Classifying and Prioritizing OTT and Non-OTT Network Traffic

Jeyadev Needhi[1], Sandhiya S[2], Latha Sri SA[3], Priyadarshni V[4], Sangeetha RK[5]

[1,2,3,4,5]Department of Computer Technology Madras Institute of Technology Chennai, India

Abstract:
The rapid expansion of digital communication and entertainment services like Over The Top (OTT) platforms have become a dominant force in the media and telecommunication industry. These platforms offer a wide range of services such as video streaming , voice calls and messaging which are delivered over the internet without the involvement of traditional service providers. Accurate Classification of network traffic into OTT and non-OTT helps in prioritizing and managing network resources such as bandwidth, server capacity and also ensures a better quality of Service(QoS). This project explores the appli- cation of machine learning techniques to classify network traffic into two primary categories: OTT and non-OTT. The primary objective is to develop a robust and accurate classification model that can automatically differentiate between these categories based on the analysis of network data packets.

IndexTerms: Over-The-Top (OTT), Logistic Regression, Pri- oritizing traffic, Deep packet Inspection, Class Based Traffic Shaping(CBTS).

1. INTRODUCTION

As mentioned in [1], 75 percent of the network traffic is video.Hence it becomes important to study the progress and the works that have been undertaken in the field of Content Delivery Networks(CDN).Content Delivery Network is defined as a large and distributed network, which stores The rapid evolution of modern networking technologies, ex- emplified by the advent of fifth-generation (5G) networks, has resulted in a new era of connectivity and communication. With this transformation, the way we experience and consume network services has fundamentally changed. In particular, the proliferation of Over-The-Top (OTT) applications for voice and multimedia services has significantly disrupted traditional telecommunications models.In this context, one of the chal- lenges is the effective classification of network traffic into two key categories: OTT and non-OTT. This classification task serves as the cornerstone of our research, and its significance cannot be overstated.

Network traffic classification plays a pivotal role in ensuring the efficient and reliable operation of modern networks. By distinguishing between OTT and non-OTT traffic, we gain valuable insights into the nature and demands of the network’s data flow. This knowledge forms the basis for our ability to enhance the Quality of Service (QoS) delivered to end-users. It’s interesting to note that the OTT services represent a substantial portion of today’s network traffic. From voice and video calls to multimedia streaming, these applications have become an integral part of our digital lives. Second, the dynamic and diverse nature of OTT traffic necessitates tailored optimization strategies to ensure a

seamless user experience.
Our research endeavors to bridge this gap by not only classifying network traffic but also by applying advanced techniques for prioritization and bandwidth allocation to OTT traffic. This dual-pronged approach promises to unlock several benefits. It enhances the overall network efficiency, ensures smoother user experiences, and provides network operators with the tools to better manage and allocate resources.
In simple words our research addresses the pressing need for precise network traffic classification, with a focus on the critical distinction between OTT and non-OTT traffic. By optimizing OTT traffic through prioritization and resource allocation, we aim to improve QoS, benefiting both end- users and network providers. As we delve deeper into our study, we embark on a journey to harness the full potential of modern networks in an era defined by connectivity and digital innovation.

2. LITERATURE SURVEY

In [2] the significance of network traffic classification in modern computer science has been highlighted. It emphasizes its crucial role for Internet Service Providers (ISPs) to under- stand and manage the various types of network applications flowing in their networks.The authors of this paper used three primary approaches like Port Based classification, Payload Based classification and Machine Learning Based classifica- tion to perform the classification task.Specifically in Machine Leraning models like Support Vector Machine, C4.5 decision tree, Na¨ıve Bayes, and Bayes Net were used and their accuracy rates were compared.
The paper [3] encompasses an exploration of the evolving landscape of Over-The-Top (OTT) services in the context of rapidly advancing multiservice networks.Also [3] identifies and addresses issues associated with data transmission through telecom operators' networks, proposing prospective solutions through Deep Packet Inspection (DPI) technology to enhance service quality. Moreover, the authors introduce a classification of traffic prioritization based on application-specific require- ments, fostering flexibility in user data transmission while emphasizing compatibility with OTT services employing open solutions.
In [4] the author claims that network traffic classification plays a crucial role in network measurement and management. Existing methods often focus on application-level results, neglecting the critical aspect of network Quality of Service (QoS) requirements. This paper proposes a novel approach that addresses this limitation by introducing two QoS-aware features: "inter-APP similarity" and "intra-APP diversity." These features capture the QoS relationships between traffic flows from different Internet applications and the variability in QoS within the same application's traffic. The methodology employs a Long Short-Term Memory neural network-based Autoencoder (LSTM-AE) to extract these QoS-aware features, which are then used to classify traffic into distinct QoS classes. Real-world data from various applications validate the effectiveness of this approach, promising advancements in network measurement and management.
The paper[5] addresses the critical issue of enhancing Qual- ity of Experience (QoE) for Over-the-top (OTT) multimedia services in modern network environments. As OTT content de- livery via the Internet becomes increasingly predominant, the study highlights the growing importance of optimizing content delivery systems for improved user satisfaction. The research introduces two distinct approaches, namely the Analytic Hi- erarchy Process (AHP) and a Neural Network-based model to optimize OTT delivery transparently near the end-user. But in the scenarios where network bandwidth is constrained,

