3 DS2 E2C Security Module (SEC)

3.1 DS2 E2C Security Module (SEC)

Owner(s):	DIGI
DOA Task:	T6.2
Tier:	0
Nature:	System
Result:	Outcome

This task will enable an edge-to-cloud connectivity through applications and devices capable of collecting, processing data and interconnecting this data with the cloud infrastructure of T6.1. It will be receiving, storing, and processing enormous amounts of data and management tools need to extract insights and make data-driven decisions based on data. This task will implement the DLC ecosystem that ensures data security and privacy and implement appropriate measures such as encryption, access control and anonymization. All relevant adapters, interfaces and UIs are developed within this task that allow the use, maintenance, and full control of the ecosystem. This task will establish concepts of open data that permit public data availability and accessibility for use and reuse without restrictions. It will further ensure that individuals and organizations will stay in control of their data, allowing them to control their own data and decide how and when it is shared. Once the federated IDT environment is operational through T6.3 amongst other task, it will be necessary to monitor it continuously at execution time (aka T4.2). Based on a data quality lifecycle, tools will be developed to define and constantly detect and monitor data quality and establish a framework for automatic checks, e.g. of loss detection, breakdown of patterns, percentile and tolerance checks, anomalous values etc. ATO do the CEP processing and other streaming detection techniques will be used and detailed in the architecture.

3.1.1 Introduction

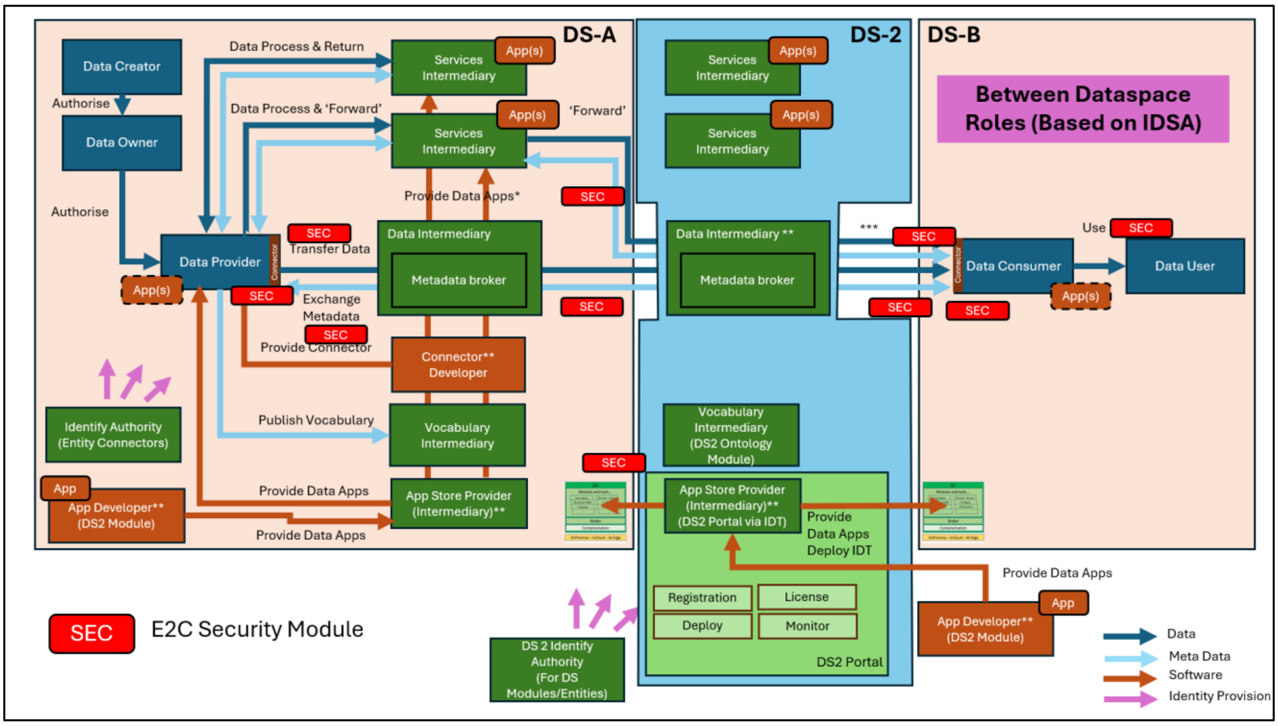
Purpose: The DS2 E2C security module (SEC) covers data security, data protection, and privacy with a focus on securing the edge-to-cloud data enablement and ensuring data quality and privacy. This involves implementing secure communication protocols, robust authentication mechanisms, encryption, anonymisation, and continuous monitoring of events in the DS2 ecosystem. The DS2 architecture is designed to handle large volumes of data, facilitating data-driven decision-making while maintaining stringent data security and data privacy standards. The proposed component works in conjunction with all DS2 architecture components.

