

The main technology gap (the problem with existing system)

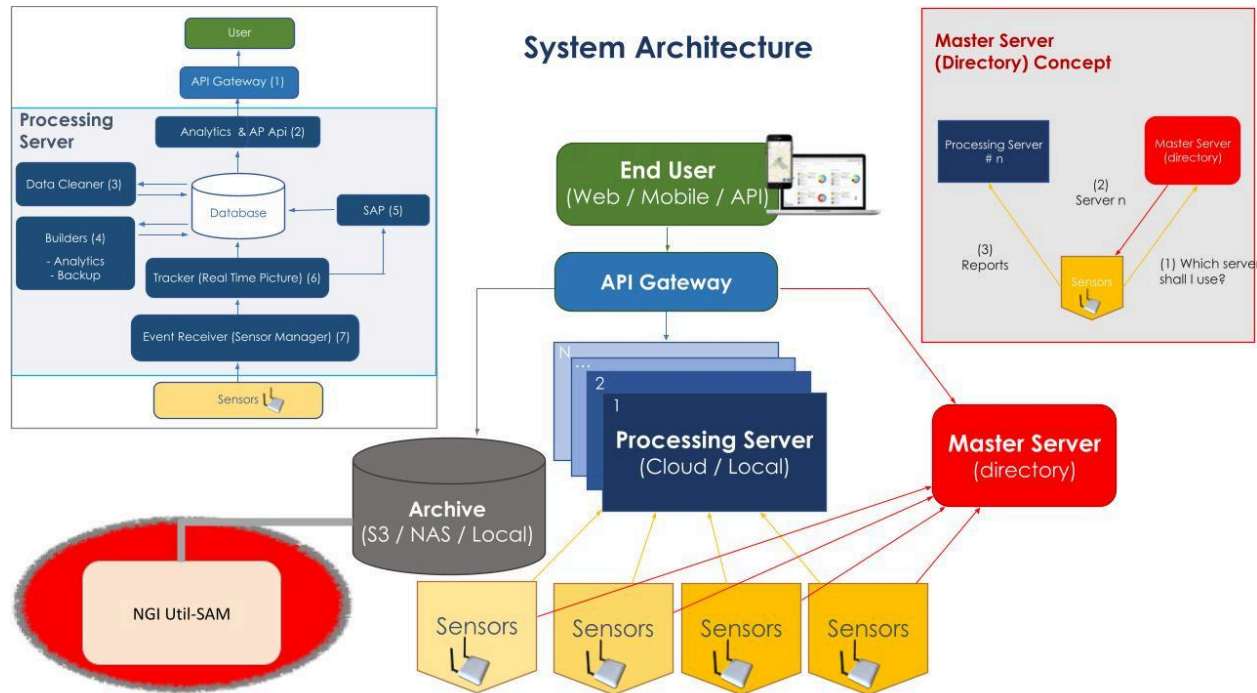
The anonymized dataset of mobile phone detection is exposed to a de-anonymization technique called profiling which is based on comparison between mobility patterns detected from 3rd party databases (e.g. facebook) and the correlation between those mobility patterns to the existing mobility dataset.

The risk in this case is that the attacker may be able to detect new mobility point (that are not part of the 3rd party dataset) in our dataset in such a way, he will be able to gain information about locations (points) visited by the person in the 3rd party dataset which are not available in the original dataset.

Behind State of the Art (What is going to be developed in the project) ?

1. An auditing tool that can scan a CSV dataset and identify mobility records that may be exposed to profiling
2. A tool that can "fix" mobility records that are exposed to profiling and anonymize those patterns.
3. A tool that can receive N points and check how many mobility patterns meet those points. ("Query Tool") - this tool will allow end users to check if their personal data is included in a dataset
4. A simple GUI that demonstrates the use of the above tools
5. Two open data datasets (Barcelona and Bangkok) that had been generated using the anonymization tools
6. Deployment of the open datasets in one of the European Open Data Platform (e.g. Gaya-X) and a tool for integration between anonymized dataset open data platform
7. A prototype of decentralized anonymization process (for EDGE sensors) including analysis of the performance of such a process compared to centralized anonymization

System Architecture with Util DAM (marked in red)



The Utip-DAM project will be connected to the archive service of the existing system. The input for the project will be the daily backup track file and file will be anonymized for mobility using the tools that had been developed at the project.

