Poster Abstract: C-Continuum: Edge-to-Cloud computing for distributed AI

Davide Aguiari*, Andrea Ferlini†, Jiannong Cao‡, Song Guo‡ and Giovanni Pau§

*Dept. of Electrical and Electronic Eng, University of Cagliari, Cagliari, IT
Email: davide.aguiari@unica.it
†Dept. of Computer Science and Technology, Cambridge University, Cambridge, UK
Email: af679@cam.ac.uk
‡Dept. of Computing, Hong Kong Polytechnic, Hong Kong, HK
Email: {csjcao, song.guo}@comp.polyu.edu.hk
§ LIP6 Sorbonne Université, Paris, FR & UCLA CS Dept., US
Email: giovanni.pau@lip6.fr

***Index Terms*—cloud computing, edge computing, name data networking, v2i, v2v**

INTRODUCTION

Mobile autonomous systems are supposed to deeply impact in manufacturing, space exploration, rescue, defense, transportation, and everyday life. Autonomous air-ground vehicles, for example, will become normal tools in the next few years, providing a natural platform for distributed artificial intelligence applications including, for example, disaster rescue and recovery, area surveying, autonomous driving, etc. The raise of autonomous cooperating robots will pose new challenges in networking, distributed systems and resource management. Heavy computational tasks will be dispatched to the closest edge node for processing and the core-cloud will be involved as last resort in an effort to reduce latency and increase the global system capacity leveraging application and resource locality. Massive amounts of data and computations will be required. For example, in the autonomous driving scenario Intel estimates that each driver-less vehicle will produce over 4 TeraBytes of data each day[1]. While most of this data is consumed in-car, cooperating autonomous vehicles will have to exchange some percentage of the 4TB and eventually offload some computation and data to the local edge or the core-cloud. This is particularly relevant when locally gathered and labeled data can be used to refine the model and ultimately increase the global "intelligence". This approach is often taken by autonomous driving automakers.

Distributed AI applications call for effective, seamless, and efficient communication and computation mechanisms across the whole computing spectrum edge, fog, and cloud. In figure 2, for example, computational tasks, AI models and relative data may be instantiated on ground vehicles, in air-drones, and/or in the local edge DC. The allocation may and will change frequently during a single execution.

[1]https://newsroom.intel.com/editorials/self-driving-cars-big-meaning-behind-one-number-4-terabytes

Here, we outrageously propose *C-Continuum*, a Computing Continuum framework targeting distributed AI in the mobile arena. C-Continuum aims to define a new generation of tools and mechanisms designed to enable fine-granularity computation, coordination and mobility management across the mobile-computing spectrum from the edge to the core. C-Continuum embraces Named Data Networking (NDN) making a case for naming any computational entity and using those names for resource location, data transfers, and computing functions as well [1].

OUR VISION

Uncertainty is one of the most significant discriminant traits of today's society. When it comes to the communication domain, uncertainty usually translates into the absence of stability and reliability. Current industrial and academics trends denote a gradual, yet steady, transition towards a ubiquitous edge computing paradigm. Sudden changes in the computing load may be extremely harmful to CPUs and GPUs. The more we move towards the Edge, the more communications become intermittent and unreliable. Changes in both the network and the availability of the services may occur both locally and in the infrastructure. Nonetheless, cybersecurity threats are increasing. These are only a few among the open-challenges that are coming along with the advent of pervasive edge devices. Fuelled by near deployment of 5G, the Research Community is putting a furious effort to tackle most of the aforementioned challenges. Nevertheless, to the best of our knowledge, the literature is not exhaustive regarding dynamic task and resource allocation approaches for mobile Edge devices.

C-Continuum targets task and resource allocation in heterogeneous mobile heterogeneous scenarios, as vehicles, drones, roadside micro data center (μDC). In particular, C-Continuum:

1) proposes a framework with mechanisms, models, and algorithms to address complex domain-specific tasks such as collaborative mission-planning, computing, and networking issues in Air-Ground Vehicular Networks;

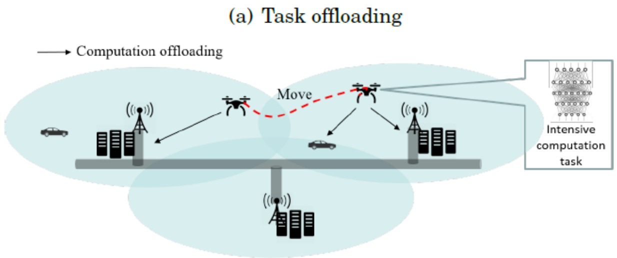

Fig. 1. Multi-Tier Computing across heterogeneous nodes.