the Analytic Hierarchy Process (AHP) approach tends to exhibit suboptimal performance that the expected enhancements in Quality of Experience (QoE) may not be achieved.

The paper [6] proposed a framework utilizing deep learning techniques like LSTM and CNN to identify encrypted OTT voice traffic in 5G networks addressing concerns about se- curity and network supervision. The experiments conducted including the evaluation of sample sizes and deep learning methods demonstrate the promise of LSTM in this context.But the paper does not explore unsupervised approaches for han- dling unknown encrypted traffic in real networks as part of its current research scope.

The significance of network traffic classification as a key tool for network management and security is emphasized in [7]. This study offers a novel method called Robust Statistical Traffic Classification (RTC), which combines supervised and unsupervised machine learning approaches, to address the issue of zero-day applications in traffic classification systems. The authors show how RTC can successfully recognize both zero-day applications and classes of preconfigured applica- tions. This study offers insightful solutions to the problem of misclassifying zero-day traffic.

The study presented in [8] emphasises the vital relevance of real-time network traffic monitoring in the face of grow- ing cybersecurity threats. Because of increased encryption and application developers' ongoing ingenuity in escaping detection, network categorization, a critical component of intrusion detection systems, remains a major issue. The study emphasises the use of Machine Learning (ML) algorithms to various categorization approaches, with a focus on the statistical features of network traffic flow. The study highlights the increased interest in this subject and notes a need for enhanced classification methods, notably in using ML for quicker recognition and detection based on sub-flows, with the ultimate objective of improving network traffic classification accuracy and security.

The authors of [9] study present a complete description of OTT services and compare their existence to conventional services. They investigate the categorization of services pro- vided by OTT platforms and find issues with data transfer across telecom operator networks.To satisfy Quality of Service standards for OTT services, the authors present potential methodologies and DPI technology procedures. They also investigate the use of open technologies to integrate DPI sys- tems with OTT services. To support its conclusions, the paper provides experimental data, signature efficiency assessments, probability estimate methods, and statistical analyses.

The publication [10] shows the significant advancements made in the application of machine learning (ML) to the categorization of IP traffic. The emphasis is shifted away from traditional payload- and port-based approaches, which is notable. This review article explores the potential of ML algorithms for offline analysis and their likelihood of being used in real-time inside functional networks. The report admits that despite these developments, there are still concerns with cross-network assessment, network problem resilience, and application adaptation for various Internet applications.

3. Tools Used

The tools include Python and popular data science libraries such as NumPy, Pandas, scikit-learn, and Matplotlib for data manipulation, machine learning, and visualization.

4. Proposed work

A. Data extraction and optimization

This is the foundational stage in our project's network traffic analysis pipeline. It begins with the collection and extraction of raw network traffic data from the dataset ”Ip network traffic flows labeled with 87 apps” from kaggle. Once obtained, the dataset undergoes a crucial process of feature selection where we carefully curate relevant attributes, effectively reducing data dimensionality while preserving essential information.Additionally, we label the data according