We are going to offer two types of services for customers:

1. For customers that like to dash data via Correlation Systems Utip-DAM the daily track data can be simple submitted to Correlation Systems instance of Utip-DAM and will become public available via this portal
2. For customers that do not like to share the data publicly, we will provide a script that prepare the dataset on a daily basis and store the data on the same local archive
3. In a case that a customer would like to offer the data via his own portal, he can select one of two options: use the anonymized dataset (item 2) and offer the data via his own portal or deploy Utip-DAM portal on his own server and offer the data with his own portal.

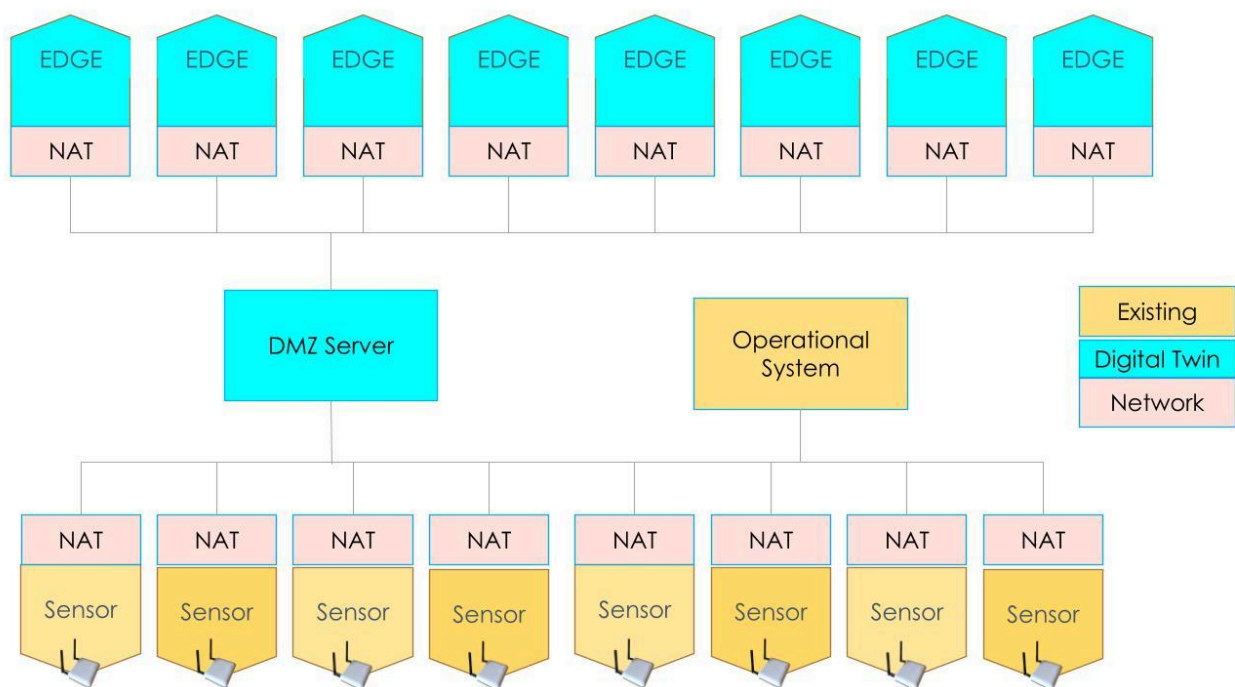
EDGE Prototype

Digital Twin concept

As we are not able to use an operational system for the development of the distributed version of Utip-DAM, we will replicate the Bangkok BTS site to a mirror site that will simulate the behavior of a large EDGE site using the real data collected by the sensors in the Bangkok BTS site.

For this we will use a feature that exists in the sensor which allow to duplicate the reporting of the sensor into multiple servers and we will prepare a small utility that will replace the actual sensor software with a software that received the data from a duplicated reporting channel

The following figure present the architecture of the EDGE prototype:



Networking issue

Typically, an EDGE sensor will be connected to the internet via a modem, in such a case by default, the sensor will be located behind NAT and it cannot be accessed from the outside. A similar architecture is also common in private installations where the local firewall prevents access from the local network to the sensor.

As this is the most common situation for the deployment of sensors, our prototype will be operated behind NAT and access to sensors will be available only via polling.

I. Software design and analysis, component specification (preliminary)

Design Considerations

In our vision, the system that we are developing is going to be deployed in multiple configurations:

1. As a part of an existing system that is collecting and managing datasets of mobility data
2. As an external tool for existing systems in order to audit existing datasets
3. As an interactive system (including user interface) mainly for testing and auditing purposes
4. As a tool for data scientist for analyzing mobility datasets

In order to meet all those requirements the system design is based on a modular approach in which the core functionalities are stored in a stand alone, self contained packages, ready to integrate in third party libraries and other components such as the interactive user interface are using the basic components