Description: Data collection and preprocessing (e.g., initial data filtering, noise reduction) often happens at the edge at devices level. Depending on the use case for DS2, an edge gateway can be attached to a data source or sensor for securely transmitting data over a secure channel, using an end-to-end encryption (E2EE) protocol. OAuth 2.0 or OpenID Connect is used for authenticating and access control. The infrastructure provides encrypted data storage and tools for privacy aware data management including anonymisation. SEC manages data encryption, both in transit and at rest. A continuous monitoring of security incidents is performed using a Security Information and Event Management (SIEM) system which allows responding to potential security threats.

3.1.2 Where this component fits

3.1.2.1 Big Picture

Given the nature and context of use of the module, it covers data security, protection, and privacy across the Dataspace components. The sub-components of the security components are applicable for components within a dataspace, as well as across dataspaces.



Where	Status
Within a single Dataspace for use between participants in that Dataspace only	Yes – dataspace components requiring data security, privacy, secure connection establishment will use this module.
Deployed and used by a single participant to enable the participant in either an In-Data space or Inter- Data space scenario	Yes - for securing the edge-to-cloud data enablement.
Across Dataspaces without Service Intermediary	Yes – for secure data sharing, data protection, and data privacy.
Across Dataspace with Intermediary	Yes - for secure data sharing, data protection, and data privacy.
Other Comments	N/A

3.1.2.2 Within a single Dataspace (where applicable)

Components requiring secure storage, anonymisation, tokenisation etc. (within a single dataspace) will utilise SEC. In addition to that, SIEM will also be used to track the events to monitor for security incidents.

3.1.2.3 Deployed and used by a single participant (where applicable)

Any data sharing from provider to consumer should always be secured through secure authentication, access control, end-to-end encryption (E2EE), and must be stored in a secured storage (even if it is for a very short time). If needed, data anonymisation, tokenisation is handled by this component for both the data provider and consumer. SIEM will also be used to track the events to monitor for security incidents and respond to if necessary.

3.1.2.4 Across Dataspaces without Intermediary(where applicable)

The same logic stated above for within a single dataspace applies for the security component as it enables data security, protection, and privacy.

3.1.2.5 Across Dataspace with Intermediary (where applicable)

A dataspace service intermediary must utilise the security component for the reasons stated above.

3.1.3 Component Definition

The figure below represents the actors, internal structure, primary sub-components, primary DS2 module interfaces, and primary other interfaces of the module.