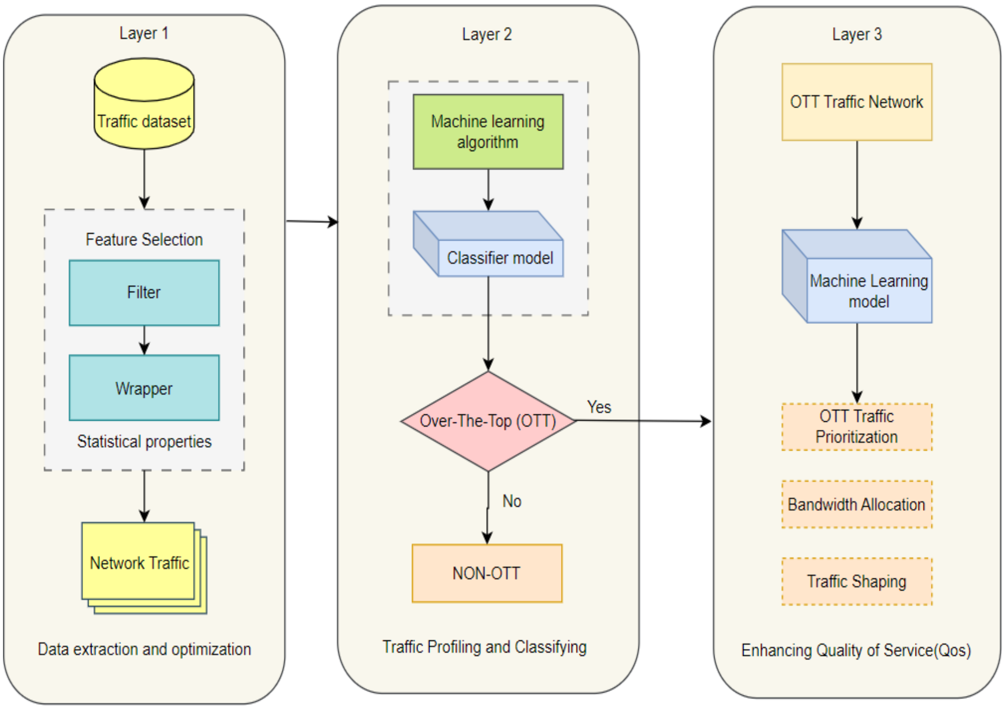

Fig. 1. Proposed architecture

to the protocol, providing a critical context for subsequent analyses.Ultimately, the outcome is a refined network traffic dataset primed for more advanced analyses such as traffic classification and prioritization in subsequent project layers.

B. Traffic Profiling and Classifying

After extracting the dataset, the dataset is split into two parts : a training set and a testing set. The training set will be used to train the logistic regression model, while the testing set will be used to evaluate its performance. The relevent features from the dataset which helps in distinguishing between OTT and NON OTT is choosen and a logistic regression model is trained using the training dataset. The model will learn to make predictions based on the selected features.After training, testing dataset is used to evaluate the model's performance.By varying the size of the training dataset, an accuracy graph is created. The size is initially small and is increased gradually like 100, 500, 1000, 5000, and 10000 samples.

Algorithm 1 Logistic Regression Model Evaluation

Data: Logistic Regression Dataset, Dataset Sizes
Result: Accuracy Scores, Accuracy vs. Dataset Size Graph Import necessary libraries for Logistic Regression
Define dataset sizes and initialize accuracy scores list
for *each nrows in dataset sizes* **do**
Load the dataset Preprocess the dataset Train a Logistic
Regression model on a dataset of size *n*rows Calculate accuracy and store it in accuracy scores

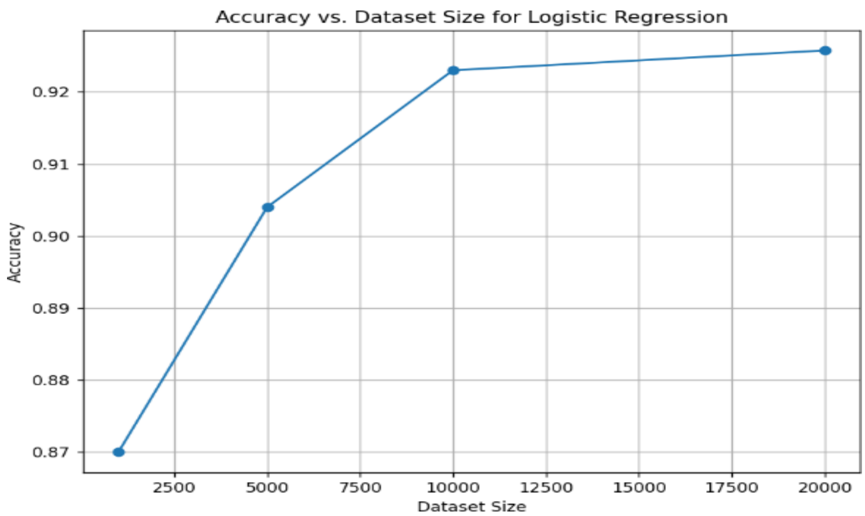

Fig. 2. Classification Accuracy using Logistic Regression

and video streaming to online gaming and business communications, relies heavily on the Quality of Service (QoS) delivered by the underlying network infrastructure. QoS, in this context, signifies the network's ability to provide data packets with predictable and consistent performance, guaranteeing low latency, high bandwidth, and minimal packet loss. Enhanced QoS not only leads to a superior user experience but also maximizes the utilization of network resources, ensuring both cost-effectiveness and scalability