2) develops a system platform that provides for the allocation of computing resources and AI models;
3) smoothly integrates unmanned aerial vehicles (UAV), ground vehicles, edge nodes, and AI-based algorithms providing a more reliable communication among them.

Indeed, C-Continuum is the first framework able to:

- request and assign tasks at any level, from in-vehicle specific tasks to data-center ones;
- offer computational resources and light-weight AI-models seamlessly addressable at, and by, any tier, independently.

System Architecture

C-Continuum leverages NDN's typical consumer/producer behaviour to offload and allocate tasks from the user to the local network and beyond. Thanks to NDN features such as multipath forwarding, in-network caching, multicast data delivery, and data authenticity, C-Continuum is able to allocate part of the workload to the closest capable node (i.e. the one which satisfies in the shortest time the initial interest, broadcasted by the user). C-Continuum assigns two different timers to every node. The first one is inversely proportional to the free CPU, whilst the latter is used for collision avoidance purposes. The user sends both the program and its inputs encapsulated into a NDN Data Packet to the best node identified by our framework. Figure 2 shows an example of the NDN workflow.

Caching and computation results reuse minimize the effectiveness of distributed denial of service attacks (DDoS) towards specific targets in the network. Yet, remote software execution comes along with other several major security concerns. In fact, even if every NDN Data Packet is digitally signed with a signature which is part of the Data Packet itself, the execution of malicious software may be fatal for the guest node. Thus, the execution program must necessarily remain inside a sandbox. Although the ultimate safe environment is still under investigation, containers, unikernels and light-weight VMs may fit the problem.

Through the straightforward process described above, C-Continuum can provide heterogeneous services, such as:

1) V2I connectivity: e.g. via Car-Fi [2], C-Continuum can enable multi-access network for vehicles, reducing by a 10 factor the time to access urban community WiFi hotspots;

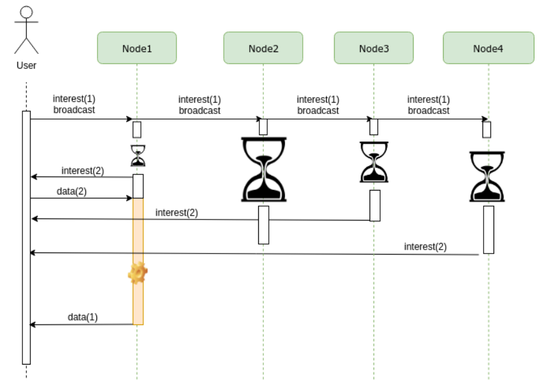

Fig. 2. *The user* asks the network to offload a task via broadcast. Node1, the fastest, replies with an interest asking for the code+input, Node2 spoofs Node1's packet and stops its flow, while Node3 and Node4 reply anyway. Node1 computes the program and sends the output to the user.

2) an ICN fashion to distribute resources on the ground: e.g. using protocols like Navigó [3], an NDN approach to retrieve any name resource by connected cars;
3) AI modules for commodity tasks such as visual analytics: e.g. ParkMaster [4], a roadside parking-spot assistance service that proves how the *C-Continuum*'s edge-computing logic works in the context of Vehicle as a Service Provider.

Acknowledgment

This work has been partially supported by the Italian Ministry of University and Research (MIUR), within the Smart City framework (project: PON04a2_00381 "CAGLIARI2020").

References

[1] C. Tschudin and M. Sifalakis. "Named functions and cached computations". In: *2014 IEEE 11th Consumer Communications and Networking Conference (CCNC)*. Jan. 2014, pp. 851–857. DOI: 10.1109/CCNC.2014.6940518.

[2] Davide Pesavento, Giulio Grassi, Giovanni Pau, Paramvir Bahl, and Serge Fdida. "Demo: Car-Fi: Opportunistic V2I by Exploiting Dual-Access Wi-Fi Networks". In: *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. MobiCom '15. Paris, France: ACM, 2015, pp. 173–175.

[3] G. Grassi, D. Pesavento, G. Pau, L. Zhang, and S. Fdida. "Navigo: Interest forwarding by geolocations in vehicular Named Data Networking". In: *2015 IEEE 16th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. June 2015, pp. 1–10.

[4] Giulio Grassi, Kyle Jamieson, Paramvir Bahl, and Giovanni Pau. "Parkmaster: An In-vehicle, Edge-based Video Analytics Service for Detecting Open Parking Spaces in Urban Environments". In: *Proceedings of the Second ACM/IEEE Symposium on Edge Computing*. SEC '17. San Jose, California: ACM, 2017, 16:1–16:14.