Muhammad, Yousuf; Yan, Zheng; Wang, Mingjun; Khan, F.

**Trusted Network Slicing among Multiple Mobile Network Operators**

*Published in:*

*DOI:*
10.1109/TrustCom50675.2020.00150

Published: 01/12/2020

**Document Version**
Peer reviewed version

*Please cite the original version:*
trusted network slicing among multiple mobile network operators

Muhammad Jehanzaib Yousuf
State Key Lab of ISN
Xidian University, Xi'an, P.R.C
yjehanzaib@hotmail.com

Mingjun Wang
State Key Lab of ISN
Xidian University, Xi'an, P.R.C
mjwang@xidian.edu.cn

Zheng Yan
State Key Lab of ISN
Xidian University, Xi'an, P.R.C
Department of Communications and Networking
Aalto University, Espoo, Finland
zyan@xidian.edu.cn

Fawad Khan
National University of Sciences & Technology
Islamabad, Pakistan
fawad.khan.xdu@gmail.com

Abstract—5G mobile networks are expected to be much bigger in size, faster in speeds and better in scalability, providing varied services to different users and businesses in contrast to previous networks. 5G will also help enabling new business models and use cases. “Network Slicing” is a driving architectural concept for multi-tenancy. Network Slicing enables Mobile Network Operators (MNOs) to deploy different services over shared physical infrastructure, increasing inter-operator resource sharing. As 5G is still in its nascent stage, inter-operator cooperation is an area that requires immediate attention of research. Traditional inter operator trust relationship models cannot fully comprehend the needs of 5G networks. In this paper, we propose an Intel SGX based multi-MNO cooperation scheme for trusted, dynamic and efficient network slice sharing in order to support inter-operator trustworthy collaboration. Furthermore, we developed a Proof of Concept of our proposed scheme using Intel SGX, flask framework and Docker containers. The obtained results indicate the applicability of the proposed scheme with little effect on performance.

Keywords—5G, Network Slicing, Intel SGX, Trust

I. INTRODUCTION

5G networks are touted to provide a multitude of services to highly diverse range of new businesses and verticals with conflicting requirements. Some customers or verticals would require ultra-high speed mobile broadband while some others would require ultra-low latency reliable communications [1]. The legacy monolithic 4G network architecture cannot support such heterogeneous network requirements. 3GPP has introduced some key architectural changes in 5G network architecture, making it a Service Based Architecture (SBA) [2]. The highlight of this SBA is transformation of the key Network Entities (NEs) into Network Functions (NFs), which make the network highly scalable, programmable and flexible. Software Defined Network (SDN), Network Function Virtualization (NFV) and cloud computing technologies are utilized in synergy to create logically isolated networks deployed over shared physical network infrastructure. A resulting new architectural concept is “Network Slicing” and corresponding logical networks are called “Network Slices”.

Network Slicing is at early stages and a hot topic in research community. It has got tremendous attention due to increasing demands of high data rates from data hungry and resource consuming applications and use cases. Afolabi et al. [1] have discussed network slicing history, key technology enablers and current state of art in detail. Operators can adopt Network Slicing-as-a-Service (NSaaS) to provide customized cellular network services as proposed in [2]. Network Slicing has existed in different forms in previous mobile network generations as network sharing concepts such as Multi-Operator Radio Access Network (MORAN), Multi-Operator Core Network (MOCN), GWCN (Gateway Core Network), Dedicated Core (DÉCOR) and enhancements for Dedicated Core Networks (eDÉCOR).

By allowing different parties to instantiate and run a softwarized architecture, 5G becomes inherently a multi-Tenant ecosystem, whereby a Tenant refers to a user or a user group with specific access rights and privileges over shared resources. Some business models and verticals, e.g., a connected car company or a social media network has a global customer base thus requires global coverage, which is beyond the capability of a single operator. Hence, to create a uniform set of services for end-customers, these businesses would need to provide the same services supported by similar slices over multiple operator networks. In such a roaming scenario between two operators, one operator can assume the role of “Renter” and the other of “Tenant”. The Tenant will deploy its own deploying network function Virtual Machines (VMs) or Containers over the Renter MNO’s rented infrastructure. The MNOs are honest-but-curious as well, indicating that they are honest to build network and provide services to users, but curious about the data being exchanged among users. Here arises key issue of integrity, confidentiality and trustworthiness of user data and traffic because of the untrusted host (Renter) infrastructure. Another challenge is trusted accounting of resources utilized by the Tenant and fairness of charging done by the renter. However, little existing work handles the above issues to ensure trustworthy inter-operator network slicing.

