Contents

[DEDICATION 3](#_Toc134714101)

[ACKNOWLEDGEMENTS 4](#_Toc134714102)

[ABSTRACT 5](#_Toc134714103)

[LIST OF TABLES 6](#_Toc134714104)

[LIST OF FIGURES 7](#_Toc134714105)

[LIST OF ABBREVIATIONS 8](#_Toc134714106)

[Chapter 1 INTRODUCTION 9](#_Toc134714107)

[1.1 Background of the Study 9](#_Toc134714108)

[1.2 Problem Statement 10](#_Toc134714109)

[1.3 Aim and Objectives 10](#_Toc134714110)

[1.4 Research Questions 10](#_Toc134714111)

[1.5 Scope of the Study 10](#_Toc134714112)

[1.6 Significance of the Study 10](#_Toc134714113)

[1.7 Structure of the Study 10](#_Toc134714114)

[Chapter 2 LITERATURE REVIEW 10](#_Toc134714115)

[2.1 Introduction 10](#_Toc134714116)

[2.2 Data Analytics in Healthcare 10](#_Toc134714117)

[2.3 Data Mining in Diagnosis of Chronic Diseases 10](#_Toc134714118)

[2.4 Data Mining in Prediction of Common Types of Cancer 10](#_Toc134714119)

[2.5 Predictive Modeling in Breast Cancer Diagnosis on Different Datasets 10](#_Toc134714120)

[2.5.1 WBC Datasets 10](#_Toc134714121)

[2.5.2 SEER Breast Cancer Datasets 10](#_Toc134714122)

[2.5.3 Other Breast Cancer Datasets 10](#_Toc134714123)

[2.6 Visual Analytics in Healthcare 10](#_Toc134714124)

[2.7 Related Research Publications 10](#_Toc134714125)

[2.8 Discussion 10](#_Toc134714126)

[2.9 Summary 10](#_Toc134714127)

[Chapter 3 RESEARCH METHODOLOGY 10](#_Toc134714128)

[3.1 Introduction 10](#_Toc134714129)

[3.2 Research Methodology 10](#_Toc134714130)

[3.2.1 Data Selection 10](#_Toc134714131)

[3.2.2 Data Pre-processing 10](#_Toc134714132)

[3.2.3 Data Transformation 10](#_Toc134714133)

[3.2.4 Interactive Visual Analytics 10](#_Toc134714134)

[3.2.5 Class Balancing 10](#_Toc134714135)

[3.2.6 Data Mining 10](#_Toc134714136)

[3.2.7 Interpretation/Evaluation 10](#_Toc134714137)

[3.3 Proposed Method (Classification) 10](#_Toc134714138)

[Chapter 4 ANALYSIS 10](#_Toc134714139)

[4.1 Introduction 10](#_Toc134714140)

[4.2 Dataset Description 10](#_Toc134714141)

[4.3 Data Preparation 10](#_Toc134714142)

[4.3.1 Elimination of Variables 11](#_Toc134714143)

[4.3.2 Transformation into Categorical Variables 11](#_Toc134714144)

[4.3.3 Identification of missing values 11](#_Toc134714145)

[4.3.4 Univariate analysis 11](#_Toc134714146)

[4.3.5 Treatment of missing values 11](#_Toc134714147)

[4.3.6 Splitting of original dataset 11](#_Toc134714148)

[4.4 Exploratory Data Analysis (Bivariate analysis) 11](#_Toc134714149)

[4.4.1 Chi-square test 11](#_Toc134714150)

[4.5 Data Visualization 11](#_Toc134714151)

[4.6 Summary 11](#_Toc134714152)

[Chapter 5 RESULTS AND DISCUSSIONS 11](#_Toc134714153)

[5.1 Introduction 11](#_Toc134714154)

[5.2 Interpretation of Visualizations 11](#_Toc134714155)

[5.3 Evaluation of Sampling Methods and Results 11](#_Toc134714156)

[5.4 Testing on Validation Dataset 11](#_Toc134714157)

[5.5 Summary 11](#_Toc134714158)

[Chapter 6 CONCLUSIONS AND RECOMMENDATIONS 11](#_Toc134714159)

[6.1 Introduction 11](#_Toc134714160)

[6.2 Discussion and Conclusion 11](#_Toc134714161)

[6.3 Contribution to knowledge 11](#_Toc134714162)

[6.4 Future Recommendations 11](#_Toc134714163)

[References 11](#_Toc134714164)

[APPENDIX A: RESEARCH PROPOSAL 11](#_Toc134714165)

[APPENDIX B: ETHICS FORMS 11](#_Toc134714166)

# DEDICATION

To my loving wife and my steadfast parents,

This thesis is dedicated to you, with heartfelt gratitude and appreciation for your unwavering support, encouragement, and love throughout my academic journey. Your sacrifices, guidance, and belief in my abilities have been the pillars upon which I have built my dreams.

