GAIA Information System Design Overview

GAIA platform architecture and design principles

A main objective of the GAIA project is environmental sustainability education and energy efficiency awareness initiatives in schools to make students aware that energy consumption is largely influenced by the sum of individual behaviors (at home, school, etc.) and that behavior changes and simple interventions in the building can have a great impact on achieving energy savings. IoT technologies allow for data production from the real world to feed a plethora of people-centric information, education and involvement initiatives conducive to effectively change the ways people live and work inside school buildings and achieve better energy efficiency. This means that by enabling a whole different set of applications, like gamifications apps that bridge the virtual world with the real one, towards the end-goal of such systems people will be better informed and capable of making more educated decisions. Also, since public buildings, and especially educational buildings, constitute a large part of the buildings in Europe. Students also constitute a sizable part of the overall population, making these technologies available and tailored specifically for that end-user group can potentially result to large benefits.

GAIA's design follows the idea that the availability of actual measurements of environmental parameters, such as energy consumption, indoor and outdoor luminosity, temperature, noise, pollution, etc., enables the conception and realization of diverse applications and scenarios. Our goal is to provide a platform which will cater to the requirements of a broad variety of applications that will facilitate the educational sector towards improving the energy efficiency of school buildings. On top of that, the educational sector has certain sensitivities that push us towards having a concrete approach with respect to privacy.

We report on the security and data privacy mechanisms employed in the development of the GAIA infrastructure and the services that will be used in the creation of the GAIA applications. We therefore start with presenting the main authentication and access control mechanisms employed in all web services of both the analytics engine and the data access modules and move to discuss issues related with data anonymization and de-identification for information related to all participants (teachers, students, administrators and platform owners).

GAIA's design philosophy in short

GAIA aims at improving energy efficiency by increasing awareness of specific target groups, all related to the educational process and community. To achieve this goal, we will utilize the infrastructure available at schools and enrich it towards gathering information. This information is used by a set of applications to guide energy efficient behavior which is assessed through the continuous monitoring of building energy consumption.

Security in GAIA APIs and Client Applications

GAIA's analytics and storage services are based on SparkWorks platform for IoT data analytics that uses the OAuth2 protocol to provide user and service authentication and authorization. OAuth2 is an open standard for authorization, commonly used as a way for Internet users to log in to third party websites using their accounts. This method is used by major Internet service providers including Google, Facebook, Microsoft, Twitter, etc. and aims at client developer simplicity as well as authorization flows for web applications, desktop applications, mobile phones, and other connected devices. By using OAuth2, users of the GAIA services can log in to all services provided by GAIA (mobile phone applications, websites, etc.) using a single account while providing only limited access, in an as needed basis, to their own personal data. The same authentication mechanism is used by the GAIA sensor data providers (school building managers) to integrate their own infrastructure to the GAIA central repository.

In more detail, SparkWorks offers a Single-Sign-On service, where all users of the GAIA system can create their account and provide any information required in the context of the project. This information can include but is not limited to their role (student, teacher, building manager), age, gender or home Institution. This account is therefore their passport to access all services of the project and the key to their participation in project's game. All communications with this portal are encrypted with high grade SSL certificates and can be access only through authorized GAIA clients of the SparkWorks services. All personal information of users (such as passwords) is also stored in an encrypted and hashed way so that they cannot be directly inspected or accessed via the database infrastructure.

Building managers also use this portal to integrate their infrastructures with the analytics and data storage service as they need to create a client application that will use a set of client identifiers and private keys to publish data. These keys are also made available through SSL encrypted connections and need to be stored safe on their bearer's responsibility. Each application can have multiple permissions on the system. Upon request these keys can be invalidated or re-issued in case they are lost or stolen. The two most common permissions are read and write. In this case, building manager applications need to have both the permissions to successfully publish data to the platform. Front-End application developers need also to register their applications as clients similarly to the building managers but in their case only the read permission will be required in most straightforward cases. In both cases, if the keys of an application are compromised they can be reissued so that the compromised keys are no longer valid.

