A Service Mesh Platform for the Mobile Network Core

Ioannis Fotis, Grigoris Gkionis, Anastasios-Stavros Charismiadis, and Dimitris Tsolkas
Fogus Innovations & Services P.C. and National and Kapodistrian University of Athens

FOGUS INNOVATIONS & SERVICES

Introduction

The current 5G Service-Based Architecture (SBA) necessitates direct interactions among the Network Functions (NFs) of the network core, in a text-based manner (using HTTP). Thus, the overhead for the control plane is becoming heavier. This increase in the signaling load is further exacerbated due to the explosion in the number of connected devices (UEs) and the need for multiplication of the NF instances to support the Network Slicing (NS) concept. In addition, each NF is currently implemented as a monolithic application, which encapsulates several functions into one entity (e.g., the NF service implementation is waved with service exposure and networking functionality). This is one of the factors promoting failure points, affecting resilience and scalability of the network core communication model. In view of these factors, the current network core architectural and implementation approaches should be revisited towards a more modular and scalable model.

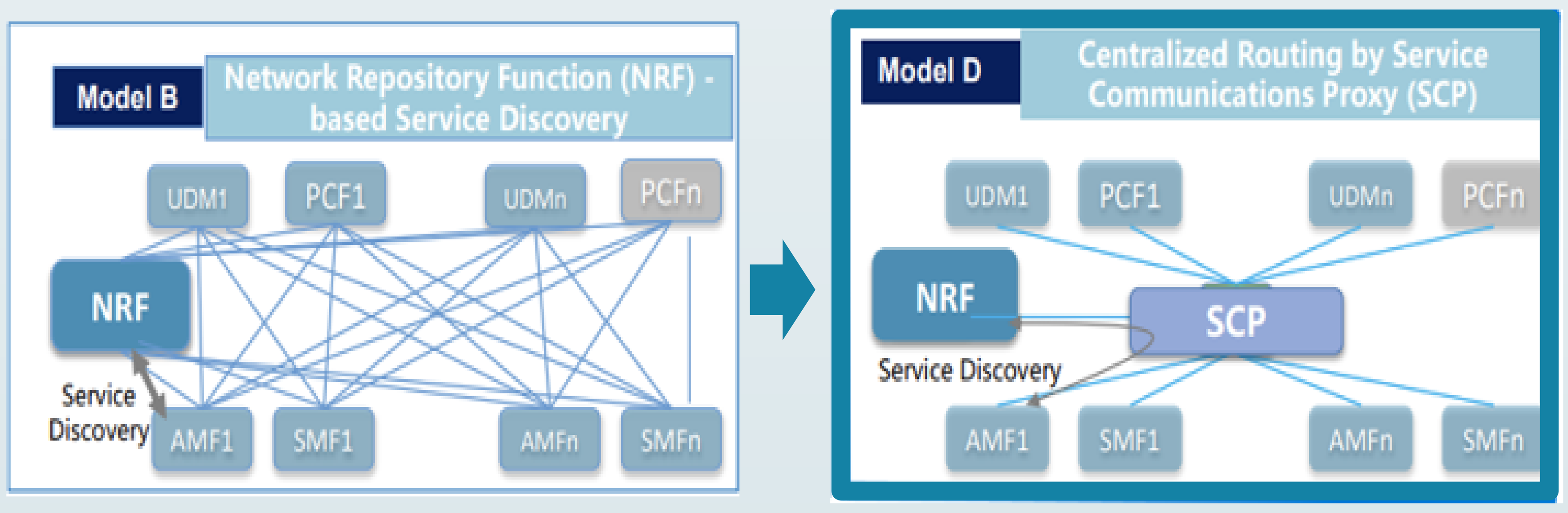
Already, 3GPP proposes an enhanced SBA approach that includes a Service Communication Proxy for indirect NF communication and may optimize NF common processes, such as the NF registration, discovery, and service provisioning. Based on the existence and the role of the SCP that four signaling Models have been defined, with Model D, being the most advanced one where the SCP realizes the full set of functions envisioned in the enhanced SBA. One step further, a decentralized implementation of Model D, e.g., through the service-mesh approach, brings new benefits since the centralized implementation of the SCP raises scalability and security concerns.

