

Semantic Programming for Device-Edge-Cloud Continuum

Anh Le-Tuan

Technical University Berlin

Berlin, Germany

David Bowden

Dell Technologies

Cork, Ireland

Danh Le-Phuoc

Technical University Berlin

Berlin, Germany

Abstract—This position paper presents ThothSP, a Semantic Programming framework with the aim of lowering the coding effort in building smart applications on the Device-Edge-Cloud continuum by leveraging semantic knowledge. It introduces a novel neural-symbolic stream fusion mechanism, which enables the specification of data fusion pipelines via declarative rules, with degrees of learnable probabilistic weights. Moreover, it includes an adaptive federator that allows the *Thoth»runtime* to be distributed across multiple compute nodes in a network, and to coordinate their resources to collaboratively process tasks by delegating partial workloads to their peers. To demonstrate ThothSP’s capability, we report a case study on a distributed camera network to show ThothSP’s behaviour against a traditional edge-cloud setup.

Index Terms—edge computing, autonomous system, distributed AI, semantic computing

I. INTRODUCTION AND MOTIVATION

The rapid advancements in machine learning, particularly *deep neural network* (DNN), have led to a significant increase in the volume of multi-modal stream data that can be processed with remarkable accuracy. However, the integration of closed-loop “Smart-X” systems, such as Smart Factories, Smart Junctions, and Advanced Driving Assistant Systems (ADAS) in the SmartEdge project [1] cannot afford the latency associated with cloud-based data ingestion and processing. Additionally, the sheer magnitude of multi-modal streaming data, such as camera and lidar data, makes the conventional approach economically infeasible, even with advancements in GPU/CPU and storage technology.

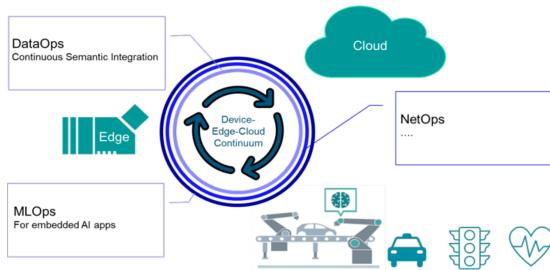


Fig. 1: Device-Edge-Cloud Continuum.

This has motivated our endeavor to tackle the pressing challenge on how to design, program and implement a highly distributed network of device-edge-cloud continuum that takes

979-8-3503-0322-3/23/\$31.00 ©2023 IEEE

part in dynamic information flows driven by unprecedented application logic. Our approach aims to facilitate the autonomous collaboration of edge devices operating in a dynamically changing environment. To accomplish this, significant enhancements are required to support the DataOps, NetOps and MLOps across the entire Device-Edge-Cloud continuum, illustrated in Figure 1. To enable this approach, we propose a Semantic Programming framework, **ThothSP**, that leverages semantic knowledge to lower effort in programming DataOps, NetOps and MLOps agnostic to hardware and infrastructure.

For DataOps, ThothSP employ Semantic Streams [2] powered by RDF stream processing (RSP) engines like CQELS engine [3] to solve the data integration problem across IoT systems. Semantic Streams facilitates the semantic interoperability among autonomous systems via shared formal semantics under standardized ontologies and vocabularies such as OGC/W3C Semantic Sensor Network Ontology (SSN) [4] and W3C WoT Thing Description (TD) Ontology [5].

Towards a seamless integration with DNN and multimodal stream data in MLOps, ThothSP integrates a number of streaming data types, such as camera video streams and LiDAR point clouds, as well as supporting additional hardware platforms, such as ARM and mobiles [6], [7]. Moreover, ThothSP also supports DNN-based data fusion operations [8], [9]. Thus, ThothSP is designed to accommodate a DNN-based data stream fusion operations, which internally generates semantic stream that can be subsequently processed by the *Thoth»runtime* in a unified data model, i.e. RDF-star [10]. To support semantic interoperability in realtime, *Thoth»runtime* uses CQELS-QL, an extension of standadized query language SPARQL with advanced query patterns such event query patterns [11], [12] and probabilistic reasoning [8], [9] so that device, edge and cloud can coordinate a single query interface and language. Semantic streams also pave the way to unify NetOps with the semantic programming paradigm following the vision of *semantic communications* [13].

