Data management, communication systems and the edge: Challenges for the future of transportation

Communications in Transportation Research 1 (2021) 100024

Contents lists available at ScienceDirect

Communications in Transportation Research

journal homepage: www.journals.elsevier.com/communications-in-transportation-research

Editorial

Data management, communication systems and the edge: Challenges for the future of transportation

ARTICLE INFO

Keywords
Edge cloud
Autonomous vehicles

Edge Clouds. Edge cloud systems have emerged as a new promising alternative to address the needs of many emerging latency-critical applications such as autonomous vehicles. These applications are ill-suited for traditional clouds due to the end-to-end latency and the limited bandwidth between the cloud's (few) data centers and these applications. In the edge cloud model, a myriad of small-scale computing clusters are brought next to the applications at the edge of the network. Many Applications in transportation engineering such as autonomous vehicle collision avoidance, and fleet management, requiring more global decision making for safety and correctness of operation, can thus make use of edge cloud systems, offloading their computations to these edge clusters. The edge can then analyze data from these vehicles to optimize the traffic flow, reduce accidents, and provide transportation systems with more autonomy. The idea of edge computing today forms a cornerstone in the design of many future systems, including, 5G networks and autonomous vehicles, among many others (see Fig. 1).

Edge in Action. Conventional wisdom is that edge provides lower latency than more remote clouds. While the edge can provide significantly lower network latency, edge resources tend to be more constrained than those in cloud data centers. Consequently, workload dynamics can cause edge server latency to exceed the cloud server latency due to higher queuing delays at edge sites compared to a similarly sized cloud. In such cases, the total end-to end response time of the edge can *exceed* the end-to-end cloud response times, since the lower network latency of the edge is offset by a higher server latency at higher utilization levels. This can be shown using queuing theory and experimentally.[1] In addition, the multiple round trips between application components can result in even worse performance.[2]

Hence, while 5G is expected to support millisecond latencies between the applications and the edge, edge utilization can play a much more decisive factor on the performance of running computations on the edge versus on a cloud with much more resources. This "hidden" cost of the edge means that for many edge applications, the resource management problem becomes a problem of where to process the data. Processing the data locally prevents the system from optimizing the overall system. Processing on the edge can result in longer tail latencies.

Transportation Systems on the Edge. While the problems above extend to almost all edge applications, transportation systems in particular face a set of unique choices. To give an example, 5G connected vehicles as a specific application not only amplifies the above challenges but also introduces a new set of challenges. The mobility of the vehicles at high speeds and the limited range of 5G signals means that vehicles can be in areas where coverage is minimal or non-existing. Hence, any data management solution for connected vehicles needs to be designed to handle network and communication failures and disruptions. In addition, one major issue with autonomous vehicles is the amount of data that needs to be transmitted, processed, and stored, often in real-time. Combined, the sensors from a single level-5 autonomous vehicle can produce 40 Gbit/s (19 TB/h).[3] Therefore, central to any edge-based data management system for 5G connected vehicles is the problem of network resource allocation and data transfer. One solution to both these issues could be to process more data locally and have some vehicles act as routers or repeaters to the channel signal, hence forming a Vehicular Ad hoc Network (VANET) between the vehicles and the edge-cloud system. However, routing data via other vehicles or offloading data processing to nearby vehicles would introduce major privacy and security issues that need to be considered in the network and data management solution.

[1] Ahmed Ali-Eldin, Bin Wang, and Prashant Shenoy. 2021. The hidden cost of the edge: a performance comparison of edge and cloud latencies. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC '21). Association for Computing Machinery, New York, NY, USA, Article 23, 1–12. DOI:https://doi.org/10.1145/3458817.3476142.

[2] Chanh Nguyen, Amardeep Mehta, Cristian Klein, and Erik Elmroth. 2019. Why cloud applications are not ready for the edge (yet). In Proceedings of the 4th ACM/IEEE Symposium on Edge Computing (SEC '19). Association for Computing Machinery, New York, NY, USA, 250–263. DOI:https://doi.org/10.1145/3318216.3363298.