Access to the Storage Services of the GAIA infrastructure is only possible using a valid client application and a user's account with permissions and a role capable of accessing the data. Additionally, each user has access that is defined by their home Institution while their access to other data is limited. Similarly, access to the semantic information of the GAIA platform, like the number of sensors in a building, sensor names and sensing properties is provided using the same method. For all interactions with the API providing this information a valid access token is required. This access token is an OAuth2 Bearer token

provided when a user logs in the platform using an application. In general, all authorization requests in the GAIA project is depicted in Figures 1 and 2.

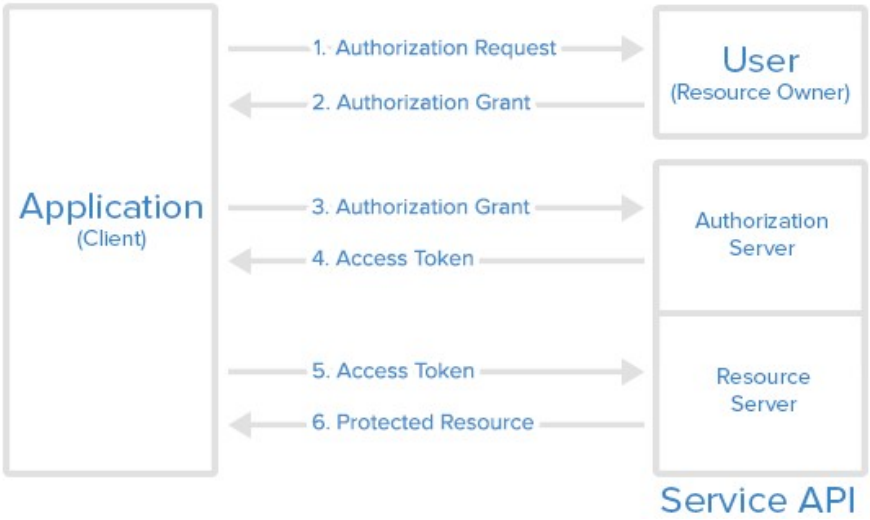

Figure 1 Implicit Auth2 Protocol Flow

This flow describes how an application requests access to a user's protected profile and resources. In general, it requires 3 steps during which:

1. The user logs in to the platform, authorizes the application and provides it with a temporary authorization grant to claim the access to the user's data,
2. The application users the temporary authorization grant to get a user-specific access token,
3. The application users use the user access token to interact with the services on behalf of the user.

In the 2nd step the application secret token is used to certify the identity of the application. To keep the application's secret key service safe, it need to be stored in a backend service, accessible only by the application itself. In order to provide client-side applications that do not use server component in their operation (e.g., a web page) a simplified flow is also available. This flow is presented in Figure 2. Access to all GAIA information and management portals is possible only through HTTPS encrypted connections.