II. SEMANTIC PROGRAMMING IN THOTHSP

This section introduces Semantic Programming (SP) paradigm for fusing multimodel data that draws inspiration from the human brain’s semantic and episodic memory systems [14]. The semantic memory refers to our brain’s repository of general world knowledge and episodic memory refers to our “episodic memory system”, which encodes, stores, and allows

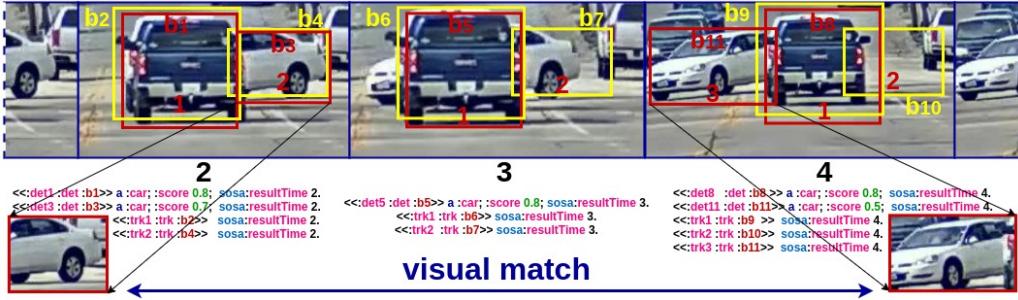


Fig. 2: Semantic representation of multimodal data stream.

access to “episodic memories”, e.g. recollection of personally experienced events situated within a unique spatial and temporal contexts. By incorporating the principles of these cognitive processes into programming, SP aims to lower coding effort by exploiting semantic knowledge paired with human cognition. In this paradigm, programs are designed to manipulate and understand information based on its *semantic symbols* rather than relying solely on rigid algorithms or explicit instructions.

To ground programming elements to semantic symbols understandable for both human and machine, we present sensory streams in RDF-star [10]. As illustrated in Figure 2, ThothSP fuses the video data stream observed by a camera (as a *Sensor*) as a *stream of symbolic observations* representing these video frames. These frames are represented as instances of the class *Image2D* that inherits from the generic *Observation* class of SSN ontology [4]. The detection of a video frame or a tracklet are represented as *SSN Sampling*, which are processed by a DNN model or a CV algorithm represented as a *Procedure*.

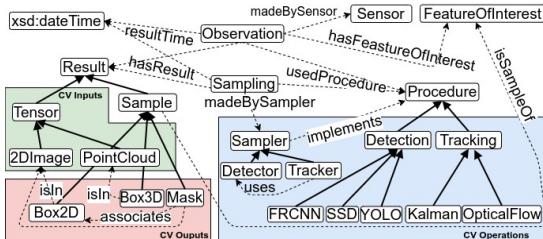


Fig. 3: Data model of Semantic Streams based on SSN Ontology.

```

1 //time point/frame 2
2 <<:image2 a :Image2D>> a sosa:Observation;
3 sosa:madeBySensor :cam1; sosa:resultTime 2.
4 //RDF-star descriptions of detection 1 and 3
5 <<:det1 :det :b1>> a :Detection; sosa:resultTime 2;
6 sosa:hasSimpleResult 'car'; :score '0.8';
7 :isDetectionOf :image2; sosa:usedProcedure :Yolo.
8 <<:det3 :det :b3>> a :Detection; sosa:resultTime 2;
9 sosa:hasSimpleResult 'car'; :score '0.7';
10 :isDetectionOf :image2; sosa:usedProcedure :Yolo.
11 //RDF-star descriptions of tracking 1 and 2
12 <<:trk1 :trk :b2>> a :Tracklet; sosa:resultTime 2;
13 sosa:usedProcedure :KalmanFilter.
14 <<:trk2 :trk :b4>> a :Tracklet; sosa:resultTime 2;
15 sosa:usedProcedure :KalmanFilter.

```

Listing 1: A Snapshot of Semantic Stream serialized with RDF-star.

For example, Listing 1 presents a snapshot of semantic stream data in Figure 2 where the red boxes are detected bounding

boxes and the yellow boxes are tracked bounding boxes. Line 5 denotes that the detection model generates an output consisting of a bounding box b_1 , object type *car*, and confidence score 0.8. Line 13 denotes that the bounding box b_2 is predicted by a Kalman filter and is tracked by tracklet 1.

