MA-Proposal-Danesh

Navid Rahimidanesh

March 2023

1 Abstract

2 Introduction

2.1 Motivation

2.2 Thesis Goal

Towards a data sharing platform for "Cyber Threat Intelligence" information focused on the Smart Grid use case, that supports privacy and sovereignty of data to increase the security of smart electricity infrastructure.

Investigate the potentials of data spaces in the context of CTI sharing and its limitations. Design of a data space for CTI sharing. Implementation of a prototype. Evaluation of the prototype.

The suggested platform is referred to as Bazaar from here.

2.3 Outline

3 Background and Related Work

3.1 Background

Smart Grid Security

A smart grid is an advanced electrical grid that uses advanced technologies to efficiently manage the generation, distribution, and consumption of electricity. Smart grid security involves protecting the system from cybersecurity threats that can disrupt or damage the grid's operations. It could be divided into three layers: physical security, network security, and data security. Physical security includes measures to protect the physical infrastructure of the grid, such as substations, transformers, and power lines. This can include fencing, security cameras, and access controls. Network security involves protecting the communication networks used by the smart grid. This can include implementing firewalls, intrusion detection systems, and encryption to prevent unauthorized access or attacks. Data security involves protecting the data generated and used

by the smart grid, including customer data, operational data, and control data. This can include implementing access controls, data encryption, and backup and recovery systems to ensure the availability and integrity of the data.

Wallis et. al. [1] mentions the following severe cyber threats for smart grids and the energy sector in general: data injection attacks on state estimation [2], distributed denial of service (DDoS) and denial of service (DoS) attacks [3], targeted attacks, coordinated attacks, hybrid attacks, and advanced persistent threats [8,9]. Moreover, in recent years, ransomware campaigns have emerged as a significant risk to the sector [10-12].

3.1.1 Threat intelligence Sharing

Cyber threat intelligence (CTI) is the process of collecting, analyzing, and disseminating information about potential or current cyber threats. CTI relies on gathering data from diverse sources, including security tools, threat feeds, honeypots, forums, social media platforms, and other relevant online and offline sources. This data can include indicators of compromise (IOCs), malware samples, network traffic logs, vulnerability information, and more. The goal is to provide organizations with a comprehensive understanding of potential cyber threats to make informed decisions. It helps identify the tactics, techniques, and procedures (TTPs) used by threat actors and vulnerabilities in an organization's security infrastructure. It is an important component of a comprehensive cybersecurity strategy to reduce the risk of a cyberattack. Sharing cyber threat intelligence allows organizations to enhance their situational awareness, proactively defend against potential threats, and improve incident response capabilities. Through collaboration and information exchange between organizations, it leads to a more robust cybersecurity posture for the entire community.

There are several approaches and frameworks for sharing CTI, including commercial and non-commercial platforms. It could include government initiatives as well as open-source communities. Commercial platforms are typically managed by cybersecurity vendors that provide CTI feeds to their customers. Non-commercial platforms include collaborative initiatives among organizations, such as Information Sharing and Analysis Organizations (ISAOs) and Information Sharing and Analysis Centers (ISACs).

Despite the benefits of CTI sharing, there are also gaps and limitations that need to be addressed. These include concerns around privacy, legal and regulatory barriers, lack of trust among participants, and difficulties in sharing information in real-time. In addition, the lack of a standardized format for CTI sharing can make it challenging for organizations to share and use CTI effectively. As such, efforts to standardize CTI sharing formats and improve trust among participants are critical for improving the effectiveness of CTI sharing initiatives.

3.1.2 Obligatory Sharing

The Directive on measures for a high common level of cybersecurity across the Union (the NIS2 Directive) provides legal measures to boost the overall level of cybersecurity in the EU. It ensures EU member states to have a national Computer Security Response Team (CSIRT) that cooperate with each other, and also a culture of information sharing between the public and private sectors in critical sectors. More specifically, organizations that are part of the critical sectors are required to share information about incidents happened to them with the national CSIRT.