Algorithm 2 Token Bucket Algorithm
Data: Token Bucket Limits, Fill Rate
Result: Data Transmission Start
Set token bucket limits and fill rate Periodically add tokens to the bucket
Verify token availability for data transmission
while *Transmitting data* **do if** *tokens suffice* **then** Transmit data
end else
Await token replenishment
end
Ensure data transmission adheres to the Committed Infor- mation Rate
end
Stop
end
Plot accuracy vs. dataset size graph
Exit

C. Enhancing Quality Of Service - QoS

Networks often contend with fluctuating and unpredictable levels of data traffic, resulting in congestion and performance degradation. The effective and reliable functioning of applications and services, spanning from internet browsing

The Token Bucket algorithm is a fundamental traffic shaping technique employed in computer networks to regulate the flow of data packets. In this approach, a virtual bucket holds tokens, each representing a fixed amount of data that can be transmitted. The tokens are replenished at a constant rate, and data packets can only be transmitted if there are sufficient tokens in the bucket. This mechanism allows bursts of traffic while maintaining an average data transmission rate, enabling efficient bandwidth utilization and preventing network conges- tion.

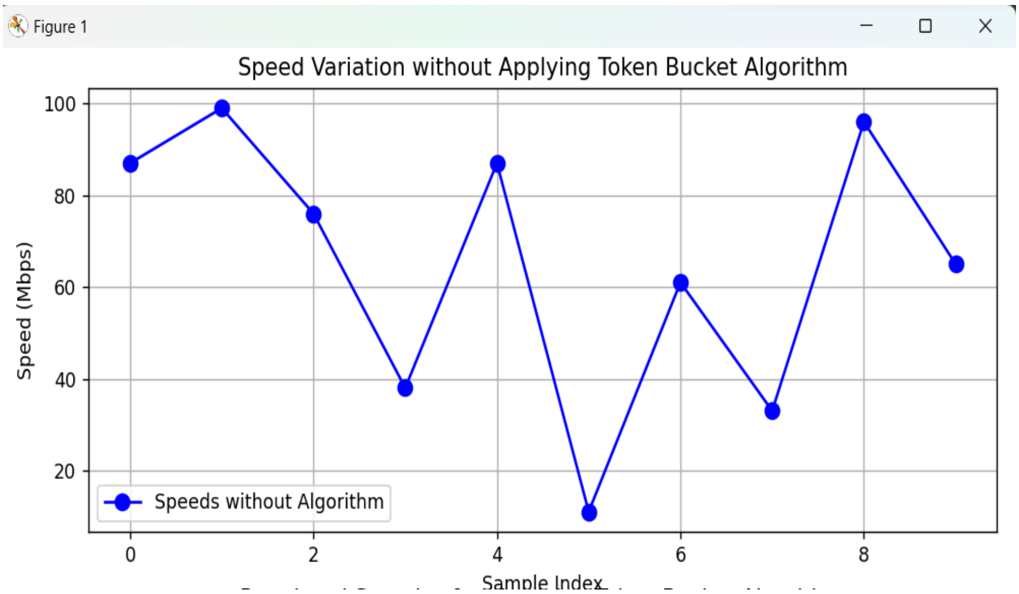

Fig. 3. Before Applying Token Bucket Algorithm on Speed

Bucket algorithm emerges as a valuable tool in optimizing network resources, elevating user experiences, and ensuring the seamless operation of OTT applications. We implemented the Token Bucket algorithm, a funda- mental technique in network traffic shaping. The algorithm was applied to regulate network speeds within a constant bandwidth of 1000 Mbps, divided into 10 intervals. Initially, random network speeds ranging from 10 to 100 Mbps were generated to mimic unregulated network traffic patterns. By applying the Token Bucket algorithm, we ensured that these speeds remained above 1 Mbps. The resulting visualizations showcase the fluctuating unregulated speeds represented by the blue line, and the stable, regulated speeds represented by the red line. This implementation provides valuable insights into how the Token Bucket algorithm effectively smoothens speed fluctuations, ensuring a consistent and optimized network performance, thereby enhancing the overall quality of service for users.

