OASEES: Leveraging DAO-based Programmable Swarms for Optimized Edge-to-Cloud Data Processing

M.A. Kourtis[1], I. Gutierrez[2], E. Areizaga[2], G. Alexandridis[3], W. Tavernier[4], A. Imeri[5], N. Tcholtchev[6], G. Xilouris[1], P. Trakadas[3], I. Chochliouros[7]

[1] National Center of Scientific Research "Demokritos", Athens, Greece
[2] Tecnalia, Bilbao, Spain
[3] National & Kapodistrian University of Athens
[4] imec, Belgium
[5] Luxembourg Institute of Science and Technology, Luxembourg
[6] Fraunhofer FOKUS, Germany
[7] Hellenic Telecommunications Organizations, OTE, Greece

Abstract. As traditional linear models stagnate decision-making and data federation, there's a pressing need for a novel, swarm-based cloud-edge computing approach to enhance European data sovereignty and foster a sustainable, circular economy across various market sectors. To that end, the EU-backed OASEES project identifies a need for an innovative, inclusive, and disruptive approach to the cloud-to-edge continuum, swarm programmability, and data federation over GAIA-X. This paper underscores the actual challenges associated with managing and orchestrating edge infrastructure and services, thereby harnessing the potential of edge processing and federated learning. Moreover, it delves into the core features of the OASEES approach, taking into account technological challenges anticipated in system development. We also explore the integration of multi-tenant, interoperable, secure, and trustworthy deployments into the cloud-to-edge paradigm, in line with the conference's scope. Briefly, we discuss several vertical edge applications with substantial market impact, demonstrating how our approach partially addresses the existing gaps and contributes to a decentralized AI ecosystem.

Keywords: AI, blockchain, cloud hosting, compute continuum, DAO, decentralized applications, edge computing, edge processing

1 Introduction

Leveraging data in a multimodal fashion [1]. Currently, the most potent data processing is centralized, primarily situated in the cloud [2]. This approach affords the capability to scale and efficiently allocate resources on-demand. Centralized data processing is typically characterized by all data being gathered into a single centralized storage area, subsequently processed by a single computer system with extensive architectural capacities in terms of memory, processor, and storage. Decisions regarding protection levels and authorized access typically fall under the purview of the system administrator [3], [4].

In this framework, centralized processing can yield significant benefits. For instance, it helps reduce cost due to the absence of the need for additional hardware, enhances data security, and ensures data and programs on each information system remain independent, thereby extending security and trust [5]. Simultaneously, cloud hosting ensures application and website accessibility using dedicated cloud resources [6]. Contrary to traditional hosting processes, solutions are not deployed on a single server but rather on a network of interconnected virtual and physical cloud servers, ensuring greater elasticity and scalability. This also enhances flexibility and reliability - cloud hosting scales to accommodate traffic spikes or seasonal demands, and hardware failures do not result in downtime as sites and applications are hosted on a network of servers [7].

However, despite the pervasive utilization of centralized processing and cloud hosting across numerous systems, these approaches inherently restrict service and application operation to a resource-constrained manner, often dependent on large single entities for authentication, data storage, data processing, connectivity, and vendor-locked environments for development and orchestration [8-13]. This significantly limits users' data governance capabilities, curtails their visibility into access privileges, and impedes the implementation of necessary controls to prevent potential inappropriate or risky access.

The current data governance paradigm of OASEES entails a sum of policies, processes, standards, metrics, and roles to ensure effective data utilization to achieve organizational objectives. It establishes responsibilities and processes that guarantee the used data's quality and security [14]. Data federation over GAIA-X and federated learning are potential avenues for improving data utilization while upholding high-quality standards [15]. Existing solutions for edge device authentication generally require a centralized entity for trust and authentication, thus creating a non-portable identification paradigm. Integrating data federation and federated learning in this context could potentially foster more robust and decentralized data management and processing solutions [16].

The paper is organized as follows: Section 2 presents in brief the OASEES concept for swarm-based systems. Section 3 covers the data federation aspects of the proposed architecture and how they leverage GAIA-X. Next, Section 4 discusses how federated learning is utilized in the current scope, and how it benefits swarm-based frameworks, and finally Section 5, concludes the paper and draws future lines.

