Deep Residual CNN-Attention Model with GMM Statistical Injection: A Forensic Approach to Mitigating Dataset Bias for Zero-Day Detection in SDN

by

Ikti Safat Anjum Soumya
(21201816)
Bokthier Bin Foysal
(21241016)
Md Nafiz Fuad Sharkar
(21301366)

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
BRAC University
January 2026.

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Ikti Safat Anjum Soumya
21201816

Bokthier Bin Foysal
21241016

Md Nafiz Fuad Sharkar
21301366

Approval

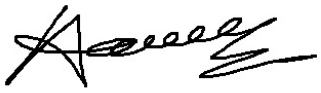
The thesis titled "Deep Residual CNN-Attention Model with GMM Statistical Injection: A Forensic Approach to Mitigating Dataset Bias for Zero-Day Detection in SDN" submitted by

1. Ikti Safat Anjum Soumya (21201816)
2. Bokthier Bin Foysal (21241016)
3. Md Nafiz Fuad Sharkar (21301366)

of Fall, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science in Fall, 2025.

Examining Committee:

Supervisor:
(Member)



Dr. Amitabha Chakrabarty, PhD

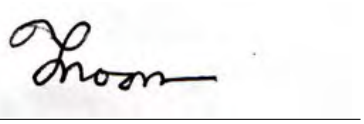
Professor
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

Mr. Zaber Mohammad

Lecturer
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)



Dr. Jannatun Noor Mukta

Director, BSDS
Associate Professor, CSE
United International University

Thesis Coordinator:
(Member)

Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi, PhD

Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Software-Defined Networks (SDN) and its centralized control plane have become highly valuable assets of the present-day infrastructure, and therefore, one of the primary targets of DDoS attacks. Although Intrusion Detection Systems (IDS) that are based on Machine Learning have potential, most of the systems experience shortcut learning. Instead of training to actually attack them, they learn fixed identifiers such as IP addresses. These models are very accurate in the laboratory, but they do not reach accuracy in real-world applications. This paper addresses the issue through one of its strategies, the Forensic-Induced Progressive Refinement Strategy, which focuses on dataset validation rather than performance measurements. First, bias-inducing qualities are located and eliminated as a result of conducting a forensic audit. Second, in physics-based constraints, a Topology-Aware Multiclass Augmentation engine creates realistic traffic profiles of twelve classes in order to solve the issue of data scarcity and class imbalance. The main contribution is a Hybrid Wide and Deep Learning Framework that is a combination of two streams: a deep stream, which is a 1D-CNN, residual connection, and Multi-Head Attention to extract spatial patterns, and a wide stream, which uses Gaussian Mixture Models (GMM) to extract statistical grounds. The integrated Zero-Day Forensics mechanism employs Integrated GMM-based Negative Log-Likelihood thresholds to identify previously unknown attack type which breaks the closed-world assumption of traditional classifiers. The experimental findings indicate that the framework is 88.96% accurate on bias-free data and generalizes well against Zero-Day threats, which are highly inaccurate compared to the baseline models SVM and KNN when trained on biased data. The piece institutionalizes forensic rigor and architectural duality as the key concepts towards strong SDN intrusion detection.

Keywords: Software-Defined Networking (SDN), Intrusion Detection System (IDS), Dataset Bias, Deep Residual Learning, Multi-Head Attention, Gaussian Mixture Model (GMM), Zero-Day Detection, Forensic Auditing.

Acknowledgement

Firstly, all praise is due to the Almighty Allah, by whose infinite mercy and blessings our thesis has been completed.

Secondly, we would like to express our deepest gratitude and sincere appreciation to our thesis supervisor, **Dr. Amitabha Chakrabarty**, for his continuous guidance, valuable suggestions, and unwavering support throughout the entire research process. His insightful feedback, constructive criticism, and encouragement greatly enhanced the quality of our work. He was always generous with his time and patience by never giving up on us and providing direction whenever we faced difficulties, and motivating us to overcome challenges with confidence. His mentorship played a vital role in shaping this thesis and in enriching our academic experience. Also, our Co-Supervisor **Mr. Zaber Mohammad** who was always there for us whenever we needed him.

Finally, we are profoundly thankful to our parents, whose constant support, prayers, and sacrifices have been our greatest source of strength. Without their encouragement and belief in us, this achievement would not have been possible. With their blessings, we are now on the verge of completing our graduation.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

AUROC Area Under the Receiver Operating Characteristic Curve

CNN Convolutional Neural Network

DDoS Distributed Denial-of-Service

DPI Deep Packet Inspection

GMM Gaussian Mixture Model

IDS Intrusion Detection Systems

KNN k-Nearest Neighbors

LSTM Long Short-Term Memory

ML Machine Learning

NLL Negative Log-Likelihood

PCG Parallel Convolutional Gaussian

ReLU Rectified Linear Unit

SDN Software Defined Networking

SMOTE Synthetic Minority Over-sampling Technique

SVM Support Vector Machines

TCAM Ternary Content-Addressable Memory

Chapter 1

Introduction

1.1 Background

Software Defined Networking (SDN) has transformed the design of modern networks by decoupling the control plane (Brain) from the underlying switches (Data Plane). This partition enables network management and coding at a central location. Nevertheless, that very design also introduces another significant vulnerability: if the Controller fails, the whole network collapses. This renders SDN an excellent source of attacks [6].
The TCP SYN flood is considered to be one of the most hazardous issues endangering this architecture. In a conventional network, such an attack attempts to shut down one server. However, in the SDN world, it is far more detrimental. The Attack overwhelms the Controller with millions of false claims (so-called Packet-In messages), and the Controller has to handle garbage information before it stops functioning. This has the potential to break the memory (TCAM tables) of the network and isolation of legitimate users [2].
Machine Learning (ML) has been embraced as a solution by researchers in order to prevent such attacks. ML models are effective at discovering attack patterns that are complex and cannot be detected by simple firewalls [7]. Nevertheless, most of the available research uses simple models that are considered Black Boxes, which identify attacks but fail to define the causes and remedies. Moreover, common models can struggle to identify new and unknown attacks (Zero-Day threats) since they just scan patterns that they have encountered previously [9].
This study is based on these results to come up with a stronger solution. This paper does not use a standard model but the hybrid approach, which integrates deep pattern recognition with statistical analysis. This enables the improved detection of these particular SDN attacks and gives the transparency required for defending in practice.

1.2 Rationale of the Study

The main reason why this study is carried out is based on a separate validation of the data presented in the reference study. On the application of standard supervised learning algorithms to this dataset, the detection accuracy, precision, and recall were close to 100% in all models. This kind of homogeneous perfection was an indication that there were possible weaknesses in the data itself and not the effectiveness of the

models. A forensic examination served to indicate critical feature leakage wherein certain features were unintentionally coded as certain predictors. This enabled models to cheat through shortcut learning of merely basic labels as opposed to the generalized behavior of an assault.
In addition to the fact that this particular dataset is limited, there is an underlying lack of suitable training data on Software-Defined Networking (SDN). Although general intrusion detection is considered through benchmark datasets, such as the CIC-IDS2017, the datasets were recorded in conventional legacy networks and do not have the architecture of SDN. SDN adds attack surfaces that are specific to Control Plane Saturation and Flow Table (TCAM) exhaustion not found in legacy datasets. As a result, the training of an SDN-specific defense model on legacy data leads to a 'Domain Mismatch' whereby the model is trained to secure end-servers, as opposed to securing the vulnerable network infrastructure. To compensate for this, the current research relies on a custom-made dataset obtained in a Mininet simulation to guarantee the measurement of the particular time dynamics of the OpenFlow protocol.
It is based on this data that the research discusses the insufficiency of generalized models to deal with SDN-specific threats. SDN designs are vulnerable to TCAM Exhaustion due to TCP SYN flooding. A generalized model can identify volumetric anomalies, but can usually not stop the rapid usage of flow-table resources, which defines such an attack. Thus, the proposed research assumes a specialized Augmented TCP SYN Defense based on a Wide-and-Deep Parallel Convolutional Gaussian (PCG)-CNN architecture. Prioritizing the asymmetric threat vectors of SDN by combining Deep Learning (to detect complex patterns) with Gaussian Mixture Models (to provide the statistical context), the given approach secures the resources of the Controller and Data Plane that cannot be covered by the generalized models.

1.3 Problem Statement

The existing Machine Learning-related solutions to SDN security have significant limitations that negatively affect their application in practice. The main problem is the failure of data validity and shortcut learning. The majority of the detection models are trained using datasets in which there is high feature leakage, and category protocol flags are a trivial predictor. It creates a false sense of strength, as models nearly identify perfect accuracy, but the retrieval of these artifacts, as opposed to the learning of the intricate temporal dynamics of an attack, is challenging. These models thus crash miserably when attackers alter the packet header or when they are used in various network contexts where such simplistic rules are not applicable.
In addition to data quality, there is the issue of the detection of Zero-Day anomalies by existing architectures. Conventional supervised classifiers are also known as Closed-Set systems, which can only identify those signatures of an attack that are available in their training set. They are not equipped with the statistical context to detect new unknown attacks. Where there is a new form of a SYN flood, or a low-rate attack not in a known pattern, these generalized models are compelled to make a binary guess, which frequently leads to high False Negative rates.
Moreover, the systems do not have a problem of interpretability, and this is commonly known as the Black Box problem. Although Deep Learning models can be successful in marking traffic as malicious, they do not usually indicate the cause of the decision.

A single label of Attack is inadequate in an SDN environment where a Controller decision can be used in the whole network infrastructure. The administrators need finer forensic details to confirm the threat and implement the appropriate mitigation policies, which the standard end-to-end learning models cannot offer.
Thus, the main issue that the study will deal with is the absence of an effective, clarifiable, and hybrid detection framework. To avoid shortcut learning on leakage-free, statistically augmented data, a system is urgently needed, which, on the one hand, should be trained on leakage-free data and, on the other hand, should also fuse Deep Learning and Probabilistic techniques to be able to identify known and zero-day threats and provide clear, actionable forensics.

1.4 Objective

Developing a reliable, comprehensible, and architecturally customized Intrusion Detection System (IDS) for Software-Defined Networks is the main objective of this project. The following particular goals are used in this study to address the limitations mentioned in the problem statement:

- Identifying the methodological errors in current benchmarks by doing a repeatability study that identifies methodological errors in existing benchmarks to establish a strict forensic baseline. In order to make sure the suggested method is based on legitimate, generalized learning rather than artifact memorizing, we link high accuracy rates to "Identity Bias" and feature leakage.

- Reducing class disparity and data scarcity by creating a multi-level augmentation framework based on physics. In order to ensure that the model is trained on a high-fidelity, balanced signal rather than statistical noise, this system creates kinetically valid synthetic traffic for minority classes (such as VoIP and Slowloris).

- Creating the innovative "PCG-CNN" Hybrid Architecture to address the "Black Box" interpretability issue. This architecture offers a dual-view analysis that performs better than single-stream classifiers by combining a Deep Learning path (1D-CNN) for feature extraction with a Probabilistic path (Gaussian Mixture Models).

- Using a statistical "Gatekeeper" approach to enable reliable Zero-Day anomaly detection. The goal of this study is to focus on employing Negative Log-Likelihood (NLL) thresholds to identify and flag unexpected traffic patterns that deviate from the learned distribution in order to efficiently convert a Closed-Set classifier into an Open-Set defensive system.

- Using t-SNE manifold analysis to validate the decision boundary in order to provide actionable forensic transparency. This ensures that network management receives physically verifiable evidence of the kinetic alignment of the traffic, allowing for precise and reliable mitigation strategies.

1.5 Methodology in Brief

The study will use a quantitative and experimental approach, designed in a systematized way to fill the mentioned gaps in SDN security. To base the performance on a standard level, a separate validation of the reference SDN dataset with standard supervised learning algorithms (SVM, KNN) was used to initiate the study. The rigorous feature forensic analysis was then done to isolate the leaked attributes that were protocol flags, which were found to have been the key contributors to artificial performance inflation and shortcut learning.

After eliminating these faulty characteristics, the study turned its attention to data rebuilding in order to address the ensuing lack of data. Statistical augmentation methods were used to sample the feature space, resulting in a rich and well-balanced multiclass dataset. This procedure was done to guarantee inclusion of both complicated and correlated traffic patterns that represent heterogeneous benign applications and diverse attack vectors, which will eradicate bias of the original source and avoid imbalance in classes.

A new architecture named Wide and Deep PCG- CNN was developed and applied to overcome the drawback of single-stream classifiers. This hybrid architecture combines two learning streams, which are complementary: a deep Convolutional Neural Network (CNN) to extract temporal patterns in traffic, and a Gaussian Mixture Model (GMM) to analyze statistical probability distributions. The suggested framework was strictly tested in a variety of conditions, such as class-balanced or imbalanced. Last, the system was put through stress tests of Zero-Day attack conditions (unseen traffic), and the decision-making process was tested by a Benign-Reference Forensic Engine to achieve explainability and trust.

1.6 Scopes and Challenges

Scope of the Study

The focus of this paper is the flow-level traffic analysis to identify DDoS in Software-Defined Networks (SDN). The study uses offline traffic data, where it employs supervised and hybrid machine learning methods to overcome the weaknesses of the current benchmarks. The test includes binary (Attack vs. benign multiclass classification tests: a particular extension to the Zero-Day attack test is considered. The paper gives importance to dataset validity, feature integrity, and forensic explainability rather than focusing on real-time controller deployment or deep packet-level inspection. Although the framework of the suggested PCG-CNN is eventually deployability-oriented, at the moment, the integration of both into an active SDN controller is not considered.

Research Challenges

This research had a number of methodological challenges. One of the main challenges was the Feature Leakage Identification, in which determining whether certain traffic patterns were really informative or not, and some attributes are due to artificial cheating, is really hard to detect without a strict forensic examination. Subsequently, Traffic Augmentation became a significant challenge; it was not easy to maintain the realistic intercorrelations of the generated traffic without adding synthetic noise. In addition, Class Imbalance management played a crucial role, as typical deep

learning models are likely to favor the majority classes, and this should be optimized to ensure that minority attack vectors receive meaningful measures. The most subtle problem was probably the Accuracy-Generalization Paradox: explaining why a decrease in raw accuracy caused by the elimination of leaked features actually signified better model robustness and generalization, rather than worse model performance. Lastly, designing a just Zero-Day Evaluation system that would reliably estimate unobservable attacks without selection bias was also a complex methodological challenge.

Chapter 2

Literature Review

2.1 Preliminaries

Software Defined Networking (SDN) is an architectural change of paradigm, the separation between the control plane (decision-making) and the data plane (packet forwarding), which allows the centralization of programmable functionality and dynamic control of the network. The SDN controller can observe the overall picture of the network and install forwarding policy proactively with such protocols as OpenFlow. Nonetheless, such centralized logic brings a single point of failure and makes the controller a valuable goal to interfere with [16]. This weakness is often used through DDoS (Distributed Denial of Service) attacks, in particular TCP SYN floods. These kinds of attacks are extremely destructive in an SDN setup because of the reactive forwarding model; switches that are sent spoofed packets with no appropriate flow rules will flood the controller with messages telling them to send packets. This saturation wastes computational resources, it drains the small tables of TCAM (Ternary Content-Addressable Memory) flow tables, and results in paralysis within a network in general [13].
High-speed networks need to use Flow-based Traffic Analysis as opposed to Deep Packet Inspection (DPI) which is computationally expensive to detect such intrusion. Systems can compute meta-features such as packet rates and inter-arrival times by summarizing the data in packets into statistical flow records, parameters that are defined by a 5-tuple (Source IP, Destination IP, Source Port, Destination Port, Protocol), and without losing privacy [30].Nevertheless, this method is highly reliant on the adequate choice of features and prevention of the so-called Feature Bias. When a model is based purely on particular protocol flags (e.g., TCP SYN) instead of behavioral correlations, it is susceptible to the so-called Shortcut Learning, and it trains well, but when applied in the real-world setting, it does not apply at all [28]. Addressing these challenges requires robust detection mechanisms capable of identifying both known and "Zero-Day" attacks—new threat patterns absent from training data. While signature-based systems fail here, Anomaly Detection succeeds by modeling the statistical probability of normal behavior and flagging significant deviations [24]. The Deep Learning (DL) models have been shown to be effective in this field, specifically Convolutional Neural Networks (CNNs). The systems can independently learn sophisticated, non-linear spatial or temporal trends by using 1D-CNNs on packet sequences or flow statistics without any feature engineering [23].
In addition, to increase the detection reliability, unsupervised models like Gaussian

Mixture Models (GMMs) are used to capture the probability density of normal traffic, and the strict limits of a supervised classifier are not required. One of the most promising developments is the so-called Hybrid Learning or Wide and Deep paradigm, that is, a combination of two paradigms: a deep component (e.g., a CNN) to match complex patterns and generalize them, and a wide component (e.g., a GMM) to memorize normal behavior. This composition makes the system identify the known attacks with a high degree of accuracy and at the same time to identify the zero-day anomalies which statistically do not fit well in the normalcy [17].

