BENCHMARKING SPACE-BASED DATA CENTER ARCHITECTURES

Michał Gumiela[1], *Alicja Musiał*[1], *Agata M. Wijata*[1,2], *Dawid Lazaj*[1], *Jonas Weiss*[3], *Patricia Sagmeister*[3], *Thomas Morf*[3], *Martin Schmatz*[3] and *Jakub Nalepa*[1,2]

[1]KP Labs, Konarskiego 18C, 44-100 Gliwice, Poland
[2]Silesian University of Technology, Akademicka 2A, 44-100 Gliwice, Poland
[3] IBM Research Europe – Zurich, Säumerstrasse 4, 8803 Rüschlikon, Switzerland

IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-2010-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IGARSS52108.2023.10281701

ABSTRACT

The rapid growth of the space industry is creating an increasing amount of data in orbit, which at the current time is not matched yet by a corresponding increase in data download capacity. Using data analytics at the edge, i.e., data processing on-board satellites, holds the promise to significantly mitigate this problem. Newly available and highly efficient artificial intelligence (AI) hardware accelerators and other off-the-shelf compute-hardware has already been successfully exploited for demonstrating satellite on-board deep learning. Aggregating such components and technologies at scale and introducing resource sharing concepts beyond individual spacecrafts, will yield the equivalent of a space data center—a space-based system that will collect, process, store, and relay data from own sensors or from "client" satellites and work in orchestration with other SDCs in a network. In this paper, we outline new opportunities for such data- and compute-sharing schemes in space and how current limitations of existing systems can be overcome.

Index Terms— Earth observation, space data center, mothership and scout, fire detection, on-board processing.

1. INTRODUCTION

Traditional EO satellite missions employ a data-producer pipeline. The satellite is therefore a single source of information. In response to telecommands, it acquires data using its sensors and transmits it to ground stations [1]. Constellations or trail-formations often consist of a number of such orbiters, working independently or in preprogrammed sequences, to provide the data-streams to the operators [2]. The main drawback and limitation of this traditional approach is bandwidth availability to transfer the data to the ground [3]. This issue is amplified by usually only few ground stations which are available along the spacecraft trajectory, specifically on lower earth orbit (LEO), where the passes above specific locations are rare and short. There is no scaling of this type of missions without additionally widely deploying a network of ground stations, without which the average data-latency of applications tied to such satellites may be high.

The concept of space data centres (SDCs) offers, besides all advantages of bringing "terrestrial cloud & edge computing" to space, to access the sparsely distributed ground stations over multi-hop-connections through a network of satellites virtually at any time, facilitating near-real-time Earth observation (EO) applications [4]. To do so, identifying the right technologies that will operate reliably in the energy, thermal and communication bandwidth constrained space-environment will be key [4]. In view of a long term SDC strategy, we break down system-level requirements to current and anticipated future technologies and identify technology gaps on current roadmaps and discuss how to address them [3, 5]. We outline new opportunities that may emerge if such data- and compute-sharing schemes in space are adopted and how current limitations of existing systems can be overcome, based on a concrete use case of an SDC. Among the many possible use case scenarios, we picked one example from the field of EO, the detection of wildfires, which covers many relevant SDC aspects also seen in other scenarios.

2. SPACE DATA CENTER: A CASE STUDY

Standard EO missions consist of pre-planned observations or continuous imaging by satellites [6]. The collected data is compressed and stored in the satellite's memory/storage. As the satellite passes over the ground station, the stored data is sent to Earth where it is further processed. This approach has several shortcomings, including a long latency (response). Another limitation is the inability to dynamically allocate resources/resolution of the on-board sensor to areas of interest, i.e., the surface is scanned "evenly" with the pre-set resolution. This produces e.g., high-resolution images of cloudy areas which are of little value [7]. An example of a solution to these limitations is the use of a duo of a mothership satellite equipped with a high-resolution sensor and a low-resolution reconnaissance satellite. This allows to reduce reaction times

This work was funded by the European Space Agency. AMW and JN were co-financed by the Silesian University of Technology grant for maintaining and developing research potential (AMW: 07/010/BK_23/1023).

in emergency situations [8], which is crucial e.g. for effective disaster management. It overcomes the bandwidth bottleneck by on-board processing and inter-satellite communication, to transmit actionable insights or highly contextualized images instead of raw data, and in turn helps to minimize implementation costs [9].