In this paper, we propose a scheme employing Intel SGX to achieve trustworthy cooperation for trusted, dynamic and efficient network slice sharing. Furthermore, our scheme ensures confidentiality of tenant user identity, enhance network configuration and policy management by Tenant over a rented network. In the end to show the applicability of our scheme with a proof of concept implementation.

The rest of the paper is structured as follows. We review related work in Section II. Section III presents the system and security models, followed by proposed scheme in Section IV.
Section V introduces proof of concept and last section is conclusion.

II. BACKGROUND AND RELATED WORK

A. Background

In a resource sharing relationship, MNOs enter into a relationship of Renter (a resource provider) and Tenant (a resource user) similar to a cloud data center and Tenant. In case of MNOs, resources are shared for creating and sharing slices over different MNOs. A Renter MNO can provide slicing services to a Tenant MNO in two different slice sharing offer types according to [3].

Offer Type 1: Renter MNO instantiates a slice on behalf of the Tenant. It provides the Tenant only a monitoring interface or some limited capabilities to configure some parameters.

Offer Type 2: Renter MNO assumes the role of an infrastructure provider and provides physical computing, storage and network resources. The Tenant deploys its own VNFs and VNF managers and orchestrators, providing connectivity between VNFs through SDN.

The trust relationship in case of these two offer types are explored below:

Trust Relationship of Offer Type 1:

The Tenant has to trust the Renter MNO that it will abide by the agreed upon Service Level Agreement (SLA) and provide the settled services to the Tenant users. The Tenant has no technical means to enforce sufficient resource allocation, confidentiality and integrity of Tenant traffic data as it is only provided with a monitoring interface or maybe not.

Trust Relationship of Offer Type 2:

The Tenant has more control over its slice in Offer Type 2, since The Tenant deploys its own network functions, so user data is not visible to the Renter. Although being an Infrastructure Provider, the Renter MNO could still access the data and traffic. In this kind of scenario, a NS (network slice) should be created in such a way that a Tenant’s devices need no extra authentication from the Renter. A device authenticated by the Tenant can access the Tenant slice, including some common functions shared with MNO. Tenant has better visibility of the resources assigned to it. Despite the NF deployment by the Tenant, it still has to trust the Renter MNO share some common functions like Radio Access Network (RAN) scheduler for assignment of radio resources or have to expose critical infrastructures like Authentication Server Function (AUSF) and Unified Data Management (UDM) towards the Renter. In event of cyberattack or security breach of Renter infrastructure, these network functions will be inadvertently vulnerable to the breach.

B. Related Work

A lot of work has been done on network slice orchestration and management, resource allocation and sharing. In [4], different novel provisioning models for 3rd-party services and their isolation properties were discussed but they lack monitoring of assigned resources and their usage. Ni et al. [5] proposed a protocol to build up network slicing for IoT services in 5G networks, however, this work cannot address secure and trusted slicing establishment between multiple mobile network operators. In [6], Tian et al. proposed a C-RAN Inter-Operator Cooperation Scheme to support the cooperation of multiple operators in C-RAN and allow them to share resources in a trustworthy and secure way based on Trusted Computing Platform (TCP), but this is only limited to RAN part only.

An interesting method for securing network slice is Trust Zoning. A security trust zone is a set of configurations and policies to ensure necessary security of a trust zone. Bian et al. [7] proposed a distributed authentication and authorization solution called Trust Zone, for an edge network. This solution is specifically useful for uRLLC network slices. This work is further extended by Schinianakis et al. [8]. Based upon different services provided by different network slices and security requirements, slices are classified into different security trust zones. To discourage repeated authentication processes, a similar concept is employed by authors in [9] called Radio Trust Zone (RTZ) or Authentication Slice (AS). These RTZ or AS can encompass multiple small cells or multiple network slices. This scheme works by radio fingerprinting a device and storing this fingerprint in a centralized database of trust zone or AS. Thus even if the device roams across multiple cells, it only needs to be authenticated only once in a given RTZ. Security Trust Zoning is useful strategy to enforce edge network trustworthiness but it lacks support on Trust Relationship Offer 2. Niu et al. proposed a model to effectively measure trust degrees of different network slices based upon granularity, integrity and dynamic nature [10]. However, this work cannot address secure and trusted slicing establishment between multiple mobile network operators.