2.2 Review of Existing Research

Early research on DDoS detection in SDN had concentrated on controller-based and statistical techniques of monitoring thresholds. Kreutz et al. [16] also pointed out the elemental security vulnerability of SDN by saying that intelligent detection is to be implemented to protect the centralized control plane. However, these primitive methods were generally reactive and could not differentiate between high rate legitimate traffic and real attack.
Subsequent efforts employed supervised machine learning techniques that include Support Vector Machines (SVMs) and k-Nearest Neighbors (KNNs) to identify flow-based DDoS attacks in SDN networks [30]. Although these methods have high accuracy, they were analyzed in terms of their precision and recall; most of them had datasets with limited diversity of attack and high-distinguishing features. Various papers have proposed deep learning models for intrusion detection, which have shown better feature representation [18]. Other architectures, such as convolutional neural networks (CNNs) [26], have demonstrated potential in non-linear traffic patterns. These models, however, tend to be very expensive to compute and are often closed-world trained, which limits their ability to adapt to new traffic patterns and counter threat vectors not seen by the model.
The validity of reported findings in DDoS detection research has been questioned increasingly in the literature. Researchers note that class imbalance, feature combinations, and the limited scope of attacks are critical sources of bias in evaluation metrics [15]. Precision and recall are shown to be inadequate when accuracy is used on its own, especially when class distributions are skewed, where recall and precision differ, creating a false sense of model robustness.

Learning paradigms: Multiple learning paradigms have been proposed to enhance robustness and generalization by using hybrid detection [25], [38]. Despite these promising entries, current literature including [35] tends to use complicated chain architectures which have not been optimized to the SDN-specific contexts. Furthermore, the gap in research, namely the intersection of hybrid detection and Explainable AI (XAI) is high because identifying attacks is not the most important goal; furthermore, it offers forensic transparency in relation to zero-day attacks. This paper aims at addressing these inefficiencies by creating an open hybrid system.
In order to support the systematic comparison of the state of the art, the main recent hybrid strategies in SDN security are summarized in 2.1, with references to their particular methodologies, and the most significant limitations that create the need to introduce the proposed architecture.

Table 2.1: Summary table of Existing Hybrid and Deep Learning Approaches in SDN Security

Study (Author & Ref)	**Methodology**	**Limitation (Research Gap)**
Idhammad et al. (2018) [25]	Semi-supervised Machine Learning	Relies on entropy and co-clustering; less effective against complex, mimicking attack patterns compared to Deep Learning.
Wang et al. (2018) [26]	1D-CNN with Flow-based Features	Uses CNN for spatial feature extraction but operates as a Closed-Set classifier (lacks zero-day rejection capability).
Xu et al. (2019) [30]	Flow-based ML (Stacking)	Demonstrates good accuracy but is heavily dependent on manual feature engineering and feature selection.
Maray et al. (2023) [35]	Optimal Deep Learning (CNN+LSTM)	Focuses on optimization in IoT-SDN; high computational overhead due to complex sequential processing.
Alshamrani et al. (2024) [38]	Hybrid Deep Learning IDS	Recent hybrid approach; acts as a "Black Box" model, lacking granular forensic explainability for network admins.
Proposed Model	**GMM + Res-CNN + Attention**	**Introduces Statistical Injection (GMM) for zero-day detection and Explainability via Feature Attention.**

2.3 Research Gaps

Although the current methods of ML-based SDN security have progressed, a critical examination of the literature has shown that the literature has some serious methodological and architectural loopholes that are filled in the study:

- **The Data Validity & Shortcut Learning Crisis:** The vast majority of the extant literature uses benchmark datasets (like CIC-IDS or NSL-KDD) and does not consider the issue of Identity Bias The models which are trained using them tend to memorize particular identifiers (IP addresses, MAC IDs) instead of learning the dynamic behavior of the attack. The literature that is devoted to Forensic Cleaning is insufficient, and depersonalizing identities in order to cause the model to learn the overall traffic physics instead of having to memorize the victim names is lacking.

- **Lack of Open-Set "Zero-Day" Robustness:** Most of the existing detection systems are Closed-Set Classifiers. They have no alternative but to put all inputs in either category, called Benign or a known Attack. It is also clearly lacking in architectures that adopt a "Rejection Mechanism" or a statistical

gatekeeper that can mark out unknown or zero-day anomalies that do not fit in any of the pre-trained categories.

- **Architectural Redundancy in Hybrid Models:** Although the hybrid models exist, they are mainly based on the Sequential Stacking(e.g., CNN → LSTM). The given method introduces a feature bottleneck, with the second model potentially constrained by the result of the former, frequently missing the raw statistical context of the traffic. The area of research of Parallel "Wide and Deep" Architectures in SDN, in which Deep Learning (Pattern Recognition) and Probabilistic Modeling (Density Estimation) can be applied independently, is unexplored to cross-verify decisions.

- **Absence of Forensic Explainability:** Lack of Forensic Explainability: The majority of Deep Learning solutions in SDN are used as black boxes and do not generate any explanations of the classification process. Network administrators require granular forensic details to justify mitigation actions. Current literature lacks frameworks that validate decision boundaries using manifold visualization techniques (like t-SNE) to prove that the model is learning valid behavioral clusters rather than noise.

2.4 Summary of Key Findings

In this section, the key findings of a thorough analysis of the available literature on DDoS detection in Software-Defined Networks are synthesized. The results identify recurring limitations, gaps in the research, and methodological issues that directly influence the design and objectives of the proposed research.

1. **SDN Control Plane Vulnerability**

   The literature consistently observes that the centralized control plane of Software Defined Networks (SDN), though useful for simplified network management and programmability, creates a serious vulnerability. The controller is one of the primary victims of Distributed Denial-of-Service (DDoS) attacks, especially those based on control-data plane interactions, due to the limited computational and memory resources that it has.

2. **Performance and weaknesses of Flow-Based machine learning**

   Flow-based machine learning is considered a popular, scalable, and controllable method for DDoS detection in SDN. Nonetheless, their performance is highly dependent on the selection of features and the quality of the datasets. Inappropriately selected features or an unbalanced data distribution heavily impair the model's reliability in real-world conditions.

3. **Problems of DatasetBias in Features**

   Another theme noted throughout the literature is that the validity of data sets significantly impacts reported model performance. Several experiments show that, due to feature leakage, biased data sharing, and artificial or overly simplified data, artificially inflated accuracy values are produced. This type of performance is hardly ever converted into strong performance in the real world.

4. **Metric Inflation and Class Imbalance**

   Several authors point out that high overall accuracy tends to mask extreme imbalance in precision, recall, and minority-class performance. This effect, known as metric inflation, leads to models that are effective on large traffic classes but fail to correctly identify unusual or changing attack behaviour.

5. **Inadequate Zero-Day Attack Testing**

   Although the idea of detection of Zero-Day attacks is becoming more popular, the current body of research is mainly based on closed-set evaluation procedures. The methods presuppose prior awareness of all types of attacks that do not lie in the open and the dynamic character of real-life network threats. As a result, most of the suggested systems have low detection rates for new attack patterns.

6. **Rational behind the proposed Methodology**

   Taken together, these results encourage the methodological decisions of this study. A special focus is placed on strict feature-level validation, realistic and correlated traffic engineering, balanced multiclass modeling, and the design of a new hybrid framework that can identify both known and Zero-Day attacks in SDN architectures.

Chapter 3

Requirements, Impacts and Constraints

3.1 Final Specifications and Requirements

This paper is firmly rooted on specifications of technical and methodological requirements to attain realistic assessment, scientific validity, and reproducibility.

Methodological Specifications:
As a research, the study is a quantitative, data-based experiment. The main criterion was the validity of datasets and integrity of features. The system is defined to work with only the flow-level network traffic characteristics (e.g. the number of packets, inter-arrival times) and not deep packet payloads to guarantee scalability and privacy compliance. Moreover, the approach requires combination of feature correlation analysis and statistical class balancing (SMOTE) in order to counter the bias which is usually present in conventional security benchmarks.

Technical and Resource specifications:
Software wise, the implementation was performed in a Python based ecosystem. The data preprocessing and classical base models (Support Vector Machines (SVM) and k-Nearest Neighbors (KNN)) were implemented using the Scikit-learn. A high-performance data processing was performed using NumPy and Pandas, whereas Deep Learning ingredients (1D-CNN) were deployed with the help of TensorFlow/Keras. Controlled offline implementation of the experiments was done to achieve consistent benchmarking. The PCG-CNN architecture proposed is efficient as opposed to large language models (LLM). Thus, the moderate financial means of computation (e.g., typical GPU-accelerated setups) were needed to be trained and evaluated, which proved that the model could be deployed to lightweight without the use of resource-intensive infrastructure.

3.2 Societal Impact

DDoS attacks are no longer a mere inconvenience, but weaponized forms of attack that can shut down major digital infrastructure, both financial systems and healthcare platforms. With the growing trend of using modern networks based on Software-Defined architecture (SDN) to enable 5G and Cloud computing, the centralized

Control Plane is becoming a single point of failure. This study plays a direct role in enhancing the stability and sustainability of intrusion detection based on SDN by making them more realistic and robust, and in this way, the digital backbone of society will stay unaffected by the shocks.
Moreover, this paper deals with the issue of Future-Proofing network security, which is very urgent. Zero-Day attack detection features are not only adequate in keeping organizations safe, but in keeping up with the ever-changing, invisible cyber-weapons of the future. Finally, regarding a privacy, ethical segment, the proposed system will be flow-based in nature where it will have the ability to analyze only metadata (ex: timing, volume) of traffic but not the content of the user. The approach shows that the high-security level can be attained without infringing on the privacy of the users and without the invasive Deep Packet Inspection (DPI) and corresponds to the responsible development of the information sharing along with the responsible data usage related to the responsible practices.

3.3 Environmental Impact

The detection framework suggested is computationally sustainable. The system can utilize lightweight flow-level functionality and an efficient hybrid architecture to make it much more computationally lightweight than resource-intensive Deep Learning ensembles that require specialized hardware acceleration (e.g., high-end GPUs) and lengthy training times.
In addition, successful DDoS protection is directly associated with environmental protection. Long network attacks stress network infrastructure to work at peak capacity over long periods, which provokes inefficient resource utilization, thermal limiting, and enormous increases in power used to cool and process data [19]. The framework that will be proposed can reduce this amount of energy waste by neutralizing these threats quickly and so network security objectives are coupled with the principles of Green Computing.

3.4 Ethical Issues

Ethical data handling and privacy procedures were followed in the course of conducting this research. Each dataset used was obtained through a publicly available repository or created through statistical augmentation, and the Personally Identifiable Information (PII) was fully absent. Moreover, the detection system uses a Privacy-by-Design model; by only analyzing flow-level metadata (e.g., packet timing, volume) and not using Deep Packet Inspection (DPI), the system would not compromise user confidentiality, although the system does not compromise the effectiveness of security.
In terms of professional responsibility, the research in this paper has given a priority to scientific integrity rather than metric inflation. An ethical decision was carried out in awareness of reporting the honest performance measures that the elimination of leaked features leads to reduced raw accuracy and increased realism. The method confronts the norm of paper-perfect outcomes in the discipline, and is consistent with accepted ethical standards of transparency and reproducibility of cybersecurity research [11].

3.5 Project Management Plan

The study was implemented as a systematically developed, recursive development process cycle. The use of a so-called Forensic Feedback Loop through which the validity of data was quantitatively measured before and after the procedure of augmentation was made into a defining feature of this workflow to find out that such a type of bias should be eliminated.

- **Phase I: Elective and Problem Definition**

  The first stage was aimed at specifying the scope of the research. These included a wide literature review in order to find out the phenomenon of Metric Inflation in SDN security studies. This stage came to a close with the purchase of the original reference data and the production of baseline models to determine the level of performance.

- **Phase II: First Forensic Audit (Feature Leakage Identification)**

  Before any such amendment, a Feature Forensic Analysis of the original data set had been carried out. Importance of features was estimated by use of tree-based estimators. This analysis affirmed this hypothesis: certain attributes (e.g., IP addresses and explicit TCP flags) acted as highly discriminative artifacts, which have an unproportionately large information gain and enable the models to attain near-perfect accuracy without learning dynamics in behavior.

- **Phase III: Augmentation and Forensic validation**

  Statistical Data Augmentation was used to correct the bias found. More importantly, an augmented dataset was subjected to secondary forensic analysis. The findings were indicative of the fact that the augmentation process was effective in terms of the dilution of the impact of the particular identifiers; the features themselves were still present, but the relative information gain became much smaller, which meant that the model needed to be based on the distributed flow statistics, as opposed to individual artifacts to classify them.

- **Phase IV: Hybrid Architecture and Novelty Development**

  This stage was concerned with the construction of the PCG-CNN, which is not similar to the standard ensembles. A Novel Conditional Forensic Logic was created as opposed to the current hybrids which merely average predictions. This involved:

  - **Deep Learning Stream:** Deep learning implementation of 1D-CNN to get accurate specifics on known attacks.
  - **Probabilistic Stream:** GMM used to map the normal density.
  - **Forensic Logic Integration:** An integration of the decision layer that ranks the probability of the GMM score higher than the label of the CNN will allow the system to override false classifications of unseen threats.

- **Phase V: Zero-Day Assessment and Stress Testing**

  This phase involved the transfer of attention to the rigorous validation. They put the Forensic Engine through Zero-Day (security scenarios not included

in training) to ensure that the conditional logic operated correctly to identify them as anomalies, but not to falsely report benign results.

- **Phase VI: Assessment and reporting**

  The last step tested the full Forensic Engine to be resistant to undetected intruders. These findings were documented in this thesis, which can now be seen as evidence of the successful implementation of the project, which took place, with a clear passage to a biased dataset to a robust and forensic-level detection system.

3.6 Risk Management

There are various critical risks identified during the project lifecycle which were mitigated systematically to make sure of the methodological rigor as well as practical feasibility.

Data and Model Validity Risks: The main risk, which was identified, is the bias in data, which can make the model learn the features of certain identifiers instead of behaviors. This was alleviated by the strict Forensic Feedback Loop that entailed the choice of features by correspondence and elimination of discriminating artifacts. Close to it was the danger of overfitting, which is widespread in deep learning applications. This was dealt with by introducing traffic diversity by means of statistical augmentation and by making sure that the model is not merely memorizing the training data and trying to apply it to the unseen Zero-Day attack vectors.

Evaluation and Operational Risks: The major methodological threat was the use of misinformed evaluation measures, namely, the paradox of the Accuracy in an unbalanced data. In response, the project required balancing the classes and instead of considering raw accuracy, Precision, Recall, and F1-Scores were put in the limelight. Lastly, the threat of implementation complexity with the model being too heavy to be practically deployed using SDN was mitigated by the choice of a streamlined hybrid architecture (1D-CNN with GMM). This design compromise had made high-performance detection and manageable computational overhead to keep the system within the resources of control planes.

