Analysis of 5G Core Network Traffic Patterns and Protocol Distribution

Harshit Agrawal
202251053
Computer Science

Malaika Varshney
202251069
Computer Science

Ameya Naik
202251012
Computer Science

***Abstract*—This paper presents a comprehensive analysis of traffic patterns in 5G core networks using the Western-OC2-Lab dataset. We analyze protocol distribution, packet length characteristics, and network topology to understand communication behaviors between different network functions. Our findings reveal that UDP dominates protocol usage (accounting for approximately 35,000 packets), with packet lengths primarily clustered between 60-200 bytes (mean: 180.03 bytes). Network topology visualization identifies two primary communication hubs (NWDAF and NRF) that orchestrate most traffic flows. The insights gained from this analysis can inform traffic forecasting models and network resource allocation strategies, ultimately contributing to improved network performance and quality of service in 5G deployments.**

***Index Terms*—5G networks, traffic analysis, protocol distribution, packet characteristics, network topology, traffic forecasting, core network**

I. Introduction

The rapid advancement of 5G technology has introduced a paradigm shift in wireless network capabilities, promising enhanced mobile broadband, ultra-reliable low-latency communications, and massive machine-type communications. These advancements necessitate a sophisticated understanding of network traffic patterns to ensure optimal resource allocation, quality of service, and network planning. As 5G networks continue to expand globally, effective traffic forecasting becomes increasingly critical for network operators and service providers.

5G core networks represent a significant architectural evolution from previous generations, adopting a service-based architecture with specialized network functions. This architectural shift, along with the diverse application scenarios supported by 5G, creates unique traffic patterns that require detailed analysis and modeling. Understanding these patterns is essential for developing accurate traffic forecasting models that can predict network behavior and enable proactive resource management.

Traffic forecasting in 5G networks faces several challenges, including the heterogeneity of services, varying quality of service requirements, and the dynamic nature of user behavior. Moreover, the increased complexity of 5G network architecture introduces new protocol interactions and communication patterns that must be considered in traffic analysis. By accurately forecasting network traffic, operators can optimize resource allocation, prevent congestion, and enhance overall network performance.

This paper focuses on analyzing traffic patterns in a 5G core network using the Western-OC2-Lab dataset. Our analysis encompasses several key aspects of network traffic, including protocol distribution, packet length characteristics, and communication patterns between different network functions. The insights gained from this analysis provide a foundation for developing effective traffic forecasting models tailored to the unique characteristics of 5G networks.

II. Related Work

Traffic analysis and forecasting in wireless networks has been an active area of research, with numerous studies focusing on different aspects of traffic characterization and prediction. As networks evolve toward 5G and beyond, the complexity and diversity of traffic patterns have prompted more sophisticated approaches to traffic modeling and forecasting.

Gao et al. [1] proposed a smoothed long short-term memory (SLSTM) traffic prediction model for 5G networks. Their approach updates the number of layers and hidden units according to prediction accuracy adaptive mechanisms and utilizes seasonal time difference methods to stabilize output feature sequences, effectively improving the accuracy of 5G traffic prediction.

Zhang et al. [2] introduced a traffic analysis and prediction system for urban wireless communication networks that combines call detail record (CDR) data analysis with multivariate prediction algorithms. Their approach employs spatial-temporal modeling for historical traffic data extraction and applies causality analysis to communication data, followed by multivariate LSTM models for traffic prediction.

Research on massive Machine-Type Communications (mMTC) traffic modeling in 5G has also gained attention. Recent work [3] derived the distribution of inter-arrival times of traffic at the Base Station from mMTC users and further extended this to model traffic patterns at the Core Network. Their findings indicate that the arrival process converges to a Poisson distribution for both homogeneous and heterogeneous traffic when a sufficient number of packets is observed.

Traditional time series modeling techniques have been applied to wireless network traffic forecasting. Gowrishankar [4] modeled wireless traffic as a nonlinear and nonstationary time series and compared neural network and statistical methods for traffic prediction across different time scales. Their work demonstrated that neural networks outperform traditional statistical approaches, particularly for capturing the nonlinear characteristics of wireless traffic.

In the domain of 5G security, research has explored the detection of malicious traffic using AI techniques. Studies have analyzed encrypted malicious traffic detection methods and their applicability to 5G networks [5], highlighting the importance of traffic analysis for security purposes in addition to network optimization.

As 5G networks continue to evolve, artificial intelligence and machine learning techniques are increasingly being leveraged for traffic management. Fu et al. [6] discussed the application of AI for managing network traffic in 5G wireless networks, emphasizing the potential of these techniques to enhance network efficiency and service quality.

