DATA MANAGEMENT PLAN

Project Number: 101030505

Project Acronym: OTIS

Project Title: Optimal Orchestration of Virtualized Network Functions for Improved Quality-of-Service

Data Management Plan

Beneficiary: University of Exeter

Department: Department of Computer Science

Name of MSCA Fellow: Dr. Haojun Huang

Name of Supervisor: Prof. Geyong Min

1. Data Summary

1.1 Background

The OTIS project participates in the Horizon 2020 Pilot on Open Research Data. The Data Management Plan (DMP) describes the types of data that will be utilized and generated during the project, the standards that will be used, how the data will be exploited and shared for verification and reuse, as well as the preservation and storage of data.

The provisions for open access to research data are described in Grant Agreement Arts 29.3 *Open access to research data and 39.2 Processing of personal data by the beneficiaries.* This document follows the template for the H2020 Data Management Plan v1.0 (13/10/2016).

1.2 Objectives of the OTIS project

OTIS aims to carry out the innovative research to develop a new suite of innovative Virtualized Network Function (VNF) orchestration approaches to resolve the two fundamental problems: SFC Parallelisms (SFCPs) and VNF Placements (VNFPs), in order to achieve higher Network Function Virtualization (NFV) performance and improve the QoS of telecom networks. The main idea is to design independent parallel Service Function Chains (SFCs), i.e., some VNF pairs are parallelized, and then orchestrate them in the virtualized equipment, by identifying independent VNFs and making them well work in a parallel and resource-efficient manner. Two specific objectives are as follows.

Objective 1: To investigate network function chain parallelism paradigms. To this end, a policy specification scheme will be introduced for operators to intuitively describe sequential or parallel VNF chain intents. Besides, a policy-based parallelism identification scheme will be proposed for identifying VNF dependencies and automatically compiling policies into VNF orchestrator.

Objective 2: To design optimal virtualized network function placement approaches. To achieve this goal, a forwarding graph and an optimal QoS-aware mapping scheme will be developed for the parallel SFCs to be mapped into the underlying networks, following the partitioning graph matching rule. Moreover, an efficient resource allocation solution will be further developed to host all VNFs with the guarantee of storage, computing, bandwidth, and latency under the cost restrictions. To maximize the impact of OTIS, the software codes of all implementations will be made available to the public as an online open source.

During this fellowship, we aim to formulate the following novel theory and further develop a prototype system based on the theory.

Theory: Aiming to optimally parallelize and orchestrate SFCs with QoS guarantee to enable NFV performance acceleration, the theories on VNF parallelism specification, identification, forwarding graph establishment, mapping and resource allocation will be presented and open to the public through the website, reputable international journal/conference publications, and public engagement activities. These theories will provide a critical solution to enhance NFV performance and further address CAPEX and OPEX issues for network stakeholders with NFV.

System: A software package with two components to achieve Objectives 1 and 2, respectively, and an NFV-based testbed centered at the University of Exeter (UNEXE) will be developed to evaluate and validate our

solutions. This testbed can be used to promote optimal NFV orchestration and support its further applications in 5G, NFV-based Internet of Things (IoT), and future Internet, etc.

1.3 Types of data utilized and generated

To achieve the above ambitious aim and evaluate the proposed approaches, OTIS will use the existing relevant data from the open-source datasets on the telecom networks, and generate new data within the project. All existing datasets will be used in accordance with any licences or access requirements, and to avoid concerns of ownership and copyright, these datasets will not be archived. Instead, the link to the original data will be added as references in the relevant publications. The generated data will not contain any personal privacy information to protect users' privacy. To this end, the evaluation of OTIS will be carried out in a pure experimental environment. In the following, the type, source, format, size of the data and the link to which objective will be elaborated in Table 1.

Table 1: Types of data utilized and generated in OTIS

Public Dataset					
Data	Description	Source	File format	Objective	Size
SNDZoo	It is an open collection of software networking data sets aiming to streamline and ease machine learning research in the software networking domain. Most of the published data sets focus on, but are not limited to, the performance of VNFs. The data is collected using fully automated NFV benchmarking frameworks, such as tng-bench or Gym.	Thesoftwarised network data zoo (SNDZoo) (https://sndzoo.github.io/#data)	.zip	Objective 1	8.5G
SNDlib 1.0	This is a data library for fixed telecommunication network design available. The library in the current version 1.0 contains data related to 22 networks which, combined with a set of selected planning parameters, leads to 830 network design problem instances.	The Survivable Network Design Library (SNDlib) (http://sndlib.zib.de)	.gz	Objective 2	2.5 GB

Generated Dataset					
Data	**Description**	**File format**	**WP**	**Objective**	**Size**
Synthetic network conditions	To obtain the better accuracy and convergence of the DRL &FL for VNF placement, we will synthesize some extra networks to train DRL/FL-based models. The data will consist of the timestamp, the IPs of the source and destination, the number of nodes and links, the network delay, and the available bandwidth, etc.	.logs	WP3	Objectives 1 &2	200M
Performance dataset	To evaluate the proposed solutions, extensive experiments will be executed in a pure experimental environment and the performance data will be collected from the networks. The data will include the timestamp, the IPs of the source and destination, the time of arriving and sending, the available bandwidth, the buffering time, and the buffer occupancy, etc.	.logs	WP3	Objectives 1 &2	300M

1.4 Data utility

The data utilized and generated by OTIS might be useful for the broader scientific community for the sake of validating results presented in OTIS's scientific publications, as well as for future network function virtualization research, especially for link VNF deployment and QoS in telecom networks.