3.7 Economic Analysis

Economically, the financial consequences of the DDoS attacks are significant, and they include loss of direct revenue due to service interruption, loss of reputation and recovery costs after the incident. The suggested detection system offers a cost-effective challenge to the expensive proprietary security devices by employing open-source architecture and tuned to run on small computational resources.
Moreover, the long-term operational risk is alleviated by the integration of Zero-Day attack detection to a large extent. The system eliminates the nightmarish financial burden that comes with extended Zero-Day attacks by preemptively eliminating the risks. As a result, the solution has a positive cost-benefit ratio, especially where

the security implementations in Software-Defined Networks need to be scalable and cost-effective [29].

Chapter 4

Proposed Methodology

4.1 Design Process and Methodology Overview

This study has a **Forensic-Driven Progressive Refinement Strategy** as its methodology. In contrast to current machine learning methods in cybersecurity, which frequently assume datasets to be an established truth [9], this study considers the dataset as a possible source of bias and it has to be forensically audited to be modeled. It was not a linear process; its design proceeded in specific steps of exploration and amendment to make the model comprehend network traffic structural dynamics instead of being memorized [28].

Figure 4.1 illustrates the iterative forensic design workflow adopted in this study. The developed methodology follows the following chronological steps:

1. **Forensic Baseline Assessment:** A stage of diagnosis during which standard classifiers (KNN, SVM) showed impractical accuracy and a forensic investigation discovered and eliminated "Feature Leakage" (e.g., IP addresses, MACs).

2. **Correlated Multiclass Augmentation:** Application of a domain-specific data engineering stage to the augmentation engine based on Topology. This modeled 12 different traffic types by following the mathematical relationships between traffic flow dynamics.

3. **Secondary Forensic Audit and Correlation Cleaning:** This is a re-evaluation step that follows augmentation. Redundancy (collinearity) was detected through a Pearson Correlation Matrix and eliminated in order to make the model dependent only on the unique behavioral kinetics.

4. **Final Class Balancing:** A stabilization step that guarantees that all traffic classes are balanced with the same number of samples to remove the bias related to the majority classes.

5. **Wide and Deep Hybrid:** It implements the main architecture that uses a dual-stream neural network (1D-CNN + Multi-Head Attention + GMM).

6. **Integrated Zero-Day Forensics:** The last layer of security working on a probabilistic gatekeeper with Negative Log-Likelihood (NLL) thresholds to Open-Set Recognition [8].

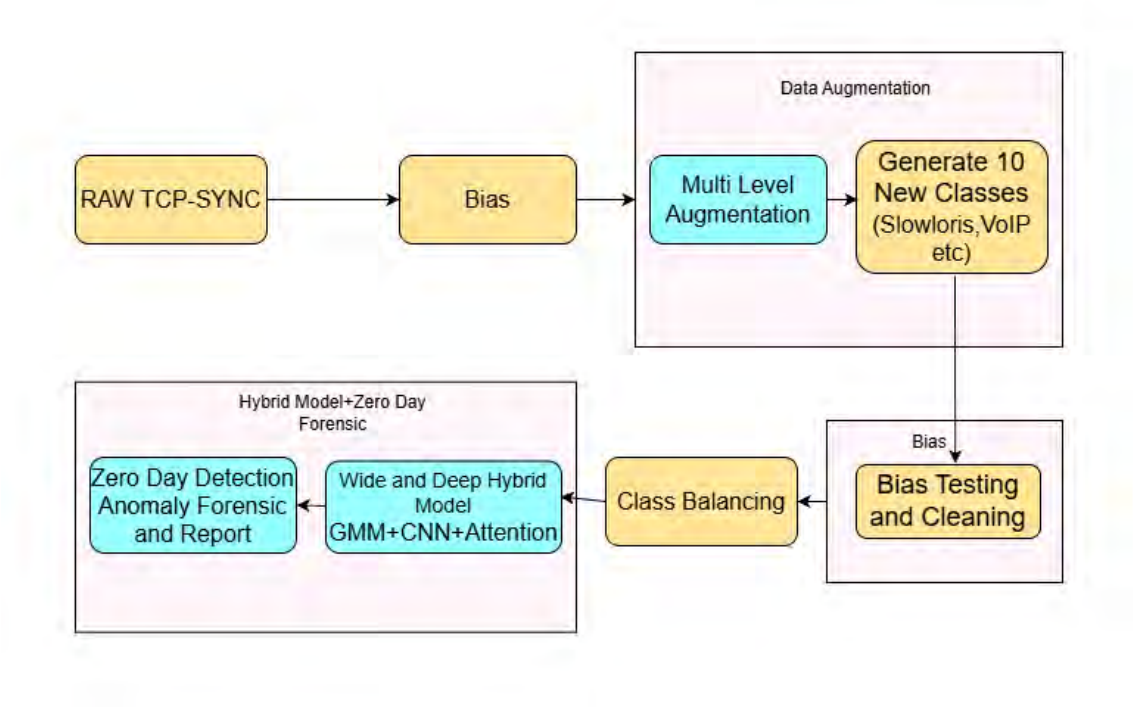


Figure 4.1: The Iterative Forensic Design Workflow

4.2 Preliminary Forensic Investigation

4.2.1 Identification of Dataset Bias (The "100%" Anomaly)

One widespread problem of research on academic IDS is "Dataset Leakage," where classifiers overfit to static identifiers rather than learning dynamic behavior [27]. The study started with the training of traditional supervised learning algorithms, i.e., k-Nearest Neighbors (KNN) and Support Vector Machine (SVM), on the raw data to provide a performance benchmark.

Observation: These baseline models had 100% Accuracy, Precision, and Recall.

Hypothesis: This ideal uniform performance is not theoretically likely in real network traffic. This was an indication of Data Leakage, whereby the models were memorizing certain identifiers and not training on attack behaviour.

Table 4.1: Forensic Feature Filtering Protocol

Identified Problem	**Proposed Solution**	**Technical Mechanism**
Identity Bias: Model memorizing attacker IPs instead of learning attack behavior.	Forensic Feature Removal	Explicit removal of Source/Dest IP, MAC Address, Flow IDs, and Timestamps.

Identified Problem	Proposed Solution	Technical Mechanism
Class Imbalance: Rare attacks (e.g., Slowloris) are statistically ignored by the model.	Manifold Augmentation	SMOTE (Synthetic Minority Over-sampling Technique) using k-Nearest Neighbors interpolation.
Vanishing Gradients: Neural network fails to learn effectively from raw, unscaled data.	Numerical Stabilization	Z-Score Standardization ($\mu = 0, \sigma = 1$) combined with Median Imputation for missing values.
Feature Redundancy: Multiple features (e.g., Total Packets vs. Subflow Packets) contain identical information.	Correlation Cleaning	Removal of features with Pearson Correlation Coefficient $r > 0.95$.
Complex Patterns: Simple models (KNN/SVM) miss subtle, non-linear variations in DDoS traffic.	Spatial Feature Extraction	1D-CNN (Convolutional Neural Network) enhanced with Multi-Head Attention.
Zero-Day Attacks: Standard classifiers misclassify unknown/new attacks as "Benign" with high confidence.	Open-Set Recognition	Gaussian Mixture Model (GMM) density estimation with Negative Log-Likelihood (NLL) thresholding.

A stringent Forensic Filtering Protocol 4.1 as applied to ensure model generalization to real-world Software-Defined Networks (SDN). This systematic approach addresses six major vulnerability categories that compromise detection integrity. Through decision weight analysis, the following feature sets were explicitly identified as Leakage-Prone:

- **Static Identifiers:** The Static Identifiers include the Source IP, Destination IP, MAC Address, and Flow ID. These characteristics require the model to be tied to a certain topology of the lab, which cannot be utilized elsewhere.
- **Contextual Artifacts:** Timestamp and Port Number. These features in deep learning make the model learn *when* an attack occurs and not *how* it occurs.
- **Protocol Flags:** Protocol flags (e.g., TCP SYN Flag) that were being used as direct labels of particular attacks.

4.3 Data Engineering and Refinement

4.3.1 Correlated Multiclass Augmentation

After eliminating the bias, the dataset was statistically augmented. This study, unlike other conventional oversampling methods (e.g., SMOTE), can be said to be blind. In contrast, this framework is a **Physics-Based Multiclass Augmentation Framework** that obeys the mathematical laws of network transmission as illustrated in Figure 4.2.

4.3.1.1 Extraction of Statistical Reference Profiles p_{95}

We extracted Statistical Reference Profiles of legitimate traffic and computed the 95^{th} Percentile (p_{95}) and Median (p_{50}) of significant characteristics such as Flow Duration and Byte Volume.
The reason why p_{95} is used in two applications:

1. **Outlier Rejection:** It establishes a ceiling parameter that can best represent the conditions of heavy traffic but leaves out extreme outliers that occur infrequently and may tend to skew the synthetic generation mechanism.

2. **Operational Realism:** It gives a statistically significant upper bound on the synthetic data it can generate, ensuring that operational limits on augmented samples are not surpassed by generated flows (e.g., that a generated flow has no throughput constraints).

4.3.1.2 Correlation-Preserving Generative Formulas

All synthetic flows are produced by imposing the mathematical relationships, followed by actual TCP/IP communications. In Equation 4.1, this retains the internal Pearson Correlation Matrix of the data:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \tag{4.1}$$

The value of the covariance is denoted by the following formula: upon division of the covariance by the product of the standard deviations of X and Y, the result gives the correlation coefficient.
The deterministic dependencies used in the generation process are:

- **Packet Rate Consistency:** Equation 4.2 describes that Packet Rate (R_p) is strictly defined as the percentage of the Packet Count (P) over Flow Duration (D):

$$\text{Packet Rate} = \frac{\text{Total Packets}}{\text{Flow Duration}} \tag{4.2}$$

- **Volume Integrity Law:** Equation 4.3 describes that the total Bytes, denoted as B, are calculated from the number of packets and the average packet size, denoted as S_{avg}:

$$\text{Total Bytes} = \text{Total Packets} \times \text{Average Packet Size} \tag{4.3}$$

- **Bidirectional Symmetry:** TCP connections are dichotomous. In our augmentation, the Total Packets are split into Forward (P_{fwd}) and Backward (P_{bwd}) counts by probabilistically distributing them with a single parameter, the handshake and acknowledgement rate, α, as discribed in Equation 4.4:

$$P_{\text{bwd}} = P_{\text{fwd}} \times \alpha \tag{4.4}$$

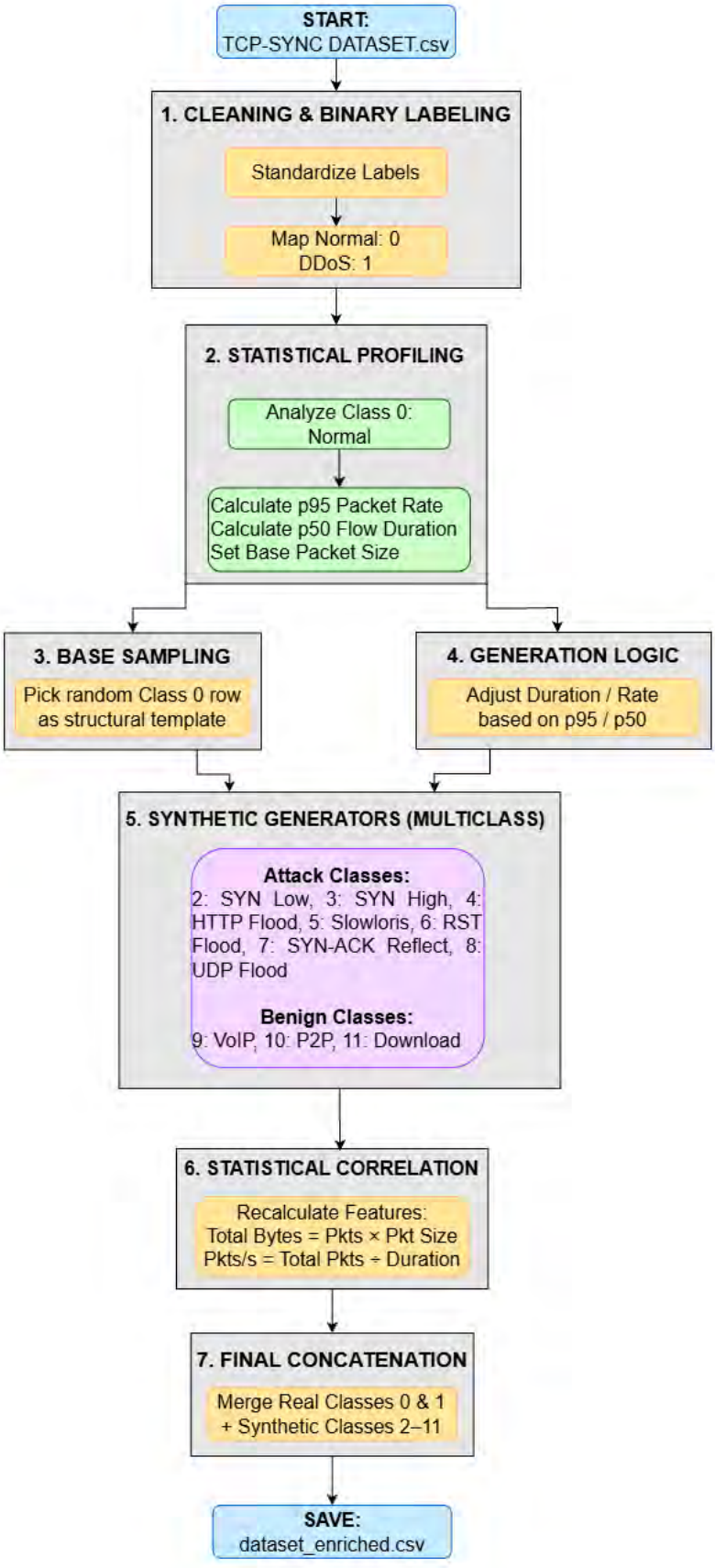


Figure 4.2: The Correlation-Preserving Augmentation Process

4.3.1.3 The 12-Class Multiclass Formulation

To obtain fine-grained behaviors, the dataset was increased through synthesis of granular traffic variants, and this led to a final taxonomy of **12 particular classes**.

- **Real Data:** realnormal (Class 0), realddos (Class 1).
- **Synthetic Malicious:** synlow (2), synhigh (3), httpflood (4), slowloris (5), rstflood (6), synackreflect (7), udpflood (8).
- **Synthetic Benign:** voip (9), p2pstream (10), downloadbenign (11).

Observation: Benign classes such as P2p streaming, had to be generated specifically because such classes are statistically similar to DDoS attacks (high volume), and the model should be trained to reject these ones.

4.3.1.4 Justification of Statistical Augmentation Parameters

Flow based features which include packet rates, number of bytes per flow and flow duration are viewed in the statistical method of network intrusion detection as legitimate indicators of a problematic traffic [31]. Kinetic vectors of legitimate and malicious behaviors have different statistical marks. These particular characteristics allow one to simulate properly the attack signatures found in the academic literature and still retain operational realism by mathematically manipulating these characteristics like intentionally increasing packet rates or stretching flow durations [4].
Table4.2 systematically documents the augmentation strategies employed for each of the 12 traffic classes, providing scientific justification grounded in established research literature. Each synthetic traffic class is designed to capture realistic attack signatures while maintaining operational validity.

Table 4.2: Augmentation Strategy and Scientific Validity for 12 Traffic Classes

Class ID	**Traffic Class**	**Augmentation Strategy**	**Scientific Validity & Justification**
0	Real Normal	None (Original Dataset)	Ground truth baseline for benign behavior [31].
1	Real DDoS	None (Original Dataset)	Ground truth baseline for malicious volumetric traffic [31].
2	SYN-Low	Variable Rate Injection: SYN flows with packet rates $1\times$ to $3\times$ of $p95$.	Attackers vary the agent rate to delay or avoid detection [4].
3	SYN-High	Pulse/Surge Injection: SYN flows with $8\times$ packet rates and pulse patterns.	Pulsed or variable rate floods maximize resource exhaustion while evading fixed thresholds [4].