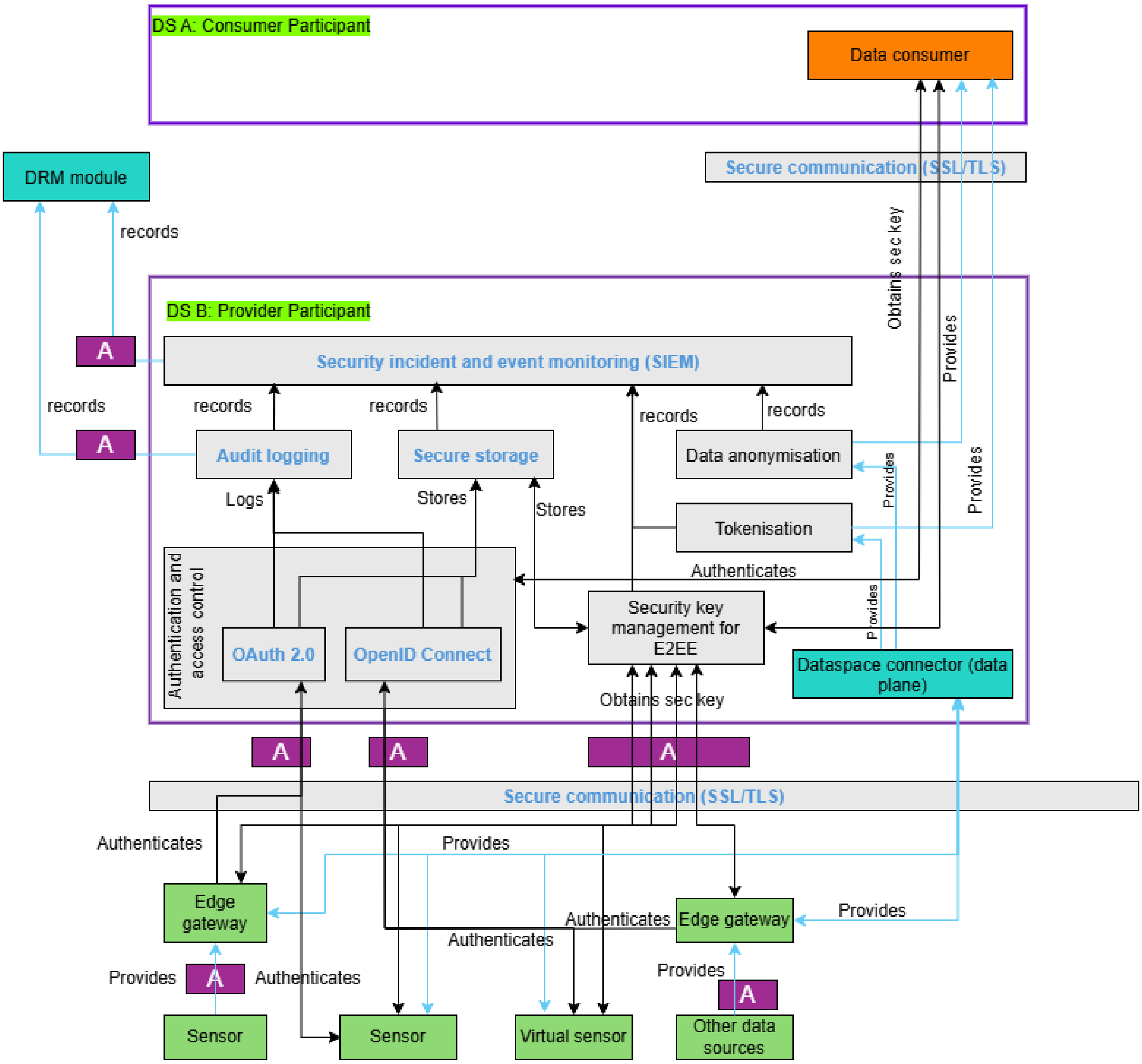


Figure 13: Schema for the DS2 E2C Security Module

This module has the following sub-components and other functions:

- **Authentication and access control**[1]:
  - **OAuth 2.0**: Provides secure authorisation for users and applications accessing the data. OAuth 2.0 allows third-party services to exchange user information securely without exposing user credentials. This module will be used as is and upon successful authentication, it will return an access token. More specifically, SEC will utilise OAuth 2.0 with Client Credentials Grant since the clients (i.e., sensors, gateways) will use it to obtain an access token outside the context of a (human) user.
  - **OpenID Connect:** OpenID Connect (OIDC), built on top of OAuth 2.0, adds an identity layer to verify user identity. This module will be used as is and upon

[1] Every successful authentication returns an access token (such as a JSON Web Token) which is used for enforcing a role-based access control (RBAC). It restricts access to data, read/write access to database based on user roles. Only users with the appropriate permissions can access or modify data and/or database.

successful authentication, it will return an access token for enforcing access control. If the DS2 ecosystem involves users owning sensors or Edge gateways, OIDC can help authenticate the owners and authorise their access to specific resources, making it more secure for managing permissions between multiple users and/or resources.

