

Decentralisation and Persistent Data Identification - the Emerging Technologies for Modern Mechanical Science

Intended length: 8 pages

Template: Springer Lecture Notes

Process: LaTeX

Deadline: **February 16, 2024**

Authors (please add yourself with ORCID and affiliation):

- Andrey Vukolov 0000-0001-6967-3236, Elettra Sincrotrone Trieste
- Erik Van Winkle, 0000-0002-7567-0311, DeSci Foundation

Abstract

Multi-agent robotic systems are complex and dynamic networks of autonomous entities that interact with each other and their environment. Such systems require effective and efficient data exchange and collaboration among agents, which can be facilitated by new technologies that support decentralized, persistent data management. In this paper, the current state-of-the-art and challenges of these technologies are reviewed, focusing on the specific domain of swarm robotics. A novel use case is also presented, which demonstrates how a new data identification and sharing technology, based on InterPlanetary File System (IPFS) and dPID, can leverage the advantages of decentralization and data persistence to ensure reliable and long-term data availability and access in swarm robotics.

Introduction

Sustainable communications are essential in the context of multi-agent robotic systems development. Previous works¹ have shown that technologies are continuously drifting to cloud robotics and heterogeneous, deeply connected environments, known as **swarm robotics**². In the swarm topology, the growth of the computational power of every single robot increases the amount of data that robots need to exchange with both the “control centre” and each other. From a distance, the multi-agent robotic systems may appear as huge clouds or clusters of independent machines, producing large and diverse amounts of

¹ A. Siriweera and K. Naruse, "Survey on Cloud Robotics Architecture and Model-Driven Reference Architecture for Decentralized Multicloud Heterogeneous-Robotics Platform," in *IEEE Access*, vol. 9, pp. 40521–40539, 2021, doi: <https://doi.org/10.1109/ACCESS.2021.3064192>

² Pacheco A, et al. 2022. Real-time coordination of a foraging robot swarm using blockchain smart contracts. In: *Swarm Intelligence – Proceedings of ANTS 2022 – Thirteenth International Conference*. LNCS, Springer; Cham https://doi.org/10.1007/978-3-031-20176-9_16

data. Handling of these data diversity, and especially the problems of connectivity within the multi-agent environment is already challenging, but it scales up to unprecedented sizes when data protection, access models, and interoperability are considered, even within the *swarm*. Currently, an abstract robot is a *computational node* that handles numerous links to other nodes, making it practically impossible to maintain traditional networks with central nodes or multiple dependent routers³⁴. Therefore, peer-to-peer networks, frameworks, and solutions have emerged with their scalability to millions of nodes⁵. The data produced in such environments must be identified, verified, and distributed reliably and uniformly across the swarm⁶. To prevent the system from becoming fragile, every participating agent must be able to verify the data on its own using standardised algorithms⁷. Thus the solutions and data framework intended for usage in the described circumstances are required to include:

- Persistent identification of data with identifiers reproducible on the client's side
- Distributed ledger and/or access control solution based on open uniform standards and cryptography
- Transport-agnostic data delivery networks, compatible with the peer-to-peer nature of the swarm
- Trustless data transportation

Summarising the above, decentralised technologies formulate a new prospective way to create modern self-controlling robotics for Industry 5.0⁸. Moreover, decentralisation looks like the only obvious solution for controlling real swarms and fleets of robots. Here and below, a brief overview of the existing decentralised control technologies for robotics will be given accompanied by the technical proposal for the usage of emerging data identification and sharing solutions.

Place of Decentralised Technologies in Robotics

Based on the features of transport-agnostic trustless delivery of the data based on persistent content-based addressing, mentioned above, two large groups of technologies can be highlighted: blockchain-based technologies and decentralised data-sharing networks. Blockchain-based technologies are useful for efficiently building webs of trust with a low rate

³ K. Khateri, M. Pourgholi, M. Montazeri and L. Sabattini, "A Comparison Between Decentralized Local and Global Methods for Connectivity Maintenance of Multi-Robot Networks," in *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 633-640, April 2019, doi: <https://doi.org/10.1109/LRA.2019.2892552>

⁴ K. Y. K. Leung, T. D. Barfoot and H. H. T. Liu, "Decentralized Localization of Sparsely-Communicating Robot Networks: A Centralized-Equivalent Approach," in *IEEE Transactions on Robotics*, vol. 26, no. 1, pp. 62-77, Feb. 2010, doi: <https://doi.org/10.1109/TRO.2009.2035741>

