Extending Actor Models in Data Spaces

Hendrik Meyer zum Felde
Fraunhofer AISEC
Garching near Munich, Germany
meyerzum@aisec.fraunhofer.de

Thomas Bellebaum
Fraunhofer AISEC
Garching near Munich, Germany
thomas.bellebaum@aisec.fraunhofer.de

Gerd Brost
Fraunhofer AISEC
Garching near Munich, Germany
gerd.brost@aisec.fraunhofer.de

Maarten Kollenstart
TNO
Groningen, Netherlands
maarten.kollenstart@tno.nl

Simon Dalmolen
TNO
Groningen, Netherlands
simon.dalmolen@tno.nl

ABSTRACT

In today's internet almost any party can share sets of data with each other. However, creating frameworks and regulated realms for the sharing of data is very complex when multiple parties are involved and complicated regulation comes into play. As solution data spaces were introduced to enable participating parties to share data among themselves in an organized, regulated and standardized way. However, contract data processors, acting as data space participants, are currently unable to execute data requests on behalf of their contract partners. Here we show that an on-behalf-of actor model can be easily added to existing data spaces. We demonstrate how this extension can be realized using verifiable credentials. We provide a sample use case, a detailed sequence diagram and discuss necessary architectural adaptations and additions to established protocols. Using the extensions explained in this work numerous real life use cases which previously could technically not be realized can now be covered. This enables future data spaces to provide more dynamic and complex real world use cases.

CCS CONCEPTS

• **Security and privacy** → **Formal security models**; **Security requirements**; **Logic and verification**; *Trust frameworks*; **Digital rights management**; **Authentication**; *Authorization.*

KEYWORDS

Data Spaces, Contract Data Processing, Actor Models, Self-Sovereign Identities, International Data Spaces, Gaia-X, On-behalf-of Model

ACM Reference Format:
Hendrik Meyer zum Felde, Thomas Bellebaum, Gerd Brost, Maarten Kollenstart, and Simon Dalmolen. 2023. Extending Actor Models in Data Spaces. In *Companion Proceedings of the ACM Web Conference 2023 (WWW '23 Companion), April 30–May 04, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3543873.3587645

WWW '23 Companion, April 30–May 04, 2023, Austin, TX, USA

ACM ISBN 978-1-4503-9419-2/23/04.
https://doi.org/10.1145/3543873.3587645

1 INTRODUCTION

Sharing data is one of the key principles to establish distributed data value chains and collaborative data processing use cases. This is true for companies, state departments and individual actors. Dynamic relationships where individual participants take varying roles such as data provider, data processing entity or data consumer create so-called data spaces. There are numerous ways how data can be shared and processed. Standards and interoperability agreements for protocols, data formats and usage agreements are necessary.

Two of the most known and influential initiatives which aim to provide such standardization are Gaia-X [6] and the International Data Spaces (IDS) [2]. These initiatives have provided baseline technology to share data and aim to achieve true data sovereignty. These technologies include among other concepts for Remote Attestation and Usage Control, which aims to keep control of data, even if the data has been sent to remote peers [5]. Furthermore, participants are also able to decide where and under what conditions to exchange data in the first place.

This allows for direct data flows between parties which have existing data exchange contracts. However, an underlying role model for delegating privileges has not been established in Gaia-X nor IDS. Here, only the directly involved gateways and services possess technical identities and act on their own behalf. In practice, numerous use cases require delegated privileges. Examples include an employee speaking on behalf of a company, a state agency issuing proof of an actor's citizenship or a software component processing data on a previously agreed upon basis for another entity. The latter case may arise due to contract data processing agreements. Here, the processing of contract data can't be realized without an actor model which provides an *on-behalf-of privilege delegation.*

In this work we provide an extension to the existing actor models of Gaia-X and IDS to enable a delegation of data processing rights to entities operating a data space connector in an on-behalf-of fashion. We contribute an analysis of the architectural requirements, technical details for the realization, necessary protocol adaptations and required organizational preparations.

2 BACKGROUND AND RELATED WORK

This section provides related initiatives and basic knowledge on verifiable credentials required to understand the proposed concepts.

2.1 Gaia-X

Gaia-X is an European initiative to create a cloud resource federation model [6]. A long-term goal is to create a data infrastructure

that allows for greater control, security, and interoperability of data within the European Union. It aims to build a federation of data infrastructures that can work together seamlessly, and to promote the use of open standards and open-source software in the development of these infrastructures. The initiative is driven by a group of companies, organizations, and governments from across Europe [8].