The service-mesh platform for the mobile network core has been developed by FOGUS to reveal the full potential of service-mesh -enabled SCP. Open-source projects (like Kubernetes, Istio, and Open5GS) have been integrated into a private cloud infrastructure which is ready to host experiments and research studies that focus on mobile network core.

Communication Models under study

Our key research question is to assess the benefits of the 3GPP signaling Model D using SCP versus Model B, which does not use SCP. The SCP acts as a focal indirect NF communication in this architecture paradigm, handling NF registration, discovery, and service provisioning. Because of the acting intermediary role for communications, the SCP is envisioned to alleviate the signaling burden by centralizing these operations so that NFs can devolve several operations to the proxy.

We focus on investigating whether deploying SCP really provides the required control signaling overhead reduction compared with the direct NF-to-NF communication model in Model B. Control signaling traffic reduction is key for scalability since the demands on the network grow during the present 5G and future 6G eras.



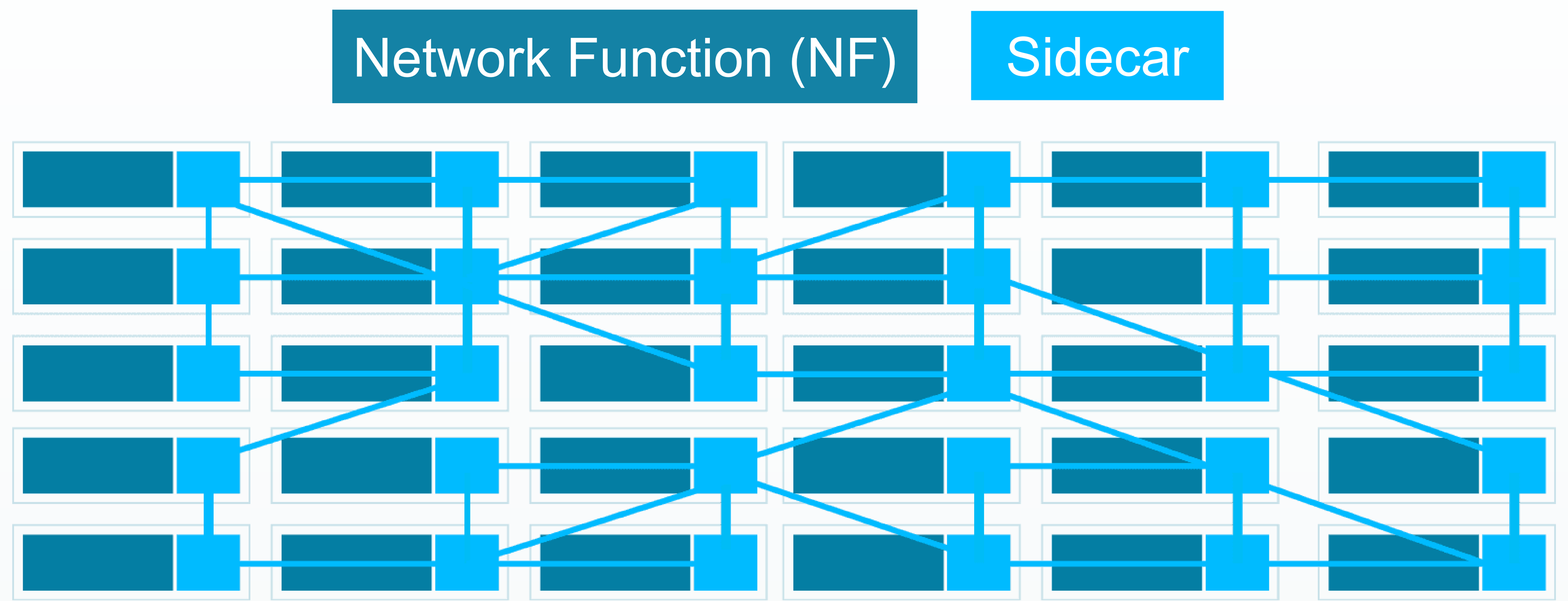
Model B and Model D architecture (Source: 'Signaling – The Critical Nerve Center of 5G Networks'