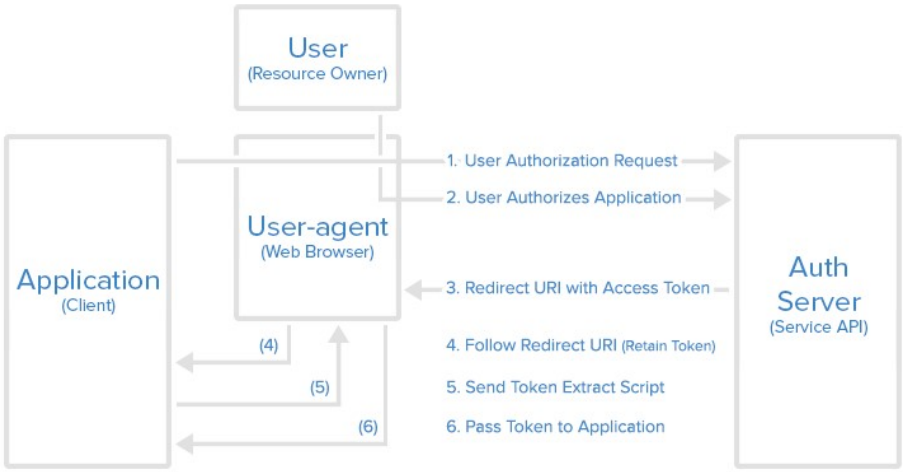

Figure 2 Implicit Auth2 Protocol Flow

Security in GAIA analytics and data storage services

Access to sensor data (historical or real time) stored in the various GAIA services is available only through the previously described APIs. The actual storage of the sensor data is performed in the appropriate databases (MongoDB, MySQL or Neo4j) with local access enabled only and the provided authentication mechanisms of the specific database. To ensure this, appropriate firewall rules and configurations were configured and respective web-based APIs allow access only to authorized users.

Data privacy in GAIA

The privacy of all GAIA participants is of great importance to all the partners of the consortium. To assure that we employ various anonymization techniques during both the data collection and data representation in every service and application provided by GAIA. The first and most important step happens during the data collection from inside the institutional buildings. All data collected are aggregated over 5 minute intervals. This step allows us to extract the required insights and analytics but also removes any possible data that could identify the actions of specific individuals.

In the next table we describe the collection and storage intervals for various sensor types already available in the GAIA platform. Some of the sensors have much lower collection times due to power restrictions on the devices.

Sensor Type	Data Collection Interval	Stored Aggregated Data Intervals
Electrical Current, Power Consumption	30 seconds	5 minutes, 1 hour
Temperature, Humidity, Luminosity, Noise	30 seconds	5 minutes, 1 hour
Movement	On event	5 minutes, 1 hour
Atmospheric Pressure	5 minutes	5 minutes, 1 hour
Pollutants	5 minutes	5 minutes, 1 hour
Radiation	5 minutes	5 minutes, 1 hour
Wind Direction, Wind Speed, Rain Level	5 minutes	5 minutes, 1 hour

Similar techniques will be used for data collected by surveys or crowdsensing applications used by participants of the system. Publicly released answers and results will be provided as answers by persons of a specific age group or inhabitants of a specific city or country. Another important aspect is the existence of multiple levels for accessing data in the context of GAIA. Students, teachers or building administrator may have only limited access to the total data collected by all the GAIA infrastructure. In an example scenario:

- Students can have access only to historical data for their own school and only charts with the rest of the buildings in the project.
- Teachers can have more detailed data for all the classrooms of their school and charts with the rest of the buildings in the project.
- Building managers can have full access to the data of the school they manage and comparison charts with the rest of the buildings in the project.
- GAIA administrators have full access to all data of the buildings in the project.

Policy regarding data after the completion of the project

All personal data that cannot immediately be anonymized and is collected for evaluation purposes, will be deleted at the latest 3 months after the end of GAIA or 9 months after collection, whatever is earlier. Data provided by students and other participants that is intended for the use within the system, will not be deleted, but all participants will be enabled to delete or correct any data produced/provided by them.

Policy regarding the use of servers and cloud infrastructure in GAIA

Although GAIA will ultimately produce solution that will be cloud-ready, the project will avoid the use of popular cloud infrastructure facilities, research of commercial, due to privacy and security concerns of the project stakeholder, as well as the end-user communities involved. As also explained in the following section of this document, the end-user data related to gamification do not belong to any personal data category. However, they will also be stored on protected servers of the involved consortium members with access to only those who absolutely need the stored data for developing GAIA software solutions and supporting the operation of the project. More specifically, for the duration of the project, relevant data for the project will be stored and processed on servers located in the headquarters of CTI Diophantus in Patras, Greece.

Gamification Design and Privacy

A significant part of the application aspects of the project will revolve around gamification and the relevant application and user interfaces, either through mobile phones/tablets, or Web interfaces. Gamification will also be one of the main engagement mechanisms implemented in the project, targeted mainly towards the students of the participating schools in Greece, Italy and Sweden. Thus, apart from the aspects described in the previous section concerning environmental sensor infrastructure reading and energy-related data, the participation of students in the gamification aspects of the project can potentially create end-user privacy concerns, if not handled in a holistic manner from the beginning of the project.