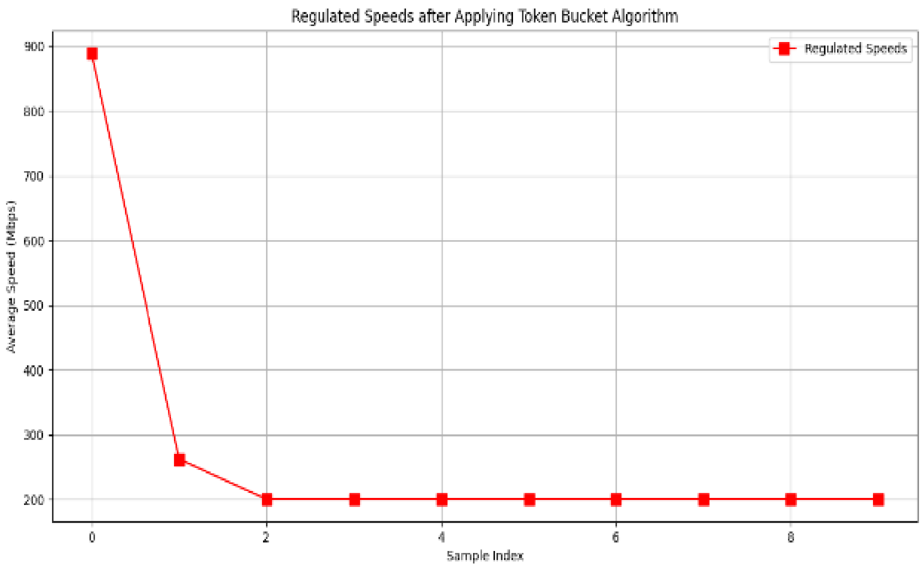

Fig. 4. After Applying Token Bucket Algorithm on Speed

OTT services, often involving video streaming and real- time data transmission, demand consistent and uninterrupted data flow to ensure seamless user experiences. By employing the Token Bucket algorithm, OTT traffic can be prioritized and regulated effectively. The algorithm allows OTT traffic to be allocated a specific portion of the available bandwidth, ensuring a steady and predictable data stream. This controlled transmission prevents network congestion, buffering issues, and latency spikes, guaranteeing uninterrupted streaming and improved video quality for OTT users. Meanwhile, non-OTT traffic coexists harmoniously by utilizing the remaining bandwidth without compromising the quality of service. By enhancing the delivery of OTT services, businesses can retain satisfied customers, foster brand loyalty, and gain a competitive edge in the digital landscape. The Token

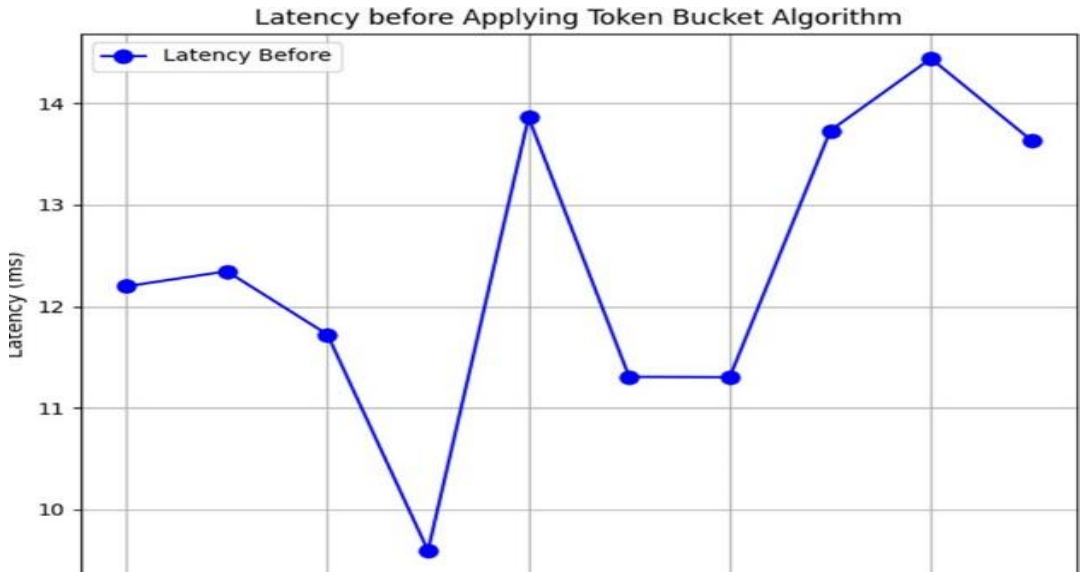

Fig. 5. Before Applying Token Bucket Algorithm on Latency

The graph visually represents the transformative impact of the Token Bucket algorithm on latency. Prior to the application of the algorithm, latency experiences substantial fluctuations due to the erratic nature of data packet arrivals. These fluctuations occur at random intervals and speeds, leading to jagged peaks in the latency profile. This unpredictability is indicative of the challenges posed by unregulated data flow within the network. In this unregulated environment, the network faces difficulties in maintaining a consistent and predictable user experience. The unpredictable nature of data arrivals introduces a level of volatility, leading to instances of delayed packet processing and communication breakdowns. Users engaging with applications or services during periods of high data demand may experience interruptions, buffering, or latency spikes, impacting the overall quality of service.

