

UNIVERSITY OF BAHRAIN
College of Information Technology



Mitigation of Distributed Denial of Service (DDoS) in 5G Virtualized Core Network

**A Thesis Submitted in Partial Fulfilment of the Requirements for the
M.Sc. Degree in Cyber Security**

Submitted by
Abdulrahman Khaled Abdulrahman
202000050

Supervised by
Dr. Reham Almesaeed
Assistant Professor
University of Bahrain

Kingdom of Bahrain
December 2022

UNIVERSITY OF BAHRAIN
College of Information Technology



Mitigation of Distributed Denial of Service (DDoS) in 5G Virtualized Core Network

A Thesis Submitted in Partial Fulfilment of the Requirements for the **M.Sc.**
Degree in Cyber Security

Submitted by
Abdulrahman Khaled Abdulrahman
202000050

Supervised by
Dr. Reham Almesaeed
Assistant Professor

University of Bahrain

Kingdom of Bahrain
December 2022



Dissertation Deposit Form

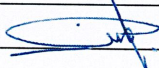
Author Name:	Abdulrahman Khaled A.Rahman Khaled A.Wahab		
Email:	202000050@stu.uob.edu.bh	Contact No:	33337077
College:	Information Technology		
Department:	Information Systems		
Thesis Title:	Mitigation of Distributed Denial of Service (DDoS) in 5G Virtualized Core Network		
Year:	2022		

Author's Declaration: I agree to the following conditions:

The above thesis will be available (in the University of Bahrain library and externally) and reproduced as necessary at the discretion of the University of Bahrain. It may also be digitized by the University of Bahrain and made available via the university's repository or via other parties on the Internet.

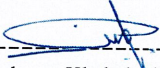
This condition shall apply to ALL copies including electronic copies.

The above thesis has been provided on the understanding that it is copyright protected material and that no quotation from this thesis may be published without proper acknowledgement.

Signature of the Author:		Date:	29-Sep-22
--------------------------	---	-------	-----------

DECLARATION

I declare that this work is the result of my own investigation and that it has not already been accepted in substance for any degree, nor is it currently submitted for any degree.

Signed:  Date: 29/Sep/2022
Abdulrahman Khaled Abdulrahman (Candidate)

The Thesis defense committee considers the thesis titled:

Mitigation of Distributed Denial of Service (DDoS) in 5G Virtualized Core Network

Submitted by Abdulrahman Khaled Abdulrahman to the College of Information Technology is satisfactory and acceptable for the Master of Science in Cyber Security.

The Defense Committee:



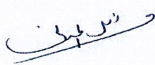
Dr. Reham Almesaeed
(Supervisor)

Department of Computer Engineering
University of Bahrain



Prof. Dr. Ayman Adel Abdel-Hamid
(External Examiner)

Arab Academy for Science,
Technology, and Maritime
Transport



Dr. Wael Elmedany
(Internal Examiner)

Department of Computer Engineering
University of Bahrain

ABSTRACT

Wireless communication technologies have evolved drastically in the past decade, with the latest evolution being the introduction of the Fifth-Generation mobile communication systems (5G). The 5G technology introduces higher data rates, ultra-low latency, high mobility, and expanded coverage. These features unlock enormous potential for Mobile Broadband user experience and Internet of Things (IoT) use cases. At the core network layer, 5G is enabled by Software Defined Networking (SDN), Network Function Virtualization (NFV), and cloud-native Service-Based Architecture (SBA). While billions of IoT devices are expected to connect to 5G, they are prone to various security weaknesses, making them an attractive platform for attackers to exploit and launch Distributed Denial of Service (DDoS) attacks from compromised 5G-IoT devices. Traditional DDoS mitigation mechanisms will not withstand 5G DDoS attacks due to the massive number of devices and dynamic nature of 5G. Signaling DDoS attacks triggered by a massive number of 5G-IoT devices is one type of DDoS in the 5G domain; it targets overloading 5G Core Network functions. This thesis introduces DDoS Kubernetes Scaling Solution (DKSS), a novel solution that utilizes the infrastructure flexibility provided by Kubernetes and Amazon Web Services (AWS) scaling in a 5G microservices architecture. The solution was demonstrated to be successful in mitigating the signaling DDoS targeting 5G Core functions with no time overhead between attack start and mitigation. In addition, the solution is highly applicable to real-world networks, as it utilizes Kubernetes built-in features and works within the boundaries of the 3rd Generation Partnership Project (3GPP) and microservices standard architectures.

Contents

APPROVAL	A
DECLARATION	B
ABSTRACT	C
TABLE OF CONTENTS	D
LIST OF FIGURES	F
LIST OF TABLES	G
ACKNOWLEDGMENTS	H
PUBLICATIONS	I
LIST OF ABBREVIATIONS	J
1 Introduction	1
1.1 Overview	1
1.1.1 Evolution to 5G Network	1
1.1.2 5G Core Network	2
1.1.3 DDoS in the 5G-IoT Domain	6
1.2 Research Problem	7
1.3 Research Objectives	7
1.4 Contribution to the Field	7
1.5 Thesis Outline	8
2 Literature Review	9
2.1 Theoretical Framework	9
2.2 Related Work	10

2.3	Summary	16
3	Mitigation of Signaling DDoS in 5GC - An Innovative Approach (DKSS)	19
3.1	Introduction	19
3.2	Containerization in 5G	20
3.3	5G Signaling	22
3.4	DDoS Attack Model	24
3.5	DKSS System Architecture	25
3.5.1	Infrastructure layer	25
3.5.2	Application layer: 5G Core and RAN	27
3.5.3	Mitigation mechanism	28
3.5.4	Summary	29
4	Experimental Evaluation and Discussion	30
4.1	Experiment Flow	30
4.2	Experimental Scenarios and Results	32
4.2.1	Unprotected Scenario	32
4.2.2	DKSS Protected Scenario	33
4.3	Experimental Evaluation and Discussion	34
4.4	Limitations	38
5	Conclusion and Future Work	39
5.1	Summary	39
5.2	Recommendations and Future Work	40
	References	41
	Abstract (in Arabic)	

List of Figures

1.1	ETSI NFV Architecture	3
1.2	5G Standalone Architecture	5
2.1	DDoS Cognitive protection solution architecture	11
2.2	Dynamic Scaling of IPS VNFs Using MANO	12
2.3	REPEL Architecture	14
3.1	5G UE registration procedure	23
3.2	Protocol stack for the Control Plane between UE and the AMF	24
3.3	DKSS architectural components	26
3.4	DKSS infrastructure and application layers	26
4.1	DKSS Operation	31
4.2	AMF memory utilization in the unprotected test scenario. . .	33
4.3	AMF memory utilization in the protected test scenario. . . .	34

List of Tables

2.1	Comparison of related work	18
4.1	Experiment summary and parameters for the unprotected scenario	32
4.2	Experiment summary and parameters for the protected sce- nario	34
4.3	Comparison between DKSS and REPEL	37

ACKNOWLEDGMENTS

First and foremost, I would like to praise and thank Allah, the Almighty, for granting me countless blessings among which I count the pursuit of knowledge and the and opportunity to accomplish this thesis.

I would like to express my sincere gratitude to my supervisor Dr. Reham Almesaeed for her kind support and guidance while writing this thesis.

To my beloved family for their endless support and encouragement.

To my friends and colleagues for their support throughout this journey.

PUBLICATIONS

The following paper was written during working on the thesis, and **submitted** to Mobile Networks and Applications Journal (Springer):

Abdulrahman Khaled, Reham Almesaeed, Abdulla Alasaadi, and Agron Demiraj. Dynamic Resource Scaling for Signaling DDoS Attack Mitigation in 5G Networks. Mobile Netw Appl (2022).

LIST OF ABBREVIATIONS

2G	Second-generation of mobile telecommunications technology
3G	Third-generation of mobile telecommunications technology
3GPP	3rd Generation Partnership Project (3GPP)
4G	Fourth-Generation mobile communication systems
5G	Fifth-Generation mobile communication systems
5G-AN	5G Access Network
5GC	5G Core
5G-SA	5G Standalone
AE	Action Enforcer
AKS	Azure Kubernetes Service
AMF	Access and Mobility Management Function
AUSF	Authentication Server Function
AWS	Amazon Web Services
AZ	Availability Zone
BSF	Binding Support Function
CAPEX	Capital expenditures
CLI	Command Line Interface
CPU	Central Processing Unit

DDoS	Distributed Denial of Service
DKSS	DDoS Kubernetes Scaling Solution
DM	Decision Maker
DNS	Domain Name System
EC2	Amazon Elastic Compute Cloud
EKS	Amazon Elastic Kubernetes Services
eMBB	Enhanced Mobile Broadband
eNodeB	Evolved Node B
EPC	Evolved Packet Core
ETSI	European Telecommunications Standards Institute
FCA	Flow Control Agent
GCC	Gulf Cooperation Council
GKE	Google Kubernetes Engine
gNB	Next Generation NodeB
GRE	Generic Routing Encapsulation
GTP	GPRS Tunnel Protocol
GUI	Graphical user interface
HSS	Home Subscriber Sub-system
HTTP	Hypertext Transfer Protocol

IDE	integrated development environment
IDS	Intrusion detection systems
IIoT	Industrial Internet of Things
IMS	IP Multimedia Subsystem
IoT	Internet of Things
IP	Internet Protocol
IPS	intrusion prevention systems
IT	Information Technology
K8s	Kubernetes
LTE	Long Term Evolution
MANO	Management and Orchestration
ML	Machine-Learning
MME	Mobility Management Entity
mMTC	Massive Machine Type Communications
NAS	Non-Access Stratum
NAS-MM	NAS Mobility Management functionality
NFV	Network Function Virtualization
NFVI	Network Functions Virtualization Infrastructure
NG-AP	NG Application Protocol

NR	New Radio
NRF	Network Repository Function
NSA	Non-Standalone
NSSF	Network Slice Selection Function
OPEX	Operating Expenses
OS	Operating System
PCF	Policy Control Function
PDN-GW	Packet Data Network Gateway
PGW	PDN gateway
RAM	Random Access Memory
RAN	Radio Access Network
RRC	Radio Resource Control
SA	Standalone
SBA	Service-Based Architecture
SCTP	Stream Control Transmission Protocol
SDN	Software Defined Networking
SGW	Serving Gateway
SLA	Service-Level Agreement
SMA	Security Monitoring Agent

SMF	Session Management Function
TCP	Transmission Control Protocol
UDM	Unified Data Management
UDP	User Datagram Protocol
UDR	Unified Data Repository
UE	User Equipment
UPF	User Plane Function
URLLC	Ultra-Reliable Low Latency Communication
vEPC	Virtual Evolved Packet Core
vHSS	Virtual Home Subscriber Sub-system
VIM	Virtualized Infrastructure Manager
VM	Virtual Machine
vMME	Virtual Mobility Management Entity
VNF	Virtual Network Function
vSGW	Virtual Serving Gateway
vSPGW	Virtual Serving Packet Data Network Gateway
VxLAN	Virtual Extensible Local Area Network

Chapter 1

Introduction

This chapter provides a background about the evolution of the telecom technology towards the 5G network, then introduces the main concepts of the 5G Core network, followed by a discussion of DDoS attacks in the context of the 5G domain, followed by introducing the research problem, objectives and contribution.