Class ID	Traffic Class	Augmentation Strategy	Scientific Validity & Justification
4	HTTP-Flood	Volume Boosting: High Total Bytes and Packet Count with valid TCP flags.	HTTP floods use valid requests at normal rates to overrun resources, making them difficult to differentiate from real traffic [32]
5	Slowloris	Time Dilation: High Flow Duration ($\gg p95$) and near-zero Pkts/s.	Attackers send partial HTTP requests and fail to respond until the timeout, occupying all available web server sockets [33].
6	RST-Flood	Flag Injection: High Total Fwd Pkts with explicit RST flag settings.	TCP reset floods impair performance by flooding targets with massive numbers of out-of-state reset packets [37].
7	SYN-ACK Reflect	Reflection Simulation: High-volume response packets without initial SYN.	Based on variable rate attack principles where reflection is utilized to amplify traffic towards a victim [4].
8	UDP-Flood	Unidirectional Surge: Extreme Fwd Pkts/s with minimal Bwd Pkts.	Flooding source packets far exceed packets sent back by the victim target, creating a high ratio function [14].
9	VoIP (Benign)	Interval Constraint: Fixed ∼20ms intervals (∼50 pkts/s) and small payloads.	Default audio packetization intervals are typically 20ms; simulating this avoids misclassifying UDP as attack traffic [1].
10	P2P Stream	High Volume/Duration: Large transfer volumes independent of ports.	P2P traffic is characterized by distinctive transport-layer connection patterns and large byte volumes [3].
11	Download Benign	Throughput Consistency: High Flow Duration and Total Bytes.	Characterizes long-lived TCP transfers where throughput is sustained, and slow start no longer dominates [12].

4.3.2 Secondary Forensic Audit & Correlation Cleaning

A Secondary Forensic Audit was conducted after the augmentation. Now that the cheating IPs were absent, we had to make sure that the augmented information did not bring about synthetic redundancy. To ensure that there was no multicollinearity, a **Pearson Correlation Matrix** was calculated.

In line with the verification workflow, **Signature Verification** audit was done together with the correlation checks. This was a qualitative test of whether the synthesized profile of attacks followed theoretically correct attack signatures, e.g.,

making sure that synthetic flows of Slowloris displayed the typical periodic time-outs and low bandwidth consumption, and not just looked like generic traffic.

Action: Correlation characteristics with extremely high correlation values were removed explicitly (defined as a value of the correlation (ρ) greater than 0.90). As an example, *Fwd Packets/s* and *Flow Packets/s* had almost the same variance in synthetic data.

Result: The identified Resolution: This "Correlation Cleaning" dimensionality reduced and constrained the model to depend on distinctive variance in the flow kinetics as opposed to recurring signals [28].
The comprehensive feature elimination process is documented in 4.3, which categorizes dropped features by their forensic justification. Features were removed based on four primary criteria: Identity Leakage (network-specific identifiers), Metadata Bias (temporal artifacts), Zero-Variance (no discriminative signal), and Redundant Collinearity (mathematical redundancy). This systematic pruning ensures that only behaviorally meaningful features remain for model training.

Table 4.3: Feature Elimination Justification

Dropped Feature Name	**Category**	**Forensic Justification (Reason for Removal)**
Flow ID	Identity	Metadata: Unique identifier; causes overfitting to specific sessions.
Src IP, Dst IP	Identity	Leakage: Model memorizes network topology instead of behavior.
Src Port, Dst Port	Identity	Leakage: Binds detection to specific services/ports.
Timestamp	Metadata	Bias: Introduces "time-of-day" artifacts irrelevant to attack kinetics.
Idle Mean, Std, Max, Min	Metadata	Irrelevant: Reflects network congestion/pauses, not malicious intent.
Fwd PSH Flags	Zero-Variance	No Signal: Constant value (0) across dataset.
Fwd URG Flags, Bwd URG Flags	Zero-Variance	No Signal: Constant value (0); URG flag rarely used in modern traffic.
URG, CWE, ECE Flag Counts	Zero-Variance	No Signal: Constant value across all samples.
Fwd/Bwd Byts/b Avg	Zero-Variance	No Signal: Metric calculation resulted in static values.
Fwd/Bwd Pkts/b Avg	Zero-Variance	No Signal: Metric calculation resulted in static values.
Fwd/Bwd Blk Rate Avg	Zero-Variance	No Signal: Metric calculation resulted in static values.
Init Fwd Win Byts	Zero-Variance	No Signal: Initial window size remained constant.
Fwd Seg Size Min	Zero-Variance	No Signal: Minimum segment size variance was zero.

Dropped Feature Name	Category	Forensic Justification (Reason for Removal)
TotLen Fwd Pkts	Redundant	Collinearity: High correlation with Total Fwd Pkts.
TotLen Bwd Pkts	Redundant	Collinearity: High correlation with Total Bwd Pkts.
PSH Flag Cnt	Redundant	Collinearity: High correlation with Bwd PSH Flags.
Pkt Size Avg	Redundant	Collinearity: Mathematically identical to Pkt Len Mean.
Fwd Seg Size Avg	Redundant	Collinearity: High correlation with Fwd Pkt Len Mean.
Bwd Seg Size Avg	Redundant	Collinearity: High correlation with Bwd Pkt Len Mean.
Subflow Fwd Pkts	Redundant	Collinearity: High correlation with Bwd Header Len.
Subflow Bwd Pkts	Redundant	Collinearity: High correlation with Bwd Header Len.
Subflow Bwd Byts	Redundant	Collinearity: High correlation with Bwd Header Len.
Fwd Act Data Pkts	Redundant	Collinearity: High correlation with Bwd Header Len.

The comprehensive feature elimination process is documented in 4.3, which categorizes dropped features by their forensic justification. Features were removed based on four primary criteria: Identity Leakage (network-specific identifiers), Metadata Bias (temporal artifacts), Zero-Variance (no discriminative signal), and Redundant Collinearity (mathematical redundancy). This systematic pruning ensures that only behaviorally meaningful features remain for model training.

Class Balancing was the last preprocessing step and has a clean and non-redundant feature set. The data was resampled in such a way that the numbers of samples within each of the classes comprising the predominant traffic- "Benign" and the rare ones - "Slowloris" were the same. This made sure that the end model would be zero-biased against the majority class.

4.4 Data Preprocessing and Transformation

In order to achieve mathematical sanity of the deep learning model, raw network values need to be mathematically sanitized [10].

- **Infinite values (Rate Limiting):** Some of the features that are used in network traffic flow analysis include Flow Packets/s, and they are calculated in the form of Equation 4.5

$$\text{rate} = \frac{\text{Count}}{\text{Duration}} \tag{4.5}$$

  In the case of $Duration = 0$, the rate tends to be infinity (inf) and this leads to gradient explosion in neural networks. We use a Clamping Function to restrict such values to the median and three standard deviations ($\mu + 3\sigma$).

- **Median Imputation Strategy:** The missing values (NaN) are treated with Median Imputation instead of Mean Imputation. The network traffic data is skewed by huge DDoS outliers to the mean because the data is heavy-tailed (power law). The median gives central tendency to IDS data with high strength [28].

- **Z-Score Standardization:** Lastly, to aid the convergence of the 1D-CNN, all features are normalized to a standard scale with a zero mean and unit variance [10], as discribed in Equation 4.6:

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{4.6}$$

4.5 Wide and Deep Hybrid Model Architecture

The proposed **Wide and Deep Hybrid Model** is a cognitive model of human decision making that processes information by two parallel streams, one of which is a **Deep stream** to analyze the complex spatial information, and the other one is a **Wide stream** to acquire the broad statistical classification [17].

4.5.1 Advantages over Current Models

Conventional models are usually susceptible to the problem of *Shortcut Learning* in which they learn only the raw feature values or the memorization of a fixed pattern (such as IP/MAC headers) of the dynamic SDN environment, which results in high error rates. Current 1D-CNN or SVM systems are closed-world systems; they are able to classify what they have observed. Our hybrid architecture is the solution to this by integrating Structural Learning (Deep Path) and Probabilistic Grounding (Wide Path) as presented in Figure 4.3. This duality enables the model both to classify known attacks with a very high degree of precision, as well as to identify statistical outliers that are crucial to the forensic detection of Zero-Day attacks.

4.5.2 The Deep Path (Spatial Stream)

The Deep Path uses a non-linear, local, and global feature extraction to extract complex patterns in the raw sequence of features.

- **1D-CNN Layer:** Why 1D-CNN? The flows of the network traffic are sequential in nature. In contrast to 2D-CNNs, which are optimized on image data, 1D-CNN is optimized on fixed-length vectors of features where the order of features (e.g., packet rate followed by flow length) is a kinetic signature. It is computationally more efficient than RNNs and more effective at localized short-range dependencies in space that are more localized between adjacent features constituting an attack fingerprint [26].

- **Multi-Head Attention:** CNNs are limited because they can only capture local patterns, but not global relations among features that are far apart. The Multi-Head Attention mechanism is the center of the architecture since it computes the weighted significance of various features irrespective of their

location. In Eqution 4.7, it splits the CNN output into Query (Q), Key (K), and Value (V) matrices using 4 heads:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{d_k}\right)V \tag{4.7}$$

What it does: This enables the model to pay attention to critical combinations, such as recognizing that a high Packet Size is only malicious when correlated with very low Flow Duration. This eliminates the chance of the model making biased decisions depending on a single feature.

4.5.3 The Wide Path (Statistical Stream)

The Wide Path offers a sound mathematical foundation to ensure that the Deep Path does not overfit the noisy data.

- **Gaussian Mixture Model (GMM):**

  Why GMM? In comparison to K-Means, which employs hard clustering, GMM employs soft probabilistic density estimation. The model used to learn the underlying normal vs. abnormal traffic distribution is a GMM with 10 components pre-fitted to the training data [27] using the Equation 4.8:

$$p(x) = \sum_{k=1}^{10} \pi_k \mathcal{N}(x \mid \mu_k, \Sigma_k) \tag{4.8}$$

- **It's benefit:** Conventional Deep Learning models have no reason to doubt they will falsely classify a Zero-Day attack into a known target. Statistical Awareness is achieved by inserting GMM probability vectors into the fusion layer of our model. When a flow is unlikely to be a part of an existing cluster (large Negative Log-Likelihood), the Wide path informs the model that this is an outlier. This *Statistical Injection* provides Zero-Day forensic functionality, making the model more robust than conventional standalone classifiers.

4.5.4 Information Fusion and Classification

The final phase of the architecture is a Concatenation Layer of the high-dimensional spatial representations of the Deep path with the 10-dimensional probability distribution of the Wide path. This interlaced representation is then passed through the Global Average Pooling layer and Softmax output layer, so that the final classification output is determined not only by what the attack looks like (CNN-Attention) but also how probable it is statistically (GMM).

4.6 Integrated Zero-Day Forensics

The last phase was the implementation of the **Zero-Day Logic**. Zero-day attacks are unknown vulnerabilities that do not have a ready-made signature.

4.6.1 Probabilistic Anomaly Scoring via NLL

To measure the extent of unknownness of a traffic flow, we use the Negative Log-Likelihood (NLL) of the GMM stream [31] using the Equation 4.9:

$$A(x) = -\ln(p(x)) \tag{4.9}$$

High-density flows (low $A(x)$) correspond to legitimate traffic, whereas Zero-Day attacks are represented by low-density flows (high $A(x)$).

4.6.2 Dynamic Thresholding and Decision Logic

A dynamic threshold, denoted as τ, is computed based on the 95th percentile (p_{95}) of the benign anomaly scores. There is a strictly hierarchical reasoning:

- **Zero-Day Check:** When $A(x)$ exceeds the threshold, the flow is tagged as a Zero-Day Anomaly.
- **Multiclass Classification:** When $A(x) < \tau$, the flow is classified by the Wide and Deep Hybrid model as one of the 12 known classes.

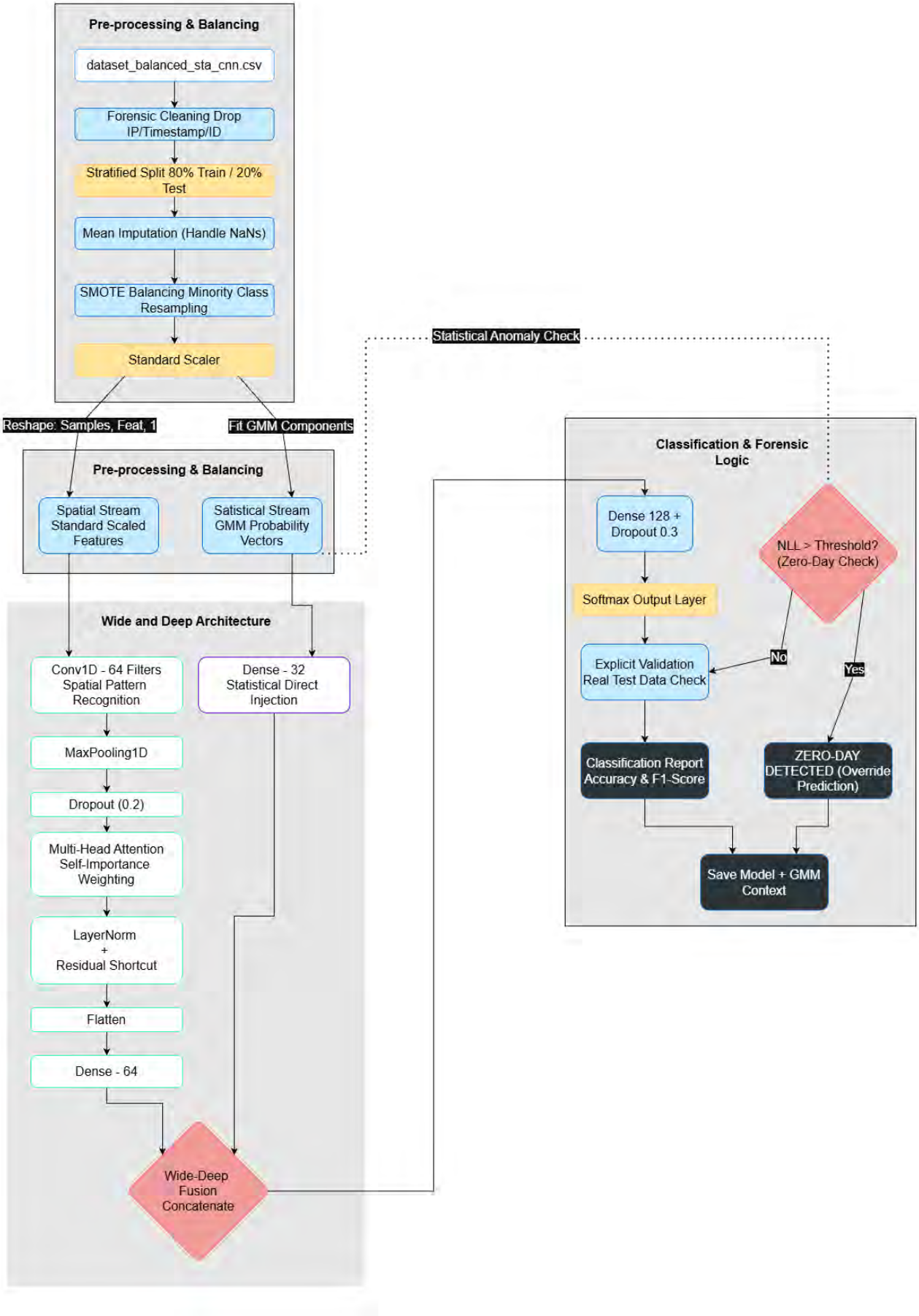


Figure 4.3: Detection Logic of Hybrid Model with Zero-Day Pipeline

4.7 Implementation of Selected Design

It was installed through **Python 3.10** and deployed the deep learning components of the TensorFlow/Keras framework and statistical modeling and data preprocessing functions of **Scikit-learn**. The implementation logic is strict Data-to-Decision forensic pipeline that guarantees that all the transformations done to the training data is also repeated in inference.