- **Audit logging**: Every authentication request and subsequent (successful and unsuccessful) authentication access are logged. Audit logs are monitored to detect and respond to block unauthorised access attempts. While this module exists in DIGI's background (as a part of both cloud-based Paradise platform and Digiotouch Edge) and will be used as is, but also it is integrated with the SIEM for a holistic event monitoring. The log information is also stored in the Blockchain of DS2 DRM module. This additional logging in blockchain is optional.
- **Secure storage:** This stores authentication tokens with access control information and security key for E2EE. It cooperates with the security key management sub-component by storing the encryption keys. Data stored in the storage is encrypted. Note it does not store sensor data.
- **Security information and event management (SIEM)**: It is a sub-component that aggregates logs and event data generated by all users, servers, edge devices, and firewalls participating in the DS2 infrastructure to monitor and analyse events for security-related incidents detection and response. Like the audit logs, SIEM also stores the incoming log and event data to the DS2 DRM module. This additional logging in blockchain is optional.
- **Secure communication:** It establishes and maintains a secure communication channel using secure virtual private cloud (VPC) peering (when communicating among cloud servers) and/or a VPN. All data in transit are protected using SSL/TLS. This module exists in Digiotouch's cloud-based Paradise platform and Digiotouch Edge (which is an Edge Computing platform) and will be used as is in this module.
- **Tokenisation**: Replaces sensitive data elements with a non-sensitive equivalent, known as a token, which can be mapped back to the original data. This process helps protect sensitive information by ensuring that tokens, rather than actual data, are used during transactions or storage.
- **Data anonymisation:** This optional sub-component provides the following functions:
    - **Differential privacy**: Techniques such as differential privacy add noise to data in a way that statistical properties of the data set are preserved while individual data points are obfuscated. This protects user privacy when data is shared or analysed.
    - **Generalisation and suppression**: These techniques modify or remove specific data points to reduce the risk of identifying individuals in a dataset.
- **Security key management for E2EE**: This covers the lifecycle of operations which are needed to ensure the security keys (e.g., AES) are created, stored, used, rotated, and destroyed securely. A key management system (KMS) is employed in this sub-component and encrypted keys are stored in the above-mentioned secure storage. These security keys are used for end-to-end data encryption.
- **Sensor**: This acts as a data source (e.g., IoT sensor, camera etc.) in the DS2 ecosystem. It is assumed that some sensors are capable of performing authentication, data encryption, and interaction with web services on its own while some sensors do not have such capabilities and are assisted by an Edge gateway.
- **Virtual sensor**: It is a software component that mimics the data behaviour of a real sensor. Synthetic data is an output of a virtual sensor.

- **Other data sources**: They include public or private repositories of data that are useful for the use case applications. For example, weather data from a region which is available in a public repository.
- **Edge gateway**: It is a device which assists sensors with limited processing power with authentication, data encryption, secure data communication etc.
- **Dataspace connector**: Enables encrypted data communication through it.
- **Data consumer**: It belongs to the consumer participant which consumes encrypted and/or anonymised data coming from the provider participant.

3.1.4 Technical Foundations and Background

The security module subcomponents will utilise foundational technologies that are well established in the cybersecurity industry. For example, secure authentication in this case will cover OAuth 2.0 which is explained in RFC 6749[2] and 8252[3] (for native apps) and OpenID 2.0[4] specs.

For tokenisation, tools like Apache Nifi, OpenPseudonymiser will be used as they support tokenisation as a part of their features. Open source Elastic SIEM[5] will be used in this module.

Subcomponent/Component	Owner	License
OAuth2.0 implementation	Open source	Apache 2.0
OpenID implementation	Open source	Apache 2.0
Apache Nifi	Open source	Apache 2.0
OpenPseudonymiser	Open source	GNU General Public license
Elastic stack	Open source	Elastic license 2.0[6]
Audit log, secure storage	Open source	Apache 2.0
SIEM	Open source	Apache 2.0
Key Management Interoperability Protocol (KMIP)[7]	Open source	Apache 2.0
Software enabling secure communication using SSL/TLS	Open source	Apache 2.0

3.1.5 Interaction of the Component

The following table specifies the primary input/output controls/data to blocks which are not part of the module.

With Module/Feature	Received From/Gives To	What
Sensor, edge gateway, other data sources	Received From	Authenticate itself to receive a security key for data encryption
Sensor, edge gateway, other data sources	Gives To	Access token after successful authentication and key for data encryption
Dataspace connector	Received From	Data to be shared through secure communication
Data consumer	Received From	Authenticate itself to receive a security key for data decryption
Data consumer	Gives to	Encrypted (and anonymised) data
DRM module	Gives to	Audit logs

[2] https://www.rfc-editor.org/rfc/rfc6749
[3] https://www.rfc-editor.org/rfc/rfc8252
[4] https://openid.net/specs/openid-authentication-2_0.html
[5] https://www.elastic.co/blog/elastic-siem-free-open
[6] https://www.elastic.co/licensing/elastic-license
[7] https://github.com/OpenKMIP/PyKMIP

3.1.6 Technical Risks

Risk	Description	Contingency Plan
Zero-day vulnerability discovered	A zero-day vulnerability is discovered and widely disclosed pertaining to a security mechanism or a tool used in this module.	Security patches will be applied as recommended while stronger mechanism/tool will be investigated and if possible implemented.
Sensitive data leakage	Incorrect implementation of the security mechanisms may lead to sensitive data leakage.	DIGI will provide a set of unit tests to ensure correct implementation of security mechanisms leading to no sensitive data leakage.