1.1 Overview

1.1.1 Evolution to 5G Network

Over the past decade, wireless communication technologies have evolved drastically, including Wi-Fi, Long Term Evolution (LTE), and recently the Fifth-Generation mobile communication systems (5G). The evolution to 5G aims to serve the following use cases: Enhanced Mobile Broadband (eMBB); providing a higher peak data rate reaching 10 Gbps, Massive Machine Type Communications (mMTC), Ultra-Reliable Low Latency Communication (URLLC), and high mobility (500 km/h). These features unlock a vast potential to IoT use cases, including machine-to-machine communication, mission-critical IoT applications, and Industrial Internet of Things (IIoT) (Ghorbani, Mohammadzadeh, & Ahmadzadegan, 2020; Wazid, Das, Shetty, Gope, & Rodrigues, 2021; De Souza Neto, Amatucci, Nassif, &

Marques Farias, 2021).

IoT and 5G are vital enablers connecting all parts of peoples' lives and activities, including transport, well-being, healthcare, smart homes, and industry. The total 5G subscriptions globally reached 620 million in 2021 and are expected to reach 4.4 billion by 2027. On a regional level, 5G adoption in the Gulf Cooperation Council (GCC) is forecasted to grow from 9% of the overall Mobile technologies in 2021 to 80% by 2027. With an expected 65 million subscriptions, the regional adoption rate would rank third globally, behind only North America (90%) and Western Europe (82%). 5G has been the fastest growing mobile technology and is expected to cover three-quarters of the world's population by 2027 (*Ericsson Mobility Report-June 2022*, 2022). This drastic increase in the number of connected devices will also expand the landscape of security risks, and vulnerabilities in the 5G-IoT domain (Ghorbani et al., 2020).

1.1.2 5G Core Network

The 5G Core (5GC) is at the heart of the 5G technology. Telecom Service Providers must embark on a journey to transform from LTE Evolved Packet Core (EPC) to 5GC. This transformation includes multiple steps; 1- Virtualizing Packet Core, IP Multimedia Subsystem (IMS), and subscription management network functions, 2- Decoupling of user plane and control plane functions, 3- Network slicing, 4- Edge cloud computing, and 5- Synergizing user plane function for mobile and fixed networks (*Ericsson: Building a new world - A guide to evolving from EPC to 5G Core*, 2020). In the 5GC, two primary technology enablers allow flexibility and simplicity. First, Software Defined Networking (SDN) is used to decouple the user plane and control plane and allow centralized management and programmability of network resources through SDN Controller. Second, Network Function Virtualization (NFV) abstracts the legacy purpose-built network hardware functions

used in the Second-generation of mobile telecommunications technology (2G), the Third-generation of mobile telecommunications technology (3G), and the Fourth-generation of mobile telecommunications technology (4G) into Virtual Network Functions (VNFs). In principle, VNFs are meant to be hardware agnostic. They can be hosted on any underlying hypervisor and hardware, enabling fast and dynamic deployment, less complex hardware lifecycle management and reduced cost (Forland, Krlevska, Garau, & Gligoroski, 2019). Refer to NFV Architecture in Figure 1.1 (*ETSI - Standards for NFV - Network Functions Virtualisation* | *NFV Solutions*, 2019).

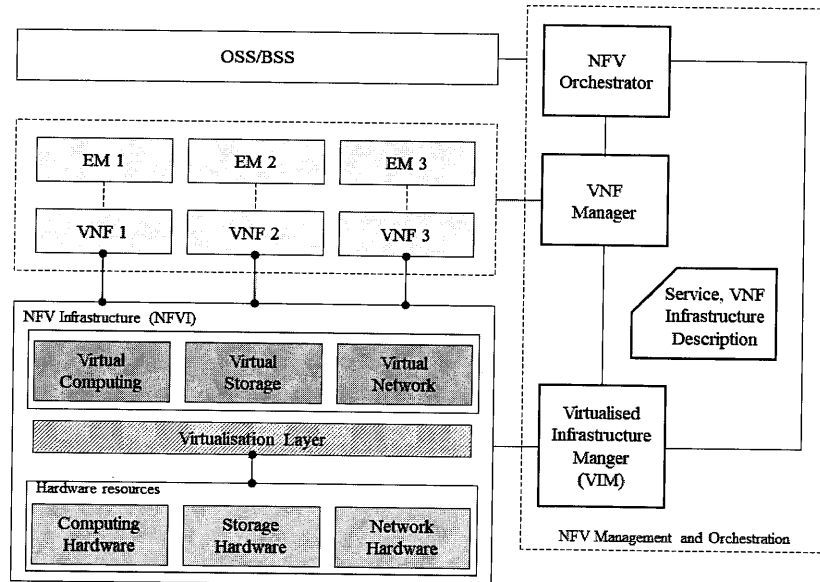


Figure 1.1: ETSI NFV Architecture

In addition to the European Telecommunications Standards Institute (ETSI), the development of the 5G Telco Cloud architecture is driven by telco, cloud, and network vendors like Samsung, Huawei, Cisco, Ericsson, Nokia, VMware, and Juniper (Gebremariam et al., 2021).

Furthermore, the Telco Cloud architecture is evolving towards cloud-native architecture by adding a layer of containerization using open-source technologies like Docker and Kubernetes (Gebremariam et al., 2021), aim-

ing to improve the agility and flexibility of the 5G network to meet various 5G use cases (*Huawei: Next-Generation Telco Cloud Based on Cloud Native*, 2020). The cloud-native 5G core is achieved by adopting microservices Service-Based Architecture (SBA), first introduced in 3GPP Release 15 (*Huawei: Next-Generation Telco Cloud Based on Cloud Native*, 2020). In SBA, the Network Functions have been defined as independent containerized network services, and the inter-function communication has been simplified over Hypertext Transfer Protocol (HTTP) protocol (Matefi & Stief, 2022). In SBA, 5G core functions can provide and request services from any other function in a flexible and modular architecture that allows faster development and more agility. Previous mobile technologies lacked this flexibility as telco-specific protocols were used in a static architecture for inter-node communications.

5GC has broadly two deployment options (*Ericsson: Building a new world - A guide to evolving from EPC to 5G Core*, 2020):

1. Non-Standalone (NSA): A hybrid core based on the New Radio (NR) Non-Standalone that serves both 4G and 5G and enables mobility and interworking between the two networks.
2. Standalone (SA) exclusively consists of 5G core functions, and is purposed for new 5G use cases, including IoT, served through the first generation of 5G New Radio.

In this thesis, 5G-SA architecture was adopted because the proposed DDoS Kubernetes Scaling Solution (DKSS) aims to protect 5GC from signaling DDoS attacks initiated by IoT devices, which will be a dominant use case in 5G-SA.

As illustrated in Figure 1.2, the 5G-SA core demarcates the user plane from the control plane. Users' payload is sent and received on the user plane layer through the User Plane Function (UPF), which is the gateway towards

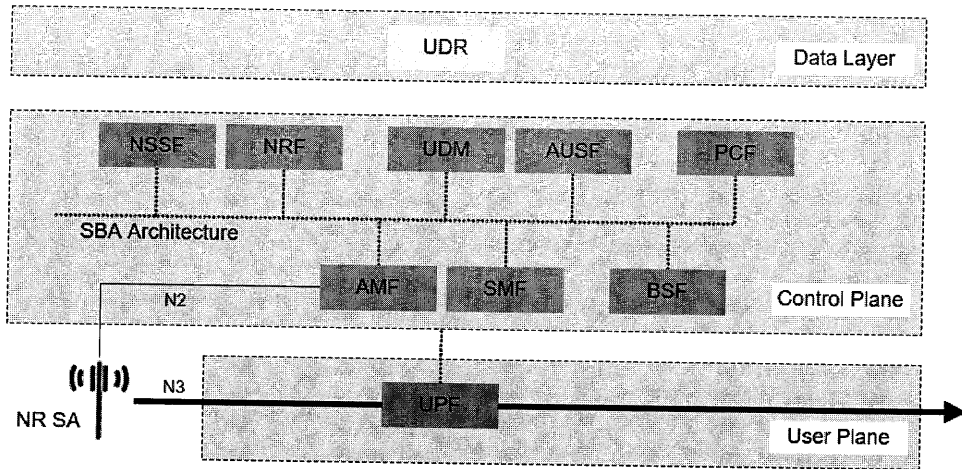


Figure 1.2: 5G Standalone Architecture

external networks, whether the internet or private clouds. The 5G New Radio (NR) SA communicates with UPF using the N3 3GPP interface. Moreover, on the Control plane layer, the radio has another signaling interface (N2) towards the Access and Mobility Management Function (AMF), the gateway for control plane communication between the Radio Access Network (RAN) and other 5GC functions. AMF is mainly responsible for the access and mobility of User Equipment (UEs). The Session Management Function (SMF) is responsible for session management and Internet Protocol (IP) address allocation. The Policy Control Function (PCF) manages the rules enforced on users based on subscriptions and profiles. The scaling of the policy and charging nodes is facilitated by the Binding Support Function (BSF). The Network Slice Selection Function (NSSF) supports selecting the appropriate network slice for UEs. The Network Repository Function (NRF) maintains the discovery, registration, and health status of 5GC Network Functions. The Management of 5G subscription, authentication, and authorization is handled by the Unified Data Management (UDM) and the Authentication Server Function (AUSF). On the Data Layer, the Unified

Data Repository (UDR) collects and stores user data from other functions like UDM and PCF. This data includes subscription, application, policy, and structured data (*3GPP-R17 5G Architecture*, 2022).

1.1.3 DDoS in the 5G-IoT Domain

IoT devices are generally prone to design, firmware, and hardware weaknesses. Common security issues in IoT devices include weak passwords, improper or no encryption, non-standard operating systems, and the lack of security support and patches (Ghorbani et al., 2020). These weaknesses make IoT an attractive platform for attackers to exploit through existing vulnerabilities and to launch DDoS attacks from a massive number of 5G-IoT devices targeting other 5G connected devices, 5G infrastructure, and other resources on the internet (Ghorbani et al., 2020; Gurusamy, Deva Priya, Yibgeta, & Bekalu, 2019).

In traditional DDoS mitigation solutions, security appliances like Firewalls, Intrusion detection systems (IDS), and intrusion prevention systems (IPS) are positioned in optimum locations in the network to mitigate the attacks, which provides adequate mitigation in traditional networks. However, with the evolution to 5G and IoT, these static DDoS mitigation mechanisms will not withstand 5G DDoS attacks due to the massive number of devices and dynamic nature of 5G and IoT environments (Sotelo Monge et al., 2018; Ahmad et al., 2019).

One of the DDoS attack types in 5G is the Signaling DDoS attack, which targets to consume the resources of the control plane elements in a 5G network. The risk of signaling storms will rise drastically in the 5G network due to the massive number of connected devices. 5G UEs use the Non-Access Stratum (NAS) protocol to perform intensive control plane signaling procedures like authentication, attach, detach, bearer activation, and location update. One of the core nodes involved in these signaling procedures is

the AMF, the successor of the Mobility Management Entity (MME) in LTE. The 5GC must be highly resilient to be protected from signaling attacks (Ahmad et al., 2019).

1.2 Research Problem

The problem of signaling DDoS attacks in the 5G domain will continue to pose a significant risk to the availability of 5G services. Its impact can extend to critical IoT services such as self-driving cars, health care applications, industrial use cases, and other mission-critical applications. This thesis aims to introduce a mitigation mechanism to absorb the DDoS attack rather than block it, utilizing the flexibility provided by Kubernetes in a cloud-native 5GC architecture.