4.7.1 Development Environment & Dependencies

The model was created and tested in a supercomputing center to meet the computing requirements of the hybrid architecture. The fundamental dependencies and their roles have been listed below:

- **TensorFlow (Keras API):** Is used in constructing the Residual 1D-CNN and Multi-Head Attention layers (the "Deep" path).
- **Scikit-learn:** It is used to perform the Gaussian Mixture Model (GMM), SMOTE augmentation, and metric evaluation (Precision/Recall/F1).
- **Pandas & NumPy:** Used to process the data manipulation of vectors and dataframes with high-speed during the forensic cleaning stage.
- **Imbalanced-learn:** This is utilized to implement the SMOTE (Synthetic Minority Over-sampling Technique) algorithm in balancing classes.
- **Matplotlib and Seaborn:** It was used to create the forensic visualizations, such as the histograms of the NLL distributions and confusion matrices.

4.7.2 Forensic Data Pipeline

In order to avoid the occurrence of Garbage In, Garbage Out, Identity Bias has been bred out due to the intensive preprocessing sequence of the raw dataset before it is fed into the model.

1. **Feature Cleaning:** All of the high-cardinality identifiers that might result in overfitting were removed. This consisted of Source IP, Destination IP, Flow ID, MAC Address and Timestamp. The other characteristics are pure kinetic flow behavior (e.g., inter-arrival time, number of packets, flag rates).
2. **Sanitization:** Although the events of flow capture errors are represented by infinite values (inf) and missing values (NaN), mean imputation was used to ensure statistical continuity.
3. **Stratified Splitting:** Stratified sampling was used to divide the dataset into 80% Training and 20% Testing components to ensure that infrequent attack classes (e.g., Slowloris) would be equally represented in both of them.
4. **SMOTE Balancing:** SMOTE was used in the training set as the number of classes is highly imbalanced. This artificial augmentation made the minority class samples equal to the majority and not to learn decision boundaries in preference to one class over the other 11.

5. **Dual-Stream Scaling:** A StandardScaler was utilized with all features standardized, with a mean-value of 0, and a variance of 1. Importantly, the scaler was only applied to the training data and then on the test data to avoid data leakage.

4.7.3 Hybrid Architecture Construction

The model was constructed as a Multi-Input Neural Network in which there are two different streams of processing which come together at a fusion layer.

Stream A: The Deep Spatial Path (CNN-Attention)

- **Input Shape:** N1 × N1 vector of the spatial distribution of flow features.
- **Convolutional Layer:** This is a 1D-CNN that has 64 filters (filter size 3) which identifies the local patterns within the flow sequence.
- **Attention Mechanism:** A Multi-Head Attention layer (4 heads, key dimension 32) is used on the CNN output. This enables the model to balance out the value of certain features (e.g., giving SYN Flag Count more weight than Packet Length when there is a flood attack).
- **Residual Connection:** The output of the attention layer is connected back to its input then Layer Norm is applied to stabilize the gradient, allowing deeper learning.

Stream B: The Wide Statistical Path (GMM)

- **Probabilistic Input:** A 10-component Gaussian Mixture Model (GMM) was trained on the training data.
- **Injection:** The posterior probabilities produced by the GMM are taken as inputs of a dense layer (32 neurons). This is a type of statistical prior that has to be obeyed by the network and requires the network to be conscious of the underlying probability distribution of the traffic.

Fusion and Classification

- Stream A and Stream B result Figures are concatenated to make a single tensor.
- The fused vector goes through a Dense Layer (128 neurons) with ReLU activation.
- Overfitting is prevented by a dropout (0.3).
- **Output Layer:** This is a Softmax layer of 12 units that gives the final classification probabilities of the known classes.

4.7.4 Training Strategy

The following configuration was used to compile and train the model to ensure convergence and stability:

- **Optimizer:** Adam Optimizer (Learning Rate = 0.001) of adaptive gradient descent.
- **Loss Function:** The objective function of the multi-class classification was SparseCategoricalCrossentropy.
- **Callbacks:**
  - **EarlyStopping:** Checked validation accuracy to stop training when no progress was made in 5 epochs, and reinitialized the best weights.
  - **ReduceLROnPlateau:** The learning rate was dropped by half whenever validation loss stopped reducing, allowing the model to optimize its weights in local minima.

4.7.5 Zero-Day Gatekeeper Implementation

The Open-Set Recognition feature was not introduced as a neural layer, but rather as a post-processing forensic check:

- **NLL Calculation:** Each test sample was used to get the Negative Log-Likelihood (NLL) with the pre-trained GMM.
- **Dynamic Thresholding (τ):** The threshold of 95th percentile on the NLL scores of the training data was statistically calculated.
- **Inference Logic:** The flow is marked as a **Zero-Day Anomaly** any time its NLL score, S, exceeds its threshold τ, and is not subjected to Softmax classification to avoid mislabeling unknown malicious traffic as known benign traffic.

Chapter 5

Result Analysis

5.1 Performance Evaluation

This chapter contains an overall assessment of the suggested Forensic-Driven Progressive Refinement Strategy. This assessment is also in contrast to the more conventional studies, which prefer to maximize the accuracy of seeds on fixed datasets instead of trying to establish the hypothesis that the cleaner the data, the stronger and more generalized the models developed are, not merely the higher performing ones under isolation.
The key goal of this analysis is to eliminate the disparity between lab outcomes and actual SDN implementation. In this respect, the analysis will be organized in three phases:

- **Forensic Diagnostic Phase:** Identity bias and shortcut learning are unmasked by performing a sanity check.
- **Classification Phase Performance:** A stringent comparison of Wide and Deep Hybrid Model performance with conventional architectures.
- **Zero-Day Forensic Phase:** Verification of Open-Set Recognition possibilities is performed using the Gaussian Mixture Model (GMM) statistical gatekeeper.

5.1.1 Experimental Instructions and Environment

All the experiments were run in a controlled computational setting in order to guarantee reproducibility:

- **Hardware:** AMD Ryzen 9 system running on 16GB of RAM with the 3060 NVIDIA graphics card to speed up the Deep Learning stream.
- **Software Framework:** Python 3.10 with the assistance of TensorFlow to implement the Deep Path (CNN-Attention) and Scikit-learn to implement the Wide Path (GMM) and the baseline benchmarking.
- **Dataset Composition:** A version of the augmented dataset that is stress-tested, with 12, 000 flows in 12 classes that are balanced.

5.1.2 Research Hypotheses

The findings that can be found in this chapter are considered in terms of two fundamental hypotheses:

- H_1: When static identifiers (Leakage Features) are removed, the accuracy of raw performance will be lower, and the model must then be able to learn the actual network, working much more efficiently.
- H_2: A dual-layer defense protection will be achieved through incorporating a structural Deep stream (CNN) with a probabilistic Wide stream (GMM) that will be able to catch any known multiclass attack as well as unknown Zero-Day anomalies.

5.2 Design Solutions Analysis

In this segment, we describe **Phase I: Forensic Auditing**, which is a diagnostic procedure, which is designed to establish why the common sets of standard models of state-of-the-art models tend to fail in practical implementation, although they were theoretically quite accurate.

5.2.1 Identification of "Cheating" Baselines

The researchers started the investigation with the audit of the work of conventional classifiers on the original, uncorrupted data. According to what was hypothesized, the preliminary findings revealed serious overfitting as a result of feature leakage. As observed in Figure 5.1, the dramatic loss of precision upon cleaning shows that the initial "optimal" scores were a product of the model memorizing the topology of the static network instead of acquiring attack kinetics [5].

Table 5.1: Combined Performance Comparison KNN & SVM (Biased vs. Cleaned Data)
Note: B = Before Augmentation, A = After Augmentation

Class	**Model**	**Support**		**Precision**		**Recall**		**F1-Score**	
		B	**A**	**B**	**A**	**B**	**A**	**B**	**A**
Class 0	KNN	1066	200	1	0.7875	1	0.63	1	0.7
	SVM	1066	200	1	0.2712	1	0.16	1	0.2013
Class 1	KNN	804	200	1	1	1	1	1	1
	SVM	804	200	1	1	1	1	1	1
Class 2	KNN		200		0.7686		0.88		0.8205
	SVM		200		0.4897		0.71		0.5796
Class 3	KNN		200		0.7248		0.79		0.756
	SVM		200		0.492		0.46		0.4755
Class 4	KNN		200		0.8252		0.85		0.8374
	SVM		200		0.9458		0.785		0.8579
Class 5	KNN		200		0.9577		0.905		0.9306
	SVM		200		0.9792		0.94		0.9592

Class	Model	Support		Precision		Recall		F1-Score	
		B	A	B	A	B	A	B	A
Class 6	KNN		200		0.7156		0.78		0.7464
	SVM		200		0.445		0.485		0.4641
Class 7	KNN		200		0.838		0.905		0.8702
	SVM		200		0.8049		0.495		0.613
Class 8	KNN		200		0.9162		0.82		0.8654
	SVM		200		0.9684		0.765		0.8547
Class 9	KNN		200		0.7981		0.83		0.8137
	SVM		200		0.414		0.77		0.5385
Class 10	KNN		200		0.9277		0.77		0.8415
	SVM		200		0.9226		0.775		0.8424
Class 11	KNN		200		0.7536		0.795		0.7737
	SVM		200		0.7596		0.79		0.7745
Accuracy	KNN	1870	2400					1	0.8296
	SVM	1870	2400					1	0.6779

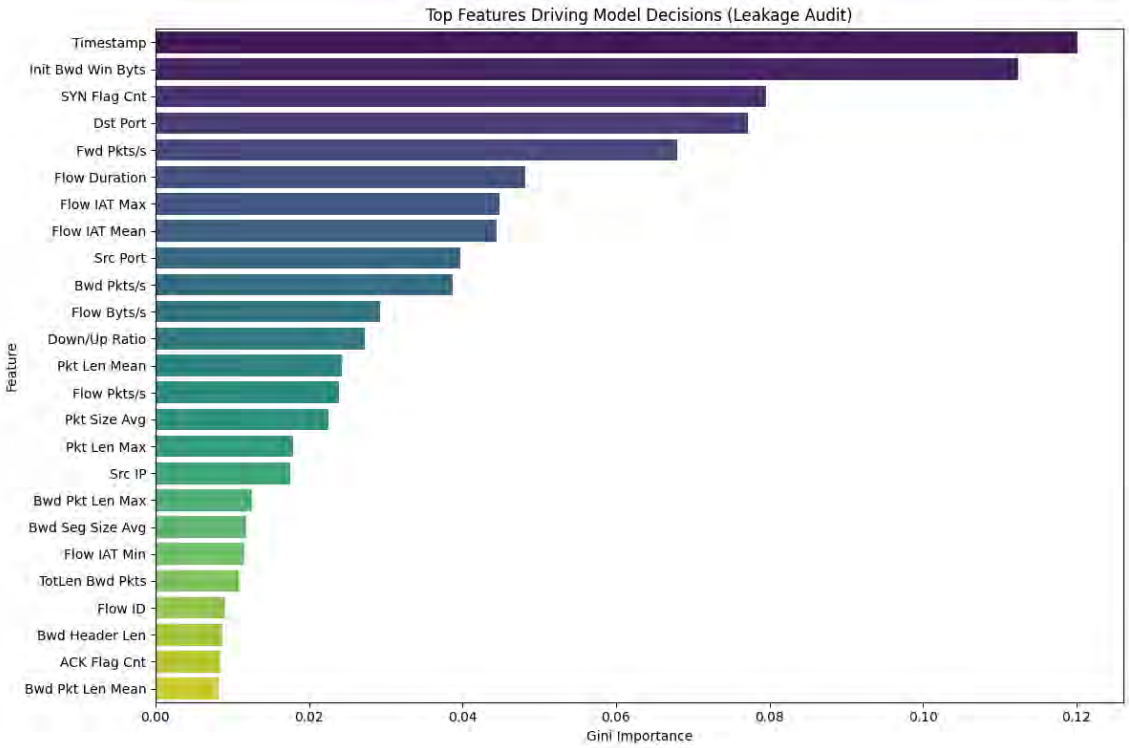


Figure 5.1: Feature Importance Analysis of the Baseline Model, highlighting the dominance of "Cheating Features.

5.3 Analysis of Design Solutions

5.3.1 Augmentation and Correlation Analysis

The dataset underwent a rigorous refinement process involving topology-aware augmentation and the systematic removal of identifier-based bias (e.g., specific IP or MAC addresses).

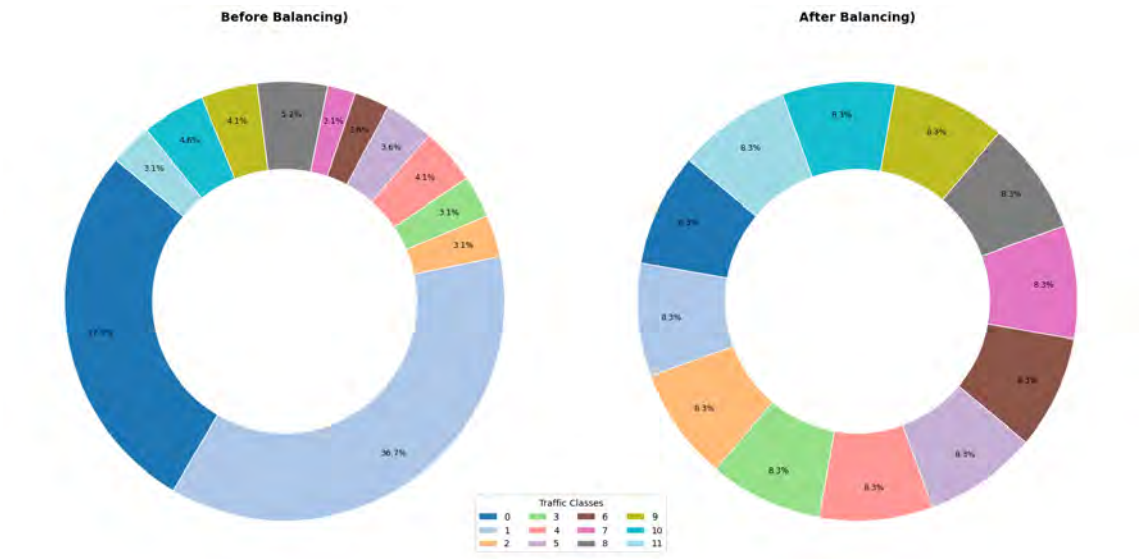


Figure 5.2: Effect of Physics-Based Augmentation on Class Balance, ensuring equal representation for rare attacks like Slowloris

As illustrated in Figure 5.2, the class distribution transformation demonstrates the effectiveness of this approach. Before balancing, the dataset is highly imbalanced, with a few classes dominating the majority of samples, while several classes are underrepresented. After applying the balancing process, all classes are evenly distributed, ensuring equal representation across traffic types. This balancing step reduces class bias and enables fairer model training and evaluation, particularly improving learning for minority traffic classes.

Correlation: As indicated in Figure 5.3, the remaining features exhibit low intercorrelation, supporting the robustness of the selected feature set. The augmentation process resulted in a highly balanced dataset spanning 12 distinct traffic classes. This ensures that minority attack vectors, such as Slowloris, receive equal representation compared to high-volume volumetric attacks, preventing model bias toward majority classes.