Based on several theoretical and empirical studies of human perception considering “*perception as hypotheses*”, e.g [15], [16], a Semantic Stream Reasoning (SSR) program is formulated as a set of rules representing hypotheses from semantic streams above. Such rules can be written in CQELS-RL (an extension from CQEL-QL and SHACL) syntax. The formal semantics of SSR is specified in [8]. In essence, there are two types of rules, *hard rules* and *soft rules*. The hard rule is used for background knowledge given by (non-monotonic) common-sense and domain knowledge that is regarded as “always true”. The soft rules express association hypotheses with weights corresponding to probability degrees of these rules. Listing 2 illustrates how a soft rule is written in CQELS-RL syntax based on the input stream illustrated in Listing 1. This rule can be interpreted in plain English as “*a vehicle enters a field of view if a bounding box of it has been detected first time in the last 5 seconds*”.

```

ssr:rule_w_1 a sh:NodeShape;
sh:rule [
  a sh:CQELSRule ;
  sh:prefixes ssr: ;
  sh:construct """
CONSTRUCT { <<?0 :enters <<ssr:FoV>> @ ?T. }
WHERE { STREAM <<:ssr>>
{ <<?Dt :det ?B >> @ ?T; :score ?S.
  ?B sosa:isSampleOf ?O ; a :car.
  FILTER (?S > 0.8)
  NAF STREAM <<:ssr>> window[5 sec]
  {?O :inFOV ssr:FoV.} }
"""
];

```

Listing 2: Soft rule to detect vehicles entering the Field of View.

The weights of the soft rules are determined by the Learning Agent component in the starting phase the learning and fusion workflow illustrated in Figure 4. For the setup, the symbolic training samples are constructed from the labeled data and are fed to the Learning Agent component. The Learning Agent component computes a vector of weights for the soft rules, which are stored as rule templates, and passes them to the Reasoning Programs Producer component. For each training iteration, the Reasoner component returns a *feedback*

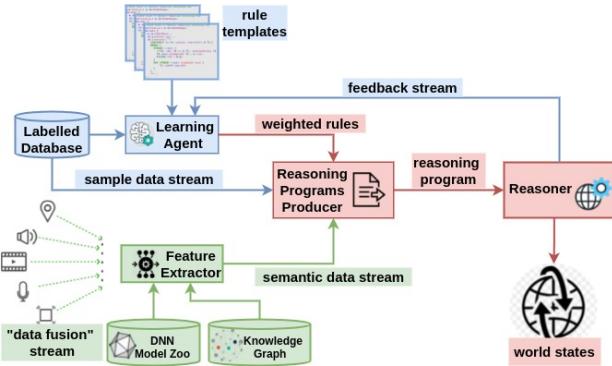


Fig. 4: Learning and Fusion Workflow in ThothSP.

stream of reasoning results back to Learning Agent. The Learning Agent then adjusts the weights of the soft rules until the answer sets, returned as a feedback streams, describe the most likely ground truth.

The Feature Extractor component detects the features of interest from the incoming data stream and maps them to their symbolic representation. The feature data is described with a *neuro-symbolic stream model* [8] and its semantics are enriched via semantic knowledge graphs.

The Reasoning Programs Producer component generates SSR programmes [8], which specify the fusion pipeline and the decision logic to choose the most likely state of the world at each evaluation. The reasoning program is evaluated by the Reasoner component, which employs an ASP solver.

III. DESIGN OF SEMANTIC PROGRAMMING FRAMEWORK

A. Swarm Architecture

The ThothSP framework takes the input from a SSR program described in Section II . In essence, the resultant reasoning programs are an amalgamation of application logic and knowledge graphs, which specify the semantics of the inputs, outputs, and behavior of smart-devices in a *Thoth**swarm. The SSR programs are then compiled and deployed to each smart-device's *Thoth*-runtime, called *Thoth*-node, which collectively form a *Thoth**swarm of devices collaborating to perform a common task. To allow flexibility in deploying a *Thoth**swarm, they can be instantiated in one of two modes.

If the devices in the swarm and their function are known at design time, the swarm can be predefined. This technique allows legacy devices, smart-devices, and clouds to be integrated into the same swarm. The characteristics of each device are stored in a common device repository in the form of a knowledge graph, which allows the function and connectivity of the devices to be statically bound into the reasoning program at design time. The reasoning programs are deployed to each smart-device's *Thoth*-runtime during swarm formation. As legacy devices do not host the *Thoth*-runtime, they are integrated into the swarm by connecting to smart-devices. Sensors on the legacy devices are statically bound to smart-devices at design time, and their data streams processed on the smart-devices' *Thoth*-runtime.