1.3 Research Objectives

1. Implement a cloud-native containerized 5G Core network on a public cloud.
2. Deploy scaling parameters in Kubernetes for memory resources to absorb the load on core functions triggered by a signaling DDoS attack.

1.4 Contribution to the Field

The thesis's contribution is to introduce a novel solution (DKSS) to mitigating signaling DDoS attacks in the 5G-SA networks. To the best of the author's knowledge, this is the first work that has tested this type of attack on 5G stand-alone functions like AMF. Furthermore, DKSS provides two levels of scaling to absorb the signaling DDoS attack, first, vertical scaling

of Kubernetes containers, and second, horizontal scaling of the underlying AWS resources. The experimental findings demonstrated successful mitigation of the signaling DDoS targeting 5G AMF function with no time overhead between attack start and mitigation. Moreover, DKSS is highly applicable to production networks, as it utilizes Kubernetes built-in features and works within the boundaries of the 5G 3GPP and microservices standard architectures.

1.5 Thesis Outline

This thesis is structured as follows. Chapter 2 presents the literature review of DDoS in the 5G domain; chapter 3 presents a novel solution to mitigating signaling DDoS attacks in cloud-native 5GC; chapter 4 illustrates the experimental evaluation and discussion; Chapter 5 concludes the thesis and presents the future work.

Chapter 2

Literature Review

This chapter outlines the theoretical framework of the literature review. Then surveys the related work in the literature, followed by a summary.

2.1 Theoretical Framework

Mitigation of DDoS attacks in the 5G domain has recently drawn increased attention from researchers; this is due to the risks introduced in 5G, including 5G-IoT vulnerabilities, the vast number of devices forecasted on 5G, and the new layers of virtualization and software components. Researchers have tried to address different types of DDoS attacks using various features of SDN and NFV. Nonetheless, the research in this domain is challenged by the lack of open-source tools and by the nature of 5G virtualized core networks, which are still evolving and have not yet matured in comparison to 4G. This research area still offers significant opportunities for contribution, especially in relation to protecting against control plane signaling DDoS

attacks, a topic which has limited contributions so far.

2.2 Related Work

This section reviews recent papers in the domain of 5G DDoS mitigation, which addressed this security risk from the virtualized core network perspective.

(Serrano Mamolar et al., 2019) proposed a solution to protect against DDoS in a 5G multi-tenant environment. They proposed a novel autonomic countermeasure with no human intervention. One of the differentiators of this solution is that it is based on a cognitive approach that decides the optimum locations in the network to mitigate an attack within a 5G traffic flow. The solution extends IDS capabilities to analyze 5G double encapsulated traffic. In 5G flow, traffic gets first encapsulated using Virtual Extensible Local Area Network (VxLAN) to achieve separation of 5G tenants, then gets encapsulated again using GPRS Tunnel Protocol (GTP) to enable mobility of devices. The suggested design provides a decentralized cognitive approach for DDoS mitigation control, where the DDoS detection happens in the core network, while the mitigation is applied in different locations in 5G core and edge segments based on the analysis of each attack. These decisions are made on flow, user, and tenant levels. The solution comprises of the following components: Security Monitoring Agent (SMA), Decision Maker (DM), Action Enforcer (AE), and multiple 5G Flow Control Agents (FCAs) distributed across the network. The simulated attack was generated from multiple 5G devices belonging to different 5G tenants. The attack traffic passes through the edge nodes, then through the Virtual Serving Gateway (vSGW) in 5G Core, reaching the targeted victim on the internet. Each element in the edge and core has an associated FCA virtual machine (VM) used to mitigate the attack. To detect the attacks, traffic gets mirrored from

the core switch that hosts the vSGW to the Security Service machine, which comprises Snort and SMA. Once Snort detects an attack using signature-based detection, it alerts SMA. SMA refines the alert with 5G tenant meta-data and sends the information to the Decision Maker (DM). The DM then analyzes the information based on multiple factors, including the type of attack, frequency of alerts, and other rules. It then it decides whether to forward or drop the traffic and where precisely in the network to implement the action. The Action Enforcer (AE) defines the action plan and instructs the respective distributed FCAs to take action accordingly. See Figure 2.1. This solution was tested against a User Datagram Protocol (UDP) flood at-

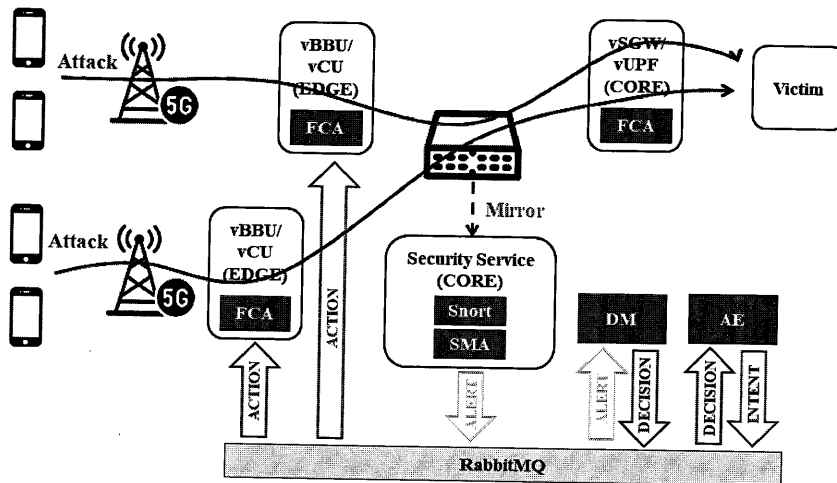


Figure 2.1: DDoS Cognitive protection solution architecture

tack, where 256 bots generated a total of 100 Mbps double encapsulated 5G traffic. The result was the effective mitigation of the attack. The authors emphasized that the solution can be generalized to protect against other types of DDoS by implementing suitable IDS within the solution to detect other attacks (Serrano Mamolar et al., 2019).

(Köksal, Dalveren, Maiga, & Kara, 2021) Proposed an approach to mitigating Domain Name System (DNS) flood attacks within the 5G NFV network by exploiting the advantage of the Network Functions Virtualization

Infrastructure (NFVI) by scaling out IPS VMs based on the analysis of their resources. The design utilizes Management and Orchestration (MANO) capabilities to scale Virtualized Network Function (VNFs) resources dynamically within the 5G core. The architecture components are as follows: 1) NFVI, an OpenStack platform to host the virtual functions. 2) Load balancer, used to distribute traffic among IPS VMs. 3) Multiple Intrusion Prevention System (IPS) VMs used to filter out malicious DNS traffic and forward only legitimate DNS traffic to the DNS server. 4) Management and orchestration (MANO) component, which collects performance metrics from NFVI/Virtualized Infrastructure Manager (VIM), including disk, memory, Central Processing Unit (CPU) utilization, and network packets. It controls the scaling of the IPS VMs accordingly. When an attack is detected, MANO dynamically decides to scale in/out the IPS VMs to ensure their resources are not depleted and can function properly to protect the DNS server. See Figure 2.2. This solution was experimented within a 5G NFV environment and illustrated successful mitigation (Köksal et al., 2021). However, one of

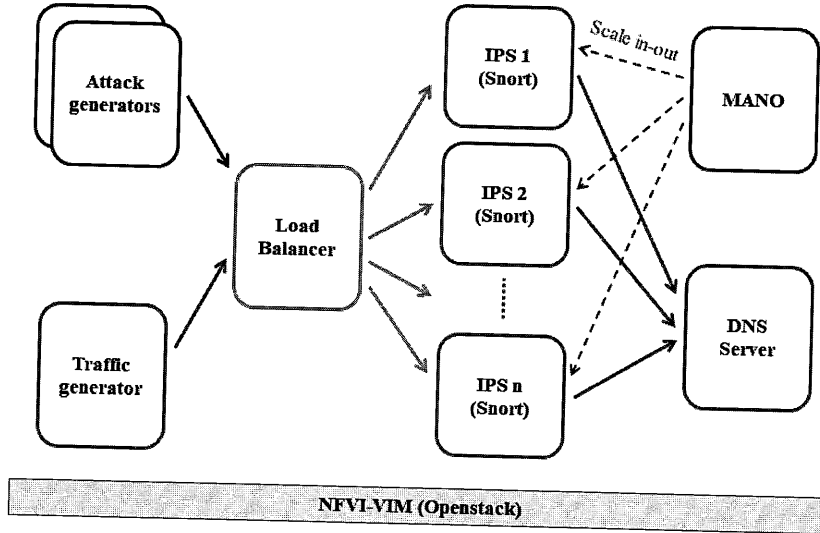


Figure 2.2: Dynamic Scaling of IPS VNFs Using MANO

the limitations of this design is the dependency on one load balancer that re-

ceives all legitimate and malicious traffic without scaling capability, which is another bottleneck that is vulnerable to DDoS if not appropriately sized.

(Silva et al., 2021) Proposed a solution (REPEL) to protect 5G Core from signaling DDoS attacks using Game theory. The attack model is based on compromised mobile devices controlled by an attacker. The attacker initiates a DDoS attack on the 5G core control plane by instructing the mobile devices to execute attach/detach signaling procedures with high frequency to overload the virtual Mobility Management Entity (vMME). The authors introduced a model based on game theory to analyze and predict the attacker's behavior. Accordingly, REPEL suggests the right time to scaling-out vMME instances to absorb all legitimate and malicious traffic. This approach helps to avoid outages, frustrate the attackers, and push them to stop the attack as the desired damage is not realized. The game model is based on two players, defender and attacker, where each one has two playing strategies, the attacker can either attack or stop, and the defender can either defend or stop. The attacker's aim is to cause the most possible damage to the network while utilizing the minimum number of bots. On the other hand, the defender's goal is to determine the optimum time to scale-out vMMEs, absorb the attack, and protect the network from damage with the minimum cost. In the proposed architecture, IDS was used to detect DDoS attacks, then reports to REPEL, which evaluates the situation based on the game theory model and decides whether additional vMME instances are required. Accordingly, it instructs the orchestrator to add extra vMMEs to the network as required. In order to provide this flexibility, the mitigation method exploits the 5G Core virtualization capabilities, where shared infrastructure is available and can be used to host the newly created VNFs. See Figure 2.3. REPEL was experimented with six bots generating signaling attach procedure DDoS attacks on a simulation testbed. Openairinterface was used to simulate vEPC, including (vMMEs, vSPGW, and vHSS)