2 OASEES Concept in the Cloud Edge Continuum

OASEES introduces a decentralized and swarm intelligence-based computing framework, employing distributed ledger technology (DLT). The foundational elements of DLT and blockchain enable trustworthy data exchange and collective collaboration. This is achieved by establishing a network where all transactions are universally accessible and validated by the network nodes. The inalterable nature of the transactions fosters transparency and trust by creating an auditable trail. This system ensures

every participant within the swarm agrees on the data's state in the ledger, leading to consensus.

DLT's capability extends beyond simple transactions to execute complex applications or smart contracts based on transaction completion. These smart contracts are essentially pre-defined, functional requirement-based computer codes, providing the foundation for decentralized applications (DApps) deployed on the blockchain. Within OASEES, these DApps function within the Service Layer, as illustrated in Figure 1. The concept of a Decentralized Autonomous Organization (DAO) for swarms also features prominently within OASEES, achieved through one or more blockchain-deployed smart contracts.

OASEES deploys these principles to instill inherent trust within swarms, where an OASEES swarm organizes its operations and intelligence via a DAO. Each device within the swarm can participate in the DAO, voting automatically on decision-making based on different parameters and conditions. This self-governance model in conjunction with the cloud-native approach of OASEES forms a flexible, agile framework for swarm architectures. Despite the inherent trust, security remains a two-pronged challenge for OASEES. OASEES's approach to this challenge includes the development of dedicated security enablers for the cloud, edge, and AI modules, integrating them seamlessly into the deployment process of the framework. Additionally, OASEES integrates various cloud and IoT security standards to incorporate root-of-trust for different cloud-native services. A Self Sovereign Identity (SSI) approach is utilized for identity management, controlling unauthorized access based on the verifiable credentials of each member.

The high-level OASEES architecture involves various actors, interacting within different layers. Figure 1 highlights these interactions and the positions of each actor in the following layers: Deployment & Execution Layer, Identity Layer, Distributed Secure Swarm Layer, Service Layer, and Programmability Layer.

Multiple stakeholders have roles in the OASEES paradigm: Infrastructure Providers (IP), Cloud Service Providers (CSP), Network Service Providers (NSP), Edge Service Providers (ESP) or Data Providers (DP), Specialists, Developers, Data Consumers (DC), and OASEES Service Providers (OSP). Each plays a crucial role in the operation and management of OASEES services, ranging from infrastructure management to service design, consumption, and provision.