2.1. Architecture of the system

The mothership and scout concept is depicted in Fig. 1. This centers around a formation consisting of at least two spaceships, namely a mothership (in the form of a microsatellite) and one or more scouts (nanosatellites) [8]. The spacecraft maintain a common orbit but are approx. 1000-2000 km apart, which yields a flight interval of 2-4 minutes. The scout satellite is equipped with a multispectral optical instrument with low spatial resolution (about 100 m/px, depending on the selected spectral range and the requirements of the application), but with a wide field of view (e.g., 300-600 km). This allows to maintain continuous imaging of the Earth's surface. The resulting multispectral images (MSI) are then processed on board the scout satellite or if enough bandwidth between the scout and the mothership is available, directly transferred in raw format to the mothership. The trade-off between the two scenarios will be on a case-by-case level, accounting for cost and power-budget in function of bandwidth and communication distance. In either case, relatively simple edge-algorithms would be deployed for removing cloudy areas [7] and for detecting regions of interest (ROI), e.g., wild fires [10]. The compressed data, e.g. in the form of segmentation maps and basic georeferencing information are then transferred to the mothership (equipped with a high resolution, e.g. 1-3 m/px and narrow FOV 5-10 km hyper- or multispectral instrument), which changes its orientation to observe the selected area at the maximum assumed viewing angle of about $\pm 30°$. The mothership may have the ability to acquire data also on different spectral bands than the scout satellite, e.g. using hyperspectral imaging (HSI), and run more sophisticated algorithms on its powerful data processing units. The performance of the data processing unit and the inter-satellite link bandwidth are the bottlenecks—they must facilitate extraction and the transfer of information between satellites in a short time (less than a few minutes), to re-orient the mother-ship's optical instrument.

Once the mothership has acquired the high-resolution observations, more potent algorithms on board the mothership may be used to extract actionable insights. The latter can be efficiently streamed to ground stations with low bandwidth requirements and thus with lower latency. If there is no ground-station in reach, data can be forwarded to other connected instances of SDCs with better ground connection to facilitate almost real-time insight availability for downstream apps.

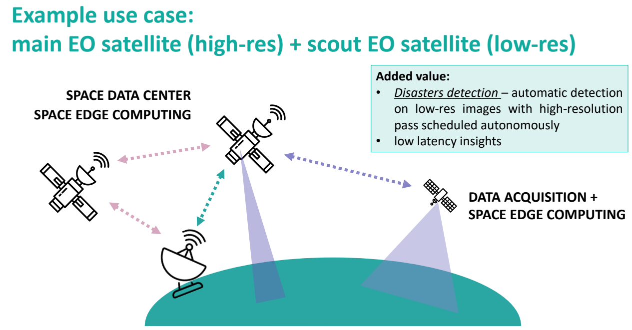

Fig. 1: A possible use case of SDCs—a mothership microsatellite and scout nanosatellite.

2.2. Application of the SDC: forest fires

Forest fires have increased over the last decade, caused by climate warming and are causing damage in the billions of Euros annually [10]. It is thus the topic of extensive academic and commercial efforts to early detect and fight them [11]. With its near-real-time (continuous) and low latency observation requirement, it is a good example use case for the previously discussed scout-mothership architecture. Besides detecting them at an early stage, this architecture also allows to (almost) continuously track and predict the development of the extent and trajectories of already detected fires, and where necessary define multiple ROIs to e.g., in great detail assess accessibility for firefighters or the condition of escape routes for the population. The consumers of these insights/information would mainly be emergency services, rescue teams and local governments and support the latter in their crisis management.

The use of satellite remote sensing, facilitates continuous detection and tracking of wildfires at scale, which would either not be possible or economical if it had to be done at this scale using terrestrial monitoring techniques. Never-the-less, it still allows to include local subject matter-experts in the decision process by using modern visualization and decision support tools. The algorithms used to detect and predict forest fires [12] employ both empirical- and machine learning models [13], often with MSI input data. Thus, MSI imaging with a large GSD allows to detect the location of a potential fire while monitoring a large area of the Earth [14]. The higher the spatial and spectral resolution, the more accurate the information about active fire locations and the level of damage, which allows inferring and predicting the direction of the fire. Fire surface segmentation using HSI imaging is successfully implemented using deep learning as well [15]. However, computational cost and time for segmenting a fire from higher-dimensional data will naturally also be higher, thus imposing higher compute requirements on the SDC or even requiring a federated compute approach including multiple SDCs. The latter applies specifically/only during times of large area droughts with increased fire risk. Outside these times, the compute and link-bandwidth of the system may be available for other applications like precision farming, (illegal) deforestation or even pouching. This also outlines one of the great benefits of sharing compute and communication capabilities in space by deploying SDCs.

While for many "close-to-real-time" apps only small data-fragments need to be transmitted, many scientific apps can profit from large "raw-data" analysis or they can be valuable for model-training too. With increased ground-station availability through a connected SDCs the raw-data can either be used for online model-training within SDCs or transmitted over the network to ground applications, whenever downlink capacity is available. In consequence constantly improved models will be available at the edge through different means of training. To assess design decisions regarding architecture and SDC implementation for use cases such as the one described above, we are currently building a simulation tool.