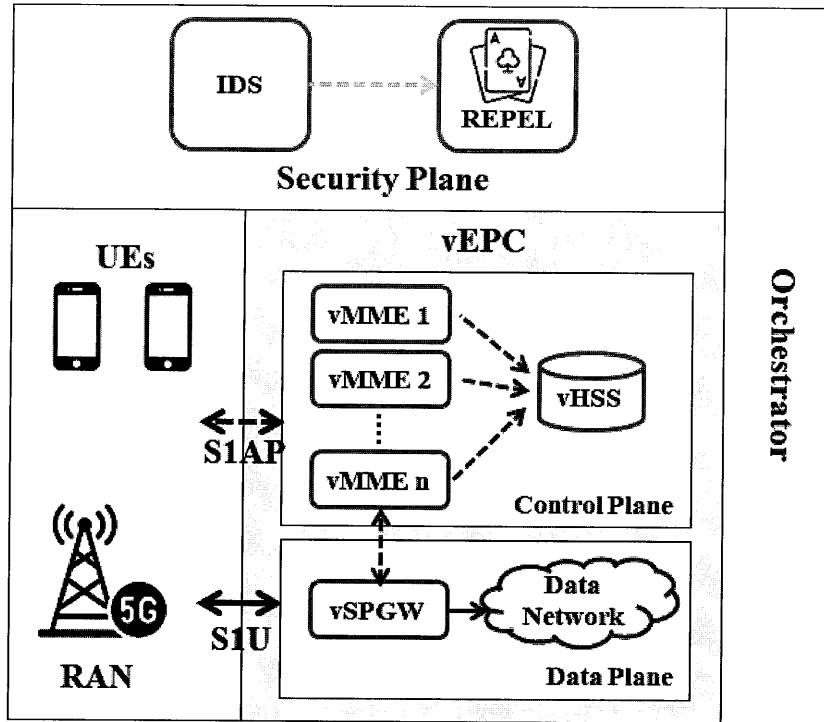


Figure 2.3: REPEL Architecture

and RAN (eNodeBs and UEs). REPEL was demonstrated to be efficient in absorbing the attack and in frustrating the adversary into stopping the attack due to failure to impose the desired damage. However, REPEL has the following limitations: scaling out vMME instances to absorb the attack consumes additional resources on the cloud and will eventually reach a limit as the available resources are finite. The authors suggested that REPEL can protect the service availability once the attack starts and provide some time for the security team to analyze the attack and apply other actions to stop it (Silva et al., 2021). Another limitation that was noticed is that the solution only scales-out vMMEs to absorb the high volume of signaling traffic while it did not consider the propelling effect on other network's signaling elements sitting behind vMME in the signaling flow, like vHSS.

The 5G slicing architecture introduced a new threat model. A DDoS at-

tack on one slice, either flooding its network links, depleting its processing capacity, or both, can impact neighboring slices that share the same physical host or network links. Sattar and Matrawi proposed a solution to protect from DDoS attacks targeting network slices. The researchers proposed a solution based on a mathematical model to mitigate the attacks using slice isolation. They have experimented with multiple isolation approaches: 1) Inter-slice isolation: The slice under attack can be shifted to another physical host to avoid impacting other slices. This approach provided the best results for protecting other slices; however, its disadvantage is the inefficient utilization of infrastructure resources. 2) Intra-slice isolation: The VNFs of one slice are distributed among multiple physical hosts. This approach provided better availability since if one VNF is attacked; the network slice does not face a complete outage. Intra-slice isolation also provided considerable mitigation of the impact of DDoS attacks, but not to the level of Inter-slice isolation. The testing was done by combining simulation and experiment. The simulation was done in ns3 tool to simulate the 5G radio where 320 mobile devices are connected and generate traffic. The output of this simulation was captured and later used as an input to the experiment testbed, where six servers were used: three servers used to allocate slices, two as DDoS generators, and one server to represent a mobile client. In order to test the impact of the DDoS, one sample 5G application (authentication) was used between the client and the Home Subscriber Sub-system (HSS) VNFs. Moreover, the authentication performance was measured before and throughout the DDoS mitigation tests. The measured metrics were available bandwidth and response time. The authors highlighted limitations of the solution, including the static methodology, which does not consider an efficient allocation of resources. Furthermore, there is no current mechanism for VNF migration between physical hosts (Sattar & Matrawy, 2019).

(Forland et al., 2019) built a 5G-like simulation model to imitate DDoS

attacks generated by 5G devices targeting any device on the internet. The model was built by integrating OpenFlow OMNet++, which was used to simulate 5G SDN-Core including SDN controller and SDN switch, with SimuLTE to simulate the Access network including UE, Evolved Node B (eNodeB), and the Packet Data Network Gateway (PDN-GW). The model was designed to protect specifically against Transmission Control Protocol (TCP) SYN Flood DDoS attacks. This attack exploits a flaw in the TCP handshake, where the receiver receives a SYN packet, sends back an ACK packet, and keeps an open session waiting for a SYN-ACK from the sender. This flaw leads to high utilization of server resources and is abused by attackers to deplete the victim's resources by sending a flood of TCP SYN packets. The proposed protection method takes advantage of the SDN in 5G Core, where all traffic destined to different VNFs is passed through SDN infrastructure, centrally monitored by the SDN controller. A simple detection method was used in this simulation to detect attacks based on the rate of sent packets, assuming that the traffic generated by DDoS participants is significantly higher than that of average users. The senders are identified by source and destination IP addresses, and any flow that crosses the defined threshold gets blocked accordingly at the SDN level. This controlled POC was successful as the designed model could detect TCP SYN attacks effectively. A disadvantage of this solution is that it is solely dependent on source IP addresses, which can be bypassed using IP spoofing (Forland et al., 2019).

2.3 Summary

The reviewed papers in the previous section have approached the DDoS problem in the 5G network from different angles. In summary, they can be categorized as follows:

1. Cognitively block DDoS in the most optimum location in the network using distributed agents (Serrano Mamolar et al., 2019)
2. Scale-up core network components to absorb the attack traffic and minimize damage (Köksal et al., 2021; Silva et al., 2021)
3. Shuffle VNFs between physical hosts to contain the damage (Sattar & Matrawy, 2019)
4. Utilize SDN as a centralized control to monitor and block the attacks globally (Forland et al., 2019)

All surveyed approaches were built on the 5G core network capabilities of network virtualization. Virtualization enables new flexible ways to address DDoS, including resource scaling, live movement of workloads to absorb and contain DDoS attacks and build distributed mitigation mechanisms in the 5G Core and RAN. Table 2.1 summarized the surveyed related work. Multiple papers have highlighted a common challenge: the lack of full-fledged testbeds and simulators for 5G networks, especially with slicing support. Researchers used the available open-source tools to the best possible to satisfy their experiment or simulation objectives.

Among the surveyed papers, one addressed signaling DDoS attacks using resource scaling (Silva et al., 2021), where the attack was tested in a 4G network prototype. In the next chapter, a novel solution is introduced, extending the same attack model and mitigation principle to the 5G-SA core network, aiming to absorb the attack rather than block it.

Table 1 Comparison between reviewed papers

Paper	DDoS Type	Attack source	Attack target	Protection Method	Limitations
Autonomic protection of multi-tenant 5G networks against UDP flood DDoS	UDP	5G endpoints	Server on the Internet	Cognitive approach to mitigate the attack in the most optimum location in the network using distributed FCAs	NA
DDoS mitigation in NFV-based 5G networks using MANO	DNS	5G endpoints	DNS Server	Dynamic Scaling of IPS VNFs Using MANO	Dependency on one load balancer that receives all legitimate and malicious traffic without scaling capability, which is another bottleneck that is vulnerable to DDoS
REPEL: Defending 5G Control Plane From DDoS Signalling Attacks	Signaling attack	5G endpoints	vMME	Scale-up vMMEs using game theory to absorb the attack and frustrate attacker	Did not consider the propelling effect on other network signaling nodes
Using Slice Isolation to Mitigate DDoS Attacks on 5G Core Network Slices	Generic TCP/UDP	1) 5G endpoints in a slice 2) Slice admin	5G network slice	Mathematical model to isolate affected VNFs/slices on different physical hosts	1) Static method, that does consider efficient allocation of infrastructure resources 2) No current mechanism for VNF migration between hosts
Preventing DDoS with SDN in 5G	TCP SYN	5G endpoints	Server on the Internet	Monitor and block the attacks on SDN level based on source IP address	Relies on IP addresses, which can be bypassed using IP address spoofing

Chapter 3

Mitigation of Signaling DDoS in 5GC - An Innovative Approach (DKSS)

This chapter first provides an introduction and context around the concepts related to the proposed solution, including containerization, how it is adopted in the 5G domain, and the concept of 5G signaling. Then the DDoS attack model explains how 5G signaling can be exploited to attack 5G core functions. Finally, the DKSS system architecture and mitigation mechanism are described.

3.1 Introduction

This chapter outlines the proposed solution to protect 5G core network elements from signaling DDoS attacks. Signaling attacks target overloading 5GC resources by initiating a massive number of simultaneous signaling procedures.

The proposed solution is based on Kubernetes containerized infrastructure, where the attack is mitigated by utilizing the scaling capabilities provided by Kubernetes, orchestrating dynamic allocation of memory resources to 5GC components, ultimately absorbing the volumetric DDoS, and maintaining the 5G service level provided to customers intact.

In a 5G-SA setup, UEs connect to the Next Generation NodeBs (gNBs),

which interface to the 5G-SA core through the AMF for the control plane and the UPF for the payload. The considered attack model in this paper is a flood of user registration requests targeting to overwhelm the stateful AMF function resources, specifically the Random Access Memory (RAM) resources, to the point where it will crash and lose all connection towards RAN and Core. This can lead to severe network outages.

3.2 Containerization in 5G

Containerization technology started in the Information Technology (IT) domain and revolutionized how applications are developed and managed (Shah & Dubaria, 2019). Containerization is considered a lighter model of virtualization, where instead of virtualizing the hardware resources, it abstracts the Operating System layer. The containerization architecture is based on the microservices concept, where applications are designed as a group of independent building blocks, i.e., containers. Each container runs a specific functionality and provides its services to other containers. This modular architecture simplifies application development and provides resilience and scalability, as each service can be scaled out by adding more instances of the same container. Moreover, each container packages its application runtime environment and dependencies within itself, which makes them highly portable and able to run on any platform. Technology giants such as Microsoft, Huawei, Amazon and Netflix started using containerization for their production applications to exploit the enormous flexibility and scalability capabilities (Bentaleb, Belloum, Sebaa, & El-Maouhab, 2021; *Huawei Kubernetes Case Study*, 2020).

Orchestration platforms complement the ecosystem of containerization to better manage and orchestrate containers at scale in production environments. Kubernetes (K8s) is becoming the de-facto standard in this domain

(Shah & Dubaria, 2019). Its implementations have doubled since 2017 and are now used to manage around 50% of all containerized applications. The Kubernetes project was started by Google, before becoming an open-source platform in 2014, which further accelerated the adoption of Kubernetes by key players in the market, including AWS, Azure, and others. The value added by Kubernetes is providing a single control plane to automate the containers' service discovery, tracking resource allocation, scaling, security, self-healing, and load balancing. In addition, one of its main strengths is the abstraction of infrastructure, as it can run on public clouds, private clouds, on-prem, and multi-cloud deployments. Kubernetes can both run as a standalone environment or can be alternatively consumed as a native cloud service such as Azure Kubernetes Service (AKS), Google Kubernetes Engine (GKE), or Amazon Elastic Kubernetes Services (EKS). Moreover, it works very well with Docker, where Docker container images can be deployed in Kubernetes clusters (Sahu & Thakre, EasyChair, 2022).

The architecture of a Kubernetes cluster on a high level is composed of one Master Node at minimum, which can be increased for redundancy, and multiple Worker Nodes. The Master Node is the brain of the cluster; it orchestrates the entire cluster operations, including monitoring the load and health of the Worker Nodes and moving workloads across Worker Nodes in case of failures. On the other hand, Worker Nodes are the servers where the application containers are deployed and processed. The smallest deployment unit in Kubernetes is the Pod, an individual instance of a process in the cluster; it can include a single or multiple containers (Sahu & Thakre, EasyChair, 2022).