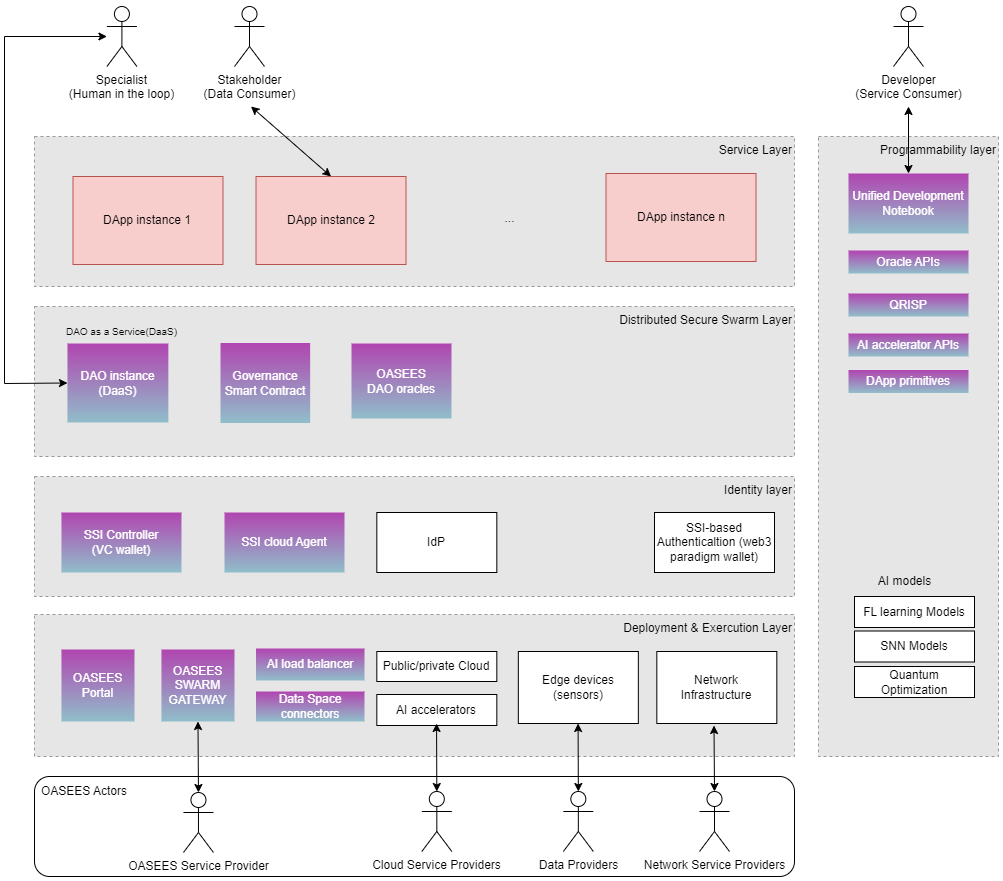

Fig. 1 OASEES High Level architecture

Swarm collaboration within OASEES is realized through a collective of edge devices and human experts, employing decentralized technology for involvement, synchronization, and oversight. A unique characteristic of this is the execution of collaboration over the GAIA-X data federation, enhancing the power of data sharing, interoperability, and sovereign data services. By using a Decentralized Autonomous Organization (DAO) coupled with Human-in-the-Loop (HITL) mechanisms, it paves the way for Federated Learning amongst disparate edge devices. Participants interface with the

DAO via digital wallets, facilitating the staking of tokens, proposal voting, and receipt of rewards. The DAO paradigm's underlying automation works in concert with human intelligence, thereby driving towards the fulfillment of intricate objectives and resolution of complex challenges.

The constituents crucial for effectuating this architecture include a blockchain platform equipped with smart contract capabilities, a robust DAO framework, a token standard for stake holding and governance, a user interface, an SSI framework for meticulous identification management, and HITL mechanisms. All these elements coalesce to foster swarm collaboration that is transparent, secure, and inherently trusted, underpinned by the power of GAIA-X data federation and the efficiency of Federated Learning.

3 Data Federation

The OASEES Data Federation aims to overcome the challenges inherent in centralizing all data in a single repository. In line with its decentralized architecture, data will also remain decentralized and distributed across different databases, while providing a unified view of data access.

OASEES is defined to operate very close to where the Data is generated and therefore operates as a "First Level Data Space", this data has to be correlated, stored and integrated. The Data Federation will allow the integration of data from different data sources (Edge Devices) without the need for physical data consolidation. The data will remain under the control of the respective source, maintaining data ownership and security. It will also improve scalability, as new data sources can be added/removed in real time, minimizing data movement and reducing latency.

GAIA-X is working on the architecture for a "Sovereign Data Exchange based on Trust between stakeholders". Gaia-X has developed a "Data Product" concept and an operational model to respond to the challenges described in its conceptual model.

OASEES will comply with the GAIA-X Data Product Description [17] for the definition of its own Data Products. These descriptions will be stored in a Federated Data Catalogue, allowing users to discover and access the Data Products. Each Data Product Description has a Data License that defines the usage policies applicable to all data within that Data Product. In addition, it contains other relevant information such as billing details, technical specifications, service levels, etc.

Data products within OASEES will also provide the Data Product Usage Contract (DPUC). This contract will be based on the Data Product Description. The DPUC will be designed as a Ricardian contract, meaning it is both human-readable and machine-readable. It is cryptographically signed to ensure tamper-proof integrity and can be verified in a decentralized manner. The DPUC is electronically linked to the data, establishing a clear connection between the contract and the subject of the contract itself. Optionally, the parties involved may choose to have the contract notarized in a federated Data Product Usage Contract Store.

