

Store Edge Networked Data (SEND): A Data and Performance Driven Edge Storage Framework¹

- **Adrian-Cristian Nicolaescu** – University College London, UK
- Spyridon Mastorakis – University of Nebraska at Omaha, USA
- Ioannis Psaras – University College London, UK

1. Adrian-Cristian Nicolaescu et al., 2021, Store Edge Networked Data (SEND): A Data and Performance Driven Edge Storage Framework, In IEEE International Conference on Computer Communications (INFOCOM 21), https://www.researchgate.net/publication/346643946_Store_Edge_Networked_Data_SEND_A_Data_and_Performance_Driven_Edge_Storage_Framework.

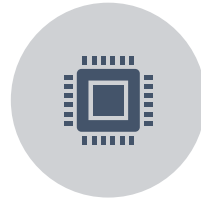
Presenting and Discussing

- Edge Data Repositories Introduction
- Store Edge Networked Data (SEND) Design
- Introducing Labels
- Data Categories and Strategies
- Packet Generation Mechanisms
- Evaluation
- Future Work and Discussion

Edge Computing and rCDN



The Edge – continuously expanding, in data diversity, heterogeneity, production capacity, device numbers and their diversity.



In-network processing extensively researched – mainly within cloud environments, but increasingly within the Edge



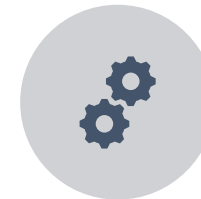
Increasing need for the development of an efficient and effective reverse and decentralized data flow management system



Efficient monitoring and management system solution – data storage and processing at or as close as possible to the Edge



Conclusion - Transitional information to be processed and fed back, discarded or offloaded to clouds¹



Development of a store-process-send system at the Edge of the Internet – a logical next step

1. Ioannis Psaras et al., 2018, Mobile Data Repositories at the Edge, In USENIX Workshop on Hot Topics in Edge Computing (HotEdge 18), USENIX Association, Boston, MA, <https://www.usenix.org/conference/hotedge18/presentation/psaras>.

EDR Initial System Development



Store-process-send – integrates processing with storage



Goal: increasing data delivery efficiency of Edge-produced data



Introduce nodes specialised in data storage and processing, called Edge Data Repositories (EDRs)



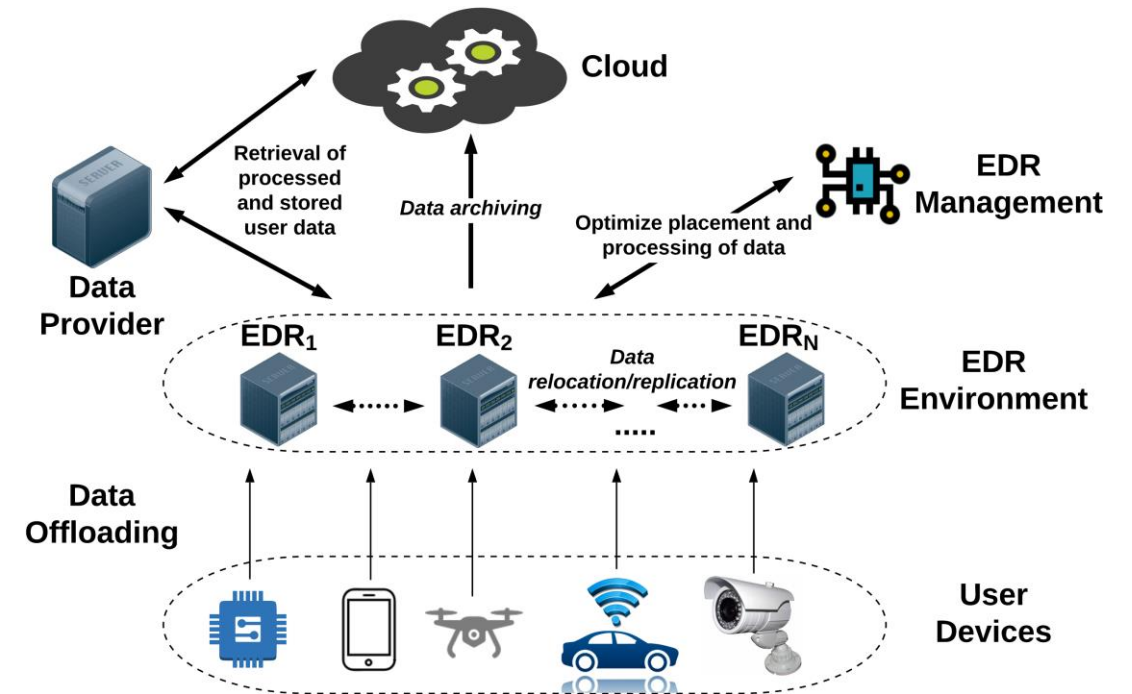
Introduction of freshness period and shelf-life