In the Telco clouds domain, 5G core networks have been designed based on the concept of NFV, which decouples the hardware and software components of the core nodes (Luong, Thieu, Outtagarts, & Ghamri-Doudane, 2018). Furthermore, building on the same cloud-native concept that started

in the IT field (*Huawei: Next-Generation Telco Cloud Based on Cloud Native*, 2020), the 3GPP standard defines the 5G architecture with modularity and segregation of services. The purpose of this architecture is to support 5G network cloudification and control the network capital expenditures (CAPEX) and operating expenses (OPEX) (Arouk & Nikaein, 2020). (Mfula, Ylä-Jääski, & Nurminen, 2021) Emphasized the importance of high availability in 5G clouds and surveyed the most crucial requirement for a unified containerization cluster management platform; the top five requirements were reported as follows: decouple infrastructure development from software development, continuous software development, vendor independence, scalability, and self-healing. These requirements make Kubernetes a perfect fit to satisfy the 5G roadmap (Arouk & Nikaein, 2020). Leading telco vendors, including Ericsson, Huawei, and Nokia, have already started adopting Kubernetes in their 5G network stacks (Bäckström & Bohra, 2022; *Huawei Kubernetes Case Study*, 2020; Csatari, 2020).

3.3 5G Signaling

5G mobile devices use 3GPP standard signaling procedures to interact with the core network, including initial registration, periodic registration, deregistration, and mobility. To illustrate one example, an initial registration procedure for a single 5G client comprises more than twenty steps to request resources, perform authentication, compute security keys, obtain user profile, policy selection and enforcement, and network slice authentication and authorization. Throughout this procedure, the gNB and multiple 5GC functions like AMF, SMF, AUSF, UDM, and PCF are involved. Figure 3.1 shows a simplified illustration of the registration procedure (*3GPP-R17 5G Procedures*, 2022). The signaling procedures contribute to the overall processing and memory load of 5GC functions. Supposing 5GC is not well designed

and sized, it will be vulnerable to signaling storms that can damage the entire network due to either attacks or operational incidents (Ahmad et al., 2019).

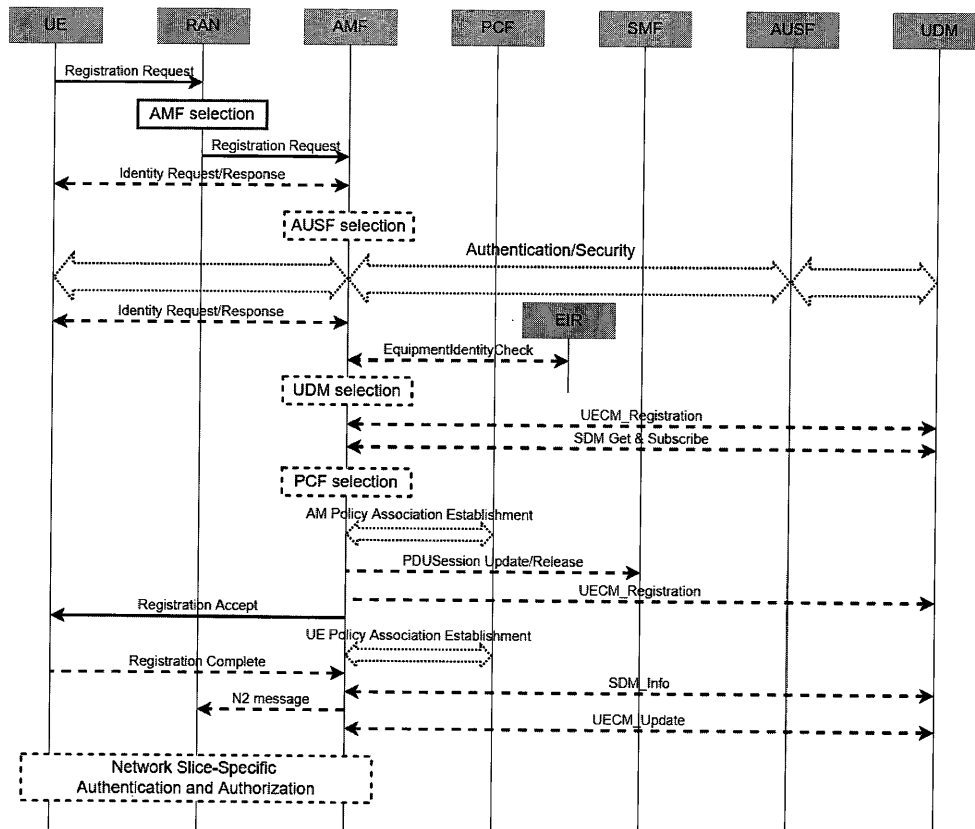


Figure 3.1: 5G UE registration procedure

AMF plays a central role in the signaling procedure; as illustrated in Figure 3.1, it represents a gateway for control plane communication between RAN and 5GC functions. AMF is mainly responsible for the access and mobility of UEs. The interaction between UE and AMF is defined by the N1 3GPP signaling interface based on the NAS protocol. NAS Mobility Management functionality (NAS-MM) handles UE registration, connection management, user plane session creation, termination, and integrity and encryption of NAS signaling (3GPP-R17 5G Architecture, 2022).

5G Access Network (5G-AN) represented by gNB relays the communication between UE and AMF, see Figure 3.2 (*3GPP-R17 5G Architecture*, 2022). Radio protocols are used for the communication leg towards the UE. In contrast, the N2 interface defines the other leg of communication toward AMF. The lower three protocol layers of the N2 interface map to the same first layers of the TCP/IP stack, i.e., the Physical layer, the Data link layer, and the IP Network layer. On the transport layer, the Stream Control Transmission Protocol (SCTP) is used; SCTP is a connection-oriented protocol intended to provide more reliable and secure communication for signaling use cases compared to TCP. One of the fundamental features of SCTP is multihoming which allows the source and destination to have multiple diversified communication channels for the same communication stream (Hercog, 2020). Next, The NG Application Protocol (NG-AP) is responsible for the signaling communication between RAN and AMF and provides transport for NAS protocol.

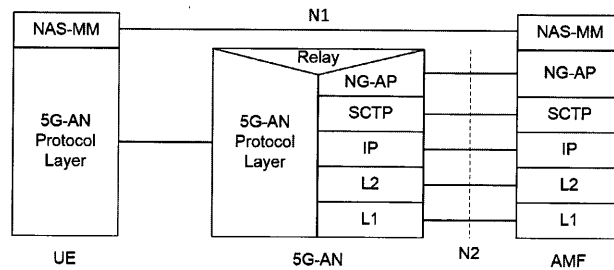


Figure 3.2: Protocol stack for the Control Plane between UE and the AMF

3.4 DDoS Attack Model

(Ernsberger, George, & Arumugam, 2019) demonstrated the impact of a signaling DDoS attack in LTE, which targets to consume MME resources. The same was simulated by (Bassil, Chehab, Elhajj, & Kayssi, 2012) using

OPNET to attack LTE networks and emphasized that this type of threat is genuine due to the spread of smartphones on LTE networks, which are vulnerable to being compromised on the Operating System (OS) level. (Silva et al., 2021) also simulated a similar attack on LTE MME and showed that the signaling caused increased memory utilization on the MME. If the same threat factor is projected onto the 5G network, the same risk will apply to the 5G AMF function as it is equivalent to MME in 4G. However, the 5G scenario's risk has a larger magnitude than 4G due to the massive number of forecasted 5G-IoT devices and their primitive unmanaged firmware, which are more prone to vulnerabilities than smartphones' OS. The DDoS attack model in this experiment borrows the same signaling attack concept demonstrated in the literature for 4G EPC, assuming an adversary has gained control over many compromised 5G-IoT devices. Furthermore, it was applied to 5GC by initiating simultaneous registration procedures from the compromised 5G-IoT devices, aiming to overwhelm 5GC functions and deny the services to legitimate users. The main target of the attack in this experiment was to increase the memory utilization of the AMF to the maximum with the aim of driving it to crash. To the author's knowledge, this is the first work to test 5G signaling DDoS attacks on genuine 5G-SA functions.

3.5 DKSS System Architecture

This section explains DKSS solution components, as illustrated logically in Figure 3.3. The components are discussed bottom-up, including infrastructure and application layers, as depicted in Figure 3.4.

3.5.1 Infrastructure layer

Public and private clouds enable significant benefits for various technology use cases. For this experiment, AWS public cloud was used to spin

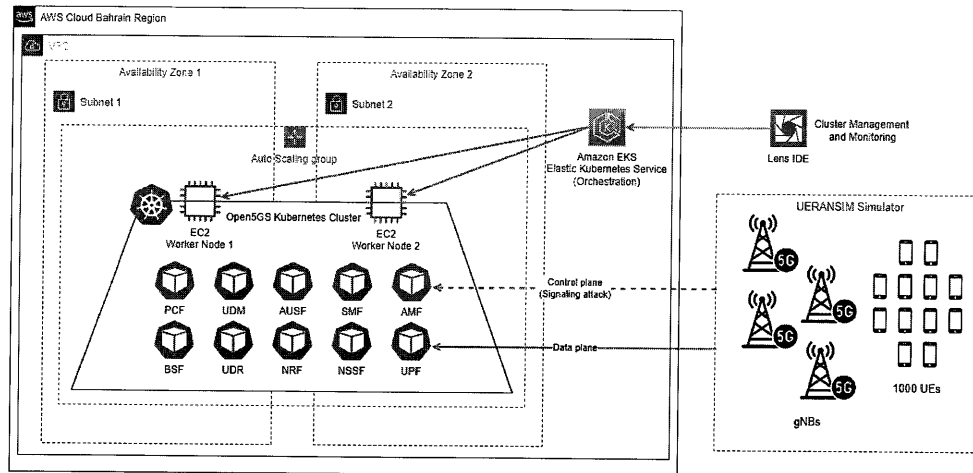


Figure 3.3: DKSS architectural components

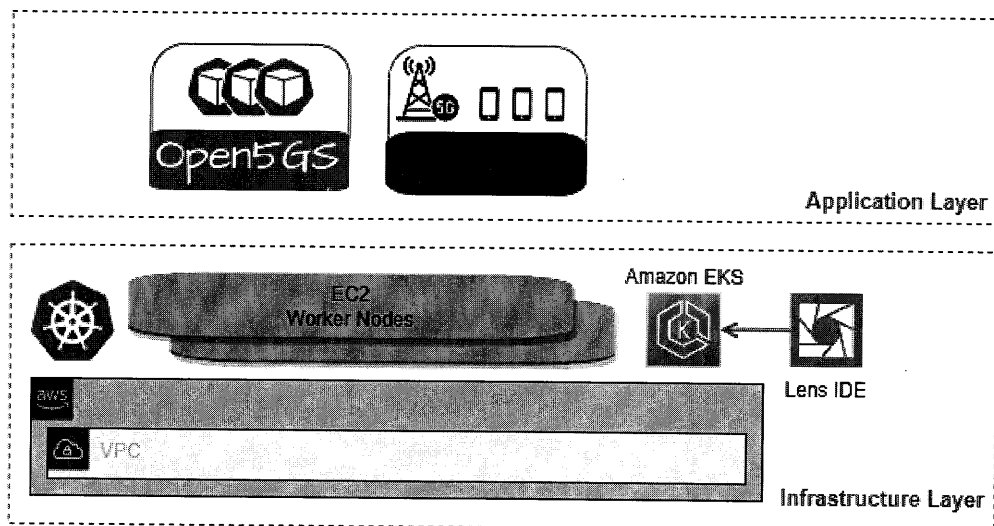


Figure 3.4: DKSS infrastructure and application layers

up the required compute resources instantly. A Kubernetes cluster (v1.22) was created using Amazon EKS, which is an off-the-shelf managed service used to create and manage Kubernetes clusters and scale applications without having to create the Kubernetes environment from scratch (*Amazon EKS - Managed Kubernetes Service*, 2022). EKS was used to deploy two Elastic

Compute Cloud (EC2) worker nodes of size t3.medium (2 vCPUs, 4 GB RAM), spanned across two availability zones (AZ) to provide geographical redundancy. The worker nodes were grouped in an AWS autoscaling group, which can be configured to scale horizontally dynamically by creating more EC2 worker node instances depending on the servers' resource utilization. Refer to Figure 3.3, which shows the detailed architecture of the implementation.