Clouds can be passive recipients of data or semantic streams, or actively participate in the swarm if they host a *Thoth*-runtime.

If the *Thoth**swarm must operate dynamically, i.e., with devices joining and leaving over its lifespan, the *Thoth*-node must act as autonomous agents that can join or leave the swarm on their own volition. All *Thoth*-nodes host a *Thoth*-runtime, which runs a basic SSR program. The *Thoth**swarm is instantiated with a single seed *Thoth*-node, which acts as the swarm coordinator (SC). The SC is running a special SSR program whose purpose is to form and manage the swarm's operation. The SC then enlists free smart-devices (SDs) into the swarm to perform its task. When the a *Thoth*-node joins the swarm it is assigned a role in performing the task. To enable the *Thoth*-node to perform this role, the required SSR program is dynamically downloaded into the ThothSP-runtime and executed. In other scenarios a free SD may request to join a swarm to make use of its services. The SDs use messages oriented middleware (MOM), such as DDS or V2X, and an customized network layer that allows them to interact with other devices in the swarm to collaborate on performing the task.

The SC maintains the primary swarm state in its dynamic knowledge graph (DKG). It updates the swarm state when smart-devices join or leave the swarm, or when significant events occur. Other SDs in the swarm may maintain a partial copy of that DKG, depending on their role in the swarm and their computational and storage capacity. At a minimum a SD will know the characteristics and skills it possesses, as well as basic swarm state information, e.g., its role in the swarm.

Listing 3 illustrates a subscription message in JSON-LD format based on the WoT TD ontology [5]. Based on the semantic description provided by the subscribed nodes, the parent node can carry the stream discovery patterns which use a variable in the stream pattern, as shown in line 4 of Listing 4. The variable *?streamURI* then can be matched in other metadata. In this example, it is used to link with the sensors that generated this stream. Recursively, the subscription process can propagate the stream information upstream hierarchically, and vice versa, the discovery process can be recursively delegated to downstream nodes via sub-queries in CQELS-QL as a simple text message.

```

"@context": "https://www.w3.org/2022/wot/td/v1.1",
"title": "Camera2-At-Helsinki",
"id": "urn:uuid:9489991a-7622-45b6-8437-f859835d4",
"description": "Traffic Camera at Junction...",
"properties":
{
  "status": {
    "description": "Stream Camera Feed at 30FPS",
    "type": "string",
    "forms": [{"op": "readproperty",
      "href": "RTSP://helsinki.fi/camera/2",
      "methodName": "RTSP",
      "contentType": "application/mp4"}],
    "readOnly": true
  }
}

```

Listing 3: Example of subscription message in JSON-LD.

To form a swarm, it would require each *Thoth*-node to operate as an autonomous agent which can collaborate with other peers to execute a distributed processing pipeline specified in CQELS-RL. An autonomous *Thoth*-node can dynamically join a network of existing peers by subscribing itself to an existing node in the network, called a parent node, and it then notifies the parent node

about the query service and streaming service it can provide to the network. These services can be semantically described by using vocabularies provided by VoCaLS [17]. Hence, a subscription can be done by sending a RDF-based message via Websocket or DDS in ROS2.

To this end, when an autonomous *Thoth · node* joins a network, it makes itself and its connected nodes discoverable and queryable to other nodes of the network. Moreover, each node can share its processing resources by executing a CQELS-QL query on it. This will help us treat a CQELS-QL query over *Thoth*swarm* as a query to a sensor network. Note that network telemetry such as network failure, status and mirrored traffic streams, can be modeled as semantic streams to seamlessly integrate them into the control flow of a *Thoth*swarm*.

With the support of the above subscription and discovery operations, a SSR program written in CQELS-RL can be deployed across several sites, e.g., traffic cameras across a road network. Each *Thothnode* gives access to data streams fed from streaming nodes connecting to it. Such stream nodes can ingest a range of sensors, such as air radar, loop detector and camera. When the stream data arrives, this node can partially process the data at its processing site, and then forward the results as tabular results or RDF stream elements to its parent node.

