Towards Autonomous Computer Networks in Support of Critical Systems

Ph.D. Thesis Summary

Candidate: Alessio Sacco
Advisor: Guido Marchetto

The soaring complexity of computer networks has opened up new opportunities, but it also raised new problems. For example, edge computing, which extends the cloud paradigm and moves it closer to the data source (i.e., to the edge of the network), can pave the way for new interactive applications, which are associated with more stringent requirements. Consider, e.g., a remote surgery operation, where the system should lead to an improvement in the accuracy and dexterity of a surgeon while minimizing trauma to the patient. Similarly, a telepathology session, which transmits delay and bandwidth-sensitive data to be processed and shared with a remote medical doctor real-time diagnosis or precomputation of digital pathology. Or also, a response to natural or human-made disaster that involves real-time video conferencing with the incident commander to recognize faces of disaster victims, or the detection of children in an attempt to reunite them with their families. Due to their mission, nature, and the intrinsic requirement for extremely low round-trip delays, these services are often referred to as critical systems. However, to meet the new requirements posed by these systems, it is desirable that the management and orchestration of networks would consider new methodologies to address the new data flow.

A first evolving factor is constituted by the softwarization and virtualization of networks, which have drastically simplified network functions' deployment and real-time reconfiguration and allow networks to quickly adapt to the events. A second factor is represented by the recent management and orchestration approaches for softwarized networks that employ Artificial Intelligence (AI) and Machine Learning (ML) to improve the accuracy of decisions and reduce reaction time. However, while the network operations can be automated to the point of realizing autonomous driving networks, we are still far from having a fully operating and efficient automated network architecture.

In this thesis, we present a novel class of software network solutions that share the goal of enabling intelligent and autonomous computer networks, exploring how to exploit the power of AI/ML to handle the growing complexity of the critical systems. We start with the definition of a new mechanism to route packets in a Software-defined network (SDN), and we demonstrated how machine learning-based models can help in dynamically avoiding over-congested paths. Our proposed system, RoPE, contributes to mitigating congestion and steering the traffic over paths that are predicted to be unloaded.

We then considered a similar problem that regards the planning of network resources. To optimize (virtual) network resource allocation, we proposed Mystique, a system that, learning from the past load on links, establishes the minimal set of active network resources to mitigate congestion, and routes packets over less congested paths. Our network management schema, using Multi-Agent Reinforcement Learning (MARL), aims at auto-scaling the underlying network topology to accommodate the traffic demand and reacts to possible failures. On the one hand, our solution unburdens network nodes that are over-congested with traffic, to preserve the high bandwidth and high availability of the applications. On the other hand, our solution leverages healing strategies to repair failing nodes and links.

With the similar aim of mitigating network congestion, we worked on a new TCP congestion-control algorithm based on the Reinforcement Learning (RL) framework. Different from the aforementioned solutions, in this new approach, the modification occurs at the end-hosts rather than on the network infrastructure. At the same time, in contrast to other Machine Learning-based approaches for transport protocols, we conduct training at the source and decide the next value of the congestion window, also using an in-network mechanism when such information is accessible from the sender.

Lastly, we considered a specific problem originated by the advent of edge computing and highly dynamic networked systems: computation offloading. In particular, we propose an edge cloud-assisted architecture for distributed and adaptive task planning management for Unmanned Aerial Vehicles (UAVs) via task offloading. We present two approaches that can be used as the constraints change. First, we focus on a set of distributed offloading decision strategies and we propose the use of Multi-Agent Reinforcement Learning (MARL) to jointly improve the energy efficiency and task completion time of edge computing enabled UAVs swarms. Then, in the case of task offloading, the model also decides the radio access technology to consume, i.e., Wi-Fi or cellular, to transmit the task from the device to the edge cloud. Second, we design a solution less hungry for computation and memory resources, and that, by predicting future device load, supports the UAV decide whether to offload incoming tasks. Two alternative methods are responsible for performing the prediction about future device load: a model belonging to time-series class and a model belonging to the class of ML regressors. Besides, using a self-learning approach, the agent not only learns how to forecast future values, but it can also learn online which type of model is more accurate.

In light of the obtained results, we can conclude that these solutions, and their combination, can lay the foundation for automated systems that better suit modern edge environments and cellular networks.