⁵ Nikita Trivedi, Bighnaraj Panigrahi, Hemant Kumar Rath, and Arpan Pal. 2018. Wireless Mesh Routing For Indoor Robotic Communications. In Proceedings of the 1st International Workshop on Internet of People, Assistive Robots and Things (IoPARTS'18). Association for Computing Machinery, New York, NY, USA, 25–30. <https://doi.org/10.1145/3215525.3215533>

⁶ Geng, Y., Zhang, D., Li, P. H., Akcin, O., Tang, A., & Chinchali, S. P. (2022, January). Decentralized sharing and valuation of fleet robotic data. In *Conference on Robot Learning* (pp. 1795-1800). PMLR. <https://proceedings.mlr.press/v164/geng22a.html>

⁷ Andrey Vukolov. (2021, июнь 18). Openly reproducible Persistent Identifiers (PIDs) as a factor of FAIRness in data sharing practices. EOSC Symposium 2021. Zenodo. <https://doi.org/10.5281/zenodo.4980522>

⁸ Y. Khaluf, E. Mathews and F. J. Rammig, "Self-Organized Cooperation in Swarm Robotics," 2011 14th IEEE International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing Workshops, Newport Beach, CA, USA, 2011, pp. 217-226, doi: <https://doi.org/10.1109/ISORCW.2011.30>

of data flow. However, blockchain-based systems face specific challenges due to the nature of the blockchain technology. They require sacrificing performance for general reliability and some implementations allow hardforking, which compromises data immutability. Moreover, blockchain systems generally do not support lazy and delayed recording because of the transaction models implemented in their smart contracts. This significantly increases the overall complexity of the system, which in this case resembles “connected clouds” (fig. 1). The image in fig. 1 is borrowed from [8] as a conscientious citation and it is copyrighted by its respective authors.

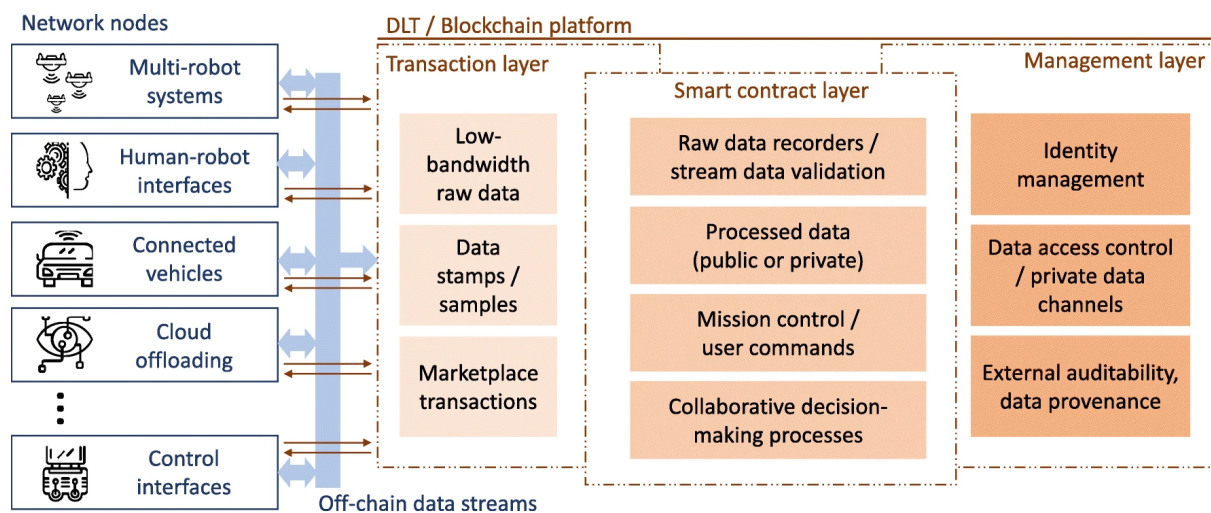


Fig. 1. Implementation model for blockchain-based data management for swarm robotics

The second class of systems, decentralised data-sharing networks, are focused on distributed data transportation, identification, and delivery⁹. These systems use decentralised data storage and retrieval solutions based on distributed hash tables (DHT). Through the DHT, the network should implement the reproducible identification of the transferred data by default, based on cryptographic hash functions. This enables the reproducibility of software, algorithms, and identifiers as a natural part of the data integrity verification process. However, the selection of implementation for robotic developers becomes more complicated because of the need to compromise between routing delays and network sustainability¹⁰¹¹.