In this context, when a query is subscribed to the top-most node, called root node, it will divide this query to sub-query fragments and deploy at one or more sites via its subscribed nodes. A query fragment consists of one or more operators, and each fragment of the same query can be deployed on different processing nodes. Recursively, a sub-query delegated to a node can be federated to its subscribed nodes. All participant nodes of a processing pipeline can synchronise their processing timeline via a timing stream that is propagated from the root node. The execution process of sub-query fragments can use resources, i.e. CPU, memory, disk space and network bandwidth of participant nodes to process incoming RDF graphs or sets of solution mappings and generate output RDF graphs/sets of solution mappings. Output streams may be further processed by fragments of the same query, until results are sent to the query issuer at the root node. For example, the sub-query of the query in below Listing 4 can be sent down to the nodes closer to the streaming nodes, then the results will be recursively sent to upper nodes to carry out the partial COUNT query until it reaches the root node to carry out final computation steps to return the expected results.

```
SELECT ?camera (COUNT(?truck) as ?truckCount)
WHERE{
  STREAM ?streamURI [RANGE 5m ON sosa:resultTime]
  {
    ?sensor a ssr:Camera; sosa:madeObservation ?obs.
    ?obs sosa:hasResult ?vFrame.
    ?truck a :Truck; ssr:detectedIn ?vFrame.
  }
  ?streamURI prov:wasGeneratedBy/a :TrafficCamera.
}
GROUP BY ?camera
HAVING (COUNT(?truck) > 1)
ORDER BY ?truckCount
```

Listing 4: Example of Federated Query accross Cameras.

This federation process can be carried out dynamically thanks to the dynamic subscription and discovery capability above. Moreover, the processing topology can be dynamically configured by changing where and how participant nodes subscribed themselves to the processing networks. The biggest advantage of this federation mechanism is the ability to dynamically push some processing operations closer to the streaming nodes to alleviate the network and processing bottlenecks which often happen at edge devices. Moreover, this mechanism can significantly improve the processing throughput by adding more processing nodes on demand.

B. Component Design

In this section, we present the design of components of the ThothSP framework as illustrated in Figure 4. ThothSP provides a hardware-independent infrastructure to implement ThothSP kernels for computing continuous queries expressed in CQELS-QL. The ThothSP-runtime accepts RDF streams as input and returns RDF streams or relational streams in the SPARQL format as output. ThothSP allows for the creation of RDF streams by annotating extracted features from sensor input data streams. The output RDF streams can then be fed into any RSP engine, and the relational streams can be used by other relational stream processing systems.

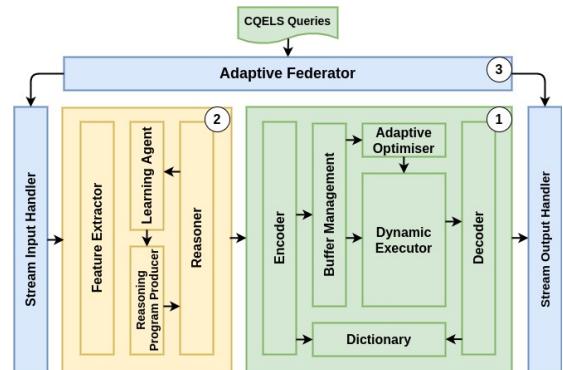


Fig. 5: Component Design of ThothSP.

Figure 5 illustrates the component design of the ThothSP framework. In general, ThothSP consists of three subsystems. Subsystem ①, the RSP processor, extends the RSP primitives of its previous version [3] to accelerate the grounding phase of a SSR. For example, multiway joins are used to accelerate the incremental ground techniques [18]. The second subsystem ② is the SSR component as presented in Section II. Finally, subsystem ③ is a stream adaptive federator, which is described in the remainder of the section.

To tailor the RDF-based data processing operations on edge devices (e.g. ARM CPU, Flash-storage), ThothSP can be integrated with RDF4Led [6], a RISC style RDF engine for lightweight edge devices, to build the Adaptive Federator based on [7]. The each edge-based *Thoth·runtime* deployment is smaller than 10MB in size and needs only 4–6 MB of RAM to process millions of RDF triples on computationally constrained devices, such as BeagleBone [19], Raspberry PI [20]. Therefore, ThothSP includes an adaptive federation mechanism to enable

the coordination of different hardware resources, which in turn enables the construction of query processing pipelines by cooperatively delegating partial workloads to their peer agents. For elastically scale processing load the cloud infrastructure, ThothSP can use Apache Flink as the underlying software stacks for coordinating parallel execution processes using the approach in CQELS Cloud [21]. This will lay the foundation for an integration of the adaptive optimizer with the cloud-based stream scheduler and operation allocations.