Blockchain is another emerging technology for achieving trust in distributed systems, such as slicing. In [11] authors proposed a blockchain and content centric networking system to ensure trusted and secure communications in vehicular network slices in 5G. Xie et al. [12] also proposed a blockchain based security and privacy preserving solution for SDN based 5G-VANETS to support dynamic and efficient trust management. Taking inspiration from recent Mirai botnet DDoS attacks, Boussard et al. [13] exploited blockchain and SDN, proposed a trust evaluation scheme for IoT devices connected in a slice. They utilized SDN to create isolated network slices and calculated trust values of different types of devices by classifying their previous interactions stored in blockchain. Taking into account these trust values, only those devices that can meet a threshold trust value are permitted by SDN controller to join a network. Yan et al. [16] proposed an adaptive trust evaluation and management framework for virtualized networks. It uses cloud computing for secure deployment of various trustworthy security services over the virtualized networks. In [17], Xu et al. put forwarded a trust insurance mechanism “TIM”, based upon Trusted Computing Platform (TCP) to provide trust authentication and trusted cooperation among VNFs. In [18], a mechanism was proposed for secure and trustworthy interactions between network applications and controller. Depending on the behavior descriptors of applications, a trust mechanism was applied to limit application privileges to run in the network.

But in all of these works, trust and privacy issues were neglected. In most of the proposed multi-tenant solutions,
schemes and architectures, a tenant has to trust a renter and share user information with the renter for authorization and authentication purposes. The tenant has to share network configurations, flow rules and policies to the renter. Also the tenant has very limited visibility of resources allocated to it. Table I compares some previously proposed works and the trust relationship they provide. It is evident from Table I that all of them provide trust relationship in Offer Type 1, but fail in Offer Type 2.

<table>
<thead>
<tr>
<th>TABLE I. COMPARSIONS OF EXISTING WORKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trust Offer Type 1</td>
</tr>
<tr>
<td>[12]</td>
</tr>
<tr>
<td>[13]</td>
</tr>
<tr>
<td>[14]</td>
</tr>
<tr>
<td>[17]</td>
</tr>
<tr>
<td>[18]</td>
</tr>
<tr>
<td>[19]</td>
</tr>
<tr>
<td>[20]</td>
</tr>
<tr>
<td>[21]</td>
</tr>
<tr>
<td>[22]</td>
</tr>
</tbody>
</table>

So the problem remains, how to establish a flexible, dynamic, efficient, secure and trusted network slice sharing relationship between multiple MNOs? The required flexibility and efficiency can be achieved by technologies like NFV, SDN and cloud computing. A secure and trusted relationship could be achieved by performing critical security processes and data exchanges in a Trusted Execution Environment (TEE). A promising technology is Intel Software Guard Extensions (SGX). Intel SGX provides a general-purpose TEE which utilizes an x86-architecture instruction set to create secure and isolated memory partitions termed as “Enclaves”. Enclaves provide confidentiality and integrity of enclosed codes and data through checksums and encryption [19]. In addition to protection, the enclaves also provide local and remote attestation services. Local attestation is used by enclaves deployed over a same machine to attest each other, where as remote attestation enables enclaves to authenticate the enclaves deployed over remote machines [20]. In [21], Goltzsche et al. proposed an Intel SGX based architecture for outsourcing middle box functions through VPN to securely execute inside enclaves on client machines. Similarly, in [22] authors proposed a privacy preserving architecture to maintain privacy and integrity of code and traffic, limiting the information exposure to an untrusted third party.