IPFS as a Data Sharing Solution for Robotics

There are many possible implementations of the DHT-based decentralized data transportation networks (BitTorrent, Freenet, Kademlia, IPFS, etc.). InterPlanetary File System (IPFS) is already named as the most promising solution for usage in robotics,

⁹ Reid, N. (2015). Literature review: Purely decentralized P2P file sharing systems and usability. <https://www.cs.ru.ac.za/research/g11R0507/LitReviewNickyg11r0507.pdf>

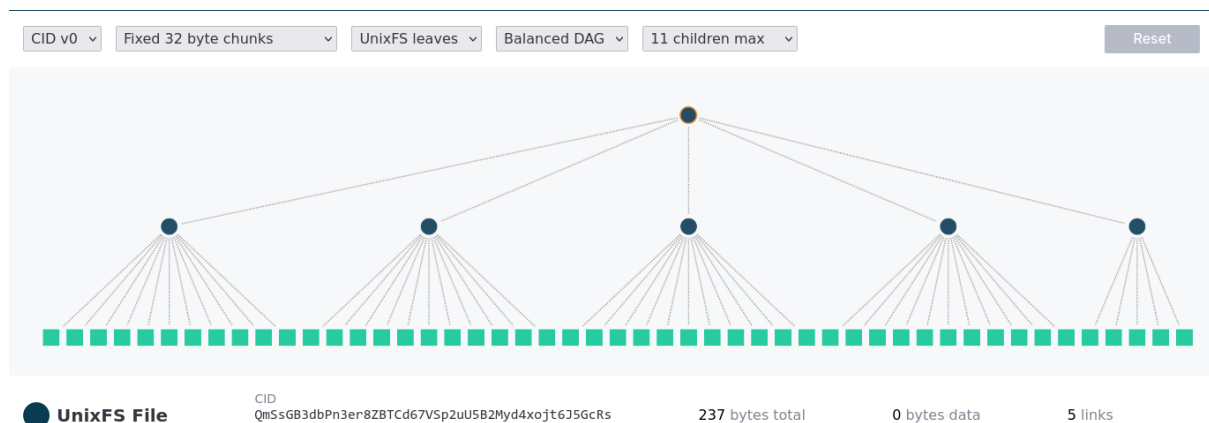
¹⁰ Shen, J., Li, Y., Zhou, Y., & Wang, X. (2019, June). Understanding I/O performance of IPFS storage: a client's perspective. In *Proceedings of the international symposium on quality of service* (pp. 1-10). <https://doi.org/10.1145/3326285.3329052>

¹¹ L. Cao and Y. Zhang, "Research on Improvement of Routing Algorithm Kademlia in IPFS," *2022 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC)*, Qingdao, China, 2022, pp. 399-403, doi: <https://doi.org/10.1109/ICFTIC57696.2022.10075293>

Internet of Things, industrial automation, and blockchain ledgers¹². In the following list, the key features of IPFS that make it useful in robotic/industrial environments, are enumerated:

- Network-wise persistence, reproducibility, and unambiguity of the identifier assigned to the specific data. The Content IDs (CIDs) used in IPFS use hash functions to identify data. These may be considered as Persistent Identifiers (PIDs) immune to content drift and link rot, implemented aside from the network.
- Pre Implemented OSI Transport Level Protocol-agnostic transport abstraction, libp2p¹³.
- Integrated cryptography-based peer addressing system, InterPlanetary Name System (IPNS)¹⁴, introducing an abstraction layer over the existing network stack-specific machine identification for the network, producing unique identifiers specific for the single network node.
- Integrated reference points definition in addition to the swarm neighbours discovery. It allows the network to define routing flexibly to obtain optimal performance.
- Support for new communication protocols, like QUIC¹⁵.
- Trustless transportation of the data with considerable resource consumption level¹⁶.
- Monolithic software design, sharing all the implemented features in the singular installation.

The IPFS storage model is designed to support a cache-on-read retrieval approach, where data, along with their metadata, are recorded into Directed Acyclic Graphs (DAGs, fig. 2) using various storage tree builders¹⁷. When some data is addressed in the IPFS network and assigned with one or more CIDs, every act of resolution of this CID on another node of the network actually may cache the data on the requesting node, increasing redundancy for the entire network or swarm.



