On the Orchestration of Dynamic Services over Distributed IT Infrastructures

Ph.D. Thesis Summary

Candidate: Gabriele Castellano
Advisor: Fulvio Risso

The arise of Network Function Virtualization and Software Defined Networking enabled the programmability of the network infrastructure, hence allowing the dynamic deployment of complex services on top of the network. Additionally, Edge Computing extends the Cloud paradigm towards the edge of the network, leading to a highly distributed infrastructure and introducing new players in the service provisioning ecosystem. This transformation leads to enhanced possibilities for service providers, which, thanks to the flexibility and new capabilities offered by the infrastructure below, may realize and deliver a plethora of new applications, such as virtual reality, remote critical tasks, and more. However, managing such a distributed infrastructure and enabling interoperability between the multiple actors involved introduces a series of challenges.

A particularly challenging problem is resource management. Since resources (such as computing, networking, and storage) should be partitioned in slices that are allocated for each service, a key component called Orchestrator is often employed to decide on the deployment and the management of each service. However, the optimality of the taken decisions may not match the actual necessity of the services, as each of them may benefit from different allocation strategies and may want to optimize on different parameters and service-specific metrics. Such metrics are often unknown to the orchestrator, which operates at the infrastructure level and based on a one-size-fits-all paradigm. Moreover, mandating the existence of centralized coordination components may not be suitable in a scenario where services are executed on scattered compute nodes, e.g., at the edge of the network, which features arbitrary and dynamic topologies. Finally, since resources are scarce and geographically distributed in different areas, service provisioning may involve multiple providers that should inter-operate coordinating the deployment of applications on top of their clusters.

Given these considerations, this Ph.D. thesis investigates new service-centric orchestration paradigms, which cover different aspects of the above problem. First, we analyze the scenario of a single Edge provider that deploys and orchestrates services on top of a multi-technological infrastructure. To cope with the problem of exploiting available features from heterogeneous clusters of resources, we propose a capability-based solution, which ensures flexibility and better optimization possibilities in service deployment. Later, we highlight the disadvantages and limitations of a centralized and monolithic orchestration model in Edge Computing. We identify the new actors

involved in service provisioning and their mutual interactions. Contextually, we propose an open and disaggregated model for the business interactions between these new edge actors and some preliminary considerations on their algorithmic optimization.

To overcome the limitations of one-size-fits-all orchestration approaches, we propose a novel Service-Defined Orchestration paradigm, where the orchestration task is distributed among the service providers. Using a formalized declarative language, they may define custom strategies to manage their own services, optimizing on metrics that are service-specific and therefore unknown to a traditional orchestrator that operates at the infrastructure layer. To coordinate such a variety of service providers that orchestrate resources on the same shared infrastructure, we design DRAGON, a Distributed Resource AssiGnment and OrchestratioN optimization algorithm that exploits a fully distributed decision process to reach a dynamic agreement on how resources should be (temporarily) partitioned among the set of involved actors. DRAGON provides guarantees on both convergence time and performance.

The solutions adopted in Edge computing lead to a highly modular and distributed infrastructure that is populated with arbitrary new facilities, generally exploited by service providers to compose the final service. In this regard, we design a configuration layer that enables interoperability between the existing actors, by mean of a model-based approach that facilitates service composition and allows monitoring and tuning of arbitrary components. At last, we overview the applicability of new generation service facilities on highly scattered infrastructures, with particular focus to the Industrial Internet of Things, where services may operate without relying on fixed networks and exploit opportunistic connections between fleets of vehicles. To enable IoT services to transparently operate on such environments we propose a service-oriented communication system that provides service management features and service-aware routing optimization on disrupted networks, with advantages over state-of-the-art algorithms.