International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 2 (2026)
https://doi.org/10.31436/ijpcc.v12i2.711

# Revisiting Layer 2 in Traditional DPI with Realist CFMO Thematic Synthesis

## Layer 2 Header field usage in DPI

Shafana M.S[1,2]*, Adamu Abubakar[2]

[1]Dept. of Information and Communication Technology, South Eastern University of Sri Lanka, Sri Lanka.
[2]*Dept. of Computer Science, International Islamic University Malaysia, Kuala Lumpur, Malaysia.*

*Corresponding Author: zainashareef@gmail.com
(Received: 4th May 2026; Accepted: 12th June, 2026; Published on-line: 30th July, 2026)

***Abstract***— Traditional Deep Packet Inspection (DPI) systems rarely emphasise data-link layer fields for anomaly detection, despite their foundational role in Ethernet-based communications. Evidence on the security contributions of individual Layer 2 fields in traditional DPI therefore remains limited and fragmented. To examine this gap, this thematic synthesis consolidates research published from 2020 to 2025 on how traditional DPI pipelines utilise Destination and Source Media Access Control (MAC) addresses, EtherType/Length, and Frame Check Sequence (FCS) for security-relevant tasks. The review aims to identify their usage patterns and map the fields to their contexts, mechanisms, outcomes, benefits, and limitations. Eight studies were selected from 2,301 records identified across eight databases through PRISMA-based screening and eligibility assessment. Evidence was extracted manually and analysed using a realist Context–Feature–Mechanism–Outcome (CFMO) lens, while ATLAS.ti 25 supported code management, thematic synthesis, visualisation, and traceability. The findings show that Layer 2 field use remains limited but purposeful. Destination and Source MAC addresses were each reported in six studies, mainly for device distinction, sender identification, session-based flow grouping, traffic-direction analysis, and spoofing detection. EtherType/Length appeared in five studies and facilitated protocol-level parsing and discrimination, while FCS contributed to frame-integrity validation in one study. The Virtual Local Area Network (VLAN) Tag (802.1Q) was absent from the reviewed corpus, indicating broader underutilisation of Layer 2 fields in mainstream DPI research. Mechanisms were largely deterministic and deployed in context-specific scenarios. Overall, Layer 2 fields provide complementary support to higher-layer inspection, but the available evidence does not independently establish high-performance anomaly detection. Future research should examine multi-layer hybrid strategies, VLAN-aware segmentation, and integration of link-layer parsing into DPI-based anomaly-detection pipelines.

***Keywords***— Anomaly Detection, Data Link Layer, EtherType, Thematic Analysis, Packet Inspection

## I. Introduction

The data-link layer header fields, such as frame type and MAC address, provide only basic network information and offer limited insight into actual network activity [1]. When deep packet inspection (DPI) incorporates both these headers and the packet payload, its capacity for threat detection and network traffic profiling is significantly enhanced. This supports both security goals and regulatory compliance. DPI enables analysts to detect anomalies and potential attacks more effectively [2]. Collaborative efforts and the sharing of detection rules through open-source platforms like Snort and Suricata ensure that practitioners remain informed about emerging threats and evolving network protocols [3]. By integrating data-link header checks into DPI, organisations can strengthen network security, meet regulatory requirements, and maintain operational efficiency [4].

### *A. Traditional Deep Packet Inspection (DPI) and Data Link Layer Field Usage*

Layer 2 fields are attributes of the Ethernet data-link frame used for local delivery, protocol identification, network segmentation, and integrity checking. In this review, they include Destination and Source Media Access Control (MAC) addresses, EtherType/Length, the optional IEEE 802.1Q Virtual Local Area Network (VLAN) tag, and the Frame Check Sequence (FCS), which is located in the frame trailer rather than the header. These fields support traffic analysis and anomaly detection within traditional DPI systems. In this review, traditional deep packet inspection (DPI) refers to deterministic, rule-, signature-, or parser-based inspection that examines packet headers and/or payload content to classify, filter, or identify anomalous traffic, without a learned model serving as the primary