3.1.7 Security

Security Issue	Description	Need
N/A since this is the security module		

3.1.8 Data Governance

Data Governance Issue	Description	Need
Handling sensitive information	Logs, access tokens represent confidential information.	Secure data transfer and encrypted storage are to be provided by this module.

3.1.9 Requirements and Functionality

This module will be used in the following use cases:

City Scape	✓
Green Deal	✓
Agriculture	✓
Inter-Sector	✓

Their requirements and functions/extensions to achieve them relative to this module, specifically extracted from the use case are as per the table below noting that in many cases further discussion might need to take placed between pilot partner and module partner to determine if a fit or the scope of the precise fit:

WHERE	WHAT	WHY	Run/Design Time	Priority
	Use Case 1: City Scape			
Section 2.2 UC1.1	“share the data on energy consumption and other patterns because I want to save money”	Confidentiality of the data, privacy of personal data, data protection	R & D	M
Section 2.2 UC1.1	“Receive consumption information to optimize the production and create more attractive packages.”		R & D	M
Section 2.2 UC1.1	“Get data on awareness and value put on energy effectiveness to create better products / fit the needs”		R & D	M

Section 2.2 all UCs	“Sharing and gathering data from multiple sources and sectors”	Secure sharing of data from multiple sources and sectors	R & D	M
	Use Case 2: Green Deal			
Section 2.2 UC2.1	“Data is shared and accessible via DIH AGRIFOOD DATA SPACE”	Inter dataspace data security for data sharing	R & D	M
Section 2.2 UC2.2	“Relevant data sources to be obtained from both data spaces within the use case. “		R & D	M
Section 2.2 UC2.3	“Data about households (including flats) to be arranged including number of participants, building area and other relevant data. Dataset shared through the MOMS DATA SPACE”	Confidentiality of the data, privacy of personal data, data protection	R & D	M
Section 2.2 UC2.4	“MOMS is using the data from two sensors in the city”	Secure communication of data coming from the sensors and data integrity	R & D	M
Section 2.2 UC2.5	“Data anonymisation should happen at the source”	This module will support such anonymisation at data source for confidentiality	R & D	S
	Use Case 3: Agriculture			
Section 2.2 UC3.1	“Retrieve data from DigiAgro DS where sensor data from crop owners is collected and stored”	Secure sharing of data from multiple sources	R & D	M
Section 2.2 UC3.1	“Use AgroScience DS for data processing services and machine learning algorithms to analyse the collected data”	Ensure data is kept in secure storage and data integrity is maintained before they are used in ML algorithms	R & D	M
Section 2.2 UC3.2 and 3.3	“Integrate environmental data and weather forecast data seamlessly between the two dataspaces”	Secure sharing of data from multiple sources	R & D	M

3.1.10 Workflows

The following sub-sections describe the sequence diagrams of the SEC Module.

3.1.10.1 Authentication using OAuth 2.0

This feature shows the sequence diagram for a sensor or an Edge gateway to authenticate itself to the OAuth 2.0 server. On a successful authentication, the access token is securely stored in the local storage. The resource server is represented with data consumer to which access is granted for data sharing operations.

The main steps/functionalities are as follows:

- Sensor/Edge gateway (i.e., the client) authenticating itself and receives an access token.
- Access token is also stored on the local secure storage.
- The client requests access to a protected resource (data consumer in the Figure 11) which validates the token and grants access.