Moreover, to manage the Kubernetes environment effectively, The AWS EKS cluster was integrated to Lens Integrated Development Environment (IDE) v6.0.0. Lens is an external management tool installed on a local machine and connected to the AWS account using AWS Command Line Interface (CLI) in the backend. Lens provided significant benefits for this experiments, as it enabled cluster management in a user-friendly Graphical user interface (GUI), and provided real-time monitoring of the resources at the cluster level, worker node level, and container level, showing the requested capacity, capacity limits, and actual usage. It also showed the health status and logs of containers, which were monitored throughout the experiment to observe the impact of the DDoS attack (*Lens | The Kubernetes IDE*, 2022).

3.5.2 Application layer: 5G Core and RAN

This section describes the implementation of 5GC and RAN on the application layer. Open5GS was used to implement the 5G core. Open5GS is an open-source project developed based on C-language following 3GPP standard Release-16; in this experiment, v2.4.8 was used. Open5GS helps researchers to implement 4G and 5G Core networks, which can be integrated to Radio networks to test complete scenarios (*Open5GS - Open source project of 5GC and EPC*, 2019). This experiment did not use the standard version, which runs on the top Linux OS; instead, A modified containerized version was deployed using a helm chart inside in the Kubernetes cluster. This con-

tainerized version has been developed by the National Network of Excellence in 5G and future technologies (Open-VERSO) leveraging Open5GS implementation (*OPEN VERSO | Excellence National Network on 5G technology*, 2020), v1.2.1 was used.

The containerized 5GC was deployed inside the AWS Kubernetes environment described in the previous section. This experiment focused on DDoS attacks generated by compromised 5G-IoT devices and was not intended to cover interworking scenarios between LTE and 5G. Therefore, the 5G-SA core model was deployed, which exclusively serves 5G devices. 5G-SA core excludes LTE EPC elements like MME, Serving Gateway (SGW), PDN gateway (PGW), and HSS, and includes only 5GC elements: AMF, SMF, UPF, PCF, UDM, UDR, AUSF, NRF, BSF, and NSSF. These components were deployed as microservices, each as a dedicated container, which can be managed and scaled individually using Kubernetes scaling capabilities.

UERANSIM was used to implement four gNBs and 1000 UEs on the RAN side to emulate the DDoS attack. UERANSIM is an open-source project developed using C++ language, and it provides essential 3GPP functionalities of 5G RAN and UEs. On the radio interface segment, UERANSIM simulates all the layers above the Radio Resource Control (RRC). In contrast, on the two segments interfacing 5GC (i.e., Control Plane and User Plane), it provides actual functionality and is integrable to operational 5GC networks based on the 3GPP standard. Version 3.2.6 was used, which is based on 3GPP Release-15 (Güngör, 2022).

3.5.3 Mitigation mechanism

In order to mitigate the memory increase on the AMF function, which is triggered by the signaling DDoS attack; the AMF Kubernetes deployment was configured with two parameters: an initial allocated amount of memory

for normal operations (memory request), and a higher limit (memory limit) which can be utilized upon requirement in the case of an attack.

3.5.4 Summary

The significance of signaling DDoS attacks in 5G networks will grow as the 5G-IoT base grows. Kubernetes capabilities were shown to be promising to mitigate these attacks. This chapter described the DKSS solution, which was built using AWS and EKS on the infrastructure layer. While Open5GS and UERANSIM were used on the application layer to simulate 5G Core and RAN. The attack in this experiment targeted to overwhelm the AMF function RAM resources by generating a large number of user registrations. Moreover, the mitigation was done by configuring initial and maximum memory limits on the container level, as elaborated in the next chapter.

Chapter 4

Experimental Evaluation and Discussion

This chapter outlines the experiment's logical flow, then describes the tested scenarios and results, the experimental evaluation and discussion, and finally, the limitations.

4.1 Experiment Flow

DKSS operates as illustrated in Figure 4.1. Initially, the AMF starts in the normal operation state with an initial amount of allocated memory resources. If the attack starts and the memory remains within the initial allocated amount, no action will be taken. However, suppose the attack persists, and the memory utilization exceeds the initial memory allocation. In that case, the AMF container is allowed to grow its memory resources beyond the initial amount and within the maximum limit; this will mitigate (absorb) the attack and allow the AMF to continue operating in a healthy condition. Then, if the attack stops and memory utilization decreases, the additional

memory will be released, and the AMF container will return to the initial normal operations state.

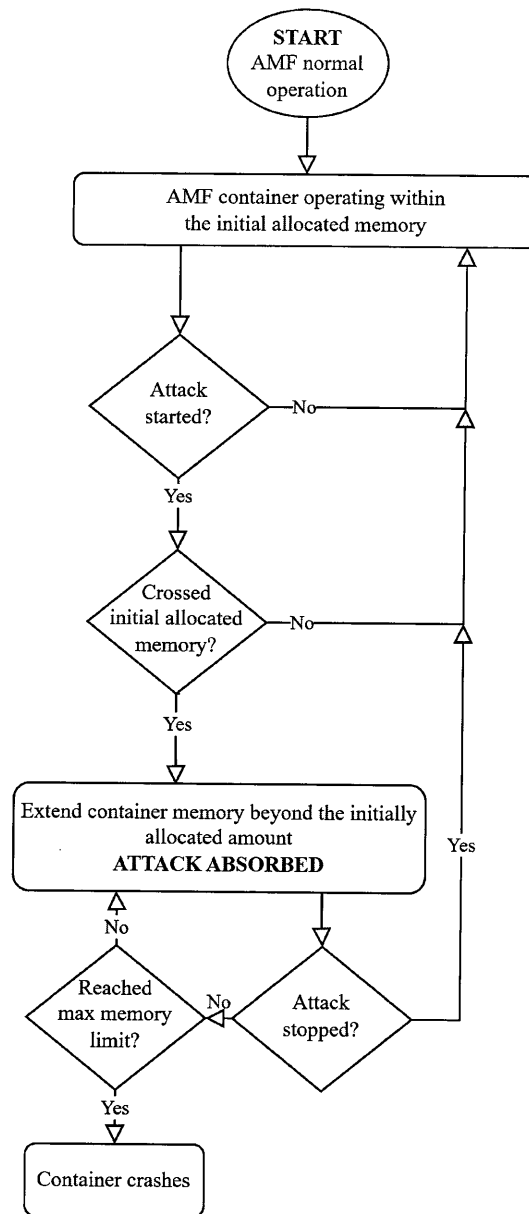


Figure 4.1: DKSS Operation

4.2 Experimental Scenarios and Results

The purpose of this experiment was to test the effectiveness of the proposed mitigation solution. Therefore, two scenarios were tested, 1- Unprotected AMF; to simulate the bare-metal setup where the core nodes do not have the flexibility of resource scaling, and 2- Protected AMF; using DKSS vertical scaling.

4.2.1 Unprotected Scenario

In the first scenario, parameters shown in Table 4.1, A fixed amount of memory (13 MB) was allocated to the AMF container to simulate the bare-metal setup. Initially, the AMF ran with no UEs attached at 5 MB RAM utilization. In the first step of the DDoS attack, 250 simultaneous user registration requests were initiated; this increased the memory utilization from 5 MB to 11.8 MB, which is within the capacity of 13MB. Then another batch of the attack was initiated, consisting of another 250 UE registration requests. As illustrated in Figure 4.2, the AMF container ran out of memory and crashed. In summary, the signaling DDoS attack was able to deplete the 5G core resources and cause damage.

Table 4.1: Experiment summary and parameters for the unprotected scenario

Parameter description	Value
Number of gNB cells	4
Number of 5G UEs	1000
Type of signaling attack	Initial registration
Number of DDoS batches	4 batches (250 UEs each)
AMF static memory allocation	13 MB
Result	AMF crashed after 2 DDoS batches

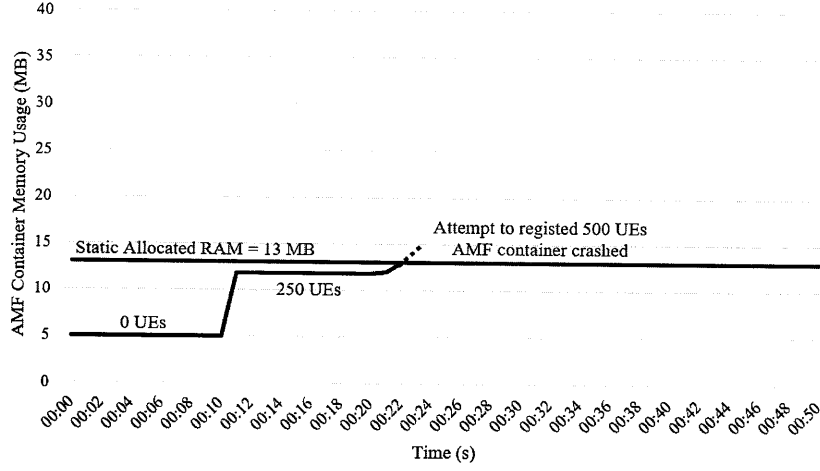


Figure 4.2: AMF memory utilization in the unprotected test scenario.

4.2.2 DKSS Protected Scenario

In the second scenario, parameters shown in Table 4.2, the proposed mitigation mechanism using Kubernetes was implemented, where an initial amount of memory (15 MB) was allocated to the AMF container. However, this amount is a soft limit, as another upper limit was configured for the container (35 MB) to allow the AMF to scale its memory resources upon requirement. The upper limit was set for experiment demonstration purposes only. In real scenarios, the container can ideally scale up to utilize all the available resources in the hosting worker node. Initially, the AMF ran with no UEs attached at 5 MB memory utilization. The DDoS attack was executed in four batches. Each batch was initiated from a different gNB, with registration requests initiated from 250 unique UEs, summing up to 1000 requests. The first batch increased the memory utilization from 5 MB to 11.8 MB, which is within the initially allocated 15 MB. Next, another batch of 250 UEs was triggered and increased the memory utilization to 15.4; the resources of AMF scaled accordingly and continued its operations without any impact. Then two more attack batches were initiated, increasing the utilization to 19.5 MB and 22.3 MB accordingly. Throughout the attack, the

AMF remained operational and healthy, see Figure 4.3. In summary, this test scenario demonstrated the flexibility provided by Kubernetes resource scaling to absorb signaling DDoS attacks.