III. Dataset and Implementation

A. Dataset Description

This study utilizes the 5G Core Network dataset provided by the Western-OC2-Lab, which is publicly available on GitHub. The dataset contains packet captures from a simulated 5G core network environment, providing comprehensive information about network operations. The dataset includes packets captured during initial UE (User Equipment) registration with the 5G Core and subsequent network operations spanning 138 minutes, making it a valuable resource for understanding communication patterns in a 5G network.

The dataset is structured to include several key packet attributes:

- Frame number (sequence number of each packet)
- Timestamp (epoch time when the packet was captured)
- Source IP address (packet sender)
- Destination IP address (packet receiver)
- IP protocol number
- Packet length in bytes
- Protocol name

The network topology represented in the dataset follows a private IP addressing scheme, with different IP addresses corresponding to specific 5G network functions:

- 192.168.0.10: MongoDB instance
- 192.168.0.11: NWDAF (Network Data Analytics Function)
- 192.168.0.12: NRF (Network Repository Function)
- 192.168.0.13: AMF (Access and Mobility Management Function)
- 192.168.0.14: SMF (Session Management Function)
- 192.168.0.15: AUSF (Authentication Server Function)
- 192.168.0.16: UDM (Unified Data Management)
- 192.168.0.17: UDR (Unified Data Repository)
- 192.168.0.18: PCF (Policy Control Function)
- 192.168.0.19: NSSF (Network Slice Selection Function)
- 192.168.0.20: BSF (Binding Support Function)
- 192.168.0.21: UPF (User Plane Function)
- 192.168.0.22: gNB (Next Generation NodeB)
- 192.168.0.23: UE (User Equipment)

This comprehensive representation of network functions provides a realistic view of a 5G core network deployment, allowing for detailed analysis of both control and user plane traffic.

B. Implementation Methodology

The analysis of the 5G core network traffic was implemented using a Python script that leverages several data analysis and visualization libraries. The implementation follows a structured approach to extract, process, and analyze the network traffic data:

1) Data Extraction: The script provides functionality to either load an existing CSV dataset or extract a sample of packets from the PCAP file using tshark (the command-line version of Wireshark).

2) Traffic Pattern Analysis: The script analyzes various aspects of traffic patterns, including:

- Identification of top source and destination IP addresses
- Protocol distribution analysis
- Basic statistical measures of traffic characteristics

3) Network Topology Visualization: Using the NetworkX library, the script creates a graph representation of the network topology based on communication patterns between different IP addresses. The visualization helps in understanding the interaction between different network functions, with node sizes representing connectivity degree and edge thickness indicating communication frequency.

4) Packet Length Analysis: The script analyzes the distribution of packet lengths across the dataset and for specific protocols, providing insights into the size characteristics of different types of traffic. This analysis includes:

- Overall packet length distribution
- Protocol-specific packet length statistics
- Comparative visualization of packet lengths across protocols

5) Time-based Analysis: The implementation includes functionality to analyze how traffic patterns evolve over time, allowing for the identification of temporal patterns in network activity:

- Time series of packet counts
- Protocol-specific activity over time
- Identification of temporal patterns and trends

IV. Experiments and Results

A. Protocol Distribution Analysis

Our analysis of protocol distribution in the 5G core network traffic reveals significant insights into the communication patterns within the network. As shown in Fig. 1, UDP dominates the protocol usage, accounting for approximately 35,000 packets in the dataset. This prevalence of UDP is consistent with its role in carrying user plane traffic and certain control plane messages that prioritize low latency over reliability.

TCP and GTP¡TCP¿ are the next most common protocols, each representing around 5,000 packets. GTP (GPRS Tunneling Protocol) is a critical protocol in mobile networks, used

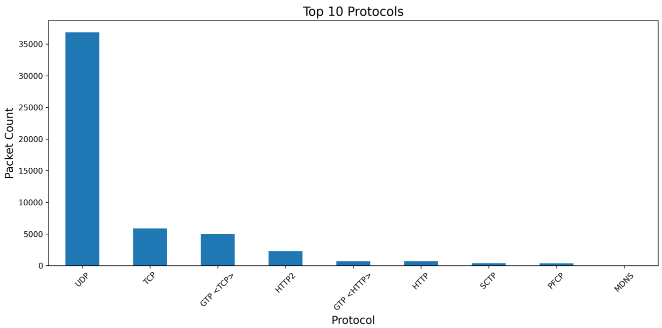

Fig. 1. Distribution of protocols in the dataset showing UDP as the dominant protocol followed by TCP and GTP¡TCP¿.

for encapsulating user data as it traverses the core network. The presence of HTTP2, GTP¡HTTP¿, HTTP, SCTP (Stream Control Transmission Protocol), PFCP (Packet Forwarding Control Protocol), and MDNS (Multicast DNS) in smaller quantities reflects the diverse protocol ecosystem required for different functions within the 5G core.