In this section, we describe the related technical aspects, as they have been formed from the design of the gamification aspect thus far. In short:

- The GAIA Consortium has taken students' privacy seriously into account for the design of the gamification aspects.
- No sensitive private data will be required from students participating.
- Related aspects will be explained through documentation on the project website and other similar means, as well as workshops or special events in schools participating in the project.
- School-specific policies with respect to e.g., online social networking platforms such as Facebook, will be respected by the consortium, and the respective changes will be implemented.

Summary of GAIA gamification activities' design

IoT real-time data from the real world can potentially make for a more interesting proposition for educational activities regarding energy and sustainability. Gamification intends to make this even more personal, adding a joyful flavor to the overall experience, along with making it more personal and challenging. Overall, achieving behavioral changes with respect to energy consumption behavior, revolves around motivation, call to action, active participation in GAIA activities and at the end, through feedback

and rewards, by incentivation and intensification to continue engagement with the project and achieve better results over time.

With respect to such data, installation in shared places and classrooms, where it is easier to engage a larger number of end-users (students) was chosen to aid the gamification aspects of the project. Instead of focusing only on smaller student groups, such as specific classes, we chose to engage larger groups to have a greater number of students involved directly with the project. This has also the added benefit regarding privacy and ethics that more end-users will operate in the same physical space, making it even further difficult to differentiate between different users and adding to their anonymity.

The overall idea of gamification in GAIA revolves around certain "tasks" and "quests" that are meant to help mainly students reach a better understanding of concepts such as sustainability and energy efficiency. To achieve this, the end-users need to:

- Register to the gamification software subsystem of GAIA.
- Login to the said system when participating in related activities.
- Interact with the real-time data provided by the system in the form that will be available via the project, such as the abovementioned sensing quests or the related educational interactive material.
- Interact with the software, and through it with other users of the system, in the same school facility or from other school facilities.

Gamification-related data

Regarding actual data required for the operation of the gamification aspects of the project, there are 2 general categories:

- Data from the real world, produced by environmental IoT sensors, examined in the previous section. Such data are use as real-world input for gamification.
- Data related strictly to the operation of the gamification software components, i.e., data related to end-user profiles and activity.

Regarding the first category of data, we have already discussed the related ethics and privacy aspects. With respect to the second category, we continue with a discussion with the 3 areas in which GAIA will store data, which are the following:

- User registration in GAIA's gamification subsystem.
- User profile in the said subsystem.
- User activity within the gamification aspects.

Regarding the end-user's registration at the challenge website, end-users will only need to provide:

- A valid username.
- The name of the school in which they attend classes.
- Name of specific class of their school.

The provision of an email address will be voluntary, to retrieve data such as a password, etc. As mentioned before, GAIA will not ask for personal data such as birth date, gender, etc., so essentially the users/students do not need to provide any sensitive user data at any point of their registration to GAIA's platform.

Furthermore, the User Profiles that will be created for each end-user participating in the gamification aspects will feature:

- An avatar picture, which can be a photo uploaded by the end-users, or an icon available from a list provided by the project. End-users will be advised against using a real-world photo.
- Experience score, which will be utilized to rank users. All users of a facility will appear in a facility-specific ranking, thus peer group users can compete with students as well. The best ranked users appear also in the Hall of Fame.
- An activity log, with respect to online activity in the gamification component of GAIA.
- A list of friends, which can include users of the same facility/school.
- Topic progress, an overview of the progress achieved with respect to GAIA's context.

In more detail, the users' activity on the gamification component of GAIA that will be monitored, will include the following aspects:

- Gained experience points by completing basic and bonus Quests provided by GAIA.
- Submission of snapshots of progress.
- Other users/friends voting on their snapshot.
- Votes on other users' snapshots.
- Addition of other users as friends.
- The user's ranking as part of a specific school facility.
- Sharing their profile on social media.
- Selection of a facility during registration (and become a contributing member of this facility for the GAIA project overall).

In all cases, students will be given guidelines to minimize any privacy issues and avoid exposure of personal data.