Data Modeling

In the context of CTI sharing, data modelling serves three purposes: (1) to provide a backbone for all relevant information, (2) to specify the data input format for further analysis, (3) to define the desired target for information gathering. [4]

Heterogeneous data formats from different incompatible security tools.

Data Spaces

The term data spaces term was first coined by Franklin et al to describe a new paradigm for data management [5]. It solves some data integration tasks by offering a consolidated perspective of data residing in diverse origins, encompassing databases, files, and web services without physically transfer the data. He proposed a DataSpace Suppport Platform (DSSP) that helps developers by enabling them to query and manipulate the data from multiple sources using a single query language with the help of this unified view of data sources.

Beside its technological definition, one could define data spaces from an economic point of view, where data spaces is a form of data exchange. In this viewpoint, dataspaces describes a situation where two or more organizations exchange data to gain a common benefit. [6]

However, there is not a single definition of data spaces. Dataspaces is a concept to fulfill several requirements. In different contexts, different requirements are more important than others.

Data sharing or integration is one requirement. Data spaces could be used to integrate data from different sources. It could also be viewed as a data exchange platform in some contexts.

Another crucial requirement, that makes data spaces interesting, is the sovereignty of data. Sovereignty can generally be defined as supreme authority. In the context of data, it denotes the right of the owner to control how and by whom will the data be used. Data spaces could fulfill this requirement by keeping the data in the data source and providing a unified view of the data to the consumers with respect to the access control policies defined by the owner of the data.

Another aspect of data spaces is its governance. It is required to define a set of policies, rules and protocols to ensure a smooth exchange of data. Therefore,

a governance body is expected to be established to define and enforce these policies. [6]

Data spaces should be open, meaning anyone complying with the policies should be able to join without restriction. This encourages a fair and non-monopolistic market. This entails an easy access, which means, anyone could be able to connect with a limited effort.

Data spaces are usually designed to be decentralized and federated. Meaning there is no entity having direct control over all data exchanges. Different participants could interact with each other directly. This emphasizes the role of interoperability. This is only possible when certain open standards are established. Consequently, data spaces complying to the same standards could be embedded inside each other enabling cross-data-space exchange [6].

"Data Spaces are defined as: A federated, open infrastructure for sovereign data sharing, based on common policies, rules and standards." [6]

3.2 Related Work

EE-ISAC

European Energy Information Sharing and Analysis Centre (EE-ISAC) is a non-profit organization that facilitates the exchange of cyber threat information between its members. Since its foundation in 2015 it acquired over 30 members from utilities, academia, governmental and non-governmental organizations. Members exchange cyber threat information through plenary meetings, working groups, and a dedicated platform (based on MISP). EE-ISAC facilitates trust based information exchange which is not present in the mandatory information sharing in the NIS directive. This trust is achieved by confidentiality agreements and regular physical meetings with the same members. [1]

IDS

The International Data Spaces (IDS) is an initiative with the goal of creating a standard for a distributed software architecture for data exchange with sovereignty. It was launched in 2015 as a Fraunhofer research project funded by the German Federal Ministry for Education and Research [7]. Shortly after that, in 2016, the IDS Association (IDSA) was founded as a non-profit organization to continue the research. It resulted in definition of the IDS Reference Architecture Model (IDS RAM). The IDS RAM is the description of IDS components and their interactions without being technology specific [7]. IDS RAM allows anyone to implement the IDS compliant components using any technology. The IDSA also provides a reference implementation of different IDS components called IDS Testbed [8].