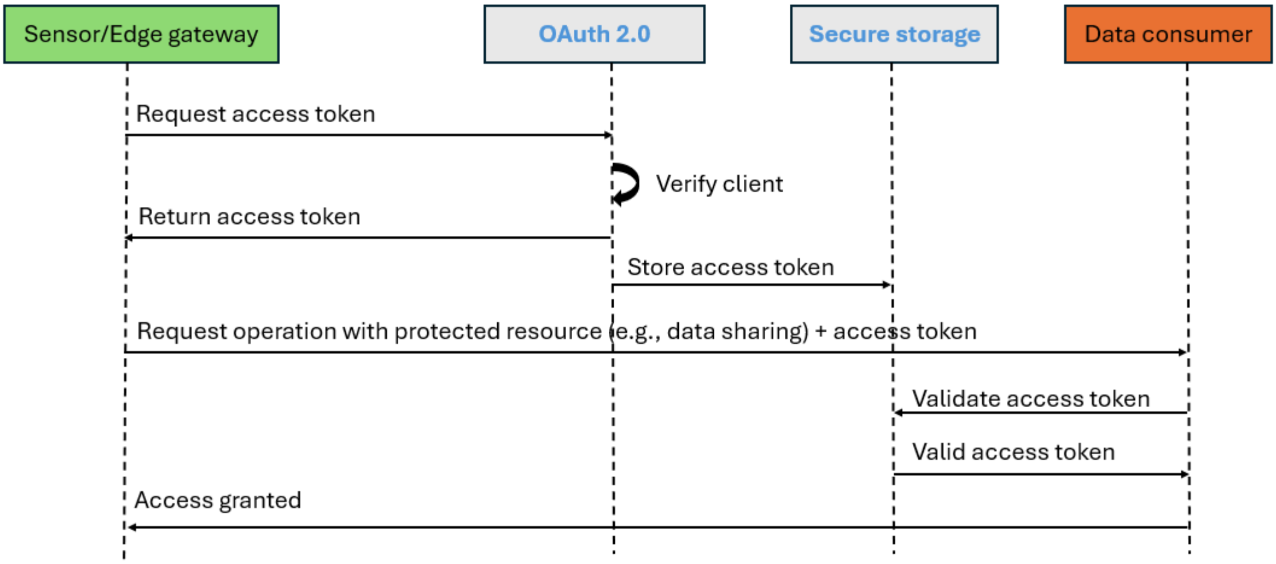


Figure 11: Authentication using OAuth 2.0 sequence diagram.

3.1.10.2 Record event information

This feature provides the capability to log events generated. For example, audit log records authentication attempts (successful and unsuccessful) while SIEM records other events. Both of them forwards the events to the DS2 DRM module for storage in its blockchain network. Figure 12 shows the sequence diagram of this feature.

The main steps/functionalities are as follows:

- Audit logging receiving authentication attempt events.
- SIEM receiving all other events.
- Both audit logging and SIEM recording the generated events into DRM module's blockchain.

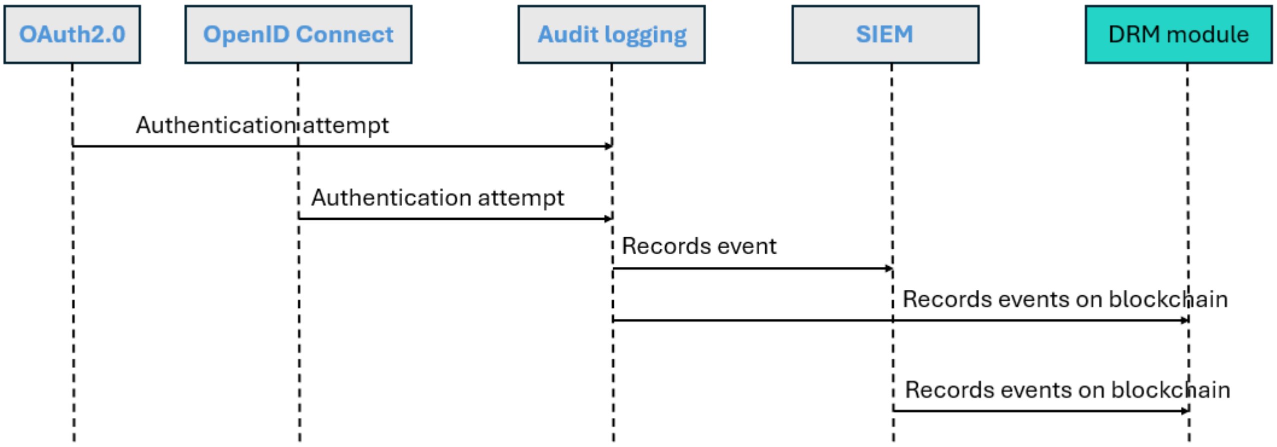


Figure 12: Record event information sequence diagram.