inspection decision mechanism. Studies combining DPI with machine learning were retained only when they described an explicit deterministic DPI component that used one or more Layer 2 fields; only evidence related to that component was extracted and coded. While early DPI implementations primarily focused on network and transport layer attributes, recent advancements have shown that including data-link layer information significantly improves visibility and the accuracy of security monitoring. These enhancements support critical functions such as device identification, session grouping, traffic classification, and the simulation of realistic network environments. They are also useful in detecting and investigating abnormal behaviours such as spoofing and unauthorised access. As security challenges have evolved, data-link layer header analysis has been incorporated into modern DPI tools and workflows. This provides foundational support for operational management and research-driven anomaly detection across varied network environments [5], [6].

### B. Data Link Layer Field Usage in DPI-Based Anomaly Detection

In traditional DPI systems, data-link layer headers including Source and Destination Media Access Control (MAC) addresses, EtherType, and Virtual Local Area Network (VLAN) tags play an important role in supporting anomaly detection. These fields assist with device identification, maintaining an accurate inventory of network assets, grouping sessions, and analysing user or endpoint behaviour. This helps uncover unusual communication patterns that may signal potential compromise. The EtherType field allows DPI systems to classify traffic by protocol efficiently, which improves the targeting and precision of anomaly detection. VLAN tags provide information about network segmentation and can indicate policy violations or unauthorised access between network zones. Current DPI methods are capable of identifying spoofing attempts, monitoring topology changes, and offering multi-layered protection against advanced threats by using the unique characteristics of data-link layer fields [5–8].

Although existing DPI research broadly examines packet inspection and whole-system performance, evidence on how individual Ethernet Layer 2 fields contribute to deterministic anomaly-detection mechanisms remains fragmented. The contexts in which these fields are used, the mechanisms they enable, and their reported benefits and limitations have not been clearly synthesised. Consequently, the operational role and research gaps of Layer 2 inspection within traditional DPI remain unclear.

### C. Objective of the Thematic Review

This thematic review examines the application of Layer 2 fields together with payload analysis for identifying network anomalies, based on research published from 2020 to 2025. By reviewing deterministic, rule-based DPI methods and the deterministic DPI components of eligible hybrid approaches, the study aims to identify the Layer 2 fields used, map their Context–Feature–Mechanism–Outcome configurations, synthesise their usage patterns, benefits, and limitations, and identify remaining evidence gaps. The guiding question is: What patterns and trends characterise the use of data link layer header fields in traditional DPI for anomaly detection between 2020 and 2025?

## II. Methods and Materials

### A. Theoretical Framework

We use a Realist Context–Feature–Mechanism–Outcome (CFMO) lens grounded in Context–Mechanism–Outcome (CMO) logic [9] and the view that mechanisms involve resources and the responses they generate [10]. We therefore make the resource explicit as Feature, similar to Context–Intervention–Mechanism–Outcome (CIMO)-style extensions [11]. Features are Layer 2 fields such as Destination MAC, Source MAC, EtherType/Length, and Frame Check Sequence (FCS), which enable mechanisms in traditional DPI, including device identification and inventory, session grouping and ingress/egress parsing, protocol discrimination and parser branching, spoofing/redirection detection, and frame-integrity validation. Context comprises the reported deployment or testing environments, such as programmable switch pipelines, software-defined networking (SDN)/Mininet testbeds, packet-analysis tools, Layer 2 protocol settings, industrial frames, and spoofing scenarios. Outcomes summarise the reported usage patterns, benefits, and limitations associated with each Feature–Mechanism relationship. Because few studies attribute quantitative performance to individual Layer 2 fields, outcomes are synthesised qualitatively [9].

In this review, CFMO thematic synthesis refers to a realist-informed qualitative process in which each reported Layer 2 field use is coded as a Context–Feature–Mechanism–Outcome configuration and recurring configurations are consolidated into themes. The unit of analysis was each reported use of a Layer 2 field in an included study. The four CFMO components were applied as deductive parent codes, while specific field uses, mechanisms, contexts, benefits, and limitations were developed as inductive subcodes and consolidated into broader themes. ATLAS.ti 25 was used to manage the codes, conduct code–document and co-occurrence analyses, generate thematic visualisations, and maintain traceability to the source studies. Study selection,

https://doi.org/10.31436/ijpcc.v12i2.711

coding decisions, and interpretation were performed manually by the researchers.