IDS RAM defines the following components [9]: - Connector: The connector is the interface between the IDS ecosystem and the data source. It is responsible for the data exchange and the enforcement of the usage control policies

as well as authentication. - Identity Provider: authentication service managing identity information - IDS Broker: Manages the metadata (description and usage policies) - Clearing House: Audits the data exchange and manages the payment - IDS Apps: Process the exchanged data. Deployed within Connector. - App Store: Provides IDS apps - Vocabulary Provider: Offers vocabularies to describe and annotate data

Furthermore, the participants could undertake different roles [9]: - Data Owner: Controls the data and defines usage policies and payment model. - Data Provider: Provides the data to the IDS ecosystem with respect to the policies defined by the data owner. - Data Consumer: Same as data provider but consumes the data. - Data User: Same as data owner but uses the data. - App Provider - ...

Usage Control:

Certification:

Gaia-X

European data strategy Towards a single market of data. + technological independence of Europe. No vendor lock-in. Launched in 2019. Sovereign digital ecosystem Gaia-X is an initiative that aims to foster generation of a data and service infrastructure by developing regulations and technical specifications which is based on European values, applicable to any existing cloud and edge technology stack. Gaia-X allows for transparency, controllability, portability and interoperability across data and services. It will ease value creation through data collection and sharing between organizations leading to a vibrant data ecosystem across Europe and beyond. Gaia-X Association deliverables include federation services, common policy rules and an architecture of standards. (Data sovereignty?) Federation services could be utilized by the ecosystem participants to achieve a global interoperability, compliance and effortless set up. This includes, "Identity and Trust", "Federated Catalog" and "Data Exchange services". [10] AISBL: Gaia-X Association Nodes, Services (Any cloud service), Service Instances (A service running on a node) and Data Assets (a data set on a node) Participants: Organizations and individuals that are part of the Gaia-X ecosystem Clearing House: Middle man in the exchange checking for compliance

Comparison of IDS and Gaia-X

Gaia X 2019, IDS 2015. IDS is more mature. it is already tested in the industry. Gaia X is still in the development phase. Gaia X offers a more holistic approach including cloud elements. Gaia X is more focused on cloud services. IDS is more focused on data exchange. Gaia X is more focused on the business and economic side. IDS is more focused on the technical side. Gaia-X could use IDS as a component [10] Gaia-X provides standards for infrastructures and cloud elements. IDS provides standards for data exchange. Federated Catalog IDS Broker + Vocabulary Provider + Information Model Identity and Trust IDS Identity Provider + Dynamic Attribute Provisioning Service (DAPS) Gaia X

Sovereign Data Exchange IDS Usage Control and Clearing House Gaia X Node IDS Connector

- Documentation: - Maturity: Gaia-X is still in the development phase. IDS is more mature. - Implementation: IDS is open source and has a reference implementation. Gaia-X is not open source and does not have a reference implementation.

Other Platforms

Platform Industrie 4.0, SWIPO: is an association that develops and safeguards codes of conduct to facilitate Switching and Porting of non-personal data between cloud providers and their customers. It follows the Free Flow of Non-Personal Data regulation.

Crowd Sourced Sharing

The sharing is caring paper [11]

4 Use case and Requirements

In this section, we will describe the use case and the requirements of the proposed platform. We will also discuss the constraints and limitations of the platform.

4.1 Overview

Bazaar can be utilized in several areas. To generalize, any set of organizations that work together to achieve a common goal who fundamentally use IT systems in their operations could benefit from Bazaar. Bazaar can accelerate the process of establishing new sector ISACs and facilitate the activities of existing ones. As mentioned in the **??**, ISACs' common use cases are critical infrastructures, including, electricity, water, finance, transportation, and healthcare.

To elicit some concrete scenarios where Bazaar can be used, we should narrow our focus on a specific use case. In this work, we will focus on the smart grid use case. Consider a situation where there are several participants representing different organizations active in the energy supply chain. The threat landscape they are exposed to is similar, because they are targeted by similar adversaries, and by using same technologies, they share the same vulnerabilities. Therefore, they can benefit from sharing CTI with each other.