3. SIMULATOR

To understand the value of our distributed edge-computing infrastructure for specific use-cases and to estimate the cost to operate it, there are key insights to convince stakeholders to invest and deploy such systems. Unfortunately, such assessments are nothing but straight forward and need to factor-in large numbers of variables, of which many may or may not be known. In either case, such assessments are always beyond a "back-of a napkin" type endeavor and require a systematic approach, which is outlined in the following. To develop a suitable model, we take a data-center- and distributed edge computing perspective, and then constrain our model(s) with satellite-constellation-, individual satellite-, inter-satellite communication- and ground-segment requirements.

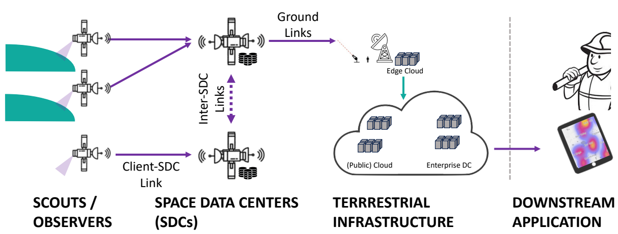

Fig. 2: SDCs as connected compute nodes in an edge-extended cloud scheme.

As depicted in Fig. 2, SDCs assume a role in the web of connected data-centers (DCs) similar to any other DC. In terrestrial settings, connectivity within the cloud can be implemented with extremely high bandwidth, in a static manner. SDCs however will be constraint by a finite number of potentially highly dynamic, i.e., available/unavailable connections to both, other SDCs but also to their client data-source, i.e., scout or observer satellites. The latter becomes obvious when considering e.g., two SDCs on orthogonal orbits. In addition to the highly constrained connectivity, the individual SDC instance will be mostly power- and thermally constrained, as a function of the satellite size and the availability of energy drawn from a limited size solar sail. To fully explore these limitations, our model will be capable of including the energy-efficiency "envelope" of different compute architectures. The models will exploit technologies such as CPUs, GPUs, TPUs, ASICS, and we will try to give an outlook into quantum computing, by mapping anticipated workloads, mostly machine learning- and AI-based, to mentioned compute platforms. A key to a broad adoption of such systems is employing open architectures and open interfaces and data/communication standards. This will allow interested players to develop their own components and plug them into an existing eco-system and tap into the value-chain. Exemplary to this platform approach, we propose to build SDCs on top of the Linux operating (eco-)system and deploy and manage apps using docker-container-like schemes, potentially with light-weight derivatives of the Kubernetes Control Plane for workload and data-stream management. This approach will facilitate creating distributed apps with dynamic workload migration to e.g., SDC nodes which are either closest to the data-source or to the ground-connection. To map all considerations into a model suitable for benchmarking different system design decisions, we are pursuing a hierarchical/sequential approach (Fig. 3), which will yield a few critical figures of merit like, cost per compute, power per compute and cost per power spent. By implementing this scheme in an openly available spreadsheet, we hope that it will be useful and accessible for anyone who wants to experiment with it, but also allow anyone to check, validate or improve and extend the model, to account for changes in future technologies or account for unforeseen use-cases.

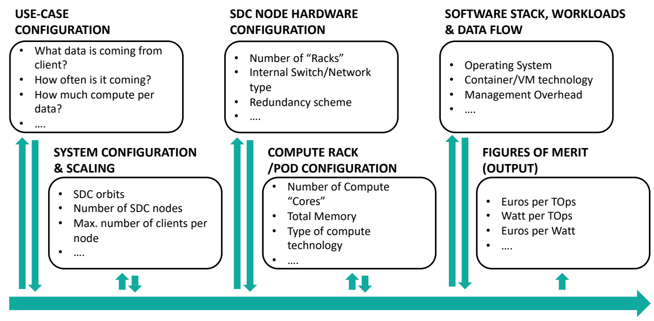

Fig. 3: Model building blocks and configuration inputs.

4. CONCLUSIONS

While SDCs will be capable of running arbitrary algorithms, we anticipate deep-learning algorithms [11], e.g., semantic segmentation [7] and related computer-vision workloads to be predominant. They are generically applicable for many EO tasks, but specifically effective for natural disaster detection, environmental monitoring or e.g., for detecting anomalies [6]. Exploring next-gen GPUs, neuromorphic and in-memory hardware AI accelerators for in-space applications will thus be a natural fit [16]. With data-aggregation from multiple sources, distributed computing schemes, federation [3], including capabilities on the ground-leg must be considered. While general edge-cloud computing schemes can be adopted for SDCs, additional constraints for communication links and available compute/storage resources must be included, requiring more energy efficient and network independent node and system-architectures [4]. With our analysis-tool, we assess system-architectural differences and technology choices to benchmark different scenarios for different use-cases. Such benchmarking will help make the right decisions when laying out the blueprints for future SDC infrastructure projects. They will also help identify new opportunities that may profit from the availability of dedicated, distributed and on-demand compute capabilities in space.