Connectors will also be used for data transfer, and while OASEES is not limited to the use of a specific connector, it will give priority to the EDC (Eclipse Data Connector)

due to the integration with Gaia-X to simplify the use of Gaia-X Verifiable Credentials for Participant Compliance in contract negotiations and access control within this ecosystem. This integration aims to improve the accessibility of Gaia-X Verifiable Credentials for Participants by allowing service providers to grant exclusive access to their services to Gaia-X compliant participants. In simpler terms, this means that service providers can limit access to their services exclusively to Gaia-X compliant participants.

4 Federated Learning

Federated learning (FL) introduces a decentralized learning paradigm that facilitates collaborative and local model training on edge devices. It eliminates the need for data exchange by sharing only model parameters for aggregation. In FL, each participating node processes data available at the edge to generate a local model, whose parameters are then shared with one or more central entities (servers), where the global model is built by applying aggregation strategies. This global model is then shared with all participating edge devices, which use it to update and refine their local models for subsequent training rounds. FL effectively addresses concerns related to data governance, privacy and scalability, as it allows for local processing of data.

In the OASEES platform, AI/ML/DL models are stored at the OASEES cloud storage. When requested, via the respective SDK, the model is pushed to edge devices, where it is either used for inference or trained with local data in a federated-learning context, thereby ensuring privacy and scalability by decoupling possibly sensitive data from the overall model training. Once training on the edge devices is completed, agents send the updated model parameters to the OASEES portal for aggregation, again using the appropriate SDK calls (Figure 1).

At the OASEES portal, AI load balancers aggregate all the updated model parameters from the edge devices, combining the respective weights according to a predefined federated learning policy (e.g. averaging) to produce the resulting global model. The aforementioned model is subsequently forwarded to all registered edge devices and is also stored at the OASEES cloud storage. Following, the edge devices replace their local models with the global one and continue with further local training, utilizing the knowledge of the deployed global model. This iterative process continues until the model is fully trained, performing well across the data of all edge nodes. Overall, this approach enables the efficient utilization of resources across the Cloud-Edge continuum, leveraging the power of edge devices while maintaining privacy and scalability.

5 Conclusion

In summation, OASEES framework introduces an innovative approach to decentralized computing, utilizing the intricacies of Distributed Ledger Technology (DLT) and blockchain principles. This methodology engenders a secure, auditable, and transparent mechanism for data transactions and collective collaboration within the constructs of a swarm architecture. Leveraging blockchain's capabilities beyond transactional

processing, OASEES deploys smart contracts, defining functional requirement-based computer codes, thereby facilitating the creation and deployment of Decentralized Applications (DApps). This seamless integration allows for the development of a Decentralized Autonomous Organization (DAO) within a swarm, fostering consensus and intrinsic trust among network nodes.

Within the OASEES architectural paradigm, several stakeholders play indispensable roles in the successful operation and management of services.

A notable attribute of the OASEES ecosystem is its symbiosis with GAIA-X data federation, serving to enhance interoperability, sovereign data services, and data sharing across the network while safeguarding data ownership and ensuring security. This collaboration facilitates Federated Learning and Human-in-the-Loop (HITL) mechanisms among the disparate edge devices, establishing a solid foundation for secure, trusted, and transparent swarm collaboration. The compliance of OASEES with GAIA-X Data Product Description demonstrates a dedicated effort to address the challenges described in its conceptual model. The utilization of Data Product Usage Contracts (DPUC), designed as Ricardian contracts, and the preference of the Eclipse Data Connector (EDC) due to its integration with Gaia-X, contributes significantly to the enhancement of data security and integrity within the system.

Moreover, the OASEES platform embraces the paradigm of federated learning, which provides a decentralized learning model facilitating collaborative local model training on edge devices. The OASEES methodology for managing AI/ML/DL models underscores the commitment to privacy and scalability by localizing potentially sensitive data processing within the overall model training scheme. The iterative process involving edge devices and OASEES cloud storage emphasizes the efficient utilization of resources across the Cloud-Edge continuum, maintaining privacy and scalability.