To my wife, you have been my rock, my inspiration, and my sanctuary. Your presence has brought joy and light into the darkest corners of this journey, and your unwavering faith in me has been a constant source of strength. Thank you for your patience, understanding, and companionship.

To my parents, your love, wisdom, and unwavering belief in the value of education have been the foundation of my aspirations. Your sacrifices and commitment to my success have not gone unnoticed, and I am forever grateful for the opportunities you have given me. I am who I am today because of your unwavering support and love.

With immense gratitude, I dedicate this thesis to you, my loving wife and my dear parents, for without you, none of this would have been possible.

# ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to everyone who has supported and guided me throughout the journey of completing this thesis.

First and foremost, I am profoundly grateful to my advisor, Dr. Prasanth for their invaluable guidance, constructive criticism, and unwavering support. Your mentorship and dedication have been instrumental in shaping my research and academic growth, and I am honored to have had the opportunity to learn from you.

# ABSTRACT

P4 programming language provides a way to configure the data plane as per the requirements. P4 programmable equipment can be configured to offload specific workloads from the application. As telecom operators embrace cloud infrastructure, the VNFs are usually hosted over COTS. One VNF can be UPF in 5G Core architecture as defined in the 3GPP standards. Our suggestion is to offload DNS packets from UPF. In this paper, we will explore the benefits and drawbacks of this approach to conclude whether it would be feasible to offload the DNS workload from UPF.

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

|  |  |
| --- | --- |
| 3GPP | 3rd Generation Partnership Project |
| 5G | 5th Generation |
| AMF | Access and Mobility Management Function |
| AVP | Attribute value pair |
| COTS | Commercial of the shelf |
| DNS | Domain name server |
| GTPU | GPRS Tunnelling Protocol |
| IP | Internet protocol |
| ISP | Internet service provider |
| NR | New Radio |
| P4 | Programming Protocol-independent Packet Processors |
| RAN | Radio Access network |
| SMF | Session Management Function |
| UE | User equipment |
| UPF | User plane function |
| URL | Uniform Resource Locator |
| VM | Virtual machine |
| VNF | virtual network function |
|  |  |

# INTRODUCTION

## Background of the Study

As the 3GPP standards have introduced a service-based architecture for 5G core, the industry is moving towards microservices over the cloud. These telecom workloads are handled mainly by COTS in a cloud environment. Since the services are running on general-purpose hardware, the performance might not be on par with the application-specific equipment. UPF can also be hosted over the cloud as VNF. As per 3GPP standards, the purpose of UPF is to handle user-plane traffic in the 5G Core mobility domain (Alain Sultan, 2022). It serves as an anchor that connects the mobility network domain with an external network. It can also serve as a policy enforcement function and enforce QoS on traffic as per the requirements (Alain Sultan, 2022). For example, if a subscriber has no quota left, the UPF would start dropping packets, including DNS packets. In this case, the UE will not be able to access the ISP's internal services that are meant to be free, such as renewal of the subscription to avail of the ISP's internet and calling services. For this purpose, DNS queries are made free of charge so that the UE can get the IP addresses of the intended free service by the ISP. To implement this, UPF would have to bypass the charging trigger or tag the traffic with the appropriate service-identified AVP, which is charged as free by the charging system for DNS queries, along with the free service’s URL, IP address, and port.

From the discussion above, marking DNS traffic as free by UPF can provide extra features and revenue chances for ISPs. Since all the DNS traffic would have to bypass the UPF, we can route DNS traffic toward the internet DNS instead of UPF. This approach can offload UPF from processing all these packets and result in efficient use of resources.

## Problem Statement

If an ISP decides not to charge DNS packets, then the UPFs processing these packets and marking them as free would be an inefficient use of resources. A P4 based traffic offloading can result in efficient use of UPF resources. The performance gains of P4 based switches as compared to other software based packet acceleration technologies has been determined in previous studies and it was observed that P4 achieved the lowest latency for small packets (Rischke *et al.*, 2022).

In previous studies, complete UPF functionality has been implemented in a P4 programmable switch (Paolucci, Scano, *et al.*, 2021). However, this might not be a feasible approach as vendors would then need to implement other UPF functionalities over the P4 switch as well. This is because key UPF features are missing such as a mechanism to trigger per UE usage reports to the SMF so that SMF can communicate with Charging Functions. Similarly, a mechanism to implement policy enforcement received by SMF from PCF is also not provided.

Considering the benefits of P4 switches P4 workload offloading from UPF has already been performed but for service function chaining which is different that routing traffic based on packet encapsulated under the GTP-U tunnel (Paolucci, Scano, *et al.*, 2021). This can be a useful approach if service chaining functionality is implemented by the operator.

So, from the above discussion, P4 seems to be a good approach however implementing complete UPF functionality of the programmable data-plane or steering traffic via service chaining might not be feasible in all the scenarios. Considering the lack of implemented features or the mandatory requirement of SFC implementation that could lead to major network design changes at the service provider’s end. Our suggestion is to bypass only the DNS packets encapsulated under the GTP tunnel and assess the feasibility of this approach.