2.2 International Data Spaces

The IDS are an initiative that aims to create a global framework for data infrastructures [2]. This framework allows for secure and controlled sharing of data between organizations and individuals. The initiative is driven by a consortium of companies, research institutions, and government agencies from various countries. The IDS aim to establish a set of common standards and protocols for data sharing, and to build a network of data spaces that can interoperate with each other. The goal of the IDS is to enable data-driven innovation and collaboration, while also ensuring data privacy and security.

The main difference between the two initiatives is the scope. In Gaia-X, the scope is the full stack of cloud resources up to the data exchange, while in IDS the scope is primarily the data exchange itself. This results in a situation where the two initiatives can complement each other [1].

2.3 Verifiable Credentials in Identity and Access Management

Verifiable Credential (VC) is a term which was coined by Self-Sovereign Identity (SSI) system developers. It broadly refers to any kind of cryptographically signed set of claims about a subject, which serve as "credential". In recent times the term refers to a JSON-LD based specification developed at the W3C [13]. In an SSI context, the owner of such a credential is typically identified by some identifier resolvable into a cryptographic signing key, which may then be used in a suitable identification scheme to prove their presence within a communication. A format combining such an identification scheme with a set of credentials is called a Verifiable Presentation (VP) [13].

Another form of VC and VP can be found in machine-to-machine communication in the form of X.509 Certificates [4] and the TLS [12] handshake signature respectively. While these also define a communicating identity, they are additionally able to establish a private communication channel. If over this channel further identification of one of the communication partners is performed and the security guarantees of the communication channel shall be conserved, these identification schemes need to be cryptographically bound to the channel.

3 DESIGN

This section provides details on the main context of the proposed advancement, required changes in architecture and protocol and a sequence diagram dealing with a sample use case to explain technical specifics.

3.1 Context Overview

An overview of the general context where the actor model extension of this work applies is given in Figure 1. Here, two companies

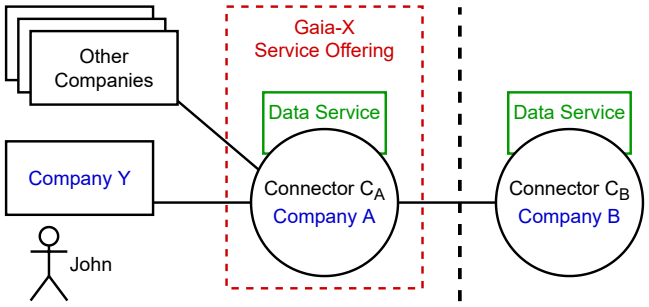

Figure 1: Architecture and context overview of the actor model extension aimed for with connectors shown in circles and involved companies shown.

denoted Y and B are trying to exchange data. Company B is able to operate their own connector while company Y is delegating this task to a subcontractor, company A. A strong security and access barrier between Company A and Company B is given which is indicated by a vertical dashed line. Connector C_A and C_B are able to connect to each other via the internet and are running services to process data.

The assumption is, an authorized employee of Company Y, named John, wants to initiate a data transfer from Connector C_B to Connector C_A. Since Company B does not have an agreement with A directly, it will be unwilling to send the data directly to A unless it knows that C_A is serving the uses of Company Y. Note that in our scenario, C_A could also be serving other companies.

This use case requires two entities to prove that they are acting on behalf of another entity. Concretely, C_A is acting on behalf of Company Y via a contract data processing agreement, and John, being an employee of Y, is also acting on behalf of that company.

To comply with both naming conventions of IDS and Gaia-X the Gaia-X Service Offering is included in the overview, denoted in a red, dashed box. According to Gaia-X a connector together with a service makes up a Service Offering (SO) [7]. The difference in the definition between IDS and Gaia-X at this point is the fact that in IDS the technical specification is defined using one connector as main technical entity. One IDS connector with only one set of authentication key material may provide multiple services or connections. Whereas in Gaia-X the corresponding entity is a set of SOs, which may each individually serve as endpoints for connections, and each authenticate themselves individually.

Please note that this difference has no impact on our sample case as it provides only one service per connector. However, with more complex cases and multiple services run by one connector, extending the actor models using the definition of the corresponding SOs may require individual workflows for each SO and not just one for the main connector in the Gaia-X perspective.