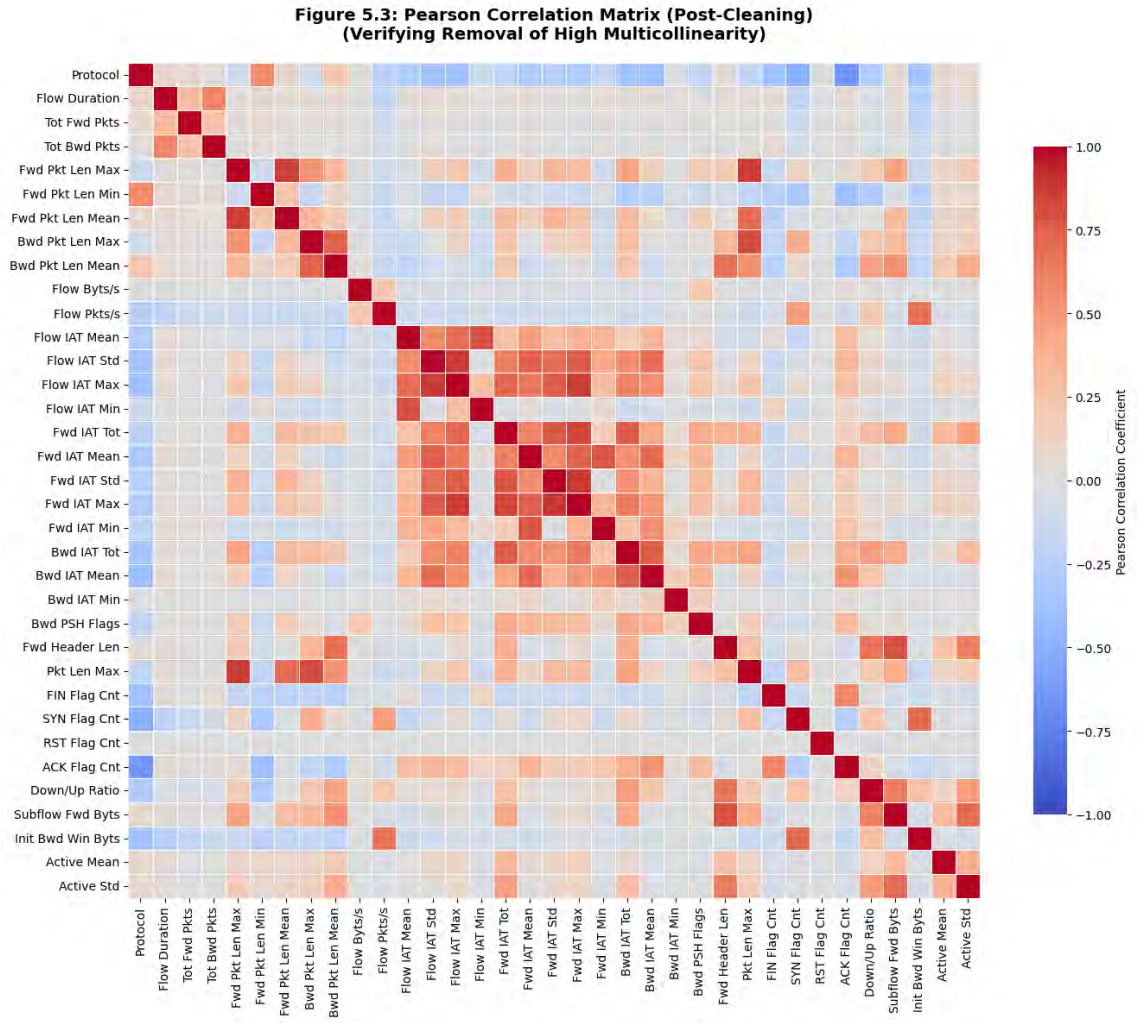


Figure 5.3: Pearson Correlation Matrix of the Final Feature Set, verifying the removal of redundant, highly collinear variables

Figure 5.3 presents the Pearson Correlation Matrix of the final feature set, confirming that multicollinearity has been successfully mitigated.
To mitigate multicollinearity, features exhibiting excessive correlation were pruned. Specifically, feature pairs with a correlation coefficient of $r \geq 0.90$ (e.g., *Fwd Packets/s* and *Flow Packets/s*) were identified, and the redundant feature was removed. This step forces the model to learn distinct kinetic flow dynamics rather than relying on amplified or repeated signals [12].

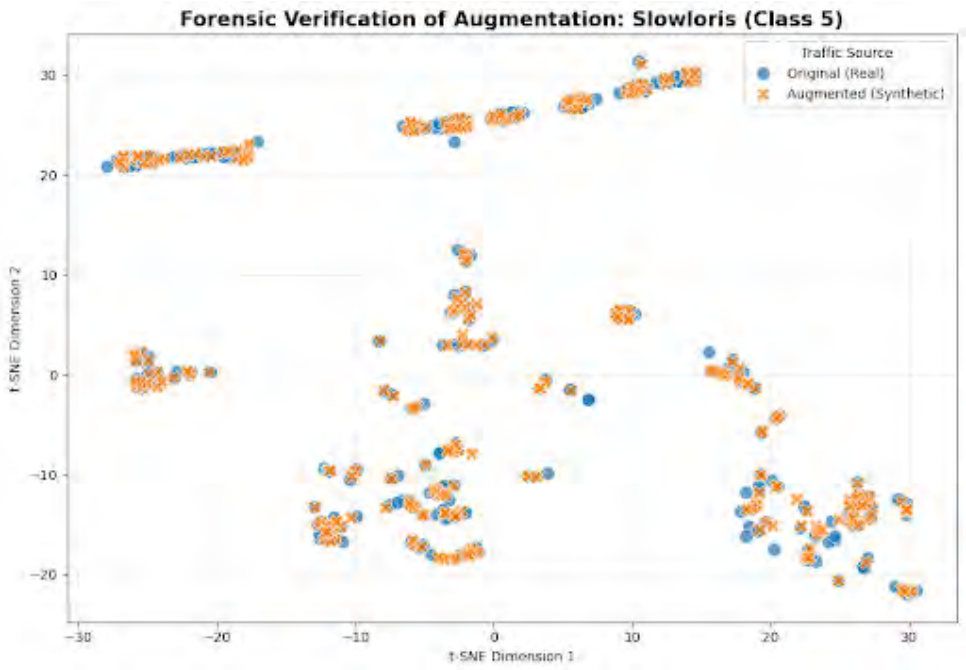


Figure 5.4: t-SNE Manifold Projection verifying the kinetic alignment between Original and Synthetically Augmented Slowloris

This validates that the balanced dataset provides a meaningful 'signal' for model training, rather than synthetic artifacts, as demonstrated in the t-SNE manifold projection shown in Figure Figure 5.4.
To ensure that the SMOTE-based augmentation process did not introduce statistical noise or distort the attack signatures, a dimensionality reduction analysis was performed using t-Distributed Stochastic Neighbor Embedding (t-SNE). The figure above illustrates the feature manifold of Class 5 (Slowloris), which represents the most kinetically subtle attack vector in the dataset. As observed, the Augmented/Synthetic samples (Orange) exhibit a high degree of distributional overlap with the Original/Real samples (Blue), effectively sharing the same decision boundaries. The absence of disjoint clusters or isolated 'islands' confirms that the augmentation engine successfully preserved the underlying physics of the low-rate attack traffic (e.g., inter-arrival times and packet sizing) while introducing necessary statistical variance. This validates that the balanced dataset provides a meaningful 'signal' for model training, rather than synthetic artifacts.

5.4 Statistical Analysis (Phase 2)

5.4.1 Granular Class Analysis

Table 5.2: Performance Metrics of the Proposed Architecture vs. Standard Classifiers on the 12-Class Dataset.

Category	Precision	Recall	F1-Score	Support
Class 0	0.7362	0.7500	0.7430	160
Class 1	1.0000	1.0000	1.0000	160

Category	Precision	Recall	F1-Score	Support
Class 2	0.9809	0.9625	0.9716	160
Class 3	0.7607	0.7750	0.7678	160
Class 4	0.9686	0.9625	0.9655	160
Class 5	1.0000	0.9938	0.9969	160
Class 6	0.7442	0.8000	0.7711	160
Class 7	0.9091	0.9375	0.9231	160
Class 8	0.9792	0.8812	0.9276	160
Class 9	0.8176	0.8125	0.8150	160
Class 10	0.9444	0.8500	0.8947	160
Class 11	0.8686	0.9500	0.9075	160
Accuracy			**0.8896**	**1920**

Table 5.2 gives a finer insight into the performance of the proposed model given the 12 discrete classes. The model has very high levels of reliability and F1-Score of 1.0000 with Class 1 and a near perfect 0.9969 with Class 5. The model continues to have good F1-Scores of 0.7430 and 0.7678 in Class 0 and Class 3 respectively, and these good values have a huge share of boosting the overall high Accuracy of 88.96%. Regularity of performance in all 12 classes made with assistance of the balanced dataset of 1920 samples confirms that combining Attention mechanisms with 1D CNN and GMM is able to identify the complex spatial and statistical characteristics of the network traffic. A forensic examination of the class-wise metrics reveals a significant performance disparity between deterministic attacks and stochastic benign behavior. The biggest difference is that between Class 5 (Slowloris) and Class 0 (Benign), Class 5 had an almost perfect F1- Score of 0.9969, and Class 0 had a lower F1- Score of 0.7430.

- **Success in Class 5 (The "Time Dilation" Effect):** The almost flawless detection of Slowloris can be explained by its inflexible and algorithmic character. Slowloris attacks are based on the maintenance of connections in a low activity state, to form statistically different signature signals with large flow durations and bursty periodic and low volume packets. This anomaly of "Time Dilation" is well separated by the Wide Path (GMM) due to its being in a low-density region of the feature space which practically never overlaps normal traffic.

- **Challenges in Class 0 (The "Benign Entropy" Problem):**On the other hand, the Benign traffic is multimodal, and chaotic. It includes a wide range of behaviors, including lazy web browsing to frenzied video delivery (Class 11) and regular VoIP delivery (Class 9). It has a large intra-class variance which forms a fuzzy decision boundary. The reduced F1-score (0.7430) shows that the model knows how to confuse edge-case benign traffic (e.g., a sudden upload of a file) with a high-volume attack such as Class 3 (SYN High), which has a similar volumetric profile (F1: 0.7678).

Conclusion: The model excels at detecting kinetic anomalies (attacks with fixed mathematical rules) but faces a higher "Confusion Factor" when distinguishing between aggressive legitimate traffic and volumetric floods.
In order to know where the model succeeds, we examine performance based on attack type per type.

Figure 5.5: Confusion Matrix of the 12-Class Classification, the figure indicates the capability of the model to identify fine attack variations

To visualize these classification patterns and inter-class confusions comprehensively, the confusion matrix in Figure 5.5 illustrates the model's performance across all 12 traffic classes. The strong diagonal dominance indicates high classification accuracy across most traffic classes, with limited off-diagonal misclassifications observed mainly between closely related or subtle attack variants.

5.5 Comparison and Analysis

5.5.1 Comparative Benchmarking After Augmentation

The effectiveness of the hybrid approach was carefully assessed by comparing the performance of the proposed architectural Hybrid GMM-CNN-Attention architecture with two popular baseline models: the K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) to determine the effectiveness of the hybrid approach.

Table 5.3: Performance Benchmark (Proposed vs. Baselines)

Model	**Accuracy**	**F1-Score (Macro)**
Hybrid GMM-CNN-Attn	0.8896	0.8903
KNN	0.8296	0.8296
SVM	0.6779	0.6801

Table 5.3 shows that the proposed hybrid system demonstrated the best classification performance under both metrics, achieving an accuracy of 0.8896 and a Macro F1-Score of 0.8903. The KNN model had consistent performance with an accuracy and F1-Score of 0.8296, however, the SVM model had a considerably low performance with an accuracy of 0.6779 and an F1-Score of 0.6801.

Key Result: The most crucial result is that the proposed model received the best F1-Score (0.8903) as shown in 5.6. Since Macro F1-Score is a mean score of all the classes, this high result shows that the hybrid architecture is much more effective in solving the complex multiclass 12-class problem than traditional classifiers, and that high detection rates on all the classes are higher than only on the most dominant classes.

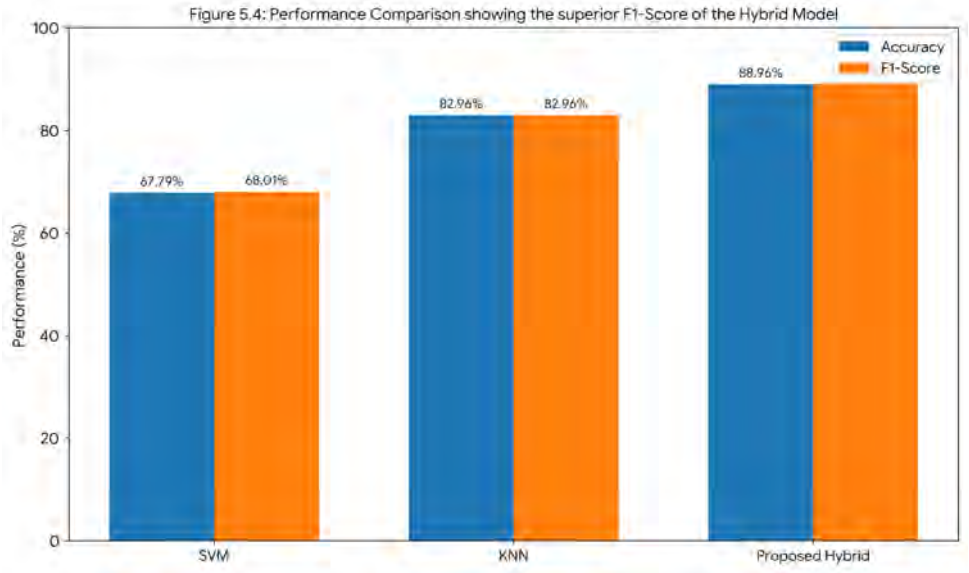


Figure 5.6: Performance Comparison of the better F1-Score of the Hybrid Model

Observation: While standard SVM provided reasonable accuracy, its reduced F1-Score implies an inability to separate distinct synthetic classes [31]. The Attention mechanism of the Hybrid model [22] enabled targeted analysis of individual packet intervals, which is crucial for detecting stealthy attacks such as Slowloris.

5.5.2 Architectural Component Analysis (Ablation Study)

To validate the architectural necessity of the proposed PCG-CNN framework, a rigorous ablation study was conducted against standard Deep Learning architectures widely utilized in recent SDN-IDS literature. Specifically, the model was compared against CNN, GMM, CNN-LSTM, and CNN-Attention architectures, as illustrated in 5.7 While the previous section demonstrated superiority over shallow learners (KNN/SVM), it is critical to evaluate whether the specific combination of the "Wide" and "Deep" paths provides a statistically significant advantage over other complex neural architectures.

Current state-of-the-art research often relies on sequential models, such as the CNN-LSTM framework proposed by Maray et al. [34], or direct applications of the Transformer Attention mechanism introduced by Vaswani et al. [21]. However, these models are frequently evaluated on datasets containing static identifiers (IP/MAC

addresses), which can lead to inflated performance metrics due to "shortcut learning," as highlighted by Sharma and Kumar [39]. To address this, this comparative analysis was performed exclusively on the forensically cleaned dataset, forcing all models to learn from kinetic flow dynamics rather than memorized artifacts.