Underlay compute Infrastructure

The platform has been built on a dedicated server with an Intel® Xeon® E-2378 CPU (8 cores, 16 threads) and 64GB RAM, which can support intensive computational workloads. The underlying infrastructure is running on Proxmox VE 8.2, a Type-1 hypervisor, which has orchestrated VMs on the self-hosted Kubernetes cluster. The cluster is composed of three Ubuntu 22.04 LTS VMs, each with 4 CPUs and 16GB of RAM, running Kubernetes version 1.29.2. The underlying cluster setup includes one Control Plane Node that concerns with the cluster state and coordinates activities between nodes and two Worker Nodes used to host the workloads that will be deployed. These include the Open5GS and UERANSIM deployments needed to emulate the 5G core network and RAN, respectively.

Methodology

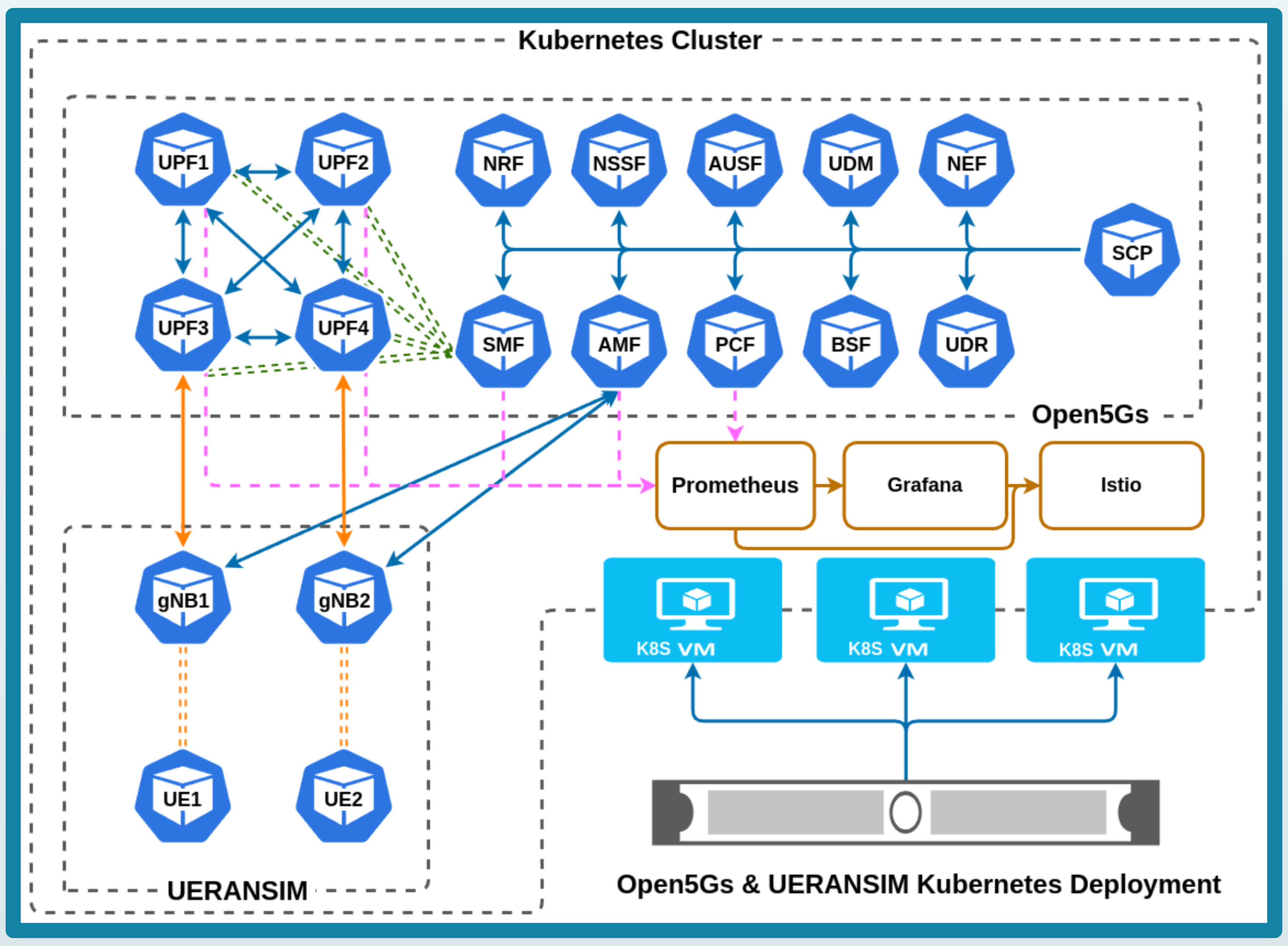
Both Model D and Model B are tested using the same Kubernetes cluster, however, in a separate namespace. This allows for independent configurations and isolated environments of both instances of the 5G core and the RAN. The same resources and policies have been set for both the model to make a fair comparison between them. Traffic is generated using custom scripts that simulate control-plane activities like the registration and de-registration of hundreds of UE instances over extended time periods. Also, Both models are serviced by Istio and complemented with observability and monitoring by Kiali, Prometheus, and Grafana. Kiali provides the user interface that displays the visualization of the Istio traffic flow, while Prometheus and Grafana serve runtime monitoring and performance metrics analysis purposes. This experiment's run keeps configurations on Istio static so that the data collected only has the inherent differences between Model B and Model D.