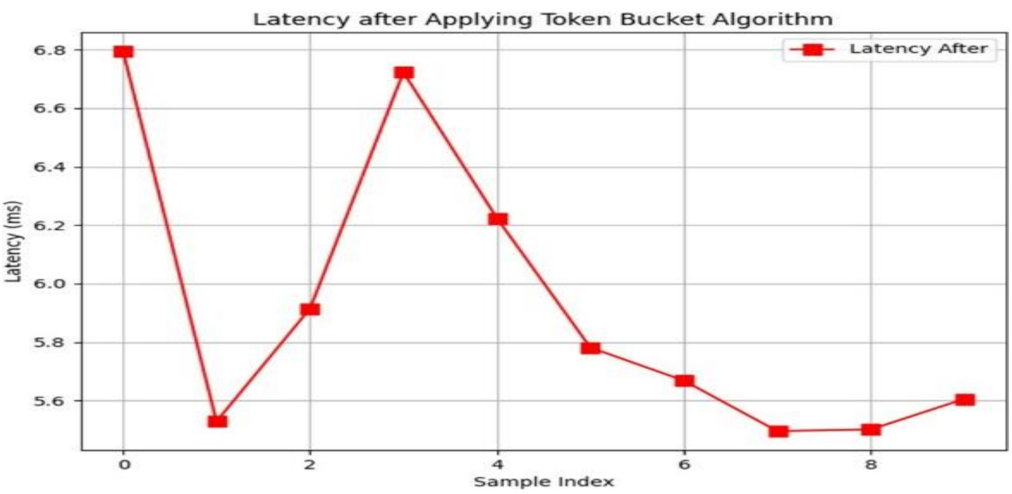

Fig. 6. After Applying Token Bucket Algorithm on Latency

Following the integration of the Token Bucket algorithm, a discernible improvement in latency becomes apparent. This reduction in latency contributes to a more responsive and predictable network performance, particularly beneficial for Over-the-Top (OTT) applications. The expectation is that the latency graph for OTT traffic, which encompasses high-bandwidth content delivery such as streaming video, will showcase a more controlled and predictable profile compared to the pre-algorithm state. The outcome of the graph is a significant reduction in latency for OTT applications, leading to an enhanced streaming experience for users. Lower latency translates to reduced buffering times, quicker response to user interactions, and an overall improvement in the quality of service. This positive impact on latency aligns with the demands of bandwidth-intensive services, contributing to improved user satisfaction and engagement. The latency graph serves as a visual representation of the algorithm's effectiveness in action. It illustrates the transition from a potentially erratic and unpredictable latency profile to a more controlled and stable state post-implementation of the Token Bucket algorithm.