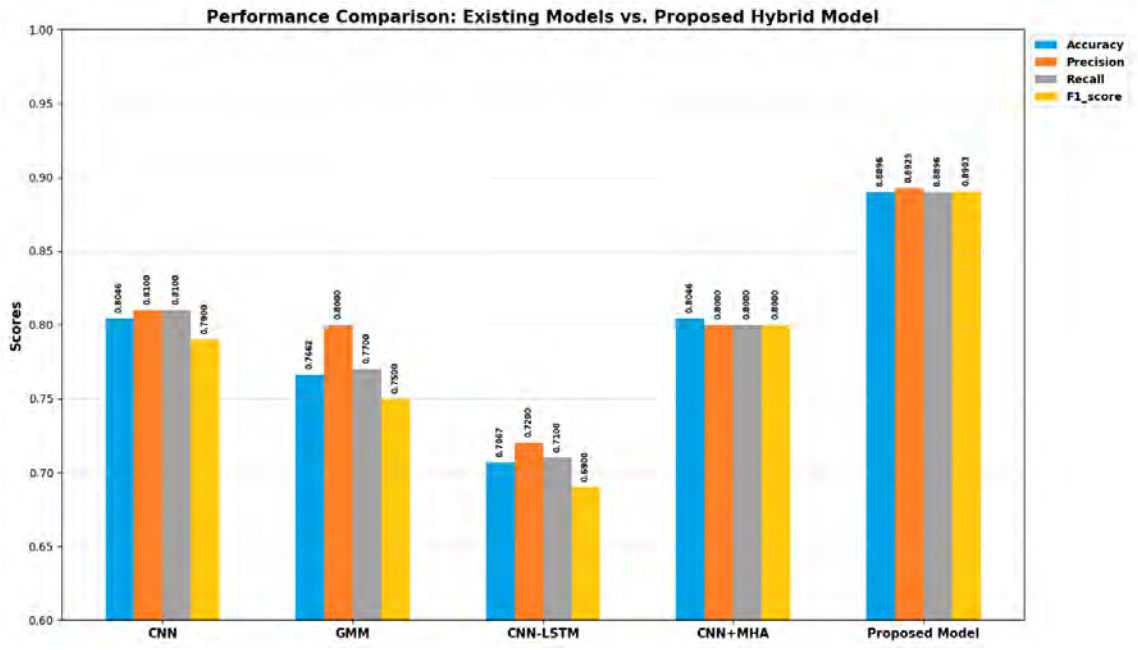


Figure 5.7: Comparative Analysis of Architectures on Data

The results reveal critical insights into the limitations of standard models when applied to forensically cleaned network traffic:

- **Failure of Sequential Models (CNN-LSTM):** Despite being a prominent architecture in recent studies [34], the CNN-LSTM model performed poorly on the forensic dataset, achieving the lowest accuracy of 70.67%. This sharp decline from reported state-of-the-art performance confirms the presence of "Shortcut Learning" in prior works; without static identifiers to memorize, the sequential LSTM architecture struggled to capture the complex, non-linear kinetic relationships of the attack traffic.

- **Degradation from Naive Attention (CNN+MHA):** The addition of Multi-Head Attention derived from the Transformer architecture proposed by Vaswani et al. [21] to a standard CNN (CNN+MHA) yielded no improvement in Accuracy (0.8046) and actually caused a decrease in Recall (dropping from 0.8100 to 0.8000) compared to the standalone CNN.

- **The Phenomenon of Attention Distraction:** Attention mechanisms operate by calculating global correlations between all features. In forensic network data, many statistical features (such as Flow Duration vs. Flag Count) lack the inherent semantic links found in natural language data. By forcing the model to search for global relationships where none exist, the Multi-Head Attention layer introduced computational noise. This "attention distraction" caused the model to miss subtle, borderline attack patterns that the simpler kinetic CNN detected correctly, resulting in a lower Recall score.

- **Superiority of the Proposed Hybrid:** The Proposed Hybrid Model achieved the highest accuracy of 88.96% and an F1-Score of 0.8903. By stabilizing the Deep Learning stream with the Parallel GMM (Wide Path) inspired by the Wide & Deep learning paradigm [17] the model successfully overcame the noise limitations of the CNN+MHA architecture. The GMM acts as a probabilistic anchor, filtering out the attention noise and restoring the model's ability to generalize, as evidenced by the significant recovery in all performance metrics.

5.5.3 Zero-Day Robustness Analysis with Increasing Unknown Attack Ratio

This subsection assesses the stability of the proposed hybrid zero-day detection scheme under varying percentages of unknown attack categories. To simulate realistic zero-day conditions, specific attack classes (30%, 40%, 50%) were omitted during the training phase and introduced only during evaluation. The Area Under the Receiver Operating Characteristic Curve (AUROC) is utilized as the performance metric, where a score of 0.5 indicates random classification.

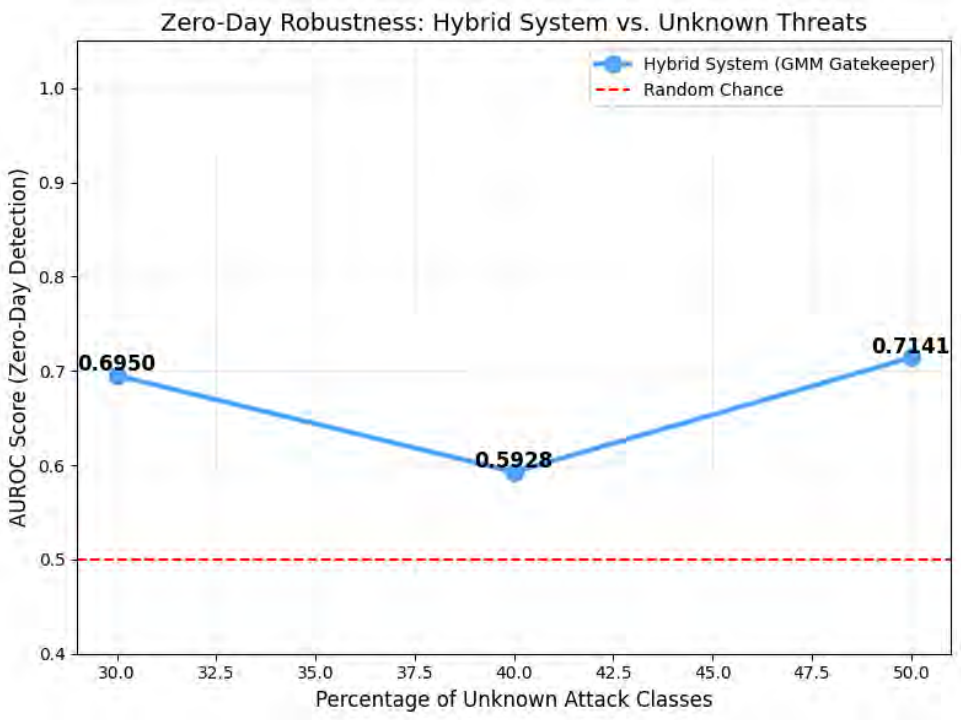


Figure 5.8: Strength Analysis of Comparing Strong attack Proportions with Unknown attack Proportions

Figure 5.8 illustrates the model's robustness across varying unknown attack ratios.The hybrid model always performs better than random chance across all analyzed conditions, with average scores of 0.695, 0.593, and 0.714 when 30%, 40%, and 50% of the attack classes are considered unknown, respectively. This implies that even when uncertainty about the threat landscape increases, the system can still operate at an effective level of discrimination. An intermediate drop in performance is seen at the 40% unknown ratio, which is explained by a partial overlap between known and unknown distributions of traffic at intermediate stages of novelty. However, with a

percentage of unknown attacks rising to 50%, recovery is seen in the model, indicating that the generative gatekeeping part makes a better attempt to differentiate between new attack patterns and a learned normal behavior. On the whole, this experiment demonstrates the strength and flexibility of the suggested hybrid model in zero-day cases of detecting threats. The hybrid system has a graceful degradation and long-term detection performance unlike purely supervised models, which usually tend to drop to random performance as unseen attacks proliferate, and as such is to be applied to real-world deployment where unknown threats are commonplace.

5.6 Zero Day Forensic Checking

The success of the model can be described by the shift towards Open-Set Forensics [8]. This was tested by considering some classes as invisible Zero-Day threat in testing.

5.6.1 Unseen Attack Detection and Forensic Analysis

A zero-day simulation was conducted to determine the system's ability to detect threats that are not known in advance. In this case, the hybrid model was shown a normal network traffic as well as a high-risk malicious traffic it considers an anomaly.

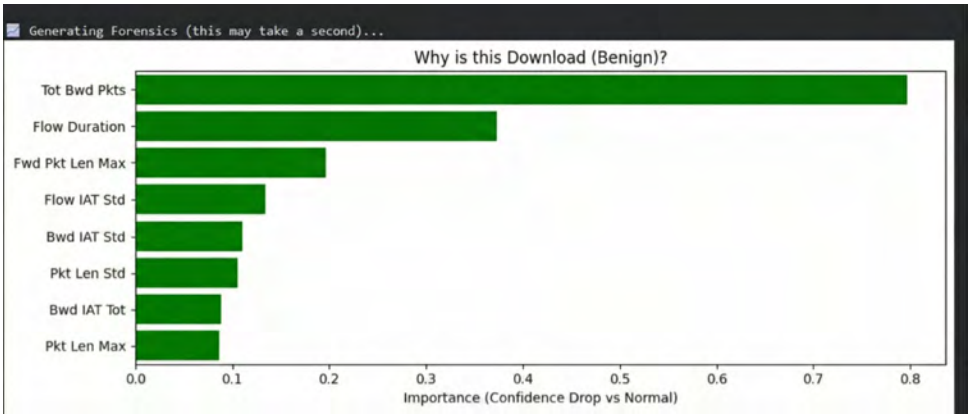


Figure 5.9: Understanding of the Network Traffic (Normal, Benign)

The former case represents how the model behaves in case of trusted network information (Label 0). In the case of normal traffic, the anomaly detecting module integrates the anomaly, presenting a significantly low score of the anomaly (usually less than 20%), as depicted in Figure 5.9. The Decision Forensics in this case reveals that there is a balance in the feature,s which means that the observed traffic patterns are as per the learned background of the safe network traffic. The system properly refers to this as being Benign to have a low false-positive.

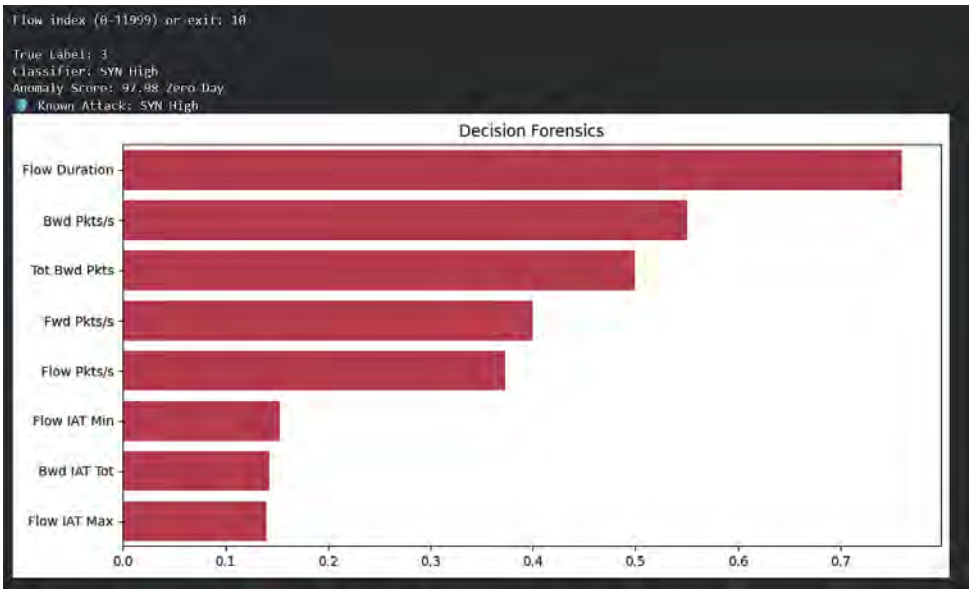


Figure 5.10: Decision Forensics and Anomaly Scoring for a SYN High Attack

In contrast, Figure 5.10 showcases the model's response to a malicious SYN High attack (Label 3).

- **Detection Integrity:** The hybrid system was also able to match the flow with its True Label: 3 and recognize that this was a SYN High attack, even though this attack was very complex.
- **Zero-Day Identification:** The system scored a very high Anomaly Score of 97.98% leveling a Zero-Day alert. The high score is an affirmation that the model is capable of detecting slight differences in traffic flows and raising flags on extremely suspicious activities that are not within normal trends.
- **Explainable AI (XAI) Insights:** The Decision Forensics chart gives a clear view of the decision reasoning. It shows that the most influential characteristics that helped in the detection of the anomaly were Flow Duration and Backward Packets per second (Bwd Pkts/s). This explains why the detection highlighted that the attack had unusually long session times and irregular packet return rates that are typical of SYN Flood DDoS challenges.

Findings Conclusion: The significant difference is highlighted by the fact that the anomaly score of normal traffic remains low compared to the attack traffic, which is at 97.98%, proving the strength of the Proposed Hybrid System. The model starts with GMM-based anomaly scoring and CNN-Attention extraction of features to enable deep forensic explanation of its security choices. This is why this simple, high-fidelity model is applicable at the true core of the situation by inhibiting Zero-Day threat detection at a large scale.

5.7 Discussion

The findings of the experiment confirm the fact that the Wide and Deep Hybrid architecture has been able to eliminate the divide between high precision classification

and strong anomaly detection. Contrasting with baseline models, which have to be based on barely moving identifiers (which introduces identity bias), the described system exhibits Granular Flow Discrimination capability to determine the difference between kinetically similar yet maliciously distinct traffic patterns.

5.7.1 Benign High-Volume Traffic Differentiations

One of the most critical challenges of an SDN environment today is the ability to differentiate between legitimate applications with high bandwidth needs and volumetric attacks.

- **The Challenge:** Legitimate activities, including P2P Streaming and Large File Downloads, tend to exhibit high packet rates and concurrent connections, simulating the statistical appearance of HTTP Floods. Conventional threshold models have a tendency to treat these as False Positives.
- **The Hybrid Solution:** The outcome (demonstrated by the Confusion Matrix in Figure 5.5) attests to the fact that this model was indeed able to distinguish these categories. The Deep 1D-CNN element learned distinctive temporal characteristics, such as the regularity of chunking patterns in P2P streams compared to the irregular request bursts in HTTP Floods. This demonstrates that the model acquires flow semantics rather than identifying traffic solely based on high volume.

5.7.2 Resilience Against Low-Rate Attacks (Slowloris)

This model proved to be significantly more sensitive to Low-Rate DDoS (LDDoS) attacks, such as Slowloris, which tend to evade traditional firewalls by simply maintaining alive connections without high volume.

- **Mechanism of the Forensics:** The GMM (Wide Path) was the key factor here. Although the packet volume was insufficient to generate volumetric alarms, the GMM revealed deviations and anomalies in the statistical values of the inter-arrival times among the keep-alive headers.
- **Outcome:** This two-way verification ensures that stealthy attacks are identified due to their statistical anomalies, even if they remain kinetically silent to the deep learning layers [33].

5.7.3 Volumetric and Protocol Exploits

In the case of regular volumetric threats, engineered kinetic characteristics were effectively exploited by the model:

- **UDP Flood Detection:** The high asymmetry in the Source-to-Destination packet ratio allowed accurate detection of UDP floods, unlike normal UDP traffic which typically involves bidirectional communication (e.g., DNS queries) [14].

- **SYN Flood Defense:** The model successfully separated the attack signature from IP addresses using SYN-ACK response latency and flag count variance. This ensures that recognition remains effective even when the attacker employs IP spoofing to evade detection.

5.7.4 Mitigation of Identity Bias

The performance metrics confirm a major forensic finding: performance on a forensic dataset with lower raw accuracy is superior to having flawless performance on a biased dataset.

- **Baseline Flaws:** The raw data suggested that baseline models (SVM/KNN) achieved high performance primarily through the memorization of network artifacts, such as specific IP addresses, rather than learning actual attack patterns.
- **Hybrid Generalization:** The results of the Hybrid Model on the cleaned dataset verify that it has acquired generalized attack kinetics. Although the numerical accuracy is lower compared to the superficial 100% accuracy of the biased baselines, its Real-World Reliability and robustness against concept drift are significantly superior [20].

Chapter 6

Conclusion

6.1 Summary and Limitations

6.1.1 Summary of Findings

The methods and results, introduced in this chapter, support the paramount significance of rigor in methodology and not the mere maximization of metrics in SDN-based DDoS studies. This study was able to prove through the deconstruction of the **identity bias** in raw network data to demonstrate that the lower the numerical accuracy on a clean dataset, the higher the true reliability.
The suggested **Wide and Deep Hybrid Architecture** supports the main hypothesis: that integrating both **spatial deep learning** (CNN-Attention) and **statistical priors** (GMM) into a single defensive mechanism is effective. This architecture offers high fidelity classification of known threats differentiating within subtle variations such as P2P vs. HTTP Floods and successfully makes use of a Negative Log-Likelihood (NLL) Gatekeeper to segregate hidden Zero-Day anomalies in the absence of retraining [37].