The Adaptive Federator acts as the query rewriter, which dynamically divides the input query into sub-queries. The rewriter then pushes the sub-queries as close to the streaming source device as possible by following the predicate push down approach used in common logical optimisation algorithms. The metadata subscribed to by the other *Thoth*-*runtime* is stored locally. Similar to [22], such metadata allows the endpoint services of a *Thoth*-*runtime* to be discovered via the Adaptive Federator. When the Adaptive Federator sends out a sub-query, it notifies the Stream Input Handler to subscribe and listens to the results returning from the sub-query. On the other hand, the Stream Output Handler sends out the sub-queries to other devices or sends back the results to the requester.

IV. A CASE STUDY ON DISTRIBUTED CAMERA NETWORK

A. Build a SSR program for Multi Object Tracking(MOT)

In computer vision, a MOT pipeline is normally programmed in C/C++ or Python. With ThothSP, the widely used tracking-by-detection approach [23] can be emulated as an \mathcal{M}^{MOT} in a declarative fashion with rules and queries, e.g. SORT in Listing 5. The key operations in tracking-by-detection approaches are as following: 1) detection of objects (using DNN-based detector), 2) propagating object states (location and velocity) into future frames, 3) associating current detection with existing.

Figure 2 illustrates the SORT [24] algorithm which is a simple object tracking algorithm based on DNN detectors such as SSD [25] or YOLO [26]. To associate resultant detections with existing targets, SORT uses a Kalman filter [27] to predict the new locations of targets in the current frame. At time point 2, the red boxes b_1 and b_3 are newly detected, and the yellow boxes b_2 and b_4 are predicted by a Kalman filter based on the tracked boxes from the previous frame. Then, the SORT algorithm computes an associative cost matrix between detections and targets based on the intersection-over-union (IOU) distance between each detection and all predicted bounding boxes from the existing tracklets. In case some detection is associated with a target, the detected bounding box is used to update the target state via the Kalman filter. As in frame 2, the tracklets trk_1 and trk_2 are set to the two new bounding box b_1 and b_2 which are associated with predicted boxes b_3 and b_4 respectively. Otherwise, the target state is simply predicted without correction using the linear velocity model. For example, at time point 3, the detector misses detecting the white car due to an occlusion, however, the tracklet 2 is till assigned to box b_7 which contains part of the white car.

To associate a detected bounding box B with an object O , we use soft rules that assert the triple $\ll B \text{sosa :isSampleOf } O \gg$.

Such rules can be used to represent *hypotheses* on temporal relations among detected objects in video frames following a tracking trajectory. Particularly, the object's movement is consistent with the constant velocity model, e.g., the Kalman filter used in SORT [24], and there is a detection associated with its trajectory, the fact $\ll B \text{sosa :isSampleOf } O \gg$ is generated. Here, $iou(B_1, B_2)$ states the IOU (intersection over union) condition of the bounding boxes B_1 and B_2 satisfy.

```

1 ssr:rule_w_2 a sh:NodeShape ;
2 sh:rule [
3   a sh:CQELSRule ;
4   sh:prefixes ssr: ;
5   sh:construct """
6   CONSTRUCT { ?B1 sosa:isSampleOf ?O . }
7   WHERE { ?STREAM <:ssr>
8     { <<?Dt :det ?B2 >> @ ?T; :score ?S .
9       <<?Trk :trk ?B1 >> @ ?T.
10      ?Trk :trklet ?O . }
11   FILTER (?S>0.8 && iou (?B1,?B2) > 0.8)}""";
12 ]

```

Listing 5: A soft rule to emulate SORT algorithm [24].

Similarly, we can also emulate DeepSORT [28] via another soft rule that can search for more supporting evidences to link a newly detected bounding boxes from an occluded tracklet using visual appearance associations, e.g. frames 2 and 4 of Figure 2. For this, one can search for pairs of bounding boxes from recently occluded tracklets w.r.t. visual appearance (more details in [8]).

B. Federating a SSR Program in *Thoth**swarm

Next, we report the implementation of our case study on a distributed camera network. This network is built based on the data provided by the AI City Challenge (AIC) [29]. The AIC has 40 cameras spanning across 10 intersections in a mid-sized US city. The baseline to compare the efficiency and scalability of *Thoth**swarm is the traditional device-cloud (DC) infrastructure that ingests data from these 40 cameras connected by 40 roadside units (RSU) into a centralized server (the red node as shown in Figure 6(b)). The red node is a powerful server (2 Intel Xeon Silver 4114 Processors, 1TB RAM, V100 GPU cards with 16GB) representing for a cloud infrastructure. In this setup, the red node will handle all workload of the above \mathcal{M}^{MOT} program including its SORT and DeepSORT algorithms. An RSU runs on a Raspberry Pi 3 Model B which is used only for encoding/decoding video streams from its attached camera in this baseline. To simulate camera streams, we used these RSUs to replay the recorded data from AIC at a speed of 10 frames per second (fps) for each camera.