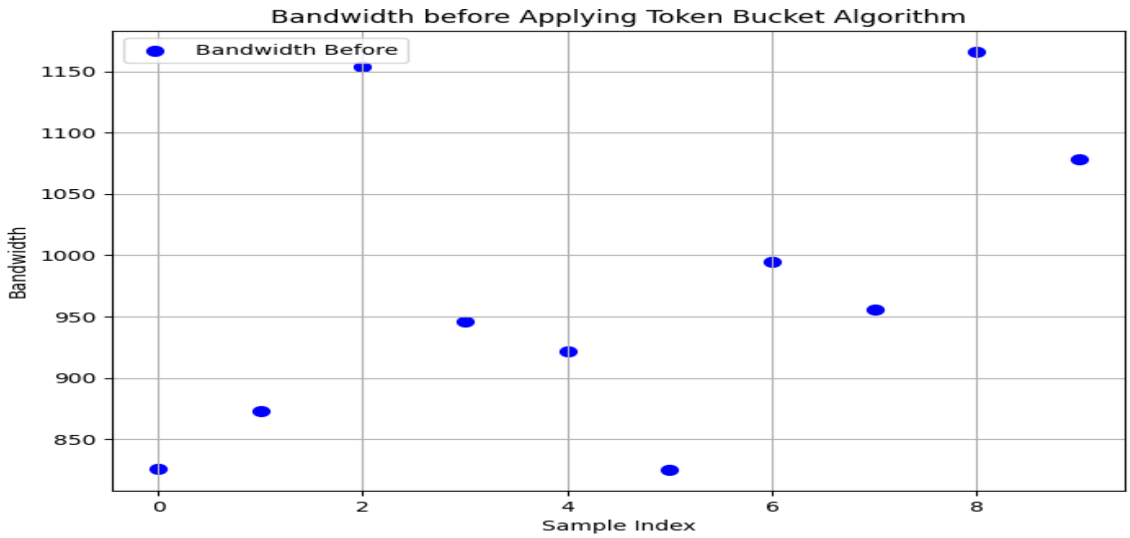

Fig. 7. Before Applying Token Bucket Algorithm on Bandwidth

In the above scatter plot, each data point on the graph corresponds to a specific sample index, and the y-axis value represents the unregulated bandwidth level before the ini- tiation of the token bucket algorithm. This scatter plot is instrumental in visually portraying the inherent randomness that characterizes the initial bandwidth values. As the sample index progresses along the x-axis, the diverse y-values vividly illustrate the unpredictable nature of bandwidth distribution in this uncontrolled context. The graph effectively captures the fluctuating and erratic pattern of bandwidth values before the implementation of any regulatory measures, providing a rich and vivid depiction of the unregulated state of network traffic. This visual representation serves as a valuable tool for understanding the inherent variability and lack of structure in bandwidth under uncontrolled conditions. It emphasizes the crucial role of regulatory mechanisms in enhancing network performance and stability by imposing control and structure on the bandwidth distribution. The scatter plot highlights the necessity of implementing measures, such as the token bucket algorithm, to bring about a more controlled and predictable state in network traffic.

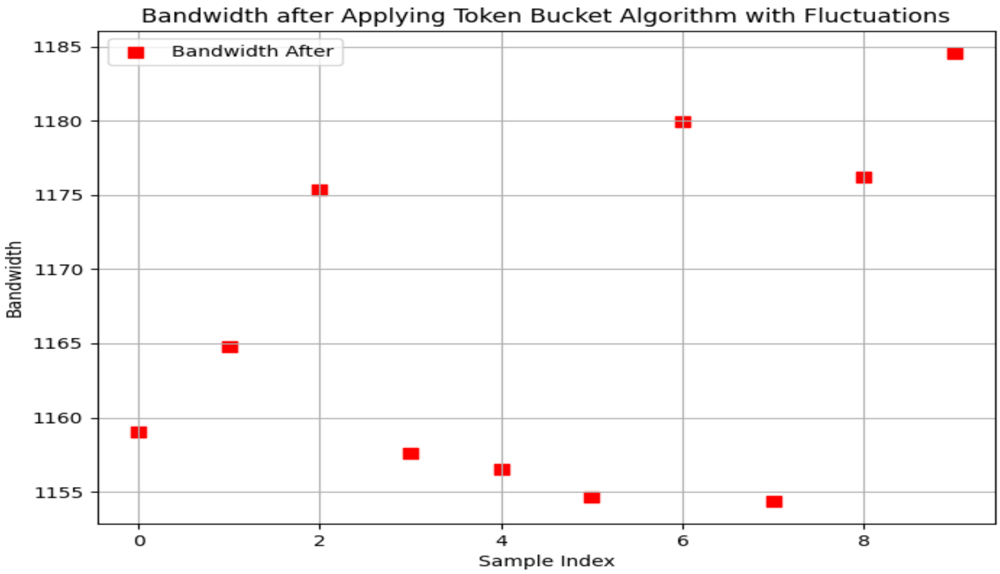

Fig. 8. After Applying Token Bucket Algorithm on Bandwidth

The scatter plot above depicts that each data point is associ- ated with a specific sample index, and the y-axis value signifies the regulated bandwidth level after the implementation of the token bucket algorithm. Unlike the initial unregulated state, this graph introduces fluctuations to the regulated bandwidth values, simulating real-world variability in network conditions. As the sample index progresses along the x-axis, the fluctu- ating y-values illustrate the controlled yet variable nature of bandwidth after the token bucket algorithm is applied.
These fluctuations serve to emulate scenarios where network conditions exhibit some degree of unpredictability, despite the regulatory measures in place. The graph provides a nuanced view of how the token bucket algorithm brings about both control and variability in the regulated bandwidth. It highlights that, even with regulatory mechanisms in effect, network dynamics can still experience fluctuations, adding a layer of realism to the model. This nuanced understanding is crucial for devising effective strategies to manage and optimize network performance under diverse and dynamic conditions.