Environment managed by Edge Repository Provider (ERP) servers

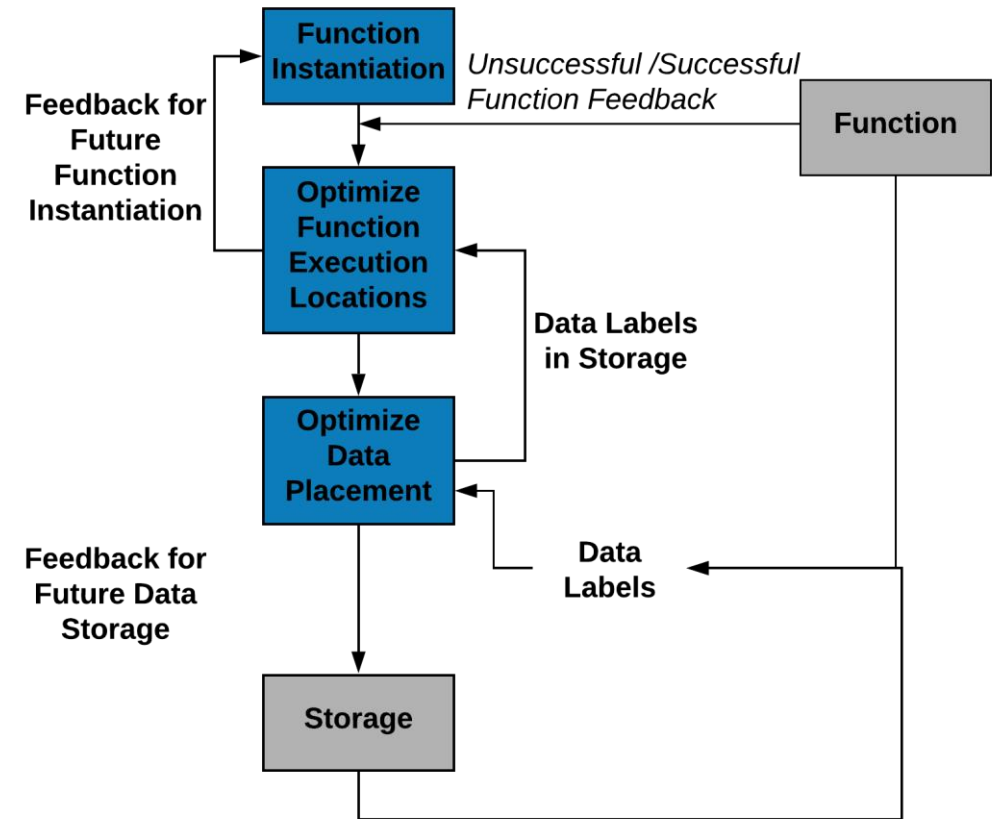
Store Edge Networked Data System Architecture and Assumptions

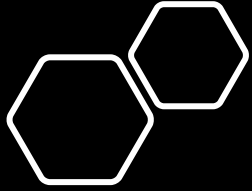
- Devices generate data based on their nature or the applications they may be running
- EDRs deployed one-hop away from user devices.
- A logically centralised EDR environment (domain) management module makes data management and placement decisions by choosing **strategies**
- Data providers may interact with the cloud and the EDRs to retrieve raw or processed data



Store Edge Networked Data Repository Architecture and Assumptions

- Different edge data or their processing results counterparts can be stored at the edge
- Data used for processing within **freshness period** – and useful at the edge for **shelf-life**
- statistics on stored, processed and served data sent periodically to management module
- Based on management decisions, strategy chosen – EDRs relocate or replicate data





Introducing Labels

- Offer the ability to track performance, the popularity of data, and data placement statistics
- Make the system aware of data context and improve its performance
- Enable accurate data placement and storage decisions.
- The most essential data attribute in SEND

SEND (Store Edge Networked Data) Simulations

Hashes and Labels

- Data identity and traceability
- Data/application/service classification
- User data production predictions

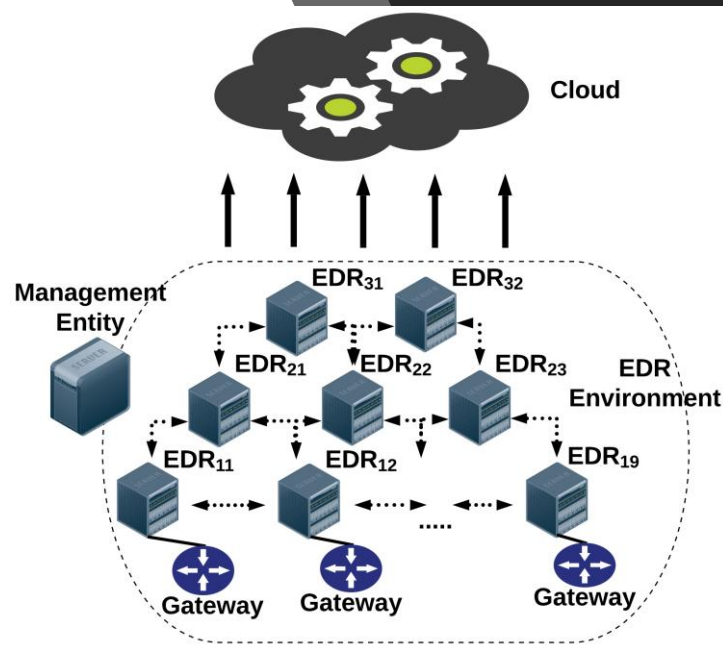
3 Different Data Categories

- Storage-related data
- Function-related data
- Data to be offloaded to the cloud

Storage Placement Strategies

- General-purpose storage strategy
- Popularity-based strategy
- Function-based strategy
- Hybrid strategy

Store Edge Networked Data Scenario-based Evaluation



- A data repository prototype on top of the Google file system, which we evaluate based on real-world datasets of images and Internet of Things device measurements.
- Scaled up to network simulation study based on synthetic and real-world datasets evaluating the performance and trade-offs of the SEND design as a whole.
- Results demonstrate that SEND achieves data insertion times of 0.06ms-0.9ms, data lookup times of 0.5ms-5.3ms, and on-time completion of up to 92% of user requests for the retrieval of raw and processed data.
- Baseline for comparison – a number- and deadlines-based strategy that optimises function placement without the use of labels¹

Future Work and Discussion

- Edge Data Repositories' network and computing efficiency
- Information retention, reliability and security in Edge Data Repositories
- Potential development into more architectural and intelligent systems with this networked storage system