Table 4.2: Experiment summary and parameters for the protected scenario

Parameter description	Value
Number of gNB cells	4
Number of 5G UEs	1000
Type of signaling attack	Initial registration
Number of DDoS batches	4 batches (250 UEs each)
AMF initial memory allocation	15 MB
AMF memory limit	35 MB
Result	DDoS absorbed (no impact)

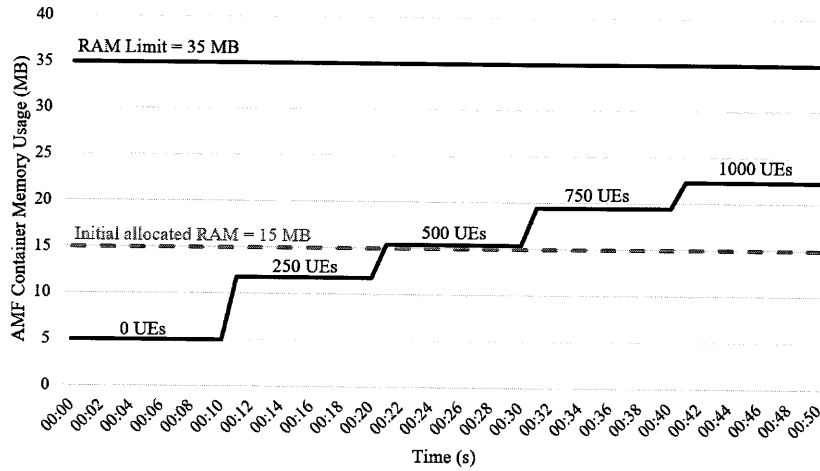


Figure 4.3: AMF memory utilization in the protected test scenario.

4.3 Experimental Evaluation and Discussion

The proposed solution to mitigate signaling DDoS in a 5G stand-alone network was experimented using Open5GS open-source 5G Core deployment

and UERANSIM simulator. The mitigation mechanism exploited the built-in flexibility capabilities within Kubernetes, which is the future of 5G core network infrastructure (Bäckström & Bohra, 2022; *Huawei Kubernetes Case Study*, 2020; Csatari, 2020). Therefore, DKSS is highly applicable to 5GC production environments at no additional complexity or cost because it does not imply changes in the 5GC architecture.

To the author's knowledge, no other work in the literature tested signaling DDoS on 5G core network elements. The closest work that could be used to benchmark the work in this thesis in principle is REPEL (Silva et al., 2021), where the researchers tested a similar DDoS scenario on a 4G core element (MME). REPEL experiment reported an increase in the memory utilization of the MME node as more users were attached to the network. MME is responsible for 4G users' access and mobility control plane. In the 5G architecture, the role of MME has been broken down into two functions, AMF and SMF. Throughout the experiment, a similar impact of the signaling attacks in the 5G network was observed, where the increase in the number of UE registrations was directly proportional to the memory utilization of the AMF function.

The observed memory increase after each batch of the DDoS, i.e., 250 registration requests, was 4.3 MB on average. The experiment was limited to 1,000 connected users due to a limitation in the simulator. However, if the memory increase is forecasted based on the average increase, the AMF would reach its maximum memory limit of 35 MB allocated in the experiment after 1,750 registration requests. Furthermore, attaching 5,500 devices would utilize around 100 MB. Nonetheless, these limits were set for experimental purposes and would depend on the available capacity in the underlying infrastructure in real scenarios, which can vary in each deployment. In addition, the recorded amount of memory increase per user is based on Open5GS Open-VERSO deployment, which can differ if tested using other

5G Core softwares.

Another comparison aspect is the type of scaling, REPEL used horizontal scaling by spinning up additional MME VM depending on the memory load, where the load gets gradually distributed among MME instances using the MME pool feature. However, the amount of RAM allocated to each VM is static and cannot be changed without restarting the VM. In contrast, DKSS is based on two levels of scaling. The first scaling level is the Kubernetes container's vertical scaling. The AMF container is allocated an initial amount of memory and allowed to scale above this amount up to a preset maximum limit or up to the available memory resources in the hosting worker node. The second scaling level is horizontal scaling, achieved by utilizing the AWS autoscaling group, where the Kubernetes cluster can scale out the EC2 worker nodes depending on their load by spinning up additional worker nodes and dynamically redistributing the containers among the nodes.

Furthermore, although REPEL approach is effective, it has a time lag between attack detection and prevention; this time is required to power up a new MME VM, bring it into service, and gradually handle part of the new attach requests initiated by UEs. If the DDoS attack rate is faster than the mitigation process, it may cause damage before the mitigation is fully implemented. On the other hand, DKSS has no time lag as the additional shared memory resources can be immediately utilized.

DKSS and REPEL both aim to defend the core networks by temporarily absorbing the attack rather than blocking it. This methodology maintains the network service-level agreement (SLA), frustrates the attacker, and allows more time for the cybersecurity engineers to analyze the attack and eventually block it.

From the attack detection and resource orchestration perspective, REPEL created an engine based on game theory to frustrate the attacker while

optimizing infrastructure resources. In contrast, in DKSS, the detection engine was out of scope, as it only focused on the mitigation part and used static thresholds to prove the concept. Refer to the comparison summary in table 4.3.

Table 4.3: Comparison between DKSS and REPEL

	REPEL	DKSS
DDoS attack model	Flood of 4G UEs attach requests	Flood of 5G UEs registrations
Attack target	Increase MME memory utilization	Increase AMF memory utilization
Defense strategy	Scale out to absorb DDoS and frustrate the attacker	Scale up/out to absorb DDoS and frustrate the attacker
Scaling type	Horizontal scaling at VM level	Vertical scaling at Kubernetes container level & Horizontal scaling at Kubernetes worker node level
Time to enable mitigation	Requires time to spin up a new MME VM, bring it into service, and gradually handle the load	Immediate
Attack detection and orchestration	Game theory	Out of scope
Common limitation	Bound by the available free infrastructure resources	Bound by the available free infrastructure resources
Test platform	OpenAirInterface prototype	Open5GS over AWS EKS

4.4 Limitations

Although vertical scaling of container resources and horizontal scaling of the Kubernetes worker nodes were achieved, an additional layer of horizontal scaling at the container level was not feasible because Open5GS does not yet support the AMF set (group) feature. AMF set horizontal scaling would have been an efficient mitigation capability, especially in the Kubernetes microservices ecosystem, to distribute the load among multiple AMF containers and avoid indefinitely increasing the load on a single AMF.

Furthermore, one of the current limitations of Kubernetes is that it cannot dynamically change the resources allocated to a container during runtime. If changed, Kubernetes will terminate and redeploy the container, which can interrupt service. For this reason, predefined memory initial allocation and maximum limit were used to avoid container redeployment. This limitation was highlighted by (Tran, Vu, & Kim, 2022) and recommended as future work to enable dynamic vertical scaling in Kubernetes during runtime.

Moreover, although DKSS is scalable, it is bound by the free shared resources in the underlying infrastructure. Otherwise, if the attack is persistent and has a large enough magnitude, it can theoretically deplete the whole infrastructure resources and cause damage if not stopped from the origin by other blocking mechanisms.

Chapter 5

Conclusion and Future Work

This chapter summarizes the thesis and proposes recommendations and future work.

5.1 Summary

In this thesis, a known cybersecurity threat was addressed: the signaling DDoS attacks targeting state-of-the-art 5G stand-alone core networks triggered by a massive number of 5G-IoT devices. DKSS was introduced, which is a novel solution that utilizes the infrastructure flexibility provided by Kubernetes and AWS scaling in a 5G microservices architecture. The solution was demonstrated to be successful in mitigating the signaling DDoS targeting 5G AMF function with no time overhead between attack start and mitigation. Moreover, the solution is highly applicable to real-world networks, as it utilizes Kubernetes built-in features and works within the boundaries of 5G 3GPP and microservices standard architectures.

5.2 Recommendations and Future Work

The proposed solution focused on the attack mitigation technique. It is recommended to integrate it with Machine-Learning (ML) based attack detection engines to improve resource allocation efficiency and ensure optimum infrastructure resource utilization.

Furthermore, supposing Open5GS introduces the AMF set (group) functionality in the future, this solution can be further enhanced by introducing a third level of scalability, i.e., Horizontal AMF scaling on the container level. The user's distribution among AMF instances can be dynamically shuffled using the AMF Load Re-Balancing feature defined in the 5G 3GPP standard. This feature allows an AMF to instruct gNBs to shift a group of users to another AMF within the AMF set with minimal service impact and without changing the static AMF balancing parameters configured on the radio side (*3GPP-R17 5G Architecture*, 2022).

Moreover, this research focused on the impact of signaling DDoS attacks on memory resources because AMF function memory is directly affected by the increased number of open sessions, as the AMF maintains each connected subscriber's connection, profile, and status. Other papers in the literature similarly focused on memory resources (Silva et al., 2021). Nonetheless, the comprehensive impact of DDoS on other resources like CPU, storage, and network utilization can be further studied.

References

- 3GPP-R17 5G Architecture*. (2022, June). 3GPP. Retrieved from <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144>
- 3GPP-R17 5G Procedures*. (2022, June). 3GPP. Retrieved 2022-09-02, from <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3145>
- Ahmad, I., Shahabuddin, S., Kumar, T., Okwuibe, J., Gurtov, A., & Ylianttila, M. (2019). Security for 5g and beyond. *IEEE Communications Surveys Tutorials*, 21(4), 3682-3722. doi: 10.1109/COMST.2019.2916180
- Amazon EKS - Managed Kubernetes Service*. (2022). Retrieved 2022-08-27, from <https://aws.amazon.com/eks/>
- Arouk, O., & Nikaein, N. (2020). 5g cloud-native: Network management automation. In *Noms 2020 - 2020 ieee/ifip network operations and management symposium* (p. 1-2). doi: 10.1109/NOMS47738.2020.9110392
- Bassil, R., Chehab, A., Elhajj, I., & Kayssi, A. (2012, October). Signaling oriented denial of service on LTE networks. In *Proceedings of the 10th ACM international symposium on Mobility management and wireless access* (pp. 153–158). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/2386995.2387024

- Bentaleb, O., Belloum, A. S. Z., Sebaa, A., & El-Maouhab, A. (2021, June). Containerization technologies: taxonomies, applications and challenges. *The Journal of Supercomputing*, 78(1), 1144–1181. doi: 10.1007/s11227-021-03914-1
- Bäckström, H., & Bohra, R. (2022, May). *Ericsson: Why Kubernetes over bare metal infrastructure for 5G*. Retrieved 2022-08-27, from <https://www.ericsson.com/en/blog/2022/5/kubernetes-over-bare-metal-cloud-infrastructure-why-its-important-and-what-you-need-to-know>
- Csatari, G. (2020). *Nokia Kubernetes Case Study*. Retrieved 2022-08-27, from <https://kubernetes.io/case-studies/nokia/>
- De Souza Neto, F. J., Amatucci, E., Nassif, N. A., & Marques Farias, P. A. (2021). Analysis for comparison of framework for 5g core implementation. In *2021 international conference on information science and communications technologies (icisct)* (p. 1-5). doi: 10.1109/ICISCT52966.2021.9670414
- Ericsson: Building a new world - A guide to evolving from EPC to 5G Core*. (2020, September). Ericsson. Retrieved 2022-09-14, from <https://www.ericsson.com/en/core-network/guide> (Last Modified: 2022-08-03T01:01:43+00:00)
- Ericsson Mobility Report- June 2022* (Tech. Rep.). (2022, June). Retrieved 2022-09-16, from <https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/june-2022> (Last Modified: 2022-07-12T01:03:35+00:00)
- Ernsberger, D., George, K. J., & Arumugam, S. (2019, January). Security Study and Monitoring of LTE Networks. *Journal of ICT Standardization*, 7(1), 43–60. (Publisher: River Publishers) doi: 10.13052/jicts2245-800X.713