¹² Jovović, I., Husnjak, S., Forenbacher, I., & Maček, S. (2018, December). 5G blockchain and IPFS: A general survey with possible innovative applications in industry 4.0. In *MMS 2018: 3rd EAI International Conference on Management of Manufacturing Systems* (Vol. 2, p. 157). European Alliance for Innovation, <https://books.google.it/books?id=Vsv2DwAAQBAJ>

¹³ libp2p. Official website. Access date 09.02.2024. <https://libp2p.io/>

¹⁴ IPNS. Official Documentation. Access date 09.02.2024 <https://docs.ipfs.tech/concepts/ipns/>

¹⁵ IPFS Documentation. Access date 21.01.2024 <https://docs.ipfs.tech/>

¹⁶ Abraham, N., Ramar, R. (2021). Secure Data Sharing with Interplanetary File System for Pharmaceutical Data. In: Misra, S., Kumar Tyagi, A. (eds) *Artificial Intelligence for Cyber Security: Methods, Issues and Possible Horizons or Opportunities*. Studies in Computational Intelligence, vol 972. Springer, Cham. https://doi.org/10.1007/978-3-030-72236-4_11

¹⁷ Sanjuan, H., Poyhtari, S., Teixeira, P., and Psaras, I., "Merkle-CRDTs: Merkle-DAGs meet CRDTs", arXiv e-prints, 2020. doi: <https://doi.org/10.48550/arXiv.2004.00107>

Fig. 2. Example visualization of the IPFS DAG

The use of balanced Merkle DAGs and OSI Transport Protocol-agnostic routing in IPFS leads to relatively reasonably fast propagation of the data over the network¹⁸. This makes IPFS a prospective middleware for open-source robotic projects¹⁹²⁰. In particular, the cryptographic nature of CIDs is attractive for developers of blockchain-based solutions. However, the perspective of using IPFS as a persistent data storage enclosure is limitedly described in the existing works. A technical proposition of the use case for an IPFS-based decentralised solution for persistent identification, versioning, and sharing of the data between robots is given below.

Use Case: Initial Technical Proposal

One of the problems regarding IPFS in robotics is that the CIDs assigned to data are immutable and not versioned due to the nature of DAGs. In this case, especially when the task is constrained by limitations of a heterogeneous system including multiple robots, or a weakly connected swarm, the solution may be found via an additional identification layer over the CID. Here the remark should be made that this identification layer should refer to the data, but not to the publishing network node. In this context, the ordinary IPNS keys intended for direct resolution in IPFS may not work as they are implemented as salted hashes of the public key owned by a *node*, and sharing a key requires reconfiguration of all involved nodes. Thus the way to solve the described problem is to place a distributed ledger over IPFS with high-level persistent identifiers forming the identification graph from CIDs. To do so, this paper introduces dPID (abbreviated from “decentralised Persistent Identifier”)²¹, a decentralised data identification and sharing system. dPID leverages a sophisticated integration of the IPFS and IPLD technologies alongside the Sidetree Protocol²². Sidetree acts as a protocol and API layer that can operate atop any data addressing system, enabling users to generate lookup databases and customised identities. dPID provides access to data that is both machine-actionable and human-readable, featuring a web interface built upon the JSON-LD specification²³ and supported by an open-source API. dPID ensures the deterministic resolution of external identifiers to the internal CIDs of IPFS and their associated content through a DAG. This process allows the content to be immediately cached on the local database of the node it is accessed from. Building upon the IPFS network, dPID software employs Ceramic²⁴, a decentralised event streaming protocol, to create a graph-based distributed lookup database. Enhancing the foundational persistence

¹⁸ E. Daniel and F. Tschorsch, "IPFS and Friends: A Qualitative Comparison of Next Generation Peer-to-Peer Data Networks," in *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 31-52, Firstquarter 2022, doi: <https://doi.org/10.1109/COMST.2022.3143147>

¹⁹ Zhang, S., Li, W., Li, X., Liu, B., Zhang, Y., & Cao, C. (2022). A Secure Data Sharing Framework for Robot Operating Systems Leveraging Ethereum. *arXiv preprint arXiv:2208.14269*
<https://doi.org/10.48550/arXiv.2208.14269>

²⁰ Varadharajan, V.S., St-Onge, D., Adams, B. *et al.* SOUL: data sharing for robot swarms. *Auton Robot* **44**, 377–394 (2020). <https://doi.org/10.1007/s10514-019-09855-2>

²¹ dPID. Official documentation. Access date: 09.02.2024
<https://docs.desci.com/technical-background/open-state-repository/pid>

²² Sidetree Protocol Documentation. Access date 22.01.2024 <https://identity.foundation/sidetree/spec/>

²³ JSON-LD Documentation. Access date 21.01.2024 <https://json-ld.org/learn.html>