5. Conclusion

This project represents a critical stride in adapting to the dynamic landscape of modern network management. The classification of network traffic into OTT and non-OTT is not merely an academic pursuit but a practical necessity in our digitally connected world. The successful development and deployment of an accurate machine learning-based clas- sification model hold the promise of revolutionizing network resource management. As we stride forward into an era where digital communication and entertainment continue to evolve, the ability to optimize network resources for enhanced QoS becomes paramount. This project stands as a testament to the intersection of innovation and network management, offering tangible solutions that can benefit both service providers and end-users. It is our hope that the insights gained from this endeavor will contribute significantly to the efficiency, performance, and ultimately, the quality of network services, ensuring a seamless and satisfying digital experience for all.

References

1. D. Kiedanski, M. Nogueira and E. Gramp´ın, ”Youtube traffic from the perspective of a

developing country: the case of Uruguay," IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Paris, France, 2019, pp. 714-719, doi: 10.1109/INFCOMW.2019.8845096.
2. M. Shafiq, X. Yu, A. A. Laghari, L. Yao, N. K. Karn and F. Abdessamia, "Network Traffic Classification techniques and comparative analysis using Machine Learning algorithms," 2016 2nd IEEE International Con- ference on Computer and Communications (ICCC), Chengdu, China, 2016, pp. 2451-2455, doi: 10.1109/CompComm.2016.7925139.
3. V. S. Elagin, B. S. Goldshtein, A. V. Onufrienko, I. A. Belozerzev and A. A. Savelieva, "Modeling OTT services in multiservice networks in order to synchronize and prioritize traffic," 2018Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO), Minsk, Belarus, 2018, pp. 1-5, doi: 10.1109/SYN- CHROINFO.2018.8456968.
4. J. Zhang, F. Li and F. Ye, "Network traffic clustering with QoS- awareness," in China Communications, vol. 19, no. 3, pp. 202-214, March 2022, doi: 10.23919/JCC.2022.03.015.
5. A. Dias, A. B. Reis and S. Sargento, "Improving the QoE of OTT Multimedia Services in Wireless Scenarios," 2019 IEEE Symposium on Computers and Communications (ISCC), Barcelona, Spain, 2019, pp. 1-6, doi: 10.1109/ISCC47284.2019.8969709.
6. Z. Qiao, L. Zhai, S. Zhang and X. Zhang, "Encrypted 5G Over- The- Top Voice Traffic Identification Based on Deep Learning," 2021 IEEE Symposium on Computers and Communications (ISCC), Athens, Greece, 2021, pp. 1-7, doi: 10.1109/ISCC53001.2021.9631458.
7. J. Zhang, X. Chen, Y. Xiang, W. Zhou and J. Wu, "Robust Network Traffic Classification," in IEEE/ACM Transactions on Networking, vol. 23, no. 4, pp. 1257-1270, Aug. 2015, doi: 10.1109/TNET.2014.2320577.
8. N. Al Khater and R. E. Overill, "Network traffic classification techniques and challenges," 2015 Tenth International Conference on Digital Infor- mation Management (ICDIM), Jeju, Korea (South), 2015, pp. 43-48, doi: 10.1109/ICDIM.2015.7381869.
9. V. S. Elagin, B. S. Goldshtein, A. V. Onufrienko, A. A. Zarubin and A. Savelieva, "The efficiency of the DPI system for identifying traffic and providing the quality of OTT services," 2018 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, Russia, 2018, pp. 1-5, doi: 10.1109/SOSG.2018.8350589.
10. T. T. T. Nguyen and G. Armitage, "A survey of techniques for internet traffic classification using machine learning," in IEEE Communications Surveys Tutorials, vol. 10, no. 4, pp. 56-76, Fourth Quarter 2008, doi: 10.1109/SURV.2008.080406.
11. Grimaudo, Luigi Mellia, Marco Baralis, Elena. (2012). Hier- archical Learning for Fine Grained Internet Traffic Classification. 10.1109/IWCMC.2012.6314248.
12. S. Xiong, A. D. Sarwate and N. B. Mandayam, "Network Traffic Shaping for Enhancing Privacy in IoT Systems," in IEEE/ACM Trans- actions on Networking, vol. 30, no. 3, pp. 1162-1177, June 2022, doi: 10.1109/TNET.2021.3140174.
13. R. Hofmann, B. Nikolic´ and R. Ernst, "Slack-based Traffic Shaping for Real-time Ethernet Networks," 2019 IEEE 25th Interna- tional Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA), Hangzhou, China, 2019, pp. 1-11, doi: 10.1109/RTCSA.2019.8864581.
14. L. K. B. Melhim, M. Jemmali and M. Alharbi, "Network Monitoring Enhancement based on Mathematical Modeling," 2019 2nd International Conference on Computer Applications

Information Security (ICCAIS), Riyadh, Saudi Arabia, 2019, pp. 1-4, doi: 10.1109/CAIS.2019.8769583.