A possible threat actor could be an advanced persistent threat (APT) supported by an enemy government, or a gang of experienced cyber criminals intended to disrupt the energy supply by attacking different actors in the supply chain. By doing so, they can reach their goal of causing a blackout, exfiltrating sensitive information, or gaining financial benefits (e.g. ransomware). To this end, they probably use some limited set of tactics, techniques, and procedures (TTP) to attack their victims over and over again. As a result, sharing these

TTPs among the participants can help them to be better prepared against these attacks.

Moreover, the participants in the energy sector are required to share information about incidents happened to them with each other and with the national CSIRT. This is required by the NIS2 directive. Therefore, they can use Bazaar to share the required information with each other.

Here, we assume that participants are the security team of the aforementioned organizations, or a managed security service if the organization does not have its own security team. That is because the security team is the one that is responsible for handling the CTI, and we do not consider other types of security information (e.g. logs) in this work.

To maximize the benefits of information sharing, we extend the possible users of Bazaar to include organizations from different countries adhering to different regulations. Example of a challenge that might arise is the GDPR. It limits the transfer of personal data outside the EU. On top of that, the language and cultural differences that might exist between the participants which will further complicate the trust building and information sharing process. As a result, we are dealing with heterogeneous organizations and different levels of trust among the participants, which the platform should be able to handle.

A use case diagram is shown in figure 1.

4.2 Scenarios

4.2.1 Scenario 1: Peer to Peer Sharing

A utility provider (Acme) has found a new malware in their network. They share the related IOC inside Bazaar. However, they set the usage policy to only allow the data to be processed in the Bazaar Connectors located within the EU (to comply with GDPR). They also set the usage policy to only allow the data to be read by the participants that have a certain minimum level on a certain trust metric. Furthermore, Acme does not want to share with competing companies. Therefore, it blacklists identified participants that compete in the same market.

Scenario 2: Paid Service and Rating

A security organization specialized in selling CTI feeds, wants to use Bazaar to sell its services. It has a set of CTI feeds that it wants to share with its customers. It wants to charge its customers based on the number of IOCs they receive. Customers are also invited to rate the quality of the feeds they receive. This ratings are used by the customers to choose between different feeds.

Scenario 3: Initiation and Explore

Another organization (Organization B) after having heard of the platform finds it useful. To join the platform, B should pass the necessary certification process. First, a committee of decision-making community members should accept the