[3] See table in: https://www.tuxera.com/blog/autonomous-cars-300-tb-of-data-per-year/.

https://doi.org/10.1016/j.commtr.2021.100024
Received 11 November 2021; Received in revised form 28 November 2021; Accepted 28 November 2021
Available online 13 December 2021

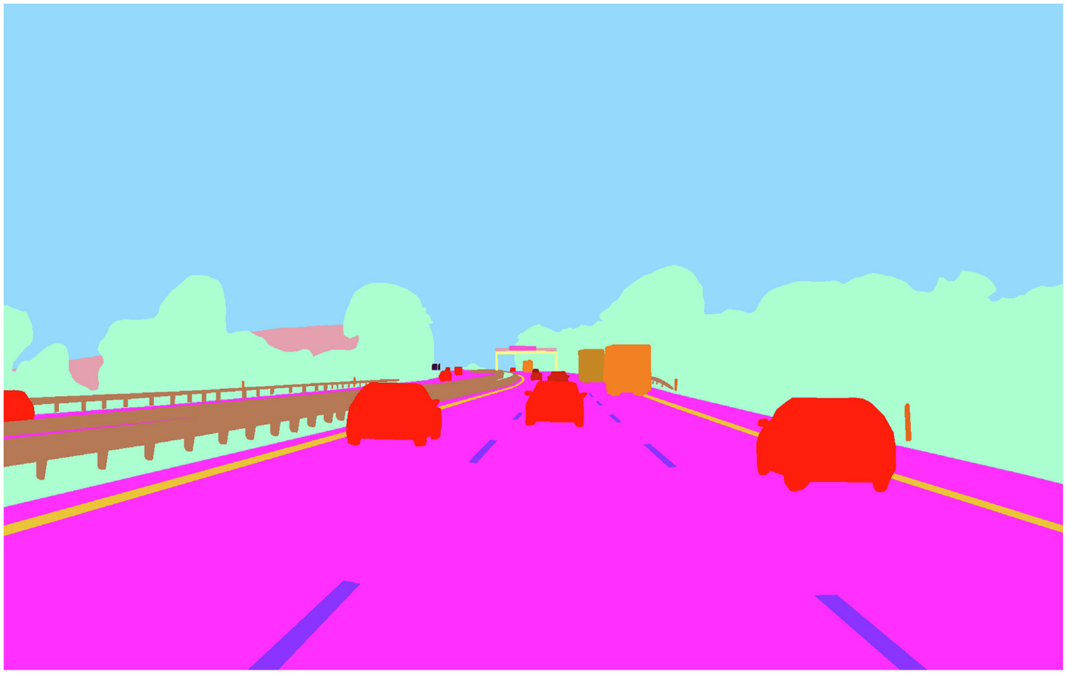


Fig. 1. Segmented camera image from the Audi Open Dataset for autonomous vehicles.

Interesting open problems

The research problems arising from using edge computing for transportation systems are unique in many ways. We believe that in order for transportation systems to reach autonomy, the following research problems need to be addressed.

1. **Where should the data be processed?** Data from transportation systems can be processed locally, on the edge, or on a distant cloud. Besides the compute resources available, the available network bandwidth will be a major factor in this decision. It is crucial for the design and implementation of a fully functioning 5G (and beyond)-based system for connected vehicles to be capable of optimizing the network resource allocation. Each vehicle needs to decide locally on how much data it can send, and possibly what path should that data be routed through, i.e., directly to a nearby base-station, or via other vehicles that act as repeaters. To answer these questions, studying how the network dynamics can affect data transfer to the edge or the cloud for processing taking into account, e.g., wireless resources allocation such as time/frequency slots and mm-wave beams, interference, handover, and mobility of the vehicles, is crucial. Analytical models and bounds on data offloading from the vehicles to the edges based on the network conditions can be used to develop machine-learning based algorithms for managing the data routing problem. One interesting development in data routing is Multi-Path routing protocols[4] which can be utilized efficiently to improve the aggregate data streaming rate, hence, improving the resiliency in the system.
2. **How to efficiently manage the compute capacity of the system?** The above research question tackles how much data can be shared given a certain network condition. However, this does not take into account the compute capabilities of the on-vehicle processing, the edge sites, and the cloud. In many cases, even though the network might have enough bandwidth for sending data, it can be more beneficial to perform most of the processing locally. Deciding on if computations should be done locally versus remotely is highly dependent on the data sensitivity, and the type of processing needed. Distributed computations across the system, with some data processed on-vehicles, some on the edge sites, and some on the clouds, can be modeled as a network of queues.[5] Such models can be further utilized to develop optimization algorithms for the data splitting problem.
3. **How can privacy and security influence the optimization of where to place date and perform processing?** In addition to the above, privacy and security constraints should be considered in all of the above algorithms. Splitting computations between the on-vehicle processors, the edge, and the cloud opens up the possibilities of many attacks and byzantine behaviours. These attacks include falsification of data, dropping data, and delaying data. It is hence important for the routing and optimization to be enhanced to include security and privacy aspects. Hence, techniques such as trust-based routing,[6] and trust-based anonymous communication[7] need to be used to influence the decisions of the data, and compute management algorithms.