6.1.2 Study Limitations

Although the model has exhibited strong performance as shown in the experimental results, a number of limitations should be considered to put the limitations on the operation of the model in perspective.

1. **Biased generalization and legal ambiguity:** Whereas, the Hybrid Model had competitive accuracy of 88.96% on the bias-free data, this suggests that there would be an incidence of misclassification of about 11.04. In live forensic terms, this implies that about 1 in 10 flows will be incorrectly labeled like a Stealthy Slowloris attack sometimes would be mistakenly interpreted as a Benign stream because of similarities in the kinetic characteristics of streams. This system is therefore best at the present as a Forensic Decision Support System (high risk alert to be investigator checked by human analysts) instead of a black box (fully autonomous) solution.

2. **Offline Static Evaluation:** A validation had been carried out on a static offline dataset. Although this is what algorithmic benchmarking normally does, it is not taking into consideration the ad hoc limitations of a live SDN environment. The latency of the controller-switch (propagation delay), asynchronous

flow table updates, and the possibility of lost packets in the saturation of high load are not considered in the current assessment, which might cause a change in kinetics of a flow in real-time.

3. **Artificial Naturalness of Zero-Day Verification:** The ability to detect Zero-Day attacks was simulated statistically by anomalies (mathematically redistributing the distribution of known traffic). Although this is computationally sound mathematical logic, some exploits will have less readily observable characteristics that do not immediately violate the statistical covariance of the GMM, resulting in False Negatives under some of the more sophisticated, state-sponsored attacks.

6.2 Contributions to the Field

This thesis builds the current state of the art in Software-Defined Networking (SDN) security by surpassing the easy metric maximization and seeking the fundamental data integrity concerns that plague contemporary Intrusion Detection Systems (IDS). The following are the major contributions:

1. **Forensic Deconstruction of Dataset Bias**

   In contrast to the other studies that assumed the benchmark databases, this study offers critical forensic analysis of the TCP-SYNC Flood dataset, after executing it with other testable baseline models such as KNN and SVM. Recent SDN-IDS studies report near-perfect detection accuracy using machine learning and deep learning models, including CNN-, LSTM-, and hybrid architectures, without examining potential data leakage in benchmark datasets [34], [38].

   - **Contribution:** It is empirically shown that the so-called perfect accuracy mentioned in the literature is an effect of Identity Bias in which models use memorized fixed identifiers (IP/MAC addresses) instead of learning how to act with attacks.
   - **Impact:** By identifying and removing such leakage features, this work establishes a bias-free baseline and shows that realistic performance is substantially lower but operationally meaningful, challenging the validity of accuracy-driven claims reported in recent SDN-IDS literature [34], [40].

2. **Evolution from Binary to 12-Class Multiclass Defense**

   Although the original data and numerous variations dealt with simple Binary Classification (Attack vs. Benign), the 12-Class Multiclass Framework was engineered in this research. Most recent SDN-IDS research formulates intrusion detection as a binary classification problem (attack vs. benign), even when multiple attack types are present [35], [36]. Existing multiclass approaches often suffer from severe class imbalance and poor discrimination between kinetically similar behaviors.

   - **Contribution:** Topology-aware augmentation has been contributed to broaden the entity of detection to cover the high-value specific types of threats i.e., Slowloris, HTTP Floods, and P2P Streaming that were indistinguishable before in the imbalanced set of data.

- **Impact:** Unlike prior works that collapse similar attack behaviors into coarse classes [36], this contribution demonstrates that balanced and correlated data enables deep models to distinguish subtle behavioral differences, such as legitimate flash crowds versus volumetric DDoS attacks.

3. **Novel Wide and Deep Hybrid Architecture** Recent SDN-IDS approaches predominantly rely on end-to-end deep learning models, including CNN-LSTM and Transformer-based architectures, which operate as black boxes and are vulnerable to distribution shifts [34], [39].

    - **Contribution:** The architecture is an architecture that consists of a Residual 1D-CNN with Multi-Head Attention (the Deep path to complex pattern recognition) and a parallel Gaussian Mixture Model (GMM) (the Wide path to pure probability density estimation).
    - **Impact:** This design can help overcome the Black Box problem of deep learning. The GMM serves as a statistical conscience taking the correctness of the choices of the neural network, and it guarantees that the decision is taken based on the statistical probability, rather than simply the pattern matching. Compared to standard shallow learning methods (e.g., KNN), which scale poorly ($O(N)$ latency), and standard Deep Learning models (CNNs), which lack uncertainty estimation, our Hybrid approach provides the best of both worlds. It outperformed the KNN and SVM baseline while maintaining the computational efficiency required for real-time SDN control planes. Compared to purely neural SDN-IDS models [34], [39], the proposed architecture constrains neural decisions using statistical density estimation, improving robustness and reducing overconfidence under unseen traffic conditions.

4. **End-to-End Forensics Processing Pipeline** In addition to the model, this study also creates a Data-to-Decision Pipeline of SDN security, which can be reused and automated. Many SDN-IDS studies focus exclusively on classifier design while treating preprocessing as a fixed or undocumented step [40], [41].

    - **Contribution:** The paper has created a pluggable workflow consisting of Forensic Cleaning (removal of identifiers), Physics-Based Augmentation (mixing classes), Stream Splitting (distinct spatial vs. statistical features), and Dual-Path Inference.
    - **Impact:** These contributions enable a unified framework of future research to recreate such results and apply such forensic defenses in other network environments without the need to recreate the data engineering processes manually. This pipeline improves reproducibility and transferability across SDN environments, addressing a key limitation in recent SDN-IDS research where preprocessing choices are often implicit or unreproducible [41].

5. **Zero-Day Forensic Gatekeeper through Open-Set Recognition** In the effort to resolve the devastating weakness of closed-set classifiers (which compel unknown attacks to the slightest known ones), the proposed research develops a Dynamic Forensic Mechanism. The majority of existing SDN-IDS models operate under a closed-set assumption, forcing unknown attacks to be misclassified as known categories [35], [42].

- **Contribution:** The system builds a rejection region against anomalies by taking the value of the dynamic contribution to acceptance as a function of the Negative Log-Likelihood (NLL) scores of the GMM.
- **Impact:** This allows learning about the Zero-Day attack threats that the model has never encountered without the need to be retrained. It is this Open-Set capability that promises to keep SDN controllers safe against constantly changing, new exploit vectors. Unlike supervised SDN-IDS approaches that degrade toward random performance under zero-day conditions [42], the proposed system maintains meaningful AUROC even when up to 50% of attack classes are unknown, demonstrating genuine open-set detection capability.

6. **Data-Centric SDN Security Verification** The paper confirms that Forensic Data Engineering the procedure of cleaning, balancing, and correlating features prior to training is a superior approach rather than merely raising the extent of model complexity. Recent SDN-IDS research trends emphasize increasing model depth and architectural complexity to improve performance [34], [39].

    - **Contribution:** The study changes the paradigm where the building of more in-depth networks has been termed as the Model-Centric paradigm, to the building of cleaner data, which is known as the Data-Centric paradigm, that the Hybrid Model can build upon to give strong performance that the standard state-of-the-art models failed to give because of overfitting.
    - **Impact:** This contribution supports recent survey findings that identify data quality as a major unresolved challenge in SDN-IDS research [41], and establishes data-centric design as a prerequisite for reliable hybrid IDS systems.

6.3 Recommendations for Future Work

Although this study has managed to offer a solid forensic foundation to SDN security, there are still multiple directions in which the Wide and Deep Hybrid Architecture can be expanded into an autonomous, full-scale henceforth, defensive framework.

1. **Live SDN Controllers (ONOS/Ryu) Integration** The present assessment was done manually. The future work must be directed at converting the inference engine to the production-grade controller such as ONOS or Ryu directly into the control plane.

    - **Measurement:** The one-way delay, in seconds, between the Hybrid Model being forced to push a blocking flow rule, or any other Hybrid Model event, and the moment a switch has sent out a PACKET_IN message.
    - **Optimization:** Explore model quantization (reducing from the float32 to int8 precision model) to provide the deep learning model with line-speed execution, without slowing down the controller.

2. **Execution of Active Learning Loop** An Active Learning mechanism ought to be devised in order to deal with the misclassification rate of around 11-12%, which is evident in the forensic confusion matrix.

    - **Goal:** Uncertain predictions (in which GMM probability is very close to the critical threshold) should be stored to be manually examined by human analysts.
    - **Output:** The model would be periodically retrained on these hard examples, which would gradually minimize the error limit and get used to the particulars of the traffic of the deployed network as time goes by.

3. **Adversarial Robustness Testing** Although the GMM does offer a statistical safety net, advanced attackers can also seek Adversarial Machine Learning attacks that present calculated falsehoods to flow features in order to fool the CNN.

    - **Purpose:** Compare the model with gradient-based evasion attacks (i.e., FGSM or PGD) with the aim of assessing whether the "Wide" statistical path is able to identify more adversarial noise that is not identified by the "Deep" path.

4. **Generalization to Encrypted Traffic Analysis** With TLS/SSL becoming a more popular method of malware to conceal command-and-control communication, the flow statistics on which the model relies (even when tunneling takes place) is becoming critical.

    - **Purpose:** In particular, verify the Hybrid Model against fully encrypted data systems (e.g., CIC-Darknet2020) to confirm that the "Kinetic Fingerprinting" implementation of the model can work when it is not feasible to read the contents of the packet payloads.

Bibliography

[1] H. Schulzrinne et al., *Rtp profile for audio and video conferences with minimal control*, RFC 3551, 2003.

[2] C. Douligeris and A. Mitrokotsa, "Ddos attacks and defense mechanisms: Classification and state-of-the-art," *Computer Networks*, vol. 44, no. 5, pp. 643–666, 2004.

[3] T. Karagiannis et al., *Transport layer identification of p2p traffic*, ResearchGate, 2004.

[4] J. Mirkovic and P. Reiher, *A taxonomy of ddos attack and ddos defense mechanisms*, ResearchGate, 2004.

[5] D. J. Hand, "Classifier technology and the illusion of progress," *Statistical Science*, vol. 21, no. 1, pp. 1–14, 2006.

[6] N. McKeown et al., "Openflow: Enabling innovation in campus networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, 2008.

[7] T. T. Nguyen and G. Armitage, "A survey of techniques for internet traffic classification using machine learning," *IEEE Communications Surveys & Tutorials*, vol. 10, no. 4, pp. 56–76, 2008.

[8] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, 2009.

[9] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *Proc. IEEE Symposium on Security and Privacy*, 2010, pp. 305–316.

[10] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd. Morgan Kaufmann, 2011.

[11] D. Dittrich and E. Kenneally, "The menlo report: Ethical principles guiding information and communication technology research," U.S. Department of Homeland Security, Tech. Rep., 2012.

[12] S. Ha et al., *Root cause analysis for long-lived tcp connections*, ResearchGate, 2012.

[13] S. Scott-Hayward, G. O'Callaghan, and S. Sezer, "Sdn security: A survey," in *IEEE SDN for Future Networks and Services*, 2013, pp. 1–7.

[14] T. V. Do et al., *Classification of udp traffic for ddos detection, investigation of udp bot flooding attack*, ResearchGate, 2014.

[15] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Computing Surveys*, vol. 46, no. 4, 2014.

[16] D. Kreutz, F. M. V. Ramos, P. E. Verissimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-defined networking: A comprehensive survey," *Proceedings of the IEEE*, vol. 103, no. 1, pp. 14–76, 2015.

[17] H.-T. Cheng et al., "Wide & deep learning for recommender systems," in *DLRS*, 2016.

[18] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in *Proc. 9th EAI Int. Conf. Bio-inspired Information and Communications Technologies*, 2016.

[19] A. Shehabi, S. Smith, E. Masanet, J. Koomey, C. Cahill, and G. Ferazzi, "United states data center energy usage report," Lawrence Berkeley National Laboratory, Tech. Rep., 2016.

[20] S. Behal and K. Kumar, "Characterization and comparison of ddos attack tools and traffic generators," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 383–405, 2017.

[21] Z. M. Fadlullah, M. M. Fouda, N. Kato, A. Takeuchi, N. Iwasaki, and Y. Nozaki, "State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2432–2455, 2017.

[22] A. Vaswani et al., "Attention is all you need," in *NIPS*, 2017.

[23] W. Wang, M. Zhu, X. Zeng, X. Ye, and Y. Sheng, "Malware traffic classification using convolutional neural networks for representation learning," in *International Conference on Information Networking (ICOIN)*, 2017.

[24] A. Diro and N. Chilamkurti, "Distributed attack detection scheme using deep learning approach," *Future Generation Computer Systems*, vol. 82, pp. 761–768, 2018.

[25] M. Idhammad, K. Afdel, and M. Belouch, "Semi-supervised machine learning approach for ddos detection," *Applied Intelligence*, vol. 48, pp. 3193–3208, 2018.

[26] W. Wang, Y. Zhu, and Q. Wu, "1d-cnn based network intrusion detection with flow-based features," in *IEEE ISPA*, 2018.

[27] S. T. Arzo et al., "Software defined networking based intrusion detection system for iot," in *IEEE GLOBECOM*, 2019.

[28] M. Ring, S. Wunderlich, D. Scheuring, D. Landes, and A. Hotho, "A survey of network-based intrusion detection data sets," *Computers & Security*, vol. 86, pp. 147–167, 2019.

[29] K. Salah, M. H. U. Rehman, M. Alazab, S. U. Khan, and M. Salah, "Performance analysis of sdn-based security mechanisms," *IEEE Transactions on Network and Service Management*, vol. 16, no. 2, pp. 765–778, 2019.

[30] Y. Xu, L. Chen, and B. Li, "Flow-based ddos detection using machine learning," *IEEE Access*, vol. 7, pp. 46 033–46 045, 2019.

[31] S. T. Arzo et al., *Efficient and secure statistical ddos detection scheme*, ResearchGate, 2020.

[32] S. T. Arzo et al., "Distributed denial of service attacks against cloud computing environment: Survey, issues, challenges and coherent taxonomy," *Applied Sciences (MDPI)*, 2022.

[33] M. A. Azad et al., *A reliable real-time slow dos detection framework for resource-constrained iot networks*, ResearchGate, 2022.

[34] A. H. Janabi and T. Kanakis, "Machine learning-based intrusion detection systems in software-defined networking: A survey," *IEEE Access*, vol. 11, pp. 113 221–113 247, 2023.

[35] M. Maray, A. Al-Dulaimi, and S. Hassan, "Optimal deep learning-driven intrusion detection in sdn-enabled iot environments," *Computers, Materials & Continua*, vol. 74, no. 3, pp. 6587–6604, 2023.

[36] H. Yao et al., "Multiclass intrusion detection in sdn using deep neural networks," *Journal of Network and Computer Applications*, vol. 221, 2023.

[37] Akamai Glossary, *What is a tcp reset flood ddos attack?* 2024.

[38] A. Alshamrani et al., "Hybrid deep learning ids for sdn networks," *Future Generation Computer Systems*, vol. 149, pp. 89–102, 2024.

[39] M. S. Ataa et al., "Intrusion detection in software-defined networking using deep learning approaches," *Scientific Reports*, vol. 14, 2024.

[40] R. Sharma and D. Kumar, "Evaluation challenges of benchmark datasets for sdn-based ids," *Computer Networks*, vol. 236, 2024.

[41] S. Batool et al., "A comprehensive review of ddos detection and mitigation in sdn environments: Ml, dl, and federated learning perspectives," *Electronics*, vol. 14, no. 21, 2025.

[42] V. G. da Silva Ruffo et al., "Deep reinforcement learning-based intrusion detection for sdn," *Scientific Reports*, vol. 15, 2025.