III. SYSTEM MODEL AND SECURITY MODEL

A. System Model

Fig. 1 shows two mobile network operators MNO1 (Home Network) and MNO2 (Visited MNO) with their different network slices created over same physical infrastructure. The example three network slices showed in Fig.1 namely are: (1) enhanced mobile broadband (eMBB) slice, (2) ultra-reliable low latency communication (uRLLC) slice and (3) IoT slice. All three network slices are utilizing the same underlying physical resources yet providing services with varied quality of service parameters to different users and verticals. It also illustrates a scenario where a UE1 (user equipment) belonging to MNO1 (Home Network) has moved into a coverage area of MNO2 (Visited MNO). The UE1 was previously enjoying services provided by MNO1 (Home Network) but after exiting the coverage area or geographical boundary of its home network, it is out of service. To deliver services to its users or customers and to keep them connected, MNO2 agrees upon a renter-tenant agreement or SLA with another network operator MNO2. Under this agreed upon SLA, MNO2 acting as a renter, will provide resources to MNO1, which is in this case a tenant, to create its slice over its physical infrastructure to provide services and access to UE1 and other such users who would migrate into the coverage area of MNO2.

![Image](image_url)

Fig. 1. System Model

B. Security Model

We assume two threat scenarios to protect against:

1. Although MNO1 has deployed its network functions over MNO2 physical infrastructure and MNO2 might appear trustworthy to MNO1 but it is not. MNO2 is curious about the MNO1 user information and traffic generated by it.

2. In another scenario, we assume that a powerful adversary has taken over control of the MNO2 infrastructure and network stack including hypervisor. The MNO1 containers are also vulnerable to this adversary. This adversary can access MNO1 user identity and can also manipulate network configurations and misdirect traffic routes.

IV. PROPOSED SCHEME

In this section, we first overview the proposed scheme. Then, we introduce the main techniques used for deploy it. In particular, we specify UE authentication and registration protocol in inter-operator network slicing.

A. Scheme Overview

Fig. 2 shows our proposed system architecture. UE1 belongs to MNO1 (Home/Tenant MNO) and is accessing the network services through RAN1 and Core network CN1. MNO1 deploys its necessary core network functions coupled with Intel SGX enclaves over MNO2’s (Visited/Renter MNO) rented hardware infrastructure in a dynamic way when needed. The sensitive user information e.g. user slice type or slice subscription etc., is accessed through the Intel SGX enclave created on MNO2 rented infrastructure and afterwards it is placed in an encrypted file
system to reduce the load on enclaves and clear up enclave memory. When an MNO1 subscriber tries to access shared network slice, its authentication and authorization process is done by MNO1’s own network functions. The NFs communicate with Home network through SGX enclave. SGX enclave provides necessary privacy protection for subscriber data and network function code. Note that SGX enclave running at MNO2 has been previously attested and authenticated by MNO1 as trusted and can authenticate and serve MNO1’s UE, as well as securely communicate with MNO1 functions.

Fig. 2. Proposed Architecture

For creating and sharing a network slice, Tenant (MNO1) requests slice sharing to Renter (MNO2). Renter will analyze the request and its available resources. If Renter has enough resources to fulfill the requested QoS requirements from MNO1, it will share R&B (resource and billing) details for requested slice with Tenant. If Tenant accepts the resource and billing R&B details, it will share an acknowledgement with Renter. In return, Renter will send an SLA. If Tenant accepts the SLA, it will reply with acknowledgement to Renter. Renter asks Tenant MNO1 to share NF container images to create a slice for Tenant. After receiving NF container images, Renter deploys them over its infrastructure to create the slice, so the Tenant’s users will start getting services from the created slice. During this slice sharing, both MNOs will monitor the slice KPIs (key performance indicators) and utilization of resources. Since Tenant has shared its own containers so it has the direct visibility of its slice. If any of both the parties violates the SLA, the opposite party will cancel the agreement to maintain the integrity of its network. When the time period for slice sharing is completed, the Renter will send the final billing and cost details incurred to Tenant (MNO1). The Tenant can verify the billing by comparing with its own logs.

