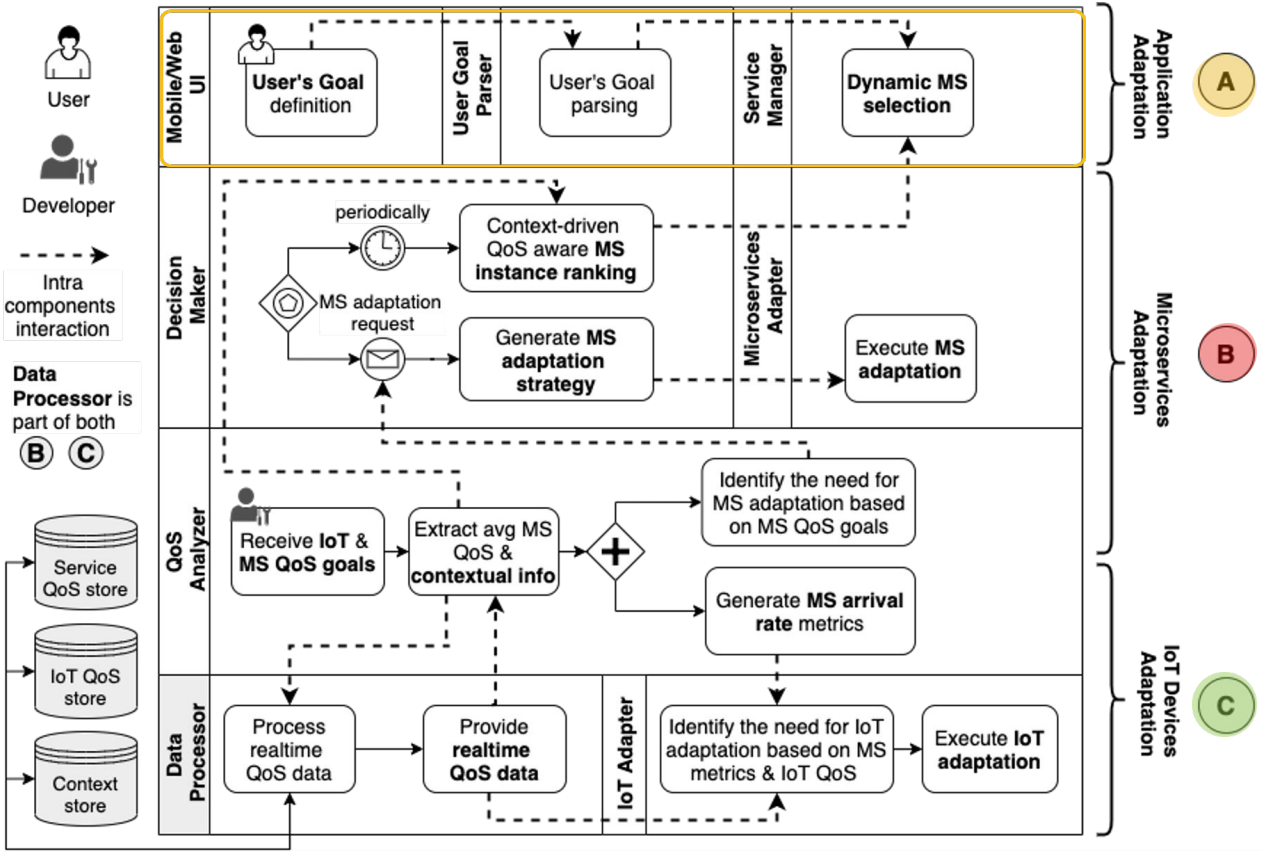
MiLA4U Architecture



M. De Sanctis, H. Muccini, K. Vaidhyanathan: **Data-drivenAdaptation in Microservice-based IoT Architectures.** IEEE ICSA 2020 - NEMI Track

M. De Sanctis, H. Muccini, K. Vaidhyanathan: **A User-driven Adaptation approach for Microservice-based IoT Applications**, IoT 2021

Overall Adaptation Process of MiLA4U



User Goal Model

Table 1: Goal model syntax.