Evaluation approach

The most important metric for evaluation is the packet ratio of each NF in both deployments, attempting to compare the signaling traffic pattern in each model. The collected data help in determining if the integration of SCP in Model D effectively reduces signaling load. Additional metrics may include Traffic prioritization under load conditions-observing whether strategies in handling traffic can maintain stability of the system by selectively dropping lower-priority traffic. Data collection is managed by Prometheus and then fed into both Grafana and Istio for in-depth visualization and analysis. Any given test iteration happens with a certain, static Istio configuration to allow comparisons to be made.

We conducted a quantitative analysis to compare the reduction in signaling traffic between the models. Multiple test executions were performed for the confirmation of the consistency and reliability of the results. Iterative testing shall help confirm that indeed SCP in Model D is robust enough for varied simulated loads and conditions; hence, the performance gain is repeatable and statistically significant. However, more emphasis on efficiency is created by highlighting configuration capabilities in Istio for the integration of 5G core network functions with service mesh in a cloud-native environment.



Architecture of the proposed Platform

Conclusion & Recommendations

- **SCP can successfully minimize signaling traffic challenges brought by increased connectivity in the beyond 5G era**. Indirect communications between NFs enabled by SCP mean network services are resilient, scalable, and maintainable, especially when the connected UE devices grow.
- **With service mesh implementation an AI-driven management of the control plane can be enabled**. Load balancing, routing, and prioritization can be applied, while the capability to monitor the control plane enables mechanisms like anomaly detection.
- **Service mesh will be integral part of the 6G network core since it provides observability and real-time adaptation.** Given that more intensive work is needed to study service mesh enabled control plane, including dynamic traffic management and service prioritization with different network loads.

Acknowledgements

This work has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under grant agreement no. 101139270 project: ORIGAMI.

6GSNS ORIGAMI