B. Main Techniques

For the two threat scenarios mentioned in previous section, firstly, MNO1 needs to deploy its own network functions over MNO2 to have effective monitoring and management. Secondly, MNO1 needs to establish a Trusted Execution Environment (TEE) to exchange confidential UE data with its network functions in MNO2. To create a TEE, we employ Intel SGX enclaves. The TEE of Intel SGX enclaves offers security and privacy protection for UE data through encryption and limited exposure. VMs can be a natural choice for network function deployment over remote hardware to offer secure and isolated execution due to virtualization. But at the same time, VMs can be quite beefy in size. VMs require to run complete Guest OS (Operating Systems) with pre-installed target applications, on emulated hardware. Hardware emulation is achieved by means of another software called Hypervisor or VM manager. The Hypervisor or VM manager is installed on a Host OS on underlying physical hardware. On the other hand, Linux containers are gaining traction in industry [23]–[26]. The containers neither have nor require a complete guest OS instead, they have binaries and libraries stacked into an image file called container image. This image becomes a fully functional container when it starts running. The containers are deployed in clusters. In place of Hypervisor or VM manager, a container engine for example Docker or Kubernetes is installed on Host OS, to orchestrate and manage the containers [27],[28]. The main motivations of implementing increased containers is their ease in encapsulation, deployment and isolation of applications, lightweight operations, efficient and flexible resource sharing. The size of a container is usually within megabytes (MB) while that of a VM can take several gigabytes (GB). A container usually takes less hardware resources since it does not need maintain an OS to run the same application as compared to VM. Thus, we select the container to implement our scheme for achieving better economy and efficiency.

To create a TEE, we employ Intel SGX enclaves at the Renter. Inside the NF containers, Intel SGX enclaves are established. The sensitive user data is accessed and exchanged between Tenant MNO’s home NFs and remotely deployed NFs in these enclaves. The TEE of Intel SGX enclaves offers security and privacy protection for UE data through encryption and limited exposure. The NF-containers for example Docker containers can communicate with each other and are interlinked with each other over bridge networks. When Docker containers are installed, three networks namely, bridge, none and host are created by default [29]. Bridge network is used for communication between containers deployed at the same host. All new deployed containers over a single host, if not specified are connected to the bridge network by default. Docker does not support automatic service discovery on the default bridge network. Alternative to default bridge network to resolve IP addresses by container name, user-defined bridges are used instead. User-defined bridge networks are used to control which containers can communicate with each other, and also to enable automatic Domain Name System (DNS) resolution of container names to IP addresses.

C. UE Authentication and Registration

Fig. 3 shows the protocol of UE1 authentication and registration at a foreign MNO2 in inter-operator network slicing.

1. UE1 to RAN2: When UE1 reaches out RAN2/MNO2, with Registration Request message. It will provide RAN2 with an 5G identifier either 5G-GUTI (Globally Unique Temporary Identifier) or SUCI (Subscription Concealed Identifier) that indicates it belongs to MNO1, while rest of Registration Request includes Registration type, last visited TAI (Tracking Area identifier), Requested NSSAI (Network Slice Selection Assistance information), PDU (Protocol Data Unit) Session status, List of PDU Sessions to Be Activated.
2. RAN2 to AMF1/MNO2: Based upon the UE1 provided 5G-GUTI or SUCI, the RAN2 directs the Registration Request to AMF1/MNO2. The AMF1/MNO2 forwards this Registration Request data to Intel SGX Enclave. Intel SGX Enclave is created by MNO2 over MNO1 provided hardware resource.

3. AMF1/MNO2 to AMF1/MNO1: Since the UE1 has provided the 5G-GUTI, Intel SGX enclave invokes the UE1 Context Transfer service operation on the AMF1/MNO1 (old AMF) to request the UE1’s SUPI and UE1 Context inside Intel SGX Enclave.

4. AMF1/MNO1 to AMF1/MNO2: The AMF1/MNO1 as queried in step 3, responds to the Intel SGX Enclave for the UE Context Transfer invocation by forwarding the UE1’s SUPI and UE1 Context, SMF information, S-NSSAI and PDU Session IDs to SGX enclave. Intel SGX Enclave saves this data into an encrypted file which can only be accessed by the enclave only.

5. AMF1/MNO2 to AUSF1/MNO1: After context transfer, AMF1/MNO2 invokes the AUSF1/MNO2 for authentication of UE1.

6. AUSF1/MNO1 to AMF1/MNO2: AUSF1/MNO1 requests UE1 authentication data from UDM1/MNO1. After authentication, AUSF1 forwards relevant security data to SGX enclave in MNO2 and later place it in the encrypted file system. Intel SGX Enclave sends signal to AMF1/MNO2 that UE1 is authenticated by MNO1.