$$
\begin{aligned}
G_{User} &::= F^{+} \\
F &::= f_{?[QoS]} \mid [F \textbf{ and } F]_{?[QoS]} \mid [F \textbf{ or } F]_{?[QoS]} \mid \textbf{one\_of } [f_1 \ldots f_n]_{?[QoS]} \mid \\
&\quad \textbf{seq } [f_1, \ldots, f_n]_{?[QoS]} \mid \top \mid \bot \\
QoS &::= \text{`rt:'} \; RT \\
RT &::= [THRESHOLD_{min} \; , \; THRESHOLD_{max}] \\
THRESHOLD &::= eRT_{min} \mid eRT_{avg} \mid eRT_{max} \\
eRT &::= NUMBER_{rt} \text{ sec} \mid NUMBER_{rt} \text{ ms} \\
NUMBER_{rt} &::= n, n \in \mathbb{N} \\
f &::= \text{`event\_booking'} \mid \text{`weather\_checking'} \mid \text{`ticket\_availability'} \mid \\
&\quad \text{`parking\_recommendation'}
\end{aligned}
$$

1. Allows system to transparently user's goal at run-time based on functionalities selected
2. Application workflow is automatically derived from the user goal

User Goal Model Examples

$\mathbf{G_{User}}$ ::= **one of** [‘weather checking'; ‘parking recommendation’]

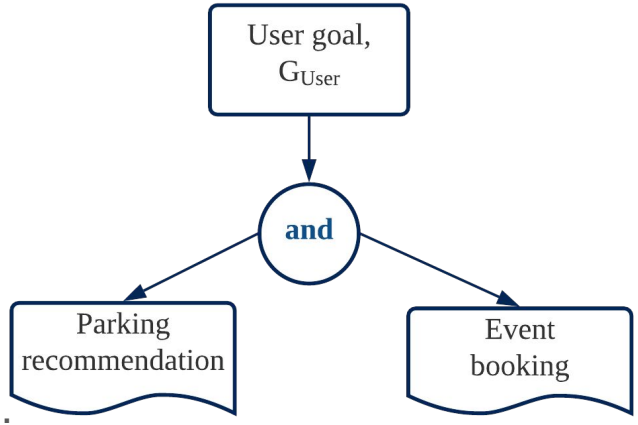
$\mathbf{G_{User}}$ ::= **seq**[‘parking recommendation’; ‘event booking’]

$\mathbf{G_{User}}$ ::= [‘ticket availability' **and** ‘event booking']$_{\text{rt:[0sec;5sec]}}$

Application Level Adaptation [Service Manager]

Input:

- *goal*, the parsed user goal
- $F_{map,}$ the mapping functionalities and microservices



Algorithm 1

1. Check the type of control-flow construct
2. Recursively refine the goal until leaf-level functionalities
3. Exploit $Rank_{map}$ to identify the optimal instance of microservice to be invoked