2. FAIR data

OTIS complies with the guiding principles for scientific data management outlined by the H2020 programme, which ensure that the research data is **Findable**, **Accessible**, **Interoperable** and **Reusable** (FAIR).

2.1. Making data findable (F)

Persistent Digital Object Identifier (DOI)

The datasets generated by OTIS will be archived in the institutional repository Open Research Exeter (ORE), which is maintained by the beneficiary, University of Exeter, to support the principles of open access to research publications and research data. ORE is fully searchable and indexed by search engines to ensure that the data are highly discoverable. Note that a unique DOI will be assigned to the dataset once it is uploaded to the data repository, which will also be linked to appropriate publication records in ORE. With the DOI, the dataset can be accessed and cited in research papers, which can further enhance the visibility of datasets.

Metadata

Each dataset deposits in ORE will be accompanied by Dublin Core metadata to describe the information regarding the owner, date accessioned, date issued, access conditions, generation methods, time references, structure and organisation of data files, file formats, variable names, labels and descriptions of variables and values. Moreover, the terms "Marie Skłodowska-Curie Actions", the name of the action (H2020-MSCA-IF-2020), acronym (OTIS) and grant number (101030505) will be included, along with the DOI and the publication date. Upon archiving the dataset, keywords will be provided to enhance the findability of the data, thus optimizing the possibility for re-use by other researchers in the related field.

Naming conventions

OTIS datasets will be named with project title (OTIS), the No. of WP to which the data is linked, the purpose of the data, and the version. Specifically, the naming rule is as follows:

OTIS -WPx-purpose-vx.

For example, the dataset entitled with "OTIS-WP01-SFCP-v1" is the dataset with an initial version (v1) for SFC Parallelisms in WP01.

The naming convention will be explained in documentation (e.g., a README file) that will accompany the data deposited to ORE so that others will be able to under-stand the data, increasing the re-use of data.

2.2. Making data openly accessible (A)

All the data generated from OTIS, including the metadata documentation and code, will be deposited in the ORE on publication of the research papers, which can be utilized to validate the results presented in the deposited scientific publications. The ORE is the University of Exeter's online repository for storing and providing access to research output, which can ensure that the research data of OTIS is securely preserved in a long-term manner. ORE can increase the visibility of publications and research data by making them available as open access and enabling compliance with open access policies of the University and funders of research. The data will be openly accessible to everyone, with the metadata being publicly searchable and discoverable, which will indicate how and on what terms the dataset can be re-used.

The existing datasets will not be deposited to ORE, but the documentation accompanying the data in ORE will include links to these existing datasets so that anyone who downloads the data from ORE will be able to replicate the analysis.

2.3. Making data interoperable (I)

Standards and formats

The data to be deposited and made available for other researchers will be archived in standard formats and in compliance with available open software applications to allow for data exchange and reuse. The file formats for different kinds of data are listed in Table 1. For better interoperability of the data, a readme file explaining the data will be created.

2.4. Increase data re-use (R)

Licencing

Data will be made available under a CC licence, which manages copyright and the terms and conditions for access and reuse. A CC licence requires attribution of the licensor (data supplier) when the use involves public sharing.

After the end of the project (30/6/2024), and as soon as project results have been published in the form of peer-reviewed publications, OTIS data will be deposited and made available for third parties. The moment, at which the data will be made available, any conditions for re-use by third parties and embargoes after the end of the project will be specified in a subsequent version of this DMP.

3. Allocation of resources

3.1 Costs

Costs related to data management are expected to be limited and will be covered by the individual beneficiary's institutional unit costs. The fellow will be responsible for setting up a budget for data gathering/collection with the help of local research and/or financial support staff.

The costs will be discussed with the beneficiary (supervisor and research manager/financial officer) and the Coordinator and Project Manager, who will make sure that the provisions of the OTIS Grant Agreement and this DMP are followed. The potential value of processing such data with the purpose of making them available for re-use by other researchers must be estimated as higher than the costs in terms of the work load, which will depend of the amount/size of the data, and sensitivity/confidentiality issues.

Costs related to the administrative coordination of data management and deposition in line with this DMP will be covered by project funds at network level administered by the coordinator.

Costs for the long-term preservation of data will not occur, as the archiving in the ORE.

3.2 Project responsibilities

The supervisor, Prof. Geyong Min, will be ultimately responsible for data management of this project, while the fellow, Dr. Haojun Huang, is responsible for the data generation, metadata production, data quality guaranteeing during the whole process of the project. Upon the publication of a paper, the fellow should also be in charge of uploading the related data to ORE, and backing up it in the server provided by the High-Performance Computing and Networking (HPCN) Research Group led by Prof. Min. The Open Research team from the University of Exeter will be responsible for the management of ORE.

4. Data security

The beneficiary will provide the fellow several options for secure storage of data. During the data collection and experimental phase, apart from his personal computer, the fellow will back up the data in the password-protected servers offered by the host re-search group, the HPCN Research Group led by Prof. Min, in the College of Engineering, Mathematics and Physical Sciences, University of Exeter.

At the end of the project, the data will be deposited for long-term storage in the institutional repository, the Exeter ORE data archive, which is maintained by the hosting beneficiary, University of Exeter.

5. Ethical aspects

There are no ethical issues concerning OTIS project. No personal data is used in OTIS project.

6. Other issues

No, OTIS does not use other national/funder/sectorial/departmental procedures for data management.

7. Further support in developing your DMP

The dataset within OTIS will be stored at the ORE, which is maintained by the hosting beneficiary, University of Exeter. No other support will be needed in developing the DMP.