Then, *Thoth**swarm setup adds 8 Jetson Nanos (JNs) [30] to DC setup for offloading heavy processing load of DNN-based object detections the DC setup, called device-edge-cloud (DEC). We use the Yolov5 pre-trained model for both DC and DEC setups. A group of RSUs will be connected to a JN via a wired network if they are geographically collocated to an intersection.

When a \mathcal{M}^{MOT} is registered to the red node of the *Thoth**swarm, it delegates some subtasks to the blue nodes representing for Jetson Nano devices. For examples, subtasks on object detections and tracking are specified as in line 9 and line 10 of Listing 5. As we can see in Figure 6(c), when the number of RSUs increases to more than 10, *Thoth**swarm starts to

perform better than a DC counterpart in terms of latency. Note that, even DEC has more processing power than DC, but adding one more communication hop to the network topology will add more delay if the red node is not overloaded in terms processing or bandwidth. However, for the most heavy operation Yolov5, a V100 can process 100-250 fps, while each Jetson nano can process 10-25 fps. Consequently, the benefit of having edge node is getting clearer when having more processing load and network demand, i.e. streaming from more than 10 cameras.

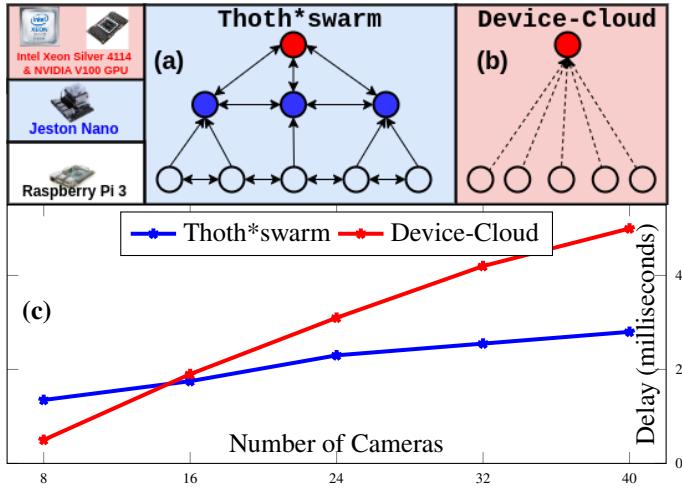


Fig. 6: Scaling behaviour of *Thoth*swarm* vs. device-cloud approach.

V. CONCLUSION AND OUTLOOK

The paper introduces the ThothSP framework, aiming to materialize the SP paradigm, wherein developers can concentrate solely on the semantic symbols abstracting away the complexities of DataOps, NetOps, and MLOps on multimodal data streams. Hence, ThothSP simplifies the development process with declarative queries and rules composed from such symbols. An initial implementation was presented, accompanied by a case study on a distributed camera network showing that a *Thoth*swarm* can dynamically offload the processing load to the edge devices closer to the sensing devices.

While the initial implementation of ThothSP within the context of the Horizon Europe project, SmartEdge, has shown a promising scaling behavior and improved programability, there are several exciting and challenging endeavors we are eager to explore to ensure ThothSP works robustly in SmartEdge use cases, such as Smart Factories, Smart Traffic, and ADAS. The first challenge is effectively managing the dynamicity of the *Thoth*swarm* when a *Thoth-node* joins or leaves the peer-to-peer network (e.g., V2X or mobile robots) without disrupting any services that do not rely on it. Addressing this issue is crucial to maintain smooth operations in dynamic environments.

The second challenge involves developing optimization algorithms to adaptively optimize the federated computing pipeline on distributed streams of multi-modal data. This is essential to ensure optimal performance and resource utilization in the face of constantly changing data patterns and processing demands. Finally, we aim to automate the process of building SSR

programs using both semantic knowledge and large language models amid the current rapid development of AI technologies. Automating this process will expedite program development and enhance the efficiency of ThothSP in handling complex and diverse data and deployment scenarios.

ACKNOWLEDGMENT

This work is supported by the German Research Foundation (DFG) under the COSMO project (grant No. 453130567), and by the European Union's Horizon RIA under the grant agreement No. 101092908 (SmartEdge) and WINDERA programme under the grant agreement No. 101079214 (AIoTwin).