To summarize, OASEES exemplifies an innovative, secure, and efficient decentralized computing framework that integrates DLT, DAO, federated learning, and GAIA-X data federation. The combination of these advanced technologies and strategies renders OASEES a promising model with the potential to significantly influence and shape the trajectory of future development in swarm-based computing. The framework merits further examination and validation in a wider range of real-world applications, promising significant potential implications for swarm intelligence, decentralized computing, and edge services..

Acknowledgments. The research leading to these results has been supported by the OASEES project (no. 101092702).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Gabriel, T., Cornel-Cristian, A., Arhip-Calin, M., and Zamfirescu, A. (2019): Cloud Storage. A comparison between centralized solutions versus decentralized cloud storage solutions

using Blockchain technology. In: Proceedings of the 54th International Universities Power Engineering Conference (UPEC), pp.1-5. IEEE (2019)
2. [2] IBM, What is cloud hosting, https://www.ibm.com/cloud/learn/what-is-cloud-hosting
3. [3] Leskinen, J.: Evaluation Criteria for Future Identity Management. In: Proceedings of the 11th International Conference on Trust, Security and Privacy in Computing and Communications, pp.801-806. IEEE (2012)
4. [4] Thakur. M.A., and Gaikwad, R.: User identity and Access Management trends in IT infrastructure - an overview. In: Proceedings of the 2015 International Conference on Pervasive Computing (ICPC), pp.1-4. IEEE (2015)
5. [5] Zareen, M.S., et al.: Artificial Intelligence/ Machine Learning in IoT for Authentication and Authorization of Edge Devices. In: Proceedings of the 2019 International Conference on Applied and Engineering Mathematics (ICAEM), pp.220-224. IEEE (2019)
6. [6] Vorakulpipat, C., et al..: Comprehensive-Factor Authentication in Edge Devices in Smart Environments: A Case Study. In: Proceedings of the 11th International Conference on Control, Automation and Information Sciences (ICCAIS), pp.391-396. IEEE (2022)
7. [7] Lu, Y., Wang, D., et al.: Edge-Assisted Intelligent Device Authentication in Cyber-Physical Systems. IEEE Internet of Things Journal 10(4), 3057--3070, (2023)
8. [8] Castellano, G., Esposito, F., and Risso, F.: A Service-Defined Approach for Orchestration of Heterogeneous Applications in Cloud/Edge Platforms. IEEE Transactions on Network and Service Management 16(4), 1404--1418 (2019)
9. [9] Taleb, T., Samdanis, K., Mada, B., Flinck, H., Dutta, S., and Sabella, D.: On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration. IEEE Communications Surveys & Tutorials 19(3), 165--1681 (2017)
10. [10] Sonmez, C., et al..: Fuzzy Workload Orchestration for Edge Computing. IEEE Transactions on Network and Service Management 16(2), 769--782 (2019)
11. [11] Ranjan, A., Guim, F., Chincholkar, M., Ramchandran, et al.: Convergence of Edge Services & Edge Infrastructure. In: Proceedings of the 2021 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), pp.96-99. IEEE (2021)
12. [12] Loghin, D., Ramapantulu, L., and Teo, Y.M.: Towards Analyzing the Performance of Hybrid Edge-Cloud Processing. In: Proceedings of the 2019 IEEE International Conference on Edge Computing (EDGE), pp.87-94. IEEE (2019)
13. [13] Risso, F.: Creating an Edge-to-Cloud Computing Continuum: Status and Perspective. In: Proceedings of the 3rd International Conference on Embedded & Distributed Systems (EDiS), pp.4-4. IEEE (2022)
14. [14] Accenture, Edge computing, https://www.accenture.com/bg-en/insights/cloud/edge-computing-index
15. [15] Shi, W., Cao, J., Zhang, Q., Li, Y., and Xu, L.: Edge computing: Vision and challenges. IEEE Internet of Things Journal 3(5), 637--646 (2016)
16. [16] Masip-Bruin, H., Marín-Tordera, E., et al.: Managing the Cloud Continuum: Lessons Learnt from a Real Fog-to-Cloud Deployment. Sensors (MDPI) 21(9), 2974 (2021).
17. [17] https://gaia-x.gitlab.io/technical-committee/architecture-document/conceptual_model/.