A way forward

Managing communication, data, and resources in 5G networks for

[4] Lucas Chaufournier, Ahmed Ali-Eldin, Prateek Sharma, Prashant Shenoy, and Don Towsley. 2019. Performance Evaluation of Multi-Path TCP for Data Center and Cloud Workloads. In Proceedings of the 2019 ACM/SPEC International Conference on Performance Engineering (ICPE ′19). Association for Computing Machinery, New York, NY, USA, 13–24. DOI:https://doi.org/10.1145/3297663.3310295.

[5] HEMANT R. Kanakia, and Fouad A. Tobagi. “On distributed computations with limited resources.” IEEE transactions on computers 36.05 (1987): 517–528.

[6] Fenye Bao, Ing-Ray Chen, Moonjeong Chang, and Jin-Hee Cho. 2011. Hierarchical trust management for wireless sensor networks and its application to trust-based routing. In Proceedings of the 2011 ACM Symposium on Applied Computing (SAC ′11). Association for Computing Machinery, New York, NY, USA, 1732–1738. DOI:https://doi.org/10.1145/1982185.1982547.

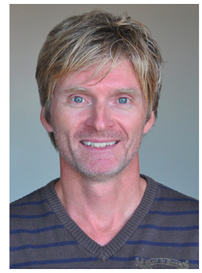
[7] Aaron M. Johnson, Paul Syverson, Roger Dingledine, and Nick Mathewson. 2011. Trust-based anonymous communication: adversary models and routing algorithms. In Proceedings of the 18th ACM conference on Computer and communications security (CCS ′11). Association for Computing Machinery, New York, NY, USA, 175–186. DOI:https://doi.org/10.1145/2046707.20467297 Available at: https://www.a2d2.audi/a2d2/en/download.html.

transportation systems requires a highly inter-disciplinary basic research approach with close collaborations with industry. Methods and techniques from queuing theory, communication systems, machine learning, and security engineering need to be combined to develop a framework for solving the problems at-hand. In this editorial, we articulate some of the problems that need to be solved. Unfortunately, much of the above research has traditionally been done in isolated research communities. We believe that the new Journal "Communications in Transportation Research" will create an interdisciplinary community that can tackle the above set of problems, and more, making use of recent advances in transportation engineering, computer systems, deep machine learning, statistical methods, and control engineering.

Acknowledgments

Partial funding for this work was provided by The Wallenberg Autonomous Systems Program, the Chalmers ICT Area-of-Advance, and the Chalmers AI Research Center.

Ahmed Ali-Eldin is an assistant Professor at the Department of Computer Science and Engineering, Chalmers University of Technology, Sweden. Ali-Eldin obtained his PhD in 2015 from Umeå University, Sweden, and spent 3 years as a postdoc at the University of Massachusetts Amherst. His research interests lies in the broad field of computer systems, with recent focus on edge computing and its application in autonomous vehicles.



Erik Elmroth is Professor in Computing Science and Head of the Department of Computing Science at Umeå University. His extensive leadership experience includes being department head and deputy head for thirteen years and being the Deputy Director for a national supercomputer center (HPC2N) for another thirteen years.He has established the Umeå University research on distributed systems, focusing on theory, algorithms, and systems for the autonomous management of ICT resources, spanning from individual servers to large scale cloud datacenters, federated clouds, highly distributed edge clouds, and software-defined infrastructures.

Ahmed Ali-Eldin*
Chalmers University of Technology, Sweden

Erik Elmroth
Umeä University, Sweden
E-mail address: elmroth@cs.umu.se.

* Corresponding author.
E-mail address: ahmed.hassan@chalmers.se (A. Ali-Eldin).