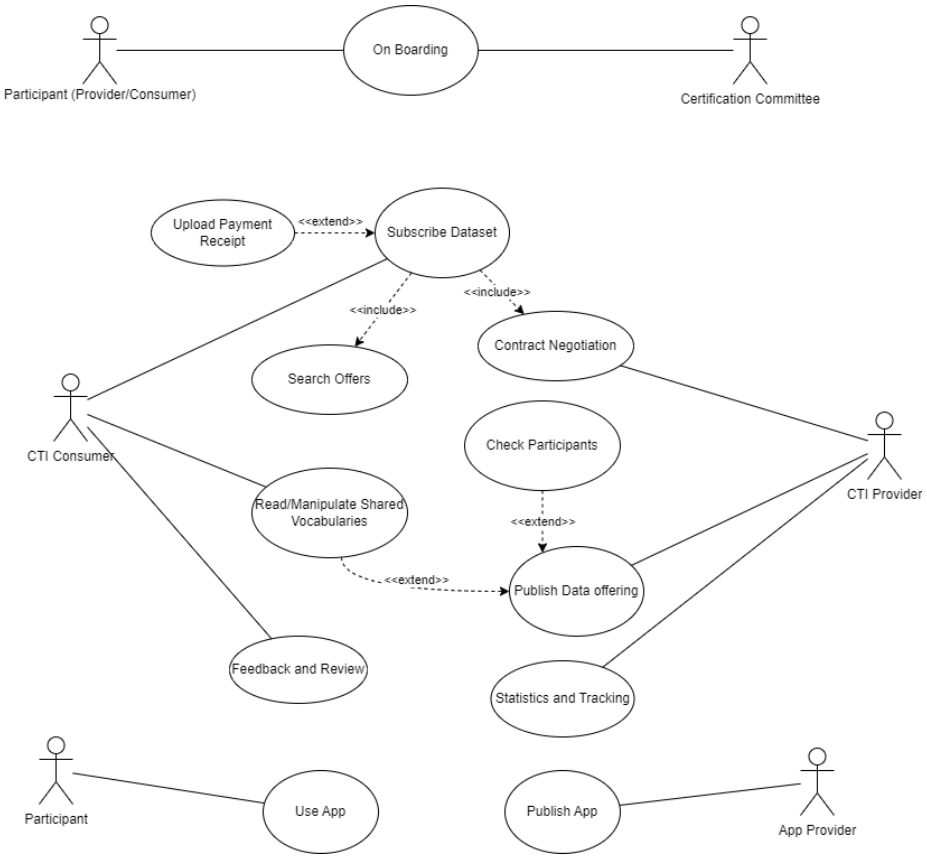

Figure 1: Use Case Diagram

request. Then, after technical certification by the platform admins, B is allowed to join.

At this point, B should be able to connect to the platform. It should be able to search for available data assets, and fetch metadata. It should explore the existing members and their reputation.

5 Conceptual Approach

Components and relation to requirements Discuss Benefits and limitations Compare with existing solutions

Important data space components that are relevant to our use case are listed below. The descriptions fully conform to the IDSA RAM 4.0 [12].

- Connector: It is the primary component involved in the data exchange acting either as data provider or consumer. Not only it performs the ac-

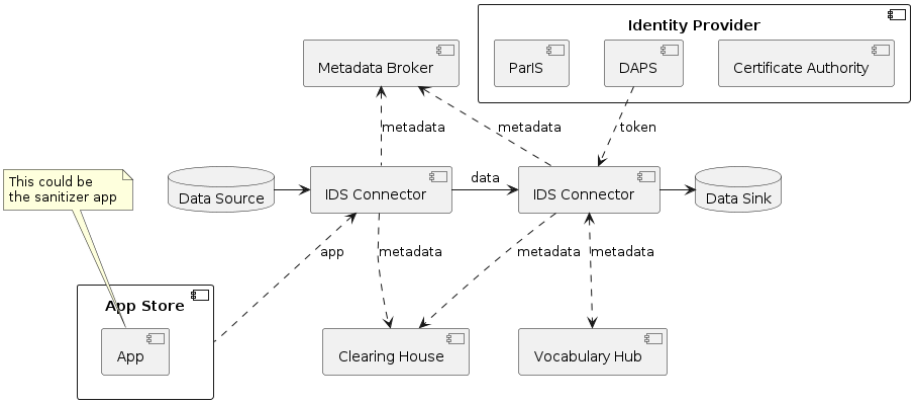

Figure 2: System Architecture Diagram

tual data exchange but also the enforcement of the usage control policies as well as authentication. It can be operated on-premises or in a cloud environment. It will run the IDS Apps that do process the data among other things (more on that later). It uses application container management technology to isolate data apps and core functionalities.

- IDS Broker: The IDS Metadata Broker is an IDS Connector, which contains an endpoint for the registration, publication, maintenance, and query of Self-Descriptions. Self-Descriptions encapsulate information about Connectors, their managing participant, the offered data assets, and the respective usage conditions. In a sense, the IDS Broker is like a phonebook.
- Identity Provider: This is the component serving as Identity and Access Management (IAM). It's responsible for assigning identities to participants, verifying identity claims and granting access based on identities. It comprises three components:
  - CAs: One or multiple certificate authorities are responsible for issuing certificates to participants upon request. They are also responsible for revoking certificates. They are the trust anchors by which all other components can be verified.
  - Dynamic Attribute Provisioning Service (DAPS): It complements the certificates issued by CAs with more volatile attributes in form of tokens. Connectors can request Dynamic Attribute Tokens (DATs) from DAPS to prove their attributes to other components.
  - Participant Information Service (ParIS): It provides business-related information about participants in the IDS that have been checked by the Support Organization. Similar to the way Broker provides metadata about data assets, ParIS provides metadata about participants.