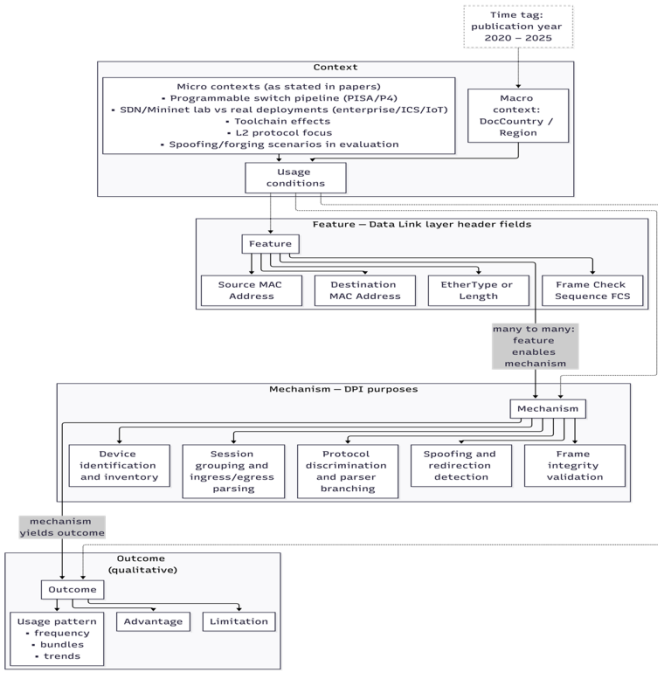


Fig. 1: CFMO-based theoretical framework used in the thematic synthesis.

### B. *Rationale for Employing a Thematic Review*

Despite the fact that systematic reviews are consistently used as a means of aggregating primary findings and assessing research designs in a systematic and repeatable way [12], the current review is not just to summarise the literature related to the use of data link layer usage in deep packet inspection (DPI). Instead, the objective is to trace common themes and new directions on how data link layer header fields are utilised in traditional DPI to detect anomalies. Since the research question explores substantive issues, such as parsing of the header field, device and session attribution, protocol discrimination and operational issues of field-level inspection, a thematic review is a broader and flexible format. This method allows grouping different studies together based upon significant analytical themes, thus bringing out patterns and nuances that may be missed by purely quantitative reviews.

### C. *Search Strategy*

This literature review aimed to capture recent developments in how data link layer header fields are used within deep packet inspection (DPI) systems for anomaly detection. The review prioritised studies that discuss the implementation, performance, and technical limitations of DPI techniques analysing both header fields and payload content.

Variation in terminology across sources posed a challenge in ensuring comprehensive coverage. To address this, the search terms were deliberately broad, with the understanding that preliminary results would be carefully refined using exclusion criteria. Notably, we did not include specific data link layer field keywords (such as "MAC Address" or "VLAN") in the search string at this stage. This approach was chosen to ensure a broader collection of articles relevant to traditional DPI methods, after which studies that explicitly addressed the use of data link layer header fields could be identified and retained during the screening and filtering process. Throughout the review, technically substantive articles were prioritised, specifically those that demonstrated practical use of DPI such as filtering, pattern matching, or protocol decoding. Theoretical works or papers with only superficial mentions of DPI or data link layer analysis were systematically excluded.

A Boolean search string was applied uniformly across databases:

("Deep Packet Inspection" OR "Traffic Inspection" OR "Packet Filtering" OR "Packet Inspection" OR "Payload Inspection" OR "Encryption Inspection") AND "Anomaly Detection".

The review was limited to peer-reviewed journal articles and conference papers published in English between 2020 and 2025. Sources included Scopus, IEEE Xplore, ACM Digital Library, ScienceDirect, ProQuest, Emerald Insight, Wiley Online Library, and Taylor & Francis Online.