3.2 Architectural Requirements

In general, any entity inside of a data space can make a statement about other entities and issue self-signed proofs, however, without an organization of underlying trust chains the proofs would provide no benefit. In such cases each component must always have access to the verification mechanisms of these services. For better

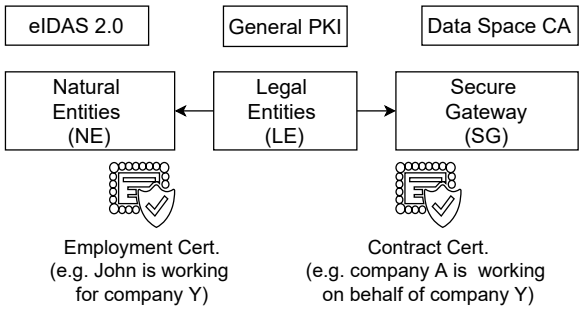

Figure 2: An overview of authentication details showing natural entities, legal entities and secure gateways and their corresponding proofs.

understanding, this section details the involved proofs and options for their verification as shown in Figure 2. The actor John, which is a Natural Entity (NE), is identified using the standard eIDAS 2.0. The company Y, which is a Legal Entity (LE), issues an Employment Certificate using classical Public Key Infrastructure (PKI). Additionally, company Y issues a contract certificate to allow company A to act on behalf of Y, also using PKI. Connector C_A serves as Secure Gateway (SG) and must authenticate itself using the TLS and its corresponding certificate issued by the Data Space CA.

The architecture of the data space at hand must provide means of authentication for every participant. Usually, this is done via a Certificate Authority. For instance, for the authentication of a NE, such as a person like John from our example, eIDAS 2.0 [3] and SSI for actors can be applied and allow for a verification. Any relation concerning a LE which was pinned down using certificates must be checked using the corresponding PKI.

The authentication of a SG in a data space is usually verified via evaluation of a token issued by a central Trusted Third Party (TTP) for the data space. During the TLS handshake of connectors, the token binds the session to the authenticated connector and thereby establishes the link between SG and authenticated party inside of a data space. In IDS ontology, this is achieved using a dynamic attribute token for authentication, called DAPS token. A simple technical solution would be to bind the transportcert's sha256 hash to the session during connection establishment to further prevent MitM-attacks.

3.3 Data Exchange Protocol

This section details a sequence of steps involved for realizing the previously explained on-behalf-of use case and provides a sequence diagram in Figure 3.

In the beginning Company Y desires to receive data from Company B delivered via Company A's Connector C_A. At this point an existing agreement to share date has already been established between Company Y and B. However, neither John's relation to the employing Company Y is known to Company A or B nor John's identity. Therefore, as first phase, John sends a Data Request message to Connector C_A with the destination of Company B's Connector C_B. Connector C_A needs to check whether this request comes from an authorized participant. For this reason, John sends along (1) his own employment credential to C_A and (2) a short-lived proof of his identity, which could be a VP. These proofs must be checked at Connector C_A.

As second phase, Connector C_A needs to forward the Data Request to Connector C_B. Similarly, to the previous step, Connector C_B is also unaware of John's authentication and relation to Company Y and Connector C_A's permission to act on behalf of Company Y and requires these proofs. Additionally, Connector C_B must authenticate Connector C_A as no previous agreements were established, yet. Therefore, Connector C_A sends the following proofs which are checked at Connector C_B. (1) John's authentication, (2) John's employment at Y. (3) A's authorization to act on behalf of Y, (4) Company A's and Connector C_A's authentication via TLS certificate. Additionally, Company B needs to check the status of its internal policy setting and the existing agreement with company Y, whether the requested data may be sent. For this sample case we assume that the policy states that employees of company Y may request a data transfer.

As third phase, company B's Connector C_B sends the requested data and a corresponding policy how the data may be processed if all checks have been passed correctly. Since John orchestrated the Data Request on behalf of Company Y, a policy which allows to process the data only for employees of Company Y is attached by Company B. Otherwise, Connector C_A would have permission to leak the data elsewhere and the delegated acting would be pointless as other companies are also interacting with Company A's Connector C_A. Company A's Connector C_A receives the data, deploys the corresponding policy and the processing of data as requested by John may begin. In Figure 3 error handling and restart routines etc. are left out of scope for simplified understanding.

4 DISCUSSION

The provided protocol description is very tailored to a particular use case, but may be generalized to a variety of use cases. We discuss a range of points to be aware of when doing so in this section. In general, the proposed protocol succeeds to give C_A access to resources accessible to company Y, provided Y can properly identify itself and company B does not require data to always be sent to company Y directly. One critical concern of the protocol is thus trust in company A by all of John, Y and B.

4.1 Reducing Required Trust

Within the IDS, several methods for reducing this need for trust in a third party by restricting the possibilities of C_A to misbehave have been further developed, such as enforced Usage Control [9] and Remote Attestation both statically at boot and dynamically during runtime. For the latter current research is done focusing on the protection of control flow integrity [11] or architectures which allow an attestation of functions as a service during runtime [10]. These techniques may further assist C_B in deciding whether to send any data and to understand how the data may be used there.