- Clearing House: It is a trusted third party in the data exchange that logs all the required information about clearing, billing, and usage control. It keeps track of the payment information and also usage information to help verify the compliance with the usage policies.
- IDS Apps: These are re-usable software components that can be deployed inside the IDS Connector. They are mainly used to transform or analyze data. However, they can also be used to connect to enterprise services, or to allow the connector to be controlled by external systems. Data apps can be chained and bundled.
- App Store: As the name suggests, it is a marketplace for IDS Apps. It contains endpoints to publish, search, and download IDS Apps. It is a Connector on its own, so it should pass the IDS certification criteria and provide a self-description.
- Vocabulary Provider: To facilitate cooperation of different IDS components, a common vocabulary is required. The vocabulary provider enables the participants to define and publish their own vocabularies. Vocabularies typically follow the RDF standard. An example usage is to reference an RDF URI in the Self-Description of a data asset.

Apart from the components described by the IDS RAM, there are other components that are not part of the IDS RAM but are relevant to our use case. These components are described below.

- Sanitization App: This is an IDS App that is responsible for sanitizing the data. It is deployed inside the IDS Connector. It is responsible for removing the sensitive information from the data before it is shared with other participants. IDS App is suitable for this task because dealing with sensitive data requires a high level of trust. Apps are certified and verified by the App Store. Furthermore, they are executed within the Connector, so they have access to the data before it is shared with other participants. This is important because it prevents the data from being leaked before it is sanitized.

6 Realization / Implementation

IDS Connector

The IDS Association publishes a monthly report of the current state of all the data connectors used for exchange of data, not limited to the IDS compliant connectors. Dam et al. [13] investigated this report and published a survey in September 2023. They found that only 4 connectors have their source code available on a public repository: 1) IDS Dataspace Connector (DSC) by sovity, Eclipse Dataspace Connector (EDC), the TRUsted Engineering (TRUE) Connector, and the Trusted Connector by Fraunhofer AISEC.

In addition to that, I found two more: First, IDS Integration Toolbox by Open Logistics Foundation which is a wrapper around the DSC. Second, TNO Security Gateway (TSG) initially developed by TNO which has implementations for many IDS components. It is used in Smart Connected Supplier Network (SCSN) data space and has a documentation. However, it has no stars on gitlab.

The overview of different conenctors is shown in table 6.

Name	Created	Stars	Commits	Last Release	Hosted
DSC	07.10.2020	27[+101]	2600	10.22	Github
EDC	13.01.2021	202	1817	10.23	Github
TRUE	30.10.2020	19	122	08.23	Github
Trusted	05.09.2017	43	2221	02.23	Github
Toolbox	31.03.2022	3	172	04.23	Self-Hosted
TSG	12.05.2021	0	243	08.23	Gitlab

Table 1: Available IDS Connectors

Most number of stars and most recent release being a deciding factor, I will choose EDC to base my implementation on.

IDS Testbed

The IDS association defines Minimum Viable Data Space (MVDS) as the minimum set of components that provide the ability to do secure and sovereign data exchange. They specify the required components as follows: Two Connectors (a data provider and a consumer), an Identity Provider (Dynamic Attribute Provisioning Service, Certificate Authority). IDSA has published an open-source project, IDS Testbed, that contains instructions to install and orchestrate these set of minimum components. It references open-source implementations of these components, see table 6 to find source code of these components.