REFERENCES

- [1] “Semantic low-code programming tools for edge intelligence.” [Online]. Available: <https://www.smart-edge.eu/>
- [2] D. Le-Phuoc and M. Hauswirth, *Semantic Stream Processing and Reasoning*. Cham: Springer International Publishing, 2020, pp. 1–10.
- [3] D. Le-Phuoc and et al., “A native and adaptive approach for unified processing of linked streams and linked data,” *ISWC’11*, 2011.
- [4] A. Haller and et al., “The modular SSN ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation,” *Semantic Web*, 2019.
- [5] World Wide Web Consortium, “Web of things (wot) – architecture.” [Online]. Available: <https://www.w3.org/TR/wot-thing-description/>
- [6] A. Le-Tuan, C. Hayes, M. Wylot, and D. Le-Phuoc, “Rdf4led: An rdf engine for lightweight edge devices,” in *IOT ’18*, 2018.
- [7] M. Nguyen-Duc and et al., “Autonomous rdf stream processing for iot edge devices,” in *JIST*, 2020.
- [8] D. Le-Phuoc, T. Eiter, and A. Le-Tuan, “A scalable reasoning and learning approach for neural-symbolic stream fusion,” *AAAI*, 2021.
- [9] M. Nguyen-Duc and et al., “Towards autonomous semantic stream fusion for distributed video streams,” in *DEBS*, 2021.
- [10] O. Hartig and et al., “RDF-star and SPARQL-star.” [Online]. Available: <https://www.w3.org/2021/12/rdf-star.html>
- [11] M. Dao-Tran and D. Le Phuoc, “Towards enriching cqels with complex event processing and path navigation.” in *HiDeSt@ KI*, 2015.
- [12] Dell’Aglio and et al, “A query model to capture event pattern matching in rdf stream processing query languages,” in *EKAW*, 2016.
- [13] Z. Qin, X. Tao, J. Lu, and G. Y. Li, “Semantic communications: Principles and challenges,” *CoRR*, vol. abs/2201.01389, 2022.
- [14] D. L. Greenberg and M. Verfaellie, “Interdependence of episodic and semantic memory: Evidence from neuropsychology,” *JINS*, 2010.
- [15] R. L. Gregory, “Perceptions as hypotheses,” *JSTOR. Series B, Biological Sciences*, vol. 290, no. 1038, pp. 181–197, 1980.
- [16] K. Friston and et al., “Perceptions as hypotheses: Saccades as experiments,” *Frontiers in Psychology*, 2012.
- [17] R. Tommasini and et al., “Vocals: Vocabulary and catalog of linked streams,” in *ISWC*, 2018.
- [18] D. Le-Phuoc, “Adaptive optimisation for continuous multi-way joins over rdf streams,” *The Web Conference*, 2018.
- [19] [Online]. Available: <https://www.beagleboard.org/>
- [20] [Online]. Available: <https://www.raspberrypi.org/>
- [21] D. Le-Phuoc and et al., “Elastic and scalable processing of linked stream data in the cloud,” in *ISWC*, 2013.
- [22] D. Dell’Aglio, D. Le Phuoc, A. Le-Tuan, M. Intizar Ali, and J.-P. Calbimonte, “On a web of data streams,” 2017.
- [23] G. Ciaparrone and et al., “Deep learning in video multi-object tracking: A survey,” *Neurocomputing*, 2019.
- [24] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, “Simple online and realtime tracking,” in *ICIP*, 2016, pp. 3464–3468.
- [25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Fu, and A. C. Berg, “SSD: single shot multibox detector,” in *ECCV*, 2016, pp. 21–37.
- [26] J. Redmon and A. Farhadi, “YOLO9000: better, faster, stronger,” in *CVPR*, 2017, pp. 6517–6525.
- [27] G. Bishop, G. Welch et al., “An introduction to the kalman filter,” *Proc of SIGGRAPH, Course*, vol. 8, no. 27599-23175, p. 41, 2001.
- [28] N. Wojke, A. Bewley, and D. Paulus, “Simple online and realtime tracking with a deep association metric,” in *ICIP*, 2017, pp. 3645–3649.
- [29] Z. Tang and et al., “Cityflow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification,” in *CVPR*, 2019.
- [30] [Online]. Available: <https://developer.nvidia.com/embedded/jetson-nano>