3.1.10.3 Obtaining security key

This feature shows the sequence diagram for obtaining a security key for data encryption for an authenticated client (e.g., sensor, Edge gateway). to authenticate itself to the OAuth

2.0 server. On a successful authentication, the access token is securely stored in the local storage. The resource server is represented with data consumer to which access is granted for data sharing operations. Figure 13 shows the sequence diagram of this feature.

The main steps/functionalities are as follows:

- Authenticated clients requests new security key.
- The key management system (KMS) generates, stores the new key, and records the event in SIEM.
- A secure key exchange takes place between the KMS and the client.

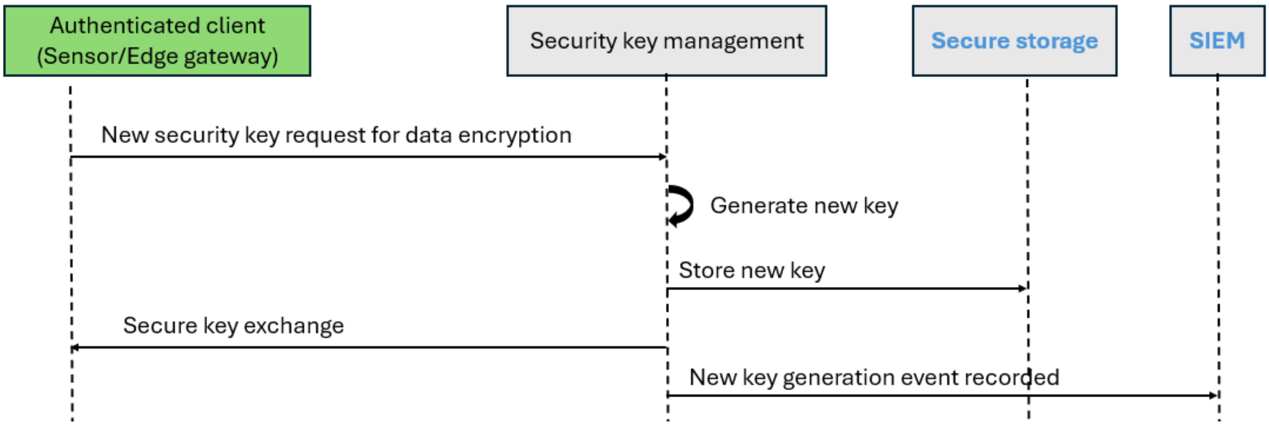

Authenticated client (Sensor/Edge gateway)
Security key management
Secure storage
SIEM
New security key request for data encryption
Generate new key
Store new key
Secure key exchange
New key generation event recorded

Figure 13: Obtaining security key sequence diagram.

3.1.10.4 Data anonymisation

It shows the sequence diagram for anonymising data before sharing with the data consumer. Figure 14 shows the sequence diagram.

The main steps/functionalities are as follows:

- Authenticated client(s) send(s) data required to be anonymised.
- Data anonymisaiton takes place with one of the mentioned techniques and this event is recorded in SIEM.
- Anonymised data is then shared with the data consumer.

