Privacy-protecting Data Management in distributed Cloud Infrastructures

Andreas Müller, Sebastian Hudert, Victor Fäßler

TWT GmbH Science & Innovation
Ernsthaldenstraße 17
70565 Stuttgart
{firstname.lastnameg}@twt-gmbh.de

Abstract: Today's cloud infrastructure platforms such as Amazon Web Services (AWS) allow to deploy and run complex services without having to worry about scalability, reliability and general maintenance tasks. However, with distributed services running on multiple instances and on potentially untrusted nodes, the protection of data and in particular the preservation of privacy has become a huge challenge: On the one hand, data storage must be achieved in a way that an attacker with access to the file system, or in some use cases also the cloud operator, is not able to retrieve critical information such as personal data. On the other hand, the combination of individual data chunks to privacy critical information at runtime must be prevented when creating distributed applications. For example, a commercial location-based service needs information about the location of a device, as well as personal data of the owner for billing issues. Today, both of these details (location and personal data) are kept at the service although a combination of the two is not necessarily needed: a location-based service can operate without knowledge about personal data of the owner and a related billing service without the exact location. This talk focuses on such scenarios and presents ongoing work that is carried out in the BMWi-funded research project Shared E-Fleet.

The research project Shared E-Fleet develops a cloud-based e-mobility platform for electric fleets in corporate environments. The main idea is to share electric vehicles between different companies to increase their efficiency and to minimize costs. For example, a technology park consisting of many individual companies may operate a fleet and host the Shared E-Fleet cloud infrastructure allowing companies to book and use vehicles in a flexible way. One of the challenges here is that on the one hand different companies share a common fleet of electric vehicles through a common platform, while on the other hand the companies might be competitors. Thus, the separation of data storage, as well as the careful aggregation of data stored in the platform is needed.

The Shared E-Fleet architecture is composed of multiple web-services, each one dedicated to a specific functionality either related to the management and booking of the fleet or to services needed at runtime. Thus, each of these services collects and stores data to operate properly: general management services hold static information about the customers and vehicles (e.g. company names, user accounts, addresses and billing

information). Additionally, on-board units (OBUs) in the vehicles periodically push status information (e.g. the current location, state of charge, error messages) to a dynamic data pool. Other services then rely on the combination of static and dynamic data to provide the essential Shared E-Fleet services: a booking service needs to know about the customers, as well as about the current electric range of available vehicles in order to make sure to assign the correct vehicle to a requested booking. Similar combinations are needed for route optimization, charging strategies and also billing. Thus, our goal is to ensure data reduction and data economy when combining such data to create new services.

To achieve this goal we have developed a concept for the data management of privacy critical data running on a private cloud infrastructure (in our case a combination of Openstack and WSO2). The exchange of data between the individual web-services is done via a trust center service that maintains authorization policies for the access and for the combination of data. After state of the art mutual authentication, authorization for the specific attributes takes place and only if the trust center policy approves the requested combination of data based on the token of the requester, the information is aggregated and forwarded.

Additionally, a more sophisticated concept, e.g. for deploying the architecture on public clouds, is envisioned. This includes the encrypted storage of data, as well as the integration of Trusted Computing technologies to ensure that only nodes that run a valid and approved configuration of the operating system and applications are allowed to access data on a specific host.