### D. *Inclusion and Exclusion Strategy*

This review used a systematic inclusion and exclusion approach to protect the rigour of the methodology and relevance of the thematic content. Specifically, it focused on research articles that covered practical DPI mechanisms involving the application of both data-link layer fields and payload analysis in the context of anomaly detection and network security. Only peer-reviewed journal articles or conference papers published in English between 2020 and 2025, inclusive, were included. Studies combining deterministic DPI with machine learning were included only when they described an explicit rule-, signature-, or parser-based DPI component involving one or more Layer 2 fields. In such cases, only evidence relating to the deterministic DPI and Layer 2 field-processing components was extracted and included in the thematic synthesis. Studies relying exclusively on machine learning without an identifiable deterministic DPI component were excluded. Articles were also excluded if they were published prior to 2020, were not published in English, did not discuss both data-link layer fields and payload inspection, were based on non-peer-reviewed materials, including blogs or white papers, or were retracted.

https://doi.org/10.31436/ijpcc.v12i2.711

The article selection involved four steps: title screening to remove irrelevant articles; abstract screening to exclude off-topic and inadequately detailed studies; diagonal (skim) reading to verify DPI relevance and inspection depth; and full-text assessment to confirm practical DPI application and analysis of data-link layer fields and payloads in accordance with the objectives of the review. Research focusing on DPI techniques, including filtering, protocol decoding, pattern matching, and content inspection involving data-link layer fields, was prioritised. A study was classified as having only a superficial mention when DPI or Layer 2 terminology appeared only in its title, abstract, keywords, introduction, or related work, without describing a practical packet-inspection operation, the use of a specific Layer 2 field, its security or anomaly-detection function, and sufficient evidence to code at least one CFMO configuration. Research focused on high-level monitoring or protocol tuning without explicitly considering packet-level inspection was excluded.