- ETSI - Standards for NFV - Network Functions Virtualisation | NFV Solutions.* (2019). Retrieved 2021-12-19, from <https://www.etsi.org/technologies/nfv>
- Forland, M. K., Krlevska, K., Garau, M., & Gligoroski, D. (2019). Preventing ddos with sdn in 5g. In *2019 ieee globecom workshops (gc wkshps)* (p. 1-7). doi: 10.1109/GCWkshps45667.2019.9024497
- Gebremariam, Y. E., Duguma, D. G., Park, H. Y., Kim, Y. N., Kim, B., & You, I. (2021). 5g and beyond telco cloud: architecture and cybersecurity challenges. In *2021 world automation congress (wac)* (p. 1-6). doi: 10.23919/WAC50355.2021.9559450
- Ghorbani, H., Mohammadzadeh, M. S., & Ahmadzadegan, M. H. (2020). Ddos attacks on the iot network with the emergence of 5g. In *2020 international conference on technology and entrepreneurship - virtual (icte-v)* (p. 1-5). doi: 10.1109/ICTE-V50708.2020.9113779
- Gurusamy, D., Deva Priya, M., Yibgeta, B., & Bekalu, A. (2019). Ddos risk in 5g enabled iot and solutions. *Int. J. Eng. Adv. Technol.(IJEAT)*, 8(5), 1574–1578.
- Güngör, A. (2022, August). *UERANSIM Open source 5G UE and RAN (gNodeB) implementation*. Retrieved 2022-08-27, from <https://github.com/aligungr/UERANSIM>
- Hercog, D. (2020). SCTP Protocol. In *Communication Protocols: Principles, Methods and Specifications* (pp. 341–342). Cham: Springer International Publishing. doi: 10.1007/978-3-030-50405-2_3
- Huawei Kubernetes Case Study.* (2020). Retrieved 2022-09-03, from <https://kubernetes.io/case-studies/huawei/>
- Huawei: Next-Generation Telco Cloud Based on Cloud Native.* (2020). Retrieved 2022-08-27, from <https://carrier.huawei.com/en/industry-perspective/5g-core-network/next-generation-telco-cloud-based-on->

cloud-native

- Köksal, S., Dalveren, Y., Maiga, B., & Kara, A. (2021). Distributed denial-of-service attack mitigation in network functions virtualization-based 5g networks using management and orchestration. *International Journal of Communication Systems*, 34(9), e4825. doi: <https://doi.org/10.1002/dac.4825>
- Lens | The Kubernetes IDE.* (2022). Retrieved 2022-08-28, from <https://k8slens.dev/>
- Luong, D.-H., Thieu, H.-T., Outtagarts, A., & Ghamri-Doudane, Y. (2018). Cloudification and autoscaling orchestration for container-based mobile networks toward 5g: Experimentation, challenges and perspectives. In *2018 ieee 87th vehicular technology conference (vtc spring)* (p. 1-7). doi: 10.1109/VTCSpring.2018.8417602
- Matefi, G., & Stief, M. (2022, July). Ericsson white paper - Achieving 5G network security with new strategies. Ericsson. Retrieved 2022-08-26, from <https://www.ericsson.com/en/reports-and-papers/white-papers/signaling-security2> (Last Modified: 2022-07-07T01:01:50+00:00)
- Mfula, H., Ylä-Jääski, A., & Nurminen, J. K. (2021). Seamless kubernetes cluster management in multi-cloud and edge 5g applications. In *International conference on high performance computing & simulation (hpcs 2020)*.
- Open5GS - Open source project of 5GC and EPC.* (2019). Retrieved 2022-08-27, from <https://open5gs.org/>
- OPEN VERSO | Excellence National Network on 5G technology.* (2020). Retrieved 2022-08-28, from <https://www.openverso.org/en/>
- Sahu, N., & Thakre, S. C. (EasyChair, 2022). *A survey on kubernetes architecture and its significance*. EasyChair Preprint no. 7391.

- Sattar, D., & Matrawy, A. (2019). Towards secure slicing: Using slice isolation to mitigate ddos attacks on 5g core network slices. In *2019 ieee conference on communications and network security (cns)* (p. 82-90). doi: 10.1109/CNS.2019.8802852
- Serrano Mamolar, A., Salvá-García, P., Chirivella-Perez, E., Pervez, Z., Alcaraz Calero, J. M., & Wang, Q. (2019). Autonomic protection of multi-tenant 5g mobile networks against udp flooding ddos attacks. *Journal of Network and Computer Applications*, 145, 102416. doi: <https://doi.org/10.1016/j.jnca.2019.102416>
- Shah, J., & Dubaria, D. (2019). Building modern clouds: Using docker, kubernetes google cloud platform. In *2019 ieee 9th annual computing and communication workshop and conference (ccwc)* (p. 0184-0189). doi: 10.1109/CCWC.2019.8666479
- Silva, R. S., Meixner, C. C., Guimarães, R. S., Diallo, T., Garcia, B. O., de Moraes, L. F. M., & Martinello, M. (2021). Repel: A strategic approach for defending 5g control plane from ddos signalling attacks. *IEEE Transactions on Network and Service Management*, 18(3), 3231-3243. doi: 10.1109/TNSM.2020.3035342
- Sotelo Monge, M. A., Lorenzo Fernández, B., Maestre Vidal, D., Rius García, G., Herranz González, A., & Maestre Vidal, J. (2018). Source-side ddos detection on iot-enabled 5g environments. In *2018 international workshop on secure internet of things (siot)* (p. 28-37). doi: 10.1109/SIoT.2018.00010
- Tran, M.-N., Vu, D.-D., & Kim, Y. (2022). A survey of autoscaling in kubernetes. In *2022 thirteenth international conference on ubiquitous and future networks (icufn)* (p. 263-265). doi: 10.1109/ICUFN55119.2022.9829572
- Wazid, M., Das, A. K., Shetty, S., Gope, P., & Rodrigues, J. J. P. C. (2021). Security in 5g-enabled internet of things communication: Is-

sues, challenges, and future research roadmap. *IEEE Access*, 9, 4466-4489. doi: 10.1109/ACCESS.2020.3047895

ملخص الدراسة

تطورت تقنيات الاتصالات اللاسلكية بشكل كبير في العقد الماضي ، وكان من آخر تطوراتها إدخال أنظمة الاتصالات المتنقلة من الجيل الخامس (5G). تقدم شبكات الجيل الخامس معدلات بيانات أعلى في زمن انتقال منخفض للغاية مع حركة عالية وتغطية أوسع. تمنح هذه الخصائص إمكانات هائلة لمستخدمي النطاق العريض للأجهزة المحمولة وفي أحوال استخدام إنترنت الأشياء (IoT).

في الشبكة المركزية، يتم تمكين 5G من خلال الشبكات المعرفة بالبرمجيات (SDN) ، والمحاكاة الافتراضية لوظيفة الشبكة (NFV) ، والهندسة القائمة على الخدمة السحابية الأصلية (SBA). في الحين الذي يُتوقع أن تتصل مليارات أجهزة إنترنت الأشياء بشبكات الجيل الخامس ، فإن أجهزة إنترنت الأشياء أصبحت معرضة لثغرات أمنية متنوعة، مما يجعلها منصة جاذبة للمهاجمين، وذلك باستغلالها في إطلاق هجمات حجب الخدمة الموزعة (DDoS) من أجهزة 5G-IoT المخترقة. إذ ولن تصمد آليات حماية DDoS التقليدية في مقاومة هجمات حجب الخدمة في شبكات الجيل الخامس نظرًا للعدد الهائل من الأجهزة والطبيعة الديناميكية لشبكات الجيل الخامس. تعد هجمات إرسال الإشارات الشبكية من عدد هائل من أجهزة 5G-IoT أحد أنواع هجمات حجب الخدمة في مجال الجيل الخامس؛ حيث تستهدف زيادة الحمل على مكونات شبكة الجيل الخامس المركزية.

في هذه الأطروحة ، يُقدم حل التخفيف من هجمات حجب الخدمة الموزعة عن طريق تقنية Kubernetes لمرونة الموارد (DKSS)، وهو حل جديد يساعد على التخفيف من هجمات حجب الخدمة في شبكات الجيل الخامس. تم إثبات فاعلية الحل في التخفيف من هجمات حجب الخدمة التي تستهدف مكونات شبكة الجيل الخامس المركزية بشكل فوري عند بدء الهجمة؛ اعتمادًا على مرونة البنية التحتية التي توفرها Kubernetes وخدمات أمازون السحابية (AWS) في بنية الخدمات المصغرة لشبكات الجيل الخامس. ومثل هذه الطريقة تقبل التنفيذ على الشبكات الحقيقية، حيث تستخدم ميزات Kubernetes المضمنة وتعمل وفق معايير مشروع شراكة الجيل الثالث (3GPP) والبنية القياسية للخدمات المصغرة.



استمارة إيداع البحث

اسم الباحث:	عبدالرحمن خالد عبدالرحمن خالد عبدالوهاب
عنوان البحث:	الحماية من هجمات حجب الخدمة الموزعة في شبكات الجيل الخامس الافتراضية
سنة النشر:	2022
الكلية:	كلية تقنية المعلومات
القسم:	هندسة نظم المعلومات
ذالبريد الإلكتروني:	202000050@stu.uob.edu.bh
رقم الاتصال:	33337077

إقرار الباحث بالموافقة على الشروط التالية:

يحق لجامعة البحرين إتاحة وتوفير الرسالة أو الأطروحة المذكورة أعلاه في مكتبة الجامعة أو من خلال موقع الإسهامات البحثية للجامعة أو على نطاق أوسع في الخارج؛ كما ويمكن إعادة طباعة الرسالة إذا اقتضت الحاجة لذلك أو رقميتها مع إتاحتها على شبكة الأنترنت وذلك حسب تقدير إدارة الجامعة.

تنطبق هذه الشروط على جميع النسخ بما في ذلك النسخ الإلكترونية

تعتبر هذه الرسالة أو الأطروحة من المواد المحمية والتي تخضع إلى حقوق النشر والتأليف، ولا يجوز نشر أي اقتباس من الرسالة بدون توثيق المصدر.

التاريخ: 29 سبتمبر 2022

توقيع الباحث:

جامعة البحرين
كلية تقنية المعلومات



الحماية من هجمات حجب الخدمة الموزعة في شبكات الجيل الخامس الافتراضية

أطروحة مقدمة كجزء من متطلبات الحصول على
درجة الماجستير في الأمن الإلكتروني

إعداد

عبدالرحمن خالد عبدالرحمن

202000050

إشراف

د. ريهام المساعيد

أستاذ مساعد

جامعة البحرين

مملكة البحرين

ديسمبر 2022