7. AMF1/MNO2 to AMF1/MNO1: After the AMF1/MNO2 has changed, the AMF1/MNO2 notifies the AMF1/MNO1 that the registration of the UE1 in the AMF1/MNO2 is completed by invoking the Registration Complete Notify service operation.

8. SMF1/MNO2 to UDM1: AMF1/MNO2 registers with the UDM1 using UE CM-Registration service operation and subscribes to be notified when the UDM1 deregisters this AMF1/MNO2. AMF1/MNO2 fetches the Access and Mobility Subscription data, SMF Selection Subscription data and UE context in SMF data using SDM_Get service operation from UDM inside enclave. After a successful data transfer to enclave from UDM1, AMF1/MNO2 configures to be notified, when the data requested is modified through SDM_Subscribe service operation. Once again the enclave stores this SMF data in encrypted file created before for future use if required.

9. AMF1/MNO1 to SMF1/MNO2: Since the "List of PDU Sessions to Be Activated" is included in the Registration Request, AMF1/MNO2 sends PDU Session Update Session Management Context Request to enclave. The enclave notifies the SMF1/MNO2, associated with the PDU Sessions to activate User Plane connections for them.

10. SMF1/MNO2 to UPF2: SMF1/MNO2 informs relative UPF2 to establish PDU session for UE1. SMF1/MNO2 selects UPF2 and establishes session.

11. SMF1/MNO2 to AMF1/MNO2: SMF1/MNO2 sends PDU Session Establishment Response to AMF1/MNO2. The message includes the tunnel information of the UPF2.

V. PROOF OF CONCEPT IMPLEMENTATION

A proof of concept of the proposed scheme was developed using Python language with “Flask” micro-framework, Docker containers and Intel SGX enclaves. First, MNO1 verifies whether the MNO2 hardware’s Intel SGX compatibility. After verification, Intel SGX Enclave is created, which has a secure channel to communicate with MNO1. After enclave creation, NFs are implemented as web application servers using Flask micro-framework. These web application servers are then deployed inside Docker containers. Containerization of the NFs enable their remote, dynamic and efficient on demand deployment. Fig. 4 illustrates the Docker containerized NFs deployed over shared physical infrastructure. It can be observed that Docker Containers utilize minimal memory resources that is less than 200 MB in comparison to the resource exhaustive VMs, which could consume memory resources in extent of several gigabytes (GB).

In order to evaluate the proposed scheme, a C++ script was written to take user data from AMF1/MNO1 and inject it into the enclave. If AMF1/MNO1 is able to map SUPI of UE1 from 5G-GUTI provided by AMF1/MNO2 then the subsequent steps are performed as described in Fig. 3. After the designed steps have completed, the UE1 is registered at MNO2, and UE1 data is stored in an encrypted file. If AMF1/MNO1 is unable to map SUPI of UE1 from 5G-GUTI provided by AMF1/MNO2, it will terminate process with an error message.

Fig.5 shows the comparison between time taken by the proposed scheme with and without enclave for per UE authentication and slice selection. The process is repeated for 10 UEs. We can observe that introduction of Intel SGX enclave in authentication process induces a time cost in UE authentication and slice selection. The delay results from cryptographic function calls. For the proposed scheme, the time taken for the first UE is about 10 times higher than that without enclave based UE authentication. This relatively large difference is due to the initialization of enclave and the creation of the encrypted file.
After the first UE authentication, the time elapsed drops to almost half for any further authentications. 

Fig. 5. Time Cost for UE Authentication Based on Enclave

VI. CONCLUSION

In this paper, we proposed an Intel SGX based multi-MNO cooperation scheme for trusted, dynamic and efficient network slice sharing. To showcase the applicability of our proposed scheme, we developed a Proof of Concept using Intel SGX, flask framework and Docker containers. The simulation results show that a slight effect on performance due to Intel SGX’s cryptogrophic function calls.

ACKNOWLEDGMENT

The work is supported in part by the National Natural Science Foundation of China under Grants 61672410 and 314203, Shaanxi Innovation Team project under grant 2018TD-61802293, the Academy of Finland under Grants 308087 and 307, and the 111 project under grant B16037.

REFERENCES