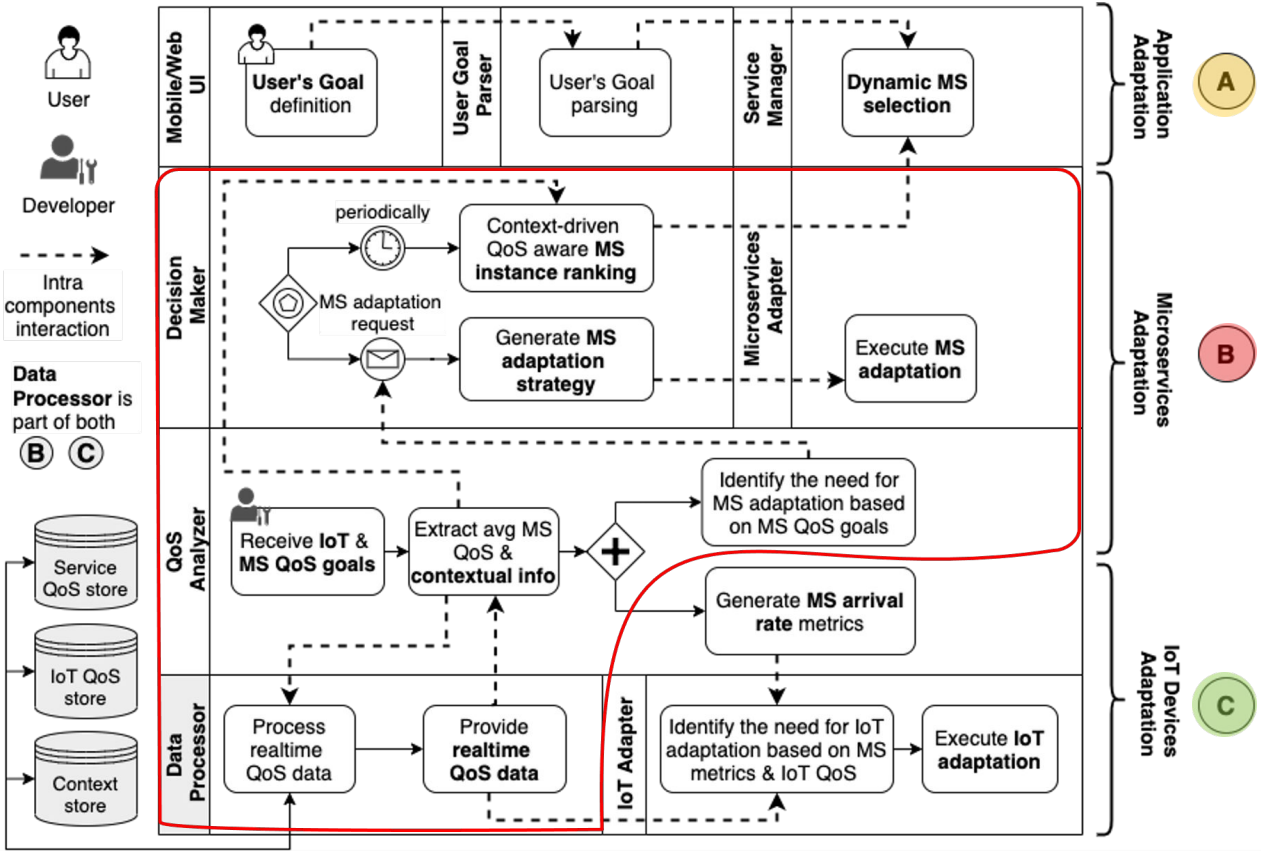
F_{map}

Functionalities	Microservices
Parking Recom	Parking
Event Booking	Event
….	….

$Rank_{map}$

Microservices	Instances
Parking	Instance_P1
	Instance_P2
Event	Instance_E1

Overall Adaptation Process of MiLA4U



Microservice Level Adaptation [Decision Maker]

Input:

- S, set of available microservices (ms)[parking]
- $QoS_{map,}$ average QoS of every microservice's instance
- C, the context, i.e., location zones [US, Europe]