SCTP's presence, although smaller in volume, is significant as it is used for communication between control plane functions in the 5G core, particularly for interfaces defined by the 3GPP standards. PFCP is used specifically for communication between the control and user planes, particularly between the SMF and UPF.

B. Packet Length Distribution

The packet length distribution provides valuable insights into the size characteristics of traffic in the 5G core network. As illustrated in Fig. 2, the distribution is heavily skewed toward smaller packet sizes, with a significant concentration below 500 bytes. The mean packet length across the entire dataset is 180.03 bytes, as indicated by the red dashed line.

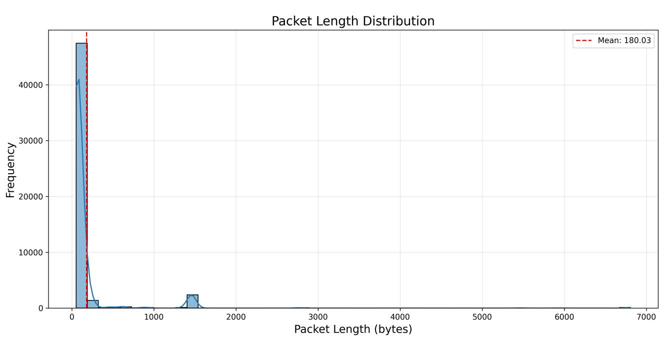

Fig. 2. Distribution of packet lengths showing a concentration of smaller packets with a mean of 180.03 bytes.

This predominance of small packets is typical of control plane traffic, which consists primarily of signaling messages rather than large data transfers. However, the distribution also shows a smaller peak around 1,400 bytes, which likely represents MTU-sized packets carrying user data or larger control messages.

The long tail of the distribution extending to around 7,000 bytes indicates occasional larger packets, which might be associated with specific operations like initial attachment, handovers, or data transfers requiring larger protocol messages.

C. Packet Lengths by Protocol

Further analysis of packet lengths categorized by protocol reveals distinct size characteristics for different protocols, as shown in Fig. 3. The box plots illustrate that HTTP, GTP¡HTTP¿, and HTTP2 protocols tend to have larger median packet sizes, which is consistent with their role in carrying application data and web content.

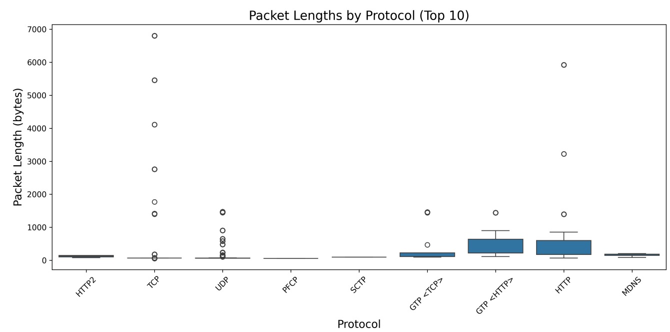

Fig. 3. Box plots showing packet length distributions by protocol. HTTP-based protocols demonstrate larger packet sizes compared to control protocols like SCTP and PFCP.

Conversely, protocols like UDP, TCP, PFCP, and MDNS generally have smaller median packet sizes, reflecting their use in control signaling and lightweight communications. TCP shows the most significant range of packet sizes, with outliers reaching up to 7,000 bytes, indicating its use for both small control messages and larger data transfers.

This variation in packet sizes across protocols has important implications for network dimensioning and quality of service management, as different types of traffic require different handling characteristics.

D. Network Communication Patterns

The network communication graph, shown in Fig. 4, provides a visual representation of the interactions between different network functions in the 5G core. The graph reveals two major communication hubs: 192.168.0.11 (NWDAF) and 192.168.0.12 (NRF), which interact with numerous other entities in the network.

The NWDAF's central role in communication patterns aligns with its function as a data analytics component that collects and processes information from various network functions. Similarly, the NRF's prominence in the communication graph reflects its role as a repository that helps network functions discover and communicate with each other.