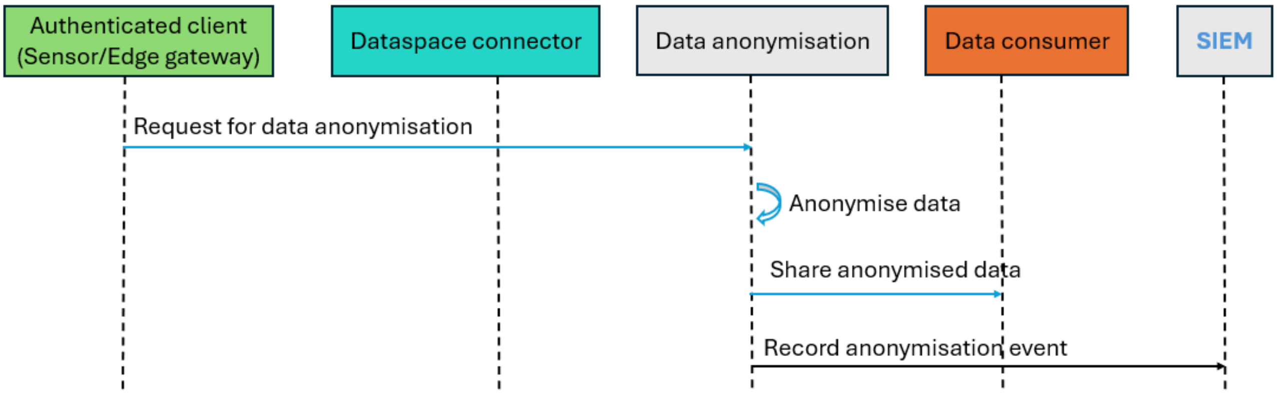

Authenticated client (Sensor/Edge gateway)
Dataspace connector
Data anonymisation
Data consumer
SIEM
Request for data anonymisation
Anonymise data
Share anonymised data
Record anonymisation event

Figure 14: Data anonymisation sequence diagram.

3.1.10.5 Tokenisation

It is dedicated to the tokenisation process of the SEC module and Figure 15 shows the sequence diagram of this process. When a request for tokenisation is obtained, the corresponding building block performs new token generation, its mapping, and token data stored in the secure storage. It is also shared with the data consumer and a corresponding event is recorded in the SIEM.

The main steps/functionalities are as follows:

- Authenticated client(s) send(s) data required to be tokenised.
- New token is generated and then token mapping take place. It is the process of assigning the created token value to its original value.
- In the next step, the original values as well as the related token values (from the mapping process) are stored in the secure storage and the tokenised data is shared with the data consumer.
- The corresponding tokenisation event is recorded in the SIEM.

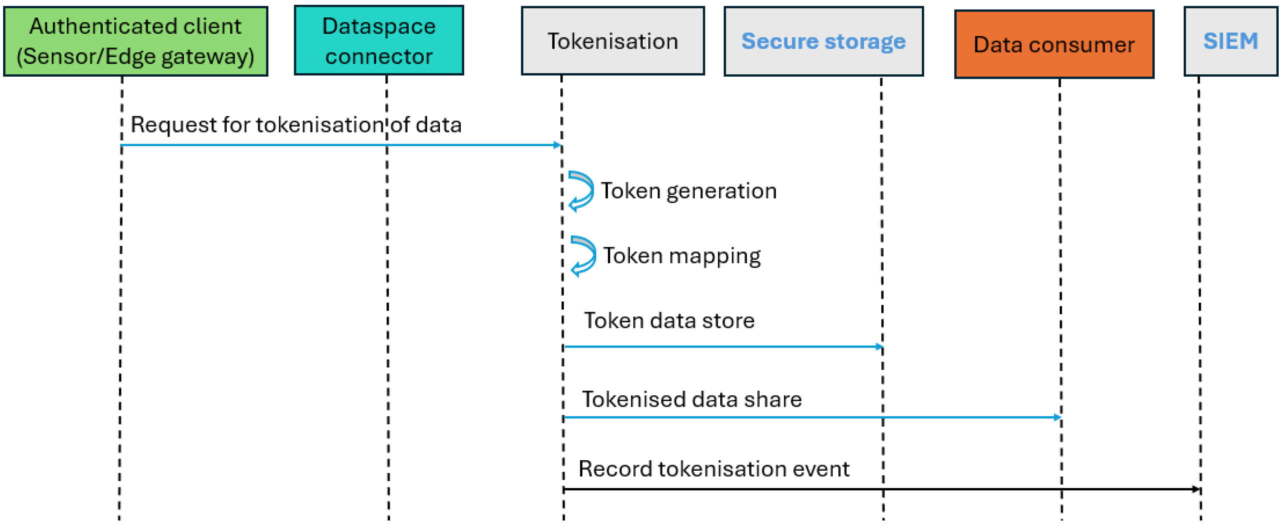


Figure 15: Tokenisation sequence diagram.

3.1.11 Role, Resourcing, and Milestones

Sub-component	Main Activity	M18	M24	M30	M36
Authentication and access control	Implement these sub-components	■			
Audit logging and secure storage	Implement these sub-components	■			
Tokenisation and data anonymisation	Implementation and ensuring their integration to other modules	■			
SIEM	Integration and deployment	■			
Security key management for E2EE	Implementing a key management system	■			
Table Total/DOA Task Total/Resilience					

3.1.12 Open Issues

The following table summarise open issues/uncertainties that need to be resolved during the next stages or implementation.

Issue	Description	Next Steps	Lead or Related Component
N/A			