²⁴ Ceramic Protocol Documentation. Access date 23.01.2024
<https://developers.ceramic.network/docs/protocol/js-ceramic/overview>

offered by IPFS, dPID introduces features like high throughput, strong consistency across the network, decentralised indexing, user-friendly URLs (on the HTTP client side), and constructing unified data access points. Consequently, dPID promises reliable persistence for stored data, making it a comprehensive solution for decentralized data management and sharing. dPIDs provide:

- Verifiable ownership with ORCID-based person identification and incremental contribution record.
- Open network participation and metadata redundancy through peer-to-peer nature of IPFS.
- “Vendor lock-in” removed in the context of data due to the removal of the singular provenance holder of the scientific record.
- Data integrity persistence with DHT, as it was described above.

Every access point in the network is intended to be identical to the others, sharing the same software and data model. Being installed on the machines connected to the swarm, dPID becomes a universal data delivery and persistence agent. It considers the identified data as a collection of versioned IPLD entries following RO-CRATE specification²⁵. The simplified example of internal linkage is presented in Fig. 3.

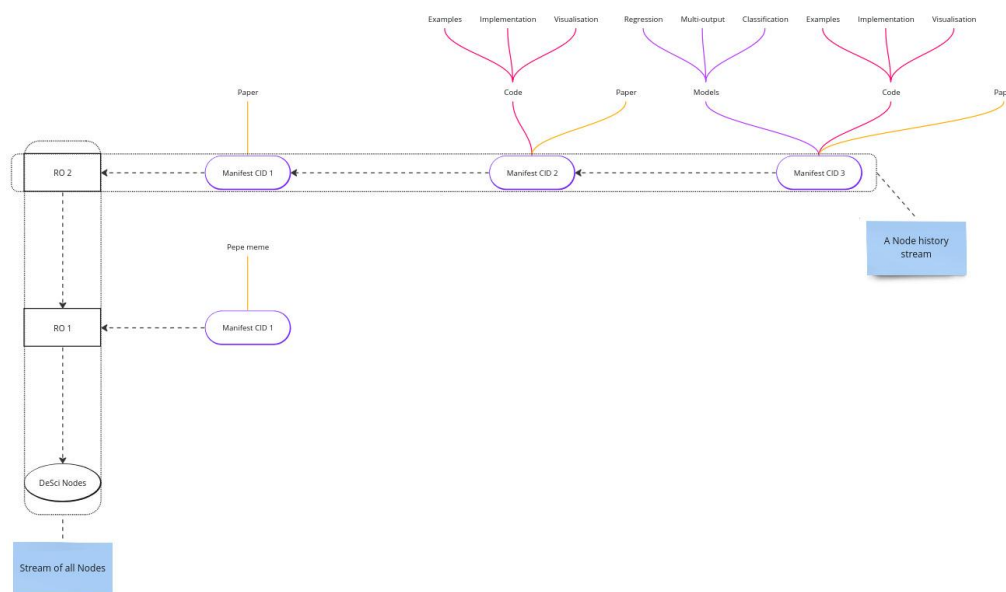


Fig. 3. Simplified internal bitstream storage diagram in dPID

From the perspective of the end user, these systems can be likened to folders that store binary data in a format-agnostic manner. The storage entries are catalogued in a distributed, redundant key-value store, each uniquely accessible via a dedicated persistent identifier that leverages underlying IPFS CIDs for addressing.

Nowadays compatibility with JSON-related specifications makes it extremely easy to share, deliver and parse the data due to high diffusion level of all flavours of JSON over the ecosystems of the popular programming languages, like Python, Rust, JavaScript, Golang.

²⁵ RO-CRATE. Official Specification. Access date 02.02.2024
<https://www.researchobject.org/ro-crate/1.1/>

All the source codes implementing dPID are open, and they are published by the leading developer DeSci Labs (<https://github.com/desci-labs/>).

Conclusion

In this paper, a brief overview of the emerging decentralised technologies of data identification and sharing is provided. IPFS is proposed as a prospective data storage and delivery network for swarm robotics, and dPID as a persistent identification and sharing solution suitable for multi-agent systems with heterogeneous connectivity. The advantages of different kinds of decentralised data sharing systems are also discussed. The results of this paper indicate that IPFS and dPID are promising technologies for decentralised data management in the area of swarm robotics, as they offer advantages such as scalability, robustness, security, and interoperability. However, some limitations and challenges still remain, such as the trade-off between data persistence and network efficiency, the integration with other middleware and frameworks. Therefore, future work will focus on addressing these issues and exploring further applications and extensions of the proposed solution.