Trust in company A by company Y can also be reduced at certain points. When data requests are forwarded by C_A, it can change the requests. Misbehavior at this point can be limited by explicitly specifying the resources which company A should have access to in the contract data processing agreement certificate. If this is

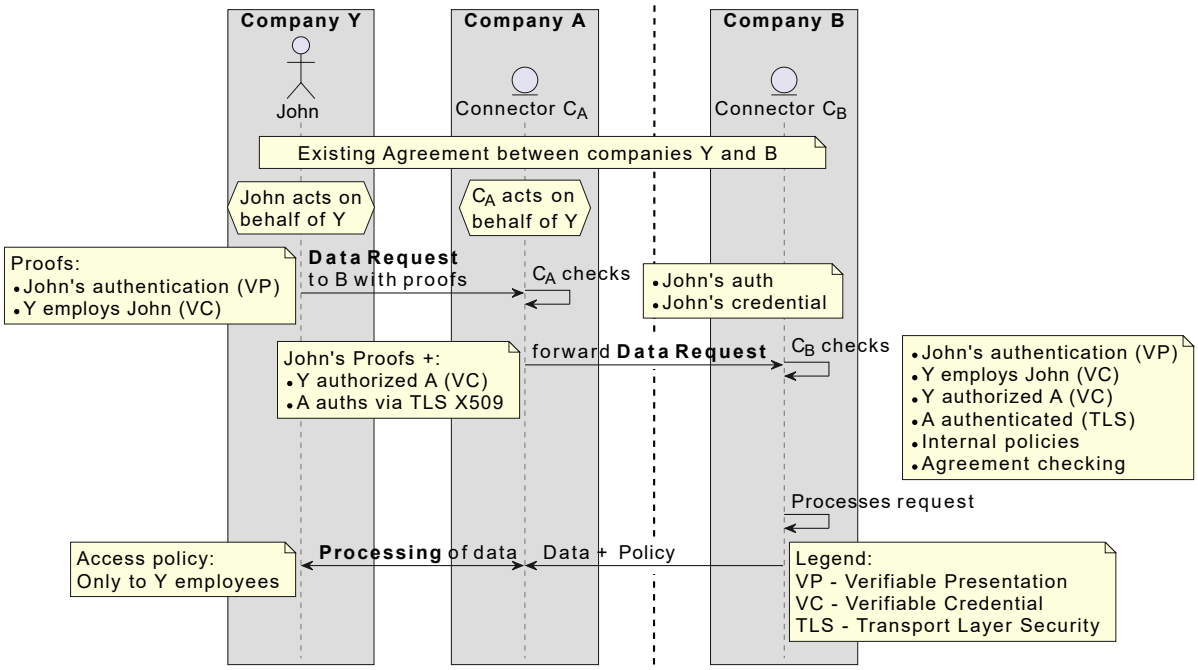

Figure 3: Sequence diagram of involved steps for acting on behalf of other entities.

impractical or undesirable, an alternative would be to have John sign their request, which allows company B to verify the original request originated from John and has not been tampered with.

4.2 Preventing Replay-Attacks

Another concern may be replay of requests by A to B. If request signing is employed, this can be countered by including the addressee of the request, as well as the current timestamp and a per-recipient nonce in the signature. It is best practice to have the nonces be chosen by each recipient, which would however require additional round trips.

4.3 Identity Data Protection

The protocol furthermore assumes that it is vital for company B to identify John as an actor directly. However, this has an impact for personal data protection. Mitigations include the use of a company-Y-wide API for data requests, which would allow for fine-grained access control by company Y, as well as zero-knowledge proof and selective disclosure techniques to attest only the relevant relations without revealing personally identifiable information.

As mentioned before, when a secure channel is to be established between participants (such as TLS), the identities established in the creation of such channels are authoritative for determining whom the data is being sent to in a first hop. Therefore, any policy decisions will need to be made for this identity. If attestations on a higher protocol level (e.g., VC) are used, the identities established there may differ from the channel identities and need to be bound explicitly if this is not desired. Note that such a binding can make a dedicated identification scheme superfluous, which is why in the presented form no VC was used for A or C_A.

4.4 Landscape of Standardization

The landscape of related initiatives, Gaia-X, IDS, and eIDAS, is rapidly developing. Therefore, the protocol must continuously be validated against advancements in these initiatives. For example, the trust framework together with the Service Definitions in Gaia-X are not in stable version yet, resulting in the fact that the protocol might need to be updated or provides input to Gaia-X to ensure compatibility. The same holds true for eIDAS, where the first implementation projects are granted, of which the results might enrich the protocol or requires adaptations. For IDS, currently there is no specification for the relations as presented in this paper. Therefore, the ideas in this paper can contribute to new versions of the specifications of IDS.