The visualization shows that most communication occurs within two clusters centered around these key functions, with fewer interactions between the clusters. This pattern suggests a modular structure in the 5G core network, where certain

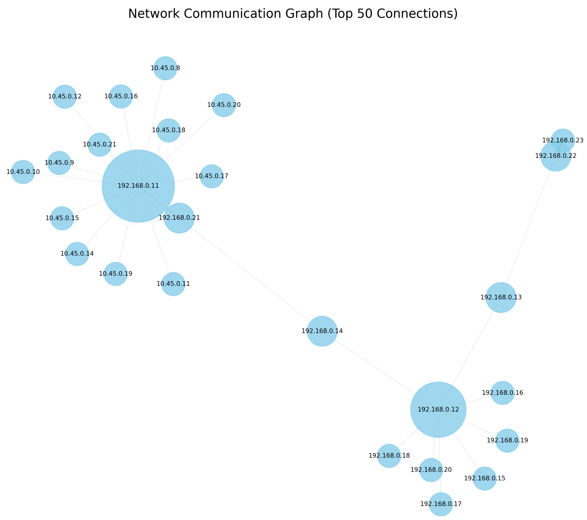

Fig. 4. Network topology visualization showing communication patterns between different network entities. Node size represents connectivity degree, and edge thickness represents communication frequency.

functions tend to communicate more frequently with specific groups of other functions.

The size of the nodes in the graph, which corresponds to the degree of connectivity, further emphasizes the central role of NWDAF and NRF in the network's operation. The thickness of the edges, representing the frequency of communication between nodes, highlights particularly active communication paths in the network.

V. Conclusion

This study has provided a comprehensive analysis of traffic patterns in a 5G core network using the Western-OC2-Lab dataset. Our analysis reveals several key characteristics of 5G core network traffic that are essential for understanding and forecasting network behavior:

- **Protocol Distribution:** UDP dominates the protocol distribution, followed by TCP and GTP-encapsulated traffic. This reflects the diverse protocol requirements for different aspects of 5G network operation, from low-latency user plane traffic to reliable control plane signaling.
- **Packet Size Characteristics:** The majority of packets in the 5G core network are small (mean: 180.03 bytes), which is consistent with the predominance of control plane signaling. However, the presence of larger packets, particularly for HTTP-based protocols, indicates the varied nature of traffic in the network.
- **Network Function Interactions:** The communication patterns between network functions reveal a structured topology with key functions like NWDAF and NRF playing central roles in network operation. This reflects the service-based architecture of 5G networks, where these functions facilitate the discovery and interaction of other components.

These findings have important implications for traffic forecasting in 5G networks. The dominance of small packets suggests that packet processing capacity, rather than raw bandwidth, might be a limiting factor in certain network functions. The varied packet size distributions across protocols indicate that forecasting models should be protocol-aware to effectively predict different traffic types. Additionally, the centralized communication patterns around certain network functions highlight potential bottlenecks that might require special attention in traffic forecasting and resource allocation.

For future work, we propose extending this analysis to develop protocol-specific forecasting models that can predict traffic patterns with greater accuracy. Incorporating machine learning techniques such as LSTM networks, as demonstrated in recent literature, could further enhance prediction accuracy by capturing the temporal dependencies in network traffic. Additionally, developing forecasting models that account for the hierarchical structure of network communications could provide more nuanced predictions of traffic patterns across different parts of the network.

References

[1] Z. Gao, "5G Traffic Prediction Based on Deep Learning," Wireless Communications and Mobile Computing, vol. 2022, Article ID 3174530, 2022.

[2] K. Zhang, G. Chuai, W. Gao, X. Liu, S. Maimaiti, and Z. Si, "A new method for traffic forecasting in urban wireless communication network," EURASIP Journal on Wireless Communications and Networking, vol. 2020, no. 1, pp. 1-14, 2020.

[3] "Modeling and Analysis of mMTC Traffic in 5G Core Networks," IEEE Transactions on Network and Service Management, vol. 22, no. 1, pp. 409-425, 2025.

[4] S. Gowrishankar, "A Time Series Modeling and prediction of wireless Network Traffic," Georgian Electronic Scientific Journal: Computer Science and Telecommunications, no. 2(16), pp. 40-52, 2008.

[5] "Exploring Emerging Trends in 5G Malicious Traffic Analysis and Incremental Learning Intrusion Detection Strategies," arXiv preprint arXiv:2402.14353, 2022.

[6] Y. Fu, S. Wang, C.-X. Wang, X. Hong, and S. McLaughlin, "Artificial Intelligence to Manage Network Traffic of 5G Wireless Networks," IEEE Network, vol. 32, no. 6, 2018.