MS Instance	Avg QoS
P1	2 sec
P2	1.5 sec
P3	3 sec

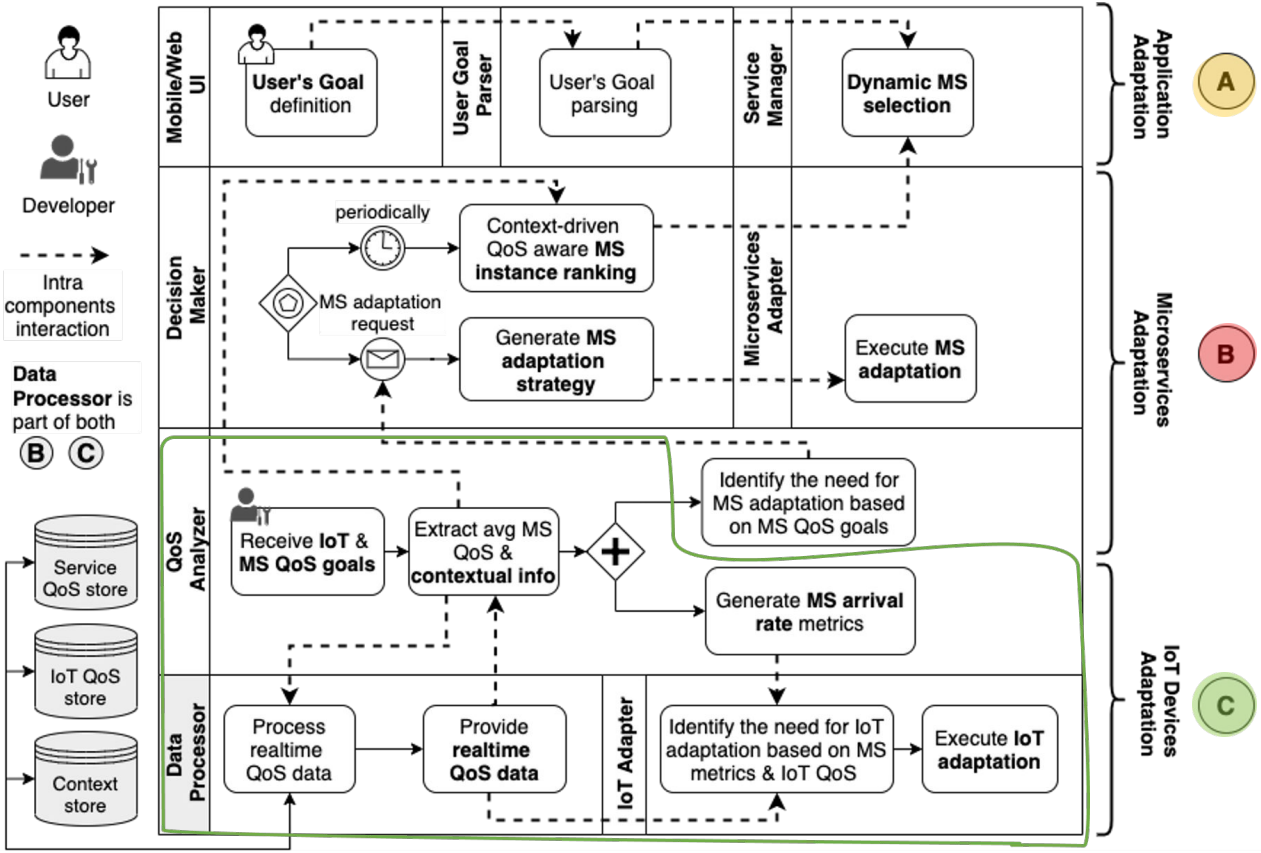
Algorithm 2

- Determine ms instances locations [p1, p3 in US, p2 in Europe]
- Identify avg QoS for each msinstance from QoS_{map}
- Sort ms instances per location base on QoS and generate $Rank_{map}$

$Rank_{map}$

Microservices,location	Instances
Parking, US	P1
	P3
Parking, Europe	P2

Overall Adaptation Process of MiLA4U



Device Level Adaptation [IoT Adapter]

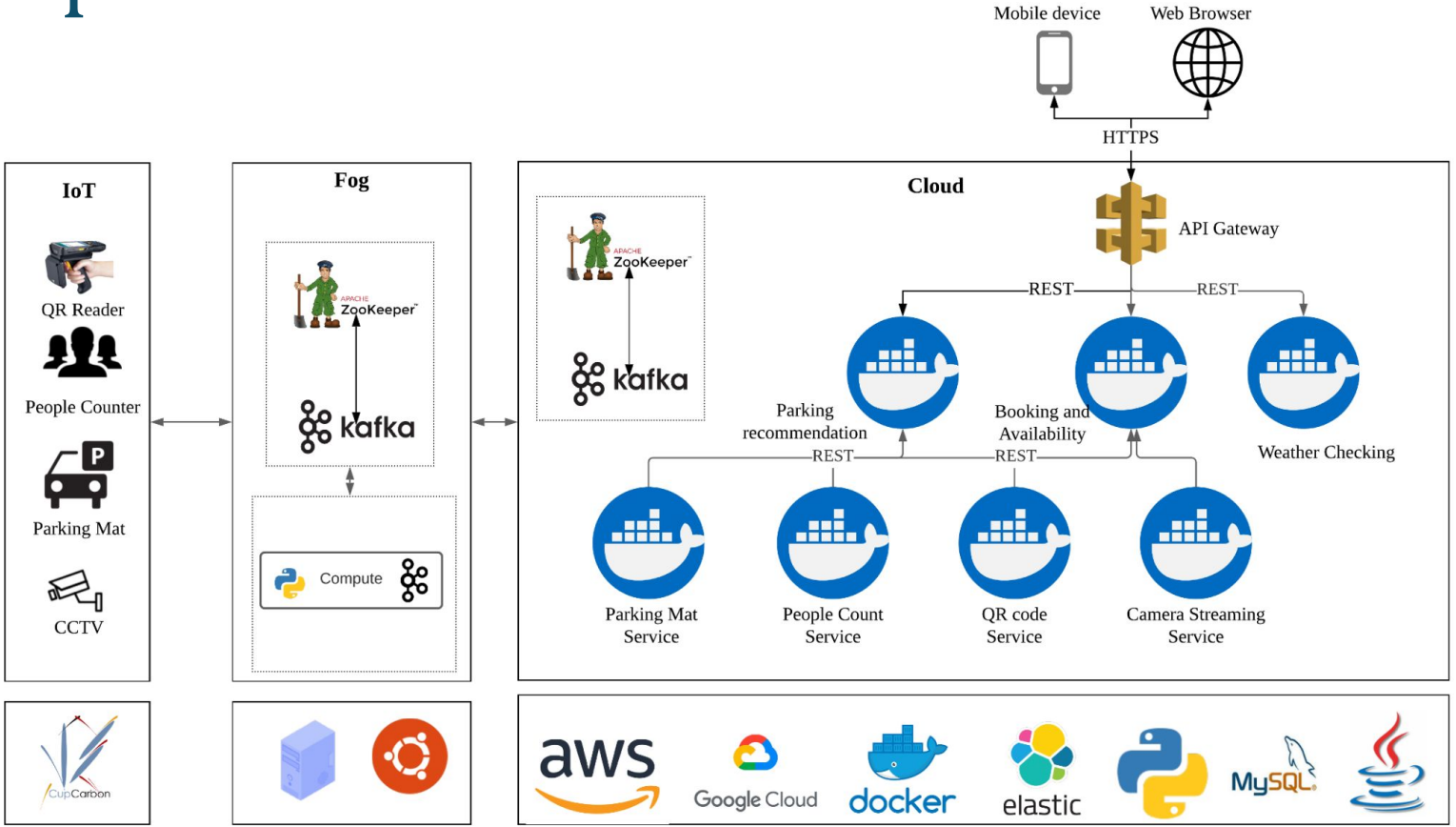
Input:

- Energy consumed(E_C), Arrival rate map (A_{Rmap}), Microservice sensor map (M_{map})
- Energy threshold (E_t), Arrival rate threshold (A_{Tmap}), Sensor reduction frequency (R_{Fmap})

Algorithm 3

- Check if $E_C > E_t$ [12.0 joules > 10.0 joules]
- Use A_{Rmap} check for each microservice, if arrival rate < T [50 req/sec < 60 req/sec]
- If yes, use: [event_booking]
  - M_{map} to identify corresponding sensors [event_booking: people counter]
  - R_{Fmap} to check data transfer frequency reduction
- Adjust sensors frequencies if they are not in critical mode [people counter]
- Once the arrival rate is above threshold, reset the sensors frequency

Implementation Architecture



Complete source: https://github.com/karthikv1392/MiLA4U

Experiment Setup and Candidates

Experimental Setup

- Deployed in Google cloud
- Two VM instances (different location, vCPU, RAM, etc)
  [different context dimensions]

Data Setup

- Each MS replicated to 4 - 28 instances
- Real time execution for 5 hours
- Goal file with 20K goals
- Web traffic based on benchmark logs of FIFA 98 worldcup site

Evaluation Candidates

Candidate	Description
SN	**Static workflow, No adaptation**
SA	**Static workflow, adaptation**
DN	**Goal model, no adaptation of MS & IoT**
DA	**MiLA4U**

Goal

- Optimize response time of MS
- Reduce energy consumed by sensors

Evaluation Metric

Given the threshold for response time, RT_{max} and Energy, E_{max}, the goal is to maximize the Utility function

$$U_\tau = x_u \cdot Q_\tau + x_e \cdot E_\tau, \text{ with}$$

$$Q_\tau = \sum_{i=1}^{n} q_i$$

$$E_\tau = \begin{cases} E_{max} - e_\tau & \text{if } e_\tau < E_{max} \\ (E_{max} - e_\tau) \cdot p_{ev} & \text{otherwise} \end{cases}$$