4.5 Anything On-Behalf-of Anything

The use case presented in this work so far was focused on a NE and a connector acting on behalf of a LE. However, in practice of course a lot more combinations for entities acting on behalf of others are at hand. For instance, a connector on behalf of a NE, or a LE acting on behalf of NE. For full flexibility in contract data processing, it is clear that anything should be able to act on behalf of anything else.

However, the following basic rules can be applied to structure and standardize possible solutions. First, whenever a NE authenticates itself, eIDAS 2.0 should be used. Second, whenever a computer system, such as a connector authenticates itself TLS PKI authentication should be used. Third, any proof required for a privilege

delegation to allow something to act on behalf of something else must be pinned down in either certificates, VC or VP.

5 CONCLUSION

We proposed an extension of actor models for data spaces which allows entities to perform operations on-behalf-of other entities. We provided a standard use case involving natural entities, legal entities, and a secure gateway. Using our extension, all these entities were capable to delegate privileges to other entities using state-of-the-art PKI, VCs and VPs. We conclude that the proposed mechanisms and concepts for delegation of privileges can serve numerous use-cases which were not covered yet and can be easily added to existing specifications of Gaia-X and IDS.

ACKNOWLEDGMENTS

This work has been funded by the Fraunhofer-Cluster of Excellence »Cognitive Internet Technologies« and the following European Union's funded research projects: EUHubs4Data (No 951771), DIH4AI (No 101017057), ZeroW (No 101036388), and IDEA4RC (No 101057048).

REFERENCES

[1] International Data Spaces Association. 2021. *Gaia-X and IDS - Position Paper version 1.0.* International Data Spaces Association. Retrieved February 02, 2023 from https://internationaldataspaces.org/download/19016/?tmstv=1675308365

[2] International Data Spaces Association. 2023. *International Data Spaces Reference Architecture Model 4.0.* International Data Spaces Association. Retrieved February 02, 2023 from https://docs.internationaldataspaces.org/ids-ram-4

[3] European Comission. 2022. *eIDAS Regulation.* European Comission. Retrieved February 15, 2023 from https://digital-strategy.ec.europa.eu/en/policies/eidas-regulation

[4] David Cooper, Stefan Santesson, Stephen Farrell, Sharon Boeyen, Russell Housley, and William Polk. 2008. *Internet X. 509 public key infrastructure certificate and certificate revocation list (CRL) profile.* Technical Report.

[5] A Eitel, C Jung, R Brandstädter, A Hosseinzadeh, S Bader, C Kühnle, P Birnstill, G Brost, M Gall, F Bruckner, et al. 2021. Usage control in the international data spaces. International Data Spaces Association, Berlin.

[6] Gaia-X. 2022. *Gaia-X Architecture Document - 22.04 Release.* Gaia-X. Retrieved February 02, 2023 from https://docs.gaia-x.eu/technical-committee/architecture-document/22.04/

[7] Gaia-X. 2022. *Gaia-X Trust Framework - main version (adb756bb).* Gaia-X. Retrieved February 21, 2023 from https://gaia-x.gitlab.io/policy-rules-committee/trust-framework

[8] Gaia-X. 2023. *Gaia-X Association.* Gaia-X European Association for Data and Cloud AISBL. Retrieved February 04, 2023 from https://gaia-x.eu/who-we-are/association/

[9] Arghavan Hosseinzadeh, Andreas Eitel, and Christian Jung. 2020. A Systematic Approach toward Extracting Technically Enforceable Policies from Data Usage Control Requirements.. In *ICISSP*. 397–405.

[10] Hendrik Meyer zum Felde, Mathias Morbitzer, and Julian Schütte. 2021. Securing Remote Policy Enforcement by a Multi-Enclave based Attestation Architecture. In *2021 IEEE 19th International Conference on Embedded and Ubiquitous Computing (EUC)*. IEEE, 102–108.

[11] Mathias Morbitzer, Benedikt Kopf, and Philipp Zieris. 2022. GuaranTEE: Introducing Control-Flow Attestation for Trusted Execution Environments. *arXiv preprint arXiv:2202.07380* (2022).

[12] Eric Rescorla. 2018. *The transport layer security (TLS) protocol version 1.3.* Technical Report.

[13] Manu Sporny, Dave Longley, and David Chadwick. 2022. *Verifiable Credentials Data Model v1.1.* Technical Report.