## Aim and Objectives

The main aim of this research is to conclude whether offloading DNS traffic from UPF would result in performance gain and if it would be feasible considering the drawbacks of this approach.

The research objectives are formulated based on the aim of this study which is as follows:

* To explore concerns of bypassing DNS traffic from UPF
* To suggest mitigation techniques if DNS traffic is bypassed
* To estimate the performance gains of this approach

## Research Questions

What would be the performance gain offloading DNS packets, and what are the significant concerns in bypassing the UPF for these packets?

## Scope of the Study

The scope of this study is limited to DNS traffic processing at the P4 programmable data plane switch and its impact on the 5G Network. For performance evaluation, two devices’ packet capture will be analyzed to estimate DNS traffic by a UE; the real-world traffic can vary depending on the usage, which the data collection methodology might not depict in this research work.

## Significance of the Study

Although UPF functionality can be offloaded completely to data plane switch, the drawbacks are significant because of the lack of key features (Paolucci, Scano, *et al.*, 2021). We are suggesting an approach that does not require any major design changes and the P4 code can simply be implemented over the transit switch or SmarNIC on the compute node hosting the UPF. In this approach, the key PF functionalities are retained with the COTS based UPF instance on VM or Container. We are suggesting to specifically offload DNS packets because DNS queries are expected to be large in number as the mobile subscriber base is huge. The total number of subscribers connected to Reliance JIO ISP in March 2022 was 403.99 million (Statista and TRAI, 2022). If all the DNS packets can bypass UPF without any disadvantages, this could save a lot of UPF resources for other important workloads.

## Structure of the Study

TBD

# LITERATURE REVIEW

## Introduction

In order to understand the need of DNS offloading, it is important to understand the 5G core network architecture. This can provide a better understanding of how packet flows from access to core network. Furthermore, we will also go through P4 based UPF implementation and explore the enhancement opportunities and drawbacks (Paolucci, Scano, *et al.*, 2021). We will go through the SFC based traffic steering via P4 switch, this can provide implementation insights on how to build a design that would be simpler and easier for the service providers to adapt (Paolucci, Cugini, *et al.*, 2021). In order to avoid missing any functionality provided by the UPF we will review the 3GPP

## Data Analytics in Healthcare

## Data Mining in Diagnosis of Chronic Diseases

## Data Mining in Prediction of Common Types of Cancer

## Predictive Modeling in Breast Cancer Diagnosis on Different Datasets

### WBC Datasets

### SEER Breast Cancer Datasets

### Other Breast Cancer Datasets

## Visual Analytics in Healthcare

## Related Research Publications

## Discussion

## Summary

# RESEARCH METHODOLOGY

## Introduction

## Research Methodology

### Data Selection

### Data Pre-processing

### Data Transformation

### Interactive Visual Analytics

### Class Balancing

### Data Mining

### Interpretation/Evaluation

## Proposed Method (Classification)

# ANALYSIS

## Introduction

## Dataset Description

## Data Preparation

### Elimination of Variables

### Transformation into Categorical Variables

### Identification of missing values

### Univariate analysis

### Treatment of missing values

### Splitting of original dataset

## Exploratory Data Analysis (Bivariate analysis)

### Chi-square test

## Data Visualization

## Summary

# RESULTS AND DISCUSSIONS

## Introduction

## Interpretation of Visualizations

## Evaluation of Sampling Methods and Results

## Testing on Validation Dataset

## Summary

# CONCLUSIONS AND RECOMMENDATIONS

## Introduction

## Discussion and Conclusion

## Contribution to knowledge

## Future Recommendations

# References

Alain Sultan, M. (2022) *5G System Overview*, *3gpp*. Available at: https://www.3gpp.org/technologies/5g-system-overview (Accessed: 13 February 2023).

Paolucci, F., Cugini, F., *et al.* (2021) ‘Enhancing 5G SDN/NFV Edge with P4 Data Plane Programmability’, *IEEE Network*, 35(3), pp. 154–160. Available at: https://doi.org/10.1109/MNET.021.1900599.

Paolucci, F., Scano, D., *et al.* (2021) ‘User Plane Function Offloading in P4 switches for enhanced 5G Mobile Edge Computing’, in *2021 17th International Conference on the Design of Reliable Communication Networks (DRCN)*. IEEE, pp. 1–3. Available at: https://doi.org/10.1109/DRCN51631.2021.9477338.

Rischke, J. *et al.* (2022) ‘Comparison of UPF acceleration technologies and their tail-latency for URLLC’, in *2022 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, pp. 19–25. Available at: https://doi.org/10.1109/NFV-SDN56302.2022.9974798.

Statista and TRAI (2022) *Number of wireless telecommunication subscribers in India as of March 2022, by service provider (in millions)*. Available at: https://www.statista.com/statistics/258794/mobile-telecom-subscribers-in-india-by-company/?locale=en (Accessed: 13 February 2023).

# APPENDIX A: RESEARCH PROPOSAL

# APPENDIX B: ETHICS FORMS