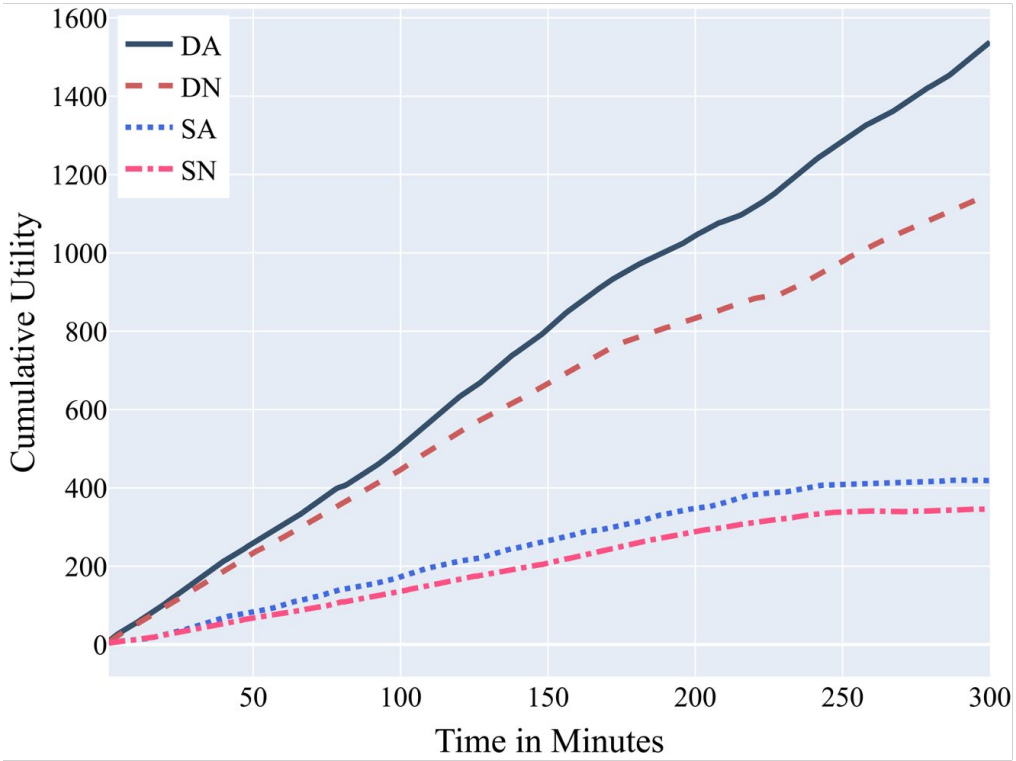
$$q_i = \begin{cases} eRT_{max} - rt(i) & \text{if } rt(i) < eRT_{max} \\ (eRT_{max} - rt(i)) \cdot p_{rt} & \text{otherwise} \end{cases}$$

Where:

- x_u, x_e are weights on user goal completion time and energy savings;
- Q_t the sum of user goals response time gain for *n* goals;
- E_t capture the energy savings
- p_{ev}, p_{rt} are penalties