5. REFERENCES

[1] Bradley Denby and Brandon Lucia, "Orbital edge computing: Nanosatellite constellations as a new class of computer system," New York, NY, USA, 2020, ASPLOS '20, p. 939–954, ACM.

[2] Brandon Lucia, Brad Denby, Zachary Manchester, Harsh Desai, Emily Ruppel, and Alexei Colin, "Computational nanosatellite constellations: Opportunities and challenges," *GetMobile: Mobile Comp. and Comm.*, vol. 25, no. 1, pp. 16–23, jun 2021.

[3] Cong Peng, Yuanzhi He, Shanghong Zhao, Lingyang Song, and Boyu Deng, "Integration of data center into the distributed satellite cluster networks: Challenges, techniques, and trends," *IEEE Network*, pp. 1–8, 2022.

[4] Gianluca Furano, Gabriele Meoni, Aubrey Dunne, David Moloney, et al., "Towards the use of artificial intelligence on the edge in space systems: Challenges and opportunities," *IEEE Aerospace and Electronic Systems Magazine*, vol. 35, no. 12, pp. 44–56, 2020.

[5] S. N. Stepanov, A. V. Korobkina, A. O. Volkov, E. E. Malikova, and A. E. Panov, "The analysis of traffic balancing for data centers serving requests of leo mobile satellite systems," in *2021 Systems of Signals Generating and Processing in the Field of on Board Communications*, 2021, pp. 1–6.

[6] Nicolas Audebert, Bertrand Le Saux, and Sebastien Lefevre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 159–173, 2019.

[7] Bartosz Grabowski, Maciej Ziaja, Michal Kawulok, Marcin Cwiek, Tomasz Lakota, Nicolas Longépé, and Jakub Nalepa, "Are cloud detection u-nets robust against in-orbit image acquisition conditions?," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 239–242.

[8] Elizabeth Gibney, "Cubesats set for deep space — if they can hitch a ride," *Nature*, vol. 535, pp. 19–20, 2016.

[9] Gianluca Giuffrida, Luca Fanucci, Gabriele Meoni, Matej Batič, et al., "The ϕ-sat-1 mission: The first on-board deep neural network demonstrator for satellite earth observation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[10] Nguyen Thanh Toan, Phan Thanh Cong, Nguyen Quoc Viet Hung, and Jun Jo, "A deep learning approach for early wildfire detection from hyperspectral satellite images," in *Proc. RiTA*, 2019, pp. 38–45.

[11] Claudio Persello, Jan Dirk Wegner, Ronny Hänsch, Devis Tuia, Pedram Ghamisi, Mila Koeva, and Gustau Camps-Valls, "Deep learning and earth observation to support the sustainable development goals: Current approaches, open challenges, and future opportunities," *IEEE Geoscience and Remote Sensing Magazine*, vol. 10, no. 2, pp. 172–200, 2022.

[12] Yongqi Pang, Yudong Li, Zhongke Feng, Zemin Feng, Ziyu Zhao, Shilin Chen, and Hanyue Zhang, "Forest fire occurrence prediction in china based on machine learning methods," *Remote Sensing*, vol. 14, no. 21, 2022.

[13] Panagiotis Barmpoutis, Periklis Papaioannou, Kosmas Dimitropoulos, and Nikos Grammalidis, "A review on early forest fire detection systems using optical remote sensing," *Sensors*, vol. 20, no. 22, 2020.

[14] Amirhossein Rostami, Reza Shah-Hosseini, Shabnam Asgari, Arastou Zarei, Mohammad Aghdami-Nia, and Saeid Homayouni, "Active fire detection from landsat-8 imagery using deep multiple kernel learning," *Remote Sensing*, vol. 14, no. 4, 2022.

[15] Kathiravan Thangavel, Dario Spiller, Roberto Sabatini, Stefania Amici, Sarathchandrakumar Thottuchirayil Sasidharan, Haytham Fayek, and Pier Marzocca, "Autonomous satellite wildfire detection using hyperspectral imagery and neural networks: A case study on australian wildfire," *Remote Sensing*, vol. 15, no. 3, 2023.

[16] Jakub Nalepa, Marek Antoniak, Michal Myller, Pablo Ribalta Lorenzo, and Michal Marcinkiewicz, "Towards resource-frugal deep convolutional neural networks for hyperspectral image segmentation," *Microprocessors and Microsystems*, vol. 73, pp. 102994, 2020.