Component	Source Code	Version	Language
IDS Testbed	Testbed Git	1.0	Docker-Compose
Connector	Connector Git	8.0.2	Java
Metadata Broker	Broker Git	5.0.3	Java
DAPS	DAPS Git	1.6.0	Ruby
Certificate Authority	Testbed Git	—	Python
App Store	App Store Git	3.0.0	Java

Table 2: Necessary Components

In addition to the aforementioned components, some new ones need to be implemented from scratch: App Store, ParIS, Clearing House, Vocabulary Hub. Of course, the required IDS Apps should be implemented, which includes, the Sanitization App. For the App Store there is a published project (App Store Git) which is not included in the testbed and might not be fully functional.

Usage Policies

To describe usage policies IDS defines its own Usage Control Language which is an extension of Open Digital Rights Language (ODRL). It is a machine readable format which is technology agnostic. There are multiple mechanism to enforce these policies automatically. The position paper on IDS Usage Control [14] lists the following mechanisms: MYDATA Control Technologies, Logic based Usage Control (LUCON), and Degree (D°). According to the paper, MYDATA Control Technologies is the most mature and comprehensive. It is also the only one that is implemented in the IDS Testbed. Therefore, I will use MYDATA Control Technologies to enforce the usage policies. The usage control is enforced in the IDS Connector possibly using a separate application container. It comprises three different components: PMP, manages policies and creating them based on some templates, PDP, evaluates the policies and decides whether to allow or deny the request, and PEP, enforces the decision made by PDP. The overview of the components is shown in figure 3.

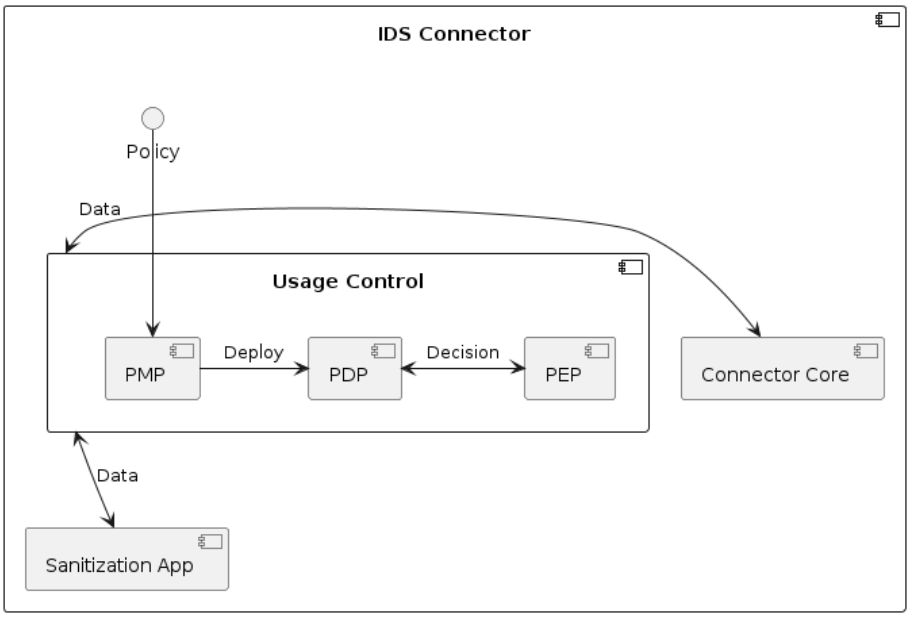

Figure 3: Connector Overview

Sanitization App

Removing all sensitive data from CTI in general can be a difficult task. There are different CTI formats and each format has its own structure. Furthermore, deciding whether a value contains confidential information or not is not straightforward. Even a sophisticated machine learning approach requires a lot of data

with different types of confidential information to be trained on. Therefore, I will focus on creating a base sanitization app that is easily extensible to detect more sensitive data. To start, it should support sanitizing JSON based CTI formats, and a configurable set of regex rules to detect sensitive data. Refining this app to support more formats and more sophisticated detection mechanisms is out of scope of this thesis.