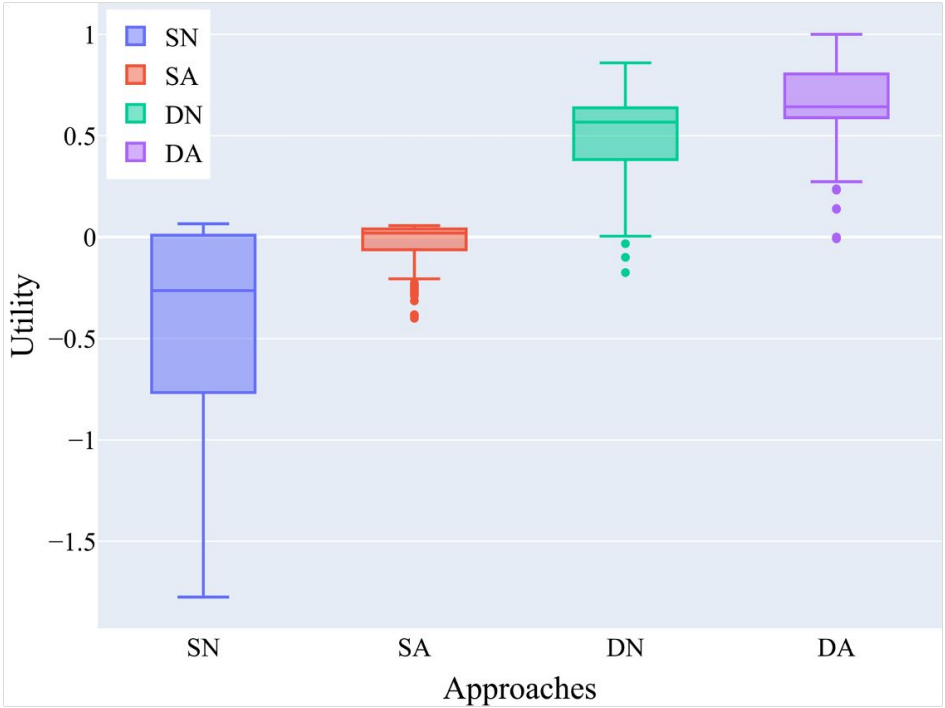
Research Questions

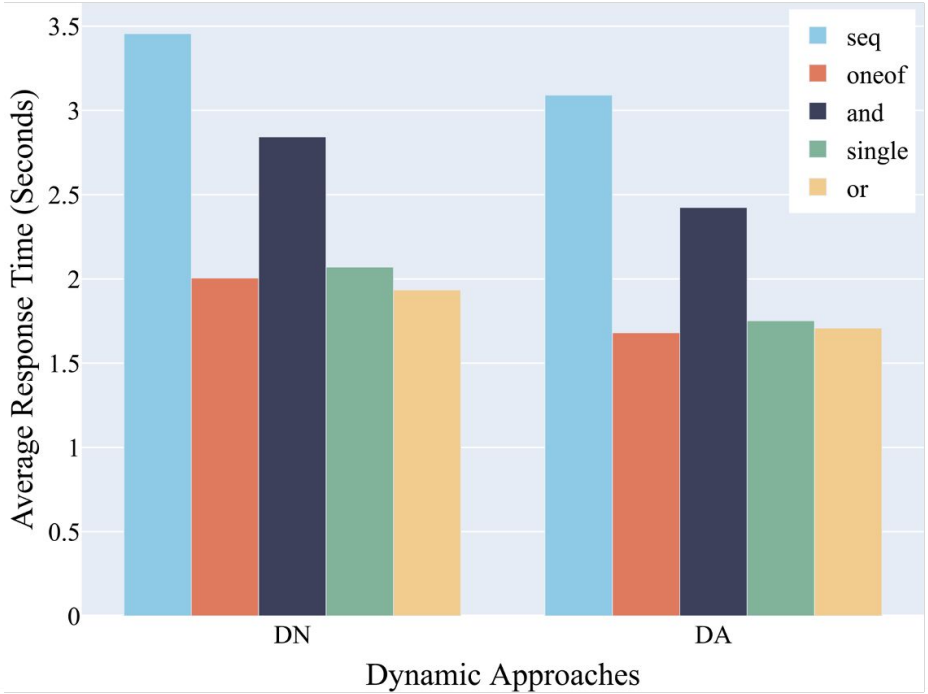
RQ 1: How does the **overall QoS** of microservice-based IoT system using MiLA4U compare to standard baselines?



Research Questions

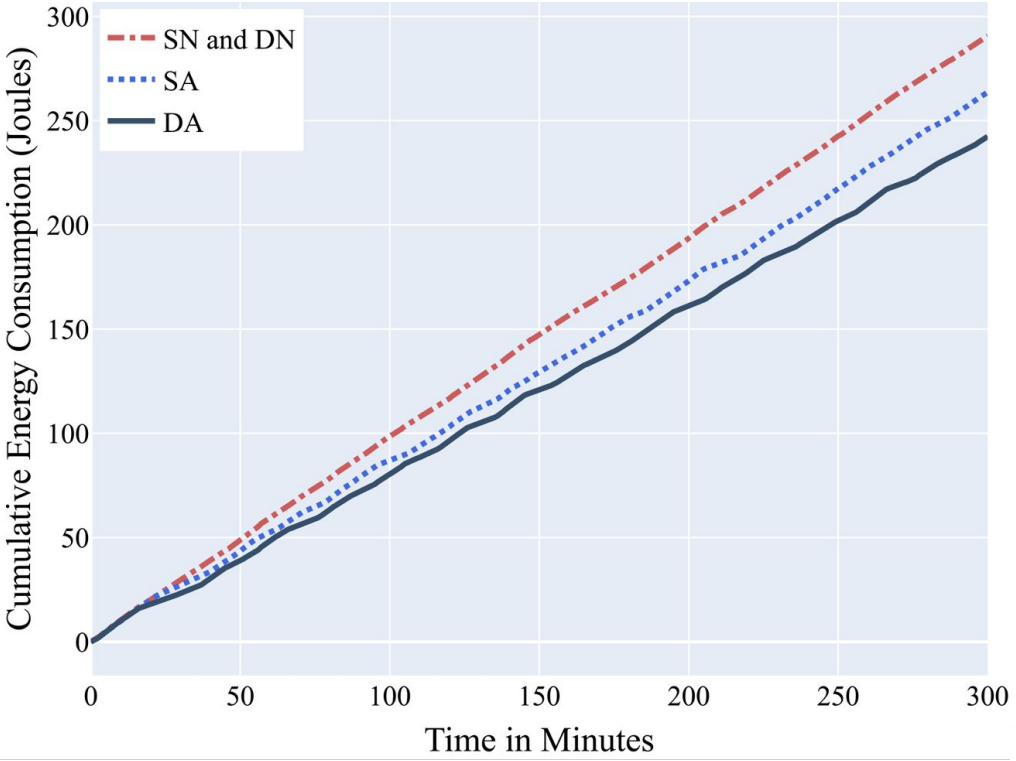
RQ 1.1: How does the **user goal satisfaction** using MiLA4U compare to the baselines and what is the impact of user goal type on the **response time**?