7 Evaluation

My contribution being design of a sharing platform for CTI data, the evaluation should measure the effectiveness of the sharing platform. The problem is that according to my research, there is no benchmark available and no defined metrics to this aim. Having no baseline to compare against, it seems hard to reach an objective evaluation. Furthermore, many performance metrics depend on the implementation, infrastructure, and the details of the scenario. Since, my main contribution is the design of the platform, not the implementation, some metrics can be misleading.

However, several options can be thought of. First, by focusing on the main effect of using data spaces for data sharing, which is the addition of usage control and changing the data flow, one can define a few specific metrics. Having defined some usage case scenarios and a sample implementation, we can measure for example the following metrics:

- Number of unnecessary participants that should have access to the data.
- The difficulty of changing the usage policies.
- How difficult it is to revoke access to the data.
- Variety of types of usage policies that are enforceable.
- How many scenarios are significantly improved using our approach.

The above-mentioned items might be subjective and change by modifying the chosen scenarios, but investigating it can provide some insights into the effectiveness of the data spaces.

Another approach which can sound more objective is to design a survey. It should use some standard templates to be comparable. The survey should be conducted on a group of experts in the field of CTI sharing after presenting them our approach. The survey should be designed to measure the effectiveness of our design in terms of privacy and security. Ideally, 10 to 20 experts should be interviewed. There is a risk of not having enough experts available. In that case, we fall back to the first approach.

8 Timeline / Milestones / Project Plan

References

[1] T. Wallis and R. Leszczyna, "EE-ISAC—practical cybersecurity solution for the energy sector," vol. 15, no. 6, p. 2170.

[2] R. Deng, P. Zhuang, and H. Liang, "False data injection attacks against state estimation in power distribution systems," vol. 10, no. 3, pp. 2871–2881.

[3] Q. Wang, W. Tai, Y. Tang, H. Zhu, M. Zhang, and D. Zhou, "Coordinated defense of distributed denial of service attacks against the multi-area load frequency control services," vol. 12, no. 13, p. 2493. Number: 13 Publisher: Multidisciplinary Digital Publishing Institute.

[4] M. Husák, L. Sadlek, S. Špaček, M. Laštovička, M. Javorník, and J. Komárková, "CRUSOE: A toolset for cyber situational awareness and decision support in incident handling," vol. 115, p. 102609.

[5] M. Franklin, A. Halevy, and D. Maier, "From databases to dataspaces: a new abstraction for information management," vol. 34, no. 4, pp. 27–33.

[6] Reiberg, A. a. Niebel, and Crispin, "What is a data space?."

[7] B. Otto, "The evolution of data spaces," in *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage* (B. Otto, M. ten Hompel, and S. Wrobel, eds.), pp. 3–15, Springer International Publishing.

[8] "IDS reference testbed."

[9] H. Pettenpohl, M. Spiekermann, and J. R. Both, "International data spaces in a nutshell," in *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage* (B. Otto, M. ten Hompel, and S. Wrobel, eds.), pp. 29–40, Springer International Publishing.

[10] H. Tardieu, "Role of gaia-x in the european data space ecosystem," in *Designing Data Spaces* (B. Otto, M. Ten Hompel, and S. Wrobel, eds.), pp. 41–59, Springer International Publishing.

[11] V. Jesus, B. Bains, and V. Chang, "Sharing is caring: Hurdles and prospects of open, crowd-sourced cyber threat intelligence," pp. 1–20.

[12] B. Otto, S. Steinbuss, A. Teuscher, and S. Lohmann, "IDS reference architecture model." Version Number: Version 3.0.

[13] T. Dam, L. D. Klausner, S. Neumaier, and T. Priebe, "A survey of dataspace connector implementations."

[14] A. Eitel, C. Jung, R. Brandstädter, A. Hosseinzadeh, S. Bader, C. Kühnle, P. Birnstill, G. Brost, Gall, F. Bruckner, N. Weißenberg, and B. Korth, "Usage control in the international data spaces." Version Number: 3.0.