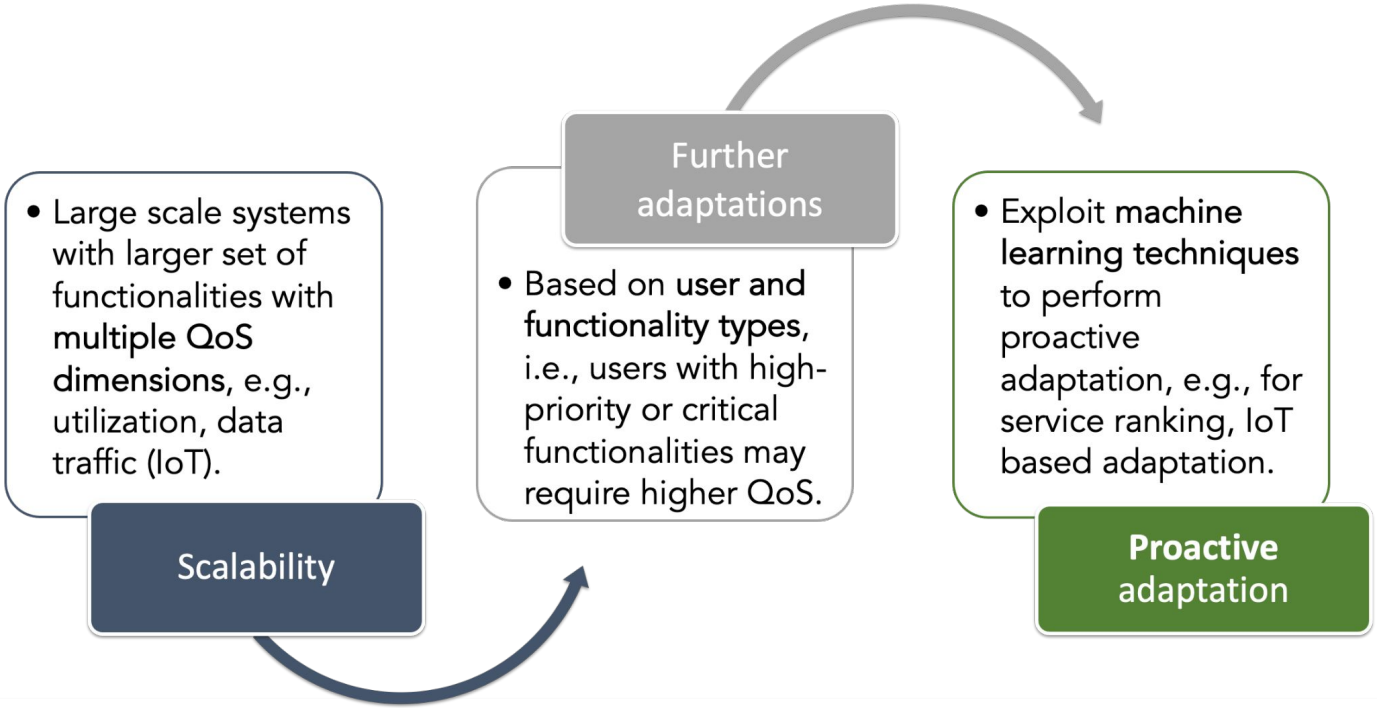
Research Questions

RQ 1.2: What is the impact of MiLA4U on the **energy consumed by IoT devices** compared to the baselines?



Conclusions and Future Work

- Adaptation centered around users can provide them with more flexibility
- Multi-level adaptation approach - Ensure optimal QoS at multiple levels



Thank you

MiLA4U Repo

Further queries:

- E-mail: karthik.vaidhyanathan@iiit.ac.in
- karthikv1392@gmail.com
- Web: https://karthikvaidhyanathan.com
- Twitter: @karthi_ishere

*"Software is **not limited by physics**, like buildings are. It is **limited by imagination**, by design, by organization. In short, it is limited by properties of people, not by properties of the world. We have **met the enemy, and he is us**"*

- Ralph Johnson, Author of Design patterns