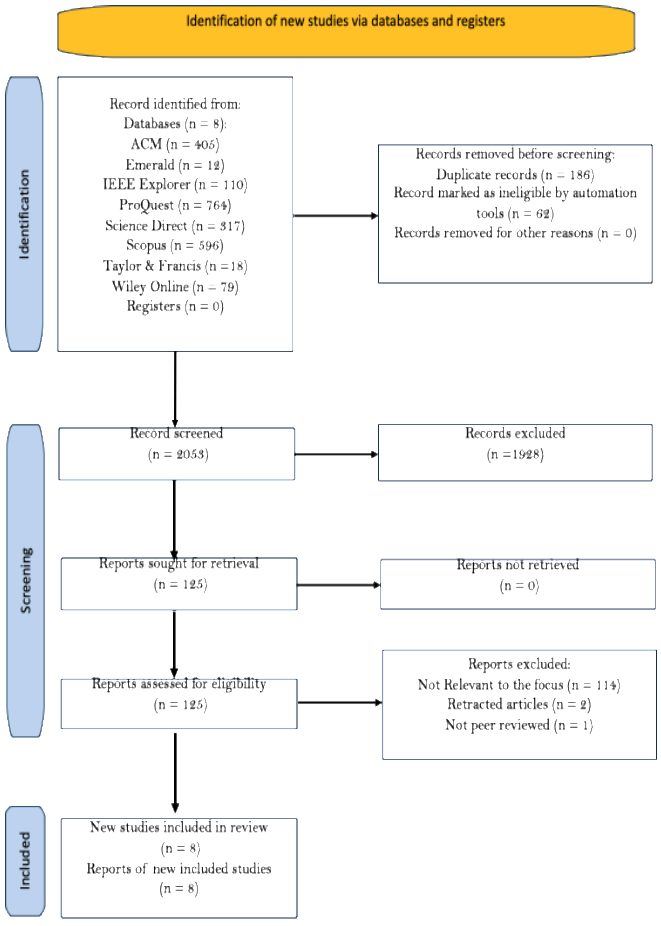


Figure 2: PRISMA Flow Diagram for Article Selection Process

The entire process of selection and screening is depicted in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram (Figure 2). Out of 2,301 records identified across eight major databases, 186 duplicate records were removed, while 62 records were excluded during Rayyan's automated record-processing stage. The remaining 2,053 records underwent manual title and abstract screening, during which 1,928 records were excluded. The full texts of 125 studies were then assessed against the inclusion criteria, including practical DPI operation, explicit use of at least one Layer 2 field, a security or anomaly-detection function, and sufficient technical detail to code at least one CFMO configuration. Following the full-text assessment, 117 studies were excluded, and eight studies were included in the qualitative synthesis and thematic analysis. The manual screening and eligibility assessment were conducted by the authors over approximately eight months.

### *E. The Thematic Review Process*

The review process began with Rayyan, a web-based tool used to filter titles and abstracts effectively and to eliminate duplicate records automatically [13]. Next, the eight studies that met the inclusion criteria were exported to Mendeley, where they were cleaned and organised for final metadata. Each included full text was manually read, and relevant evidence was recorded in a structured extraction form covering bibliographic details, context, Layer 2 feature, mechanism, outcome, benefit, and limitation. A concise evidence-based synopsis of each study was then prepared by the authors. These summaries, rather than the complete articles, were imported into and coded in ATLAS.ti 25 according to the steps outlined by [14] and [15], while the original full texts were consulted when further clarification was required. In ATLAS.ti, the documents were classified according to the year of publication (2020–2025) as well as the main analytical approach. Codes were assigned for country of publication, year, and thematic focus, enabling analysis both chronologically and by research theme. Other code groups, such as author and venue, were based on the metadata and refined during close reading. Thereafter, the coded sections were combined into larger categories, which were further distilled into the principal themes presented in this review. This structured, layered process enabled the authors to trace the evolution of DPI-based anomaly detection using data-link layer fields and identify recurring design patterns throughout the literature during 2020–2025.

Based on the ATLAS.ti-generated codes, the eight articles selected for our review were gathered onto a single thematic level that is directly related to the major question of our research. The theme labels were based on the most common header fields and analytic strategies discussed in the articles (MAC address analysis, EtherType recognition, VLAN tagging, and Frame Check Sequence parsing). The field-level themes represent important parsing logic and attribute usage across referenced studies, consistent with

https://doi.org/10.31436/ijpcc.v12i2.711

well-established Ethernet frame structures, namely, destination and source MAC addresses, EtherType, optional VLAN tag, and the FCS trailer as detailed in foundational Ethernet documentation.

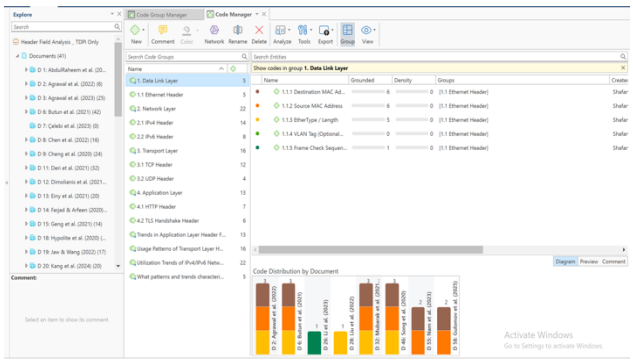

Figure 3: Established Code Group in ATLAS.ti

Notably, although this review is a subset of a broader qualitative study that will examine the utilisation of the header field framework across all four protocol layers in traditional DPI, the present analysis deals specifically with the data link layer usage, as shown in the depicted taxonomy in Figure 4.

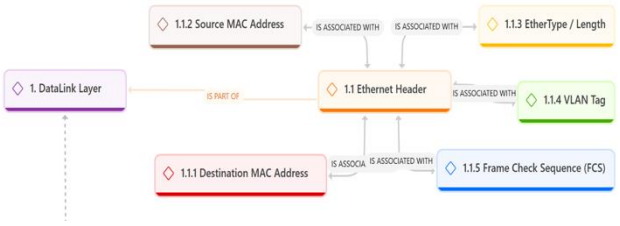


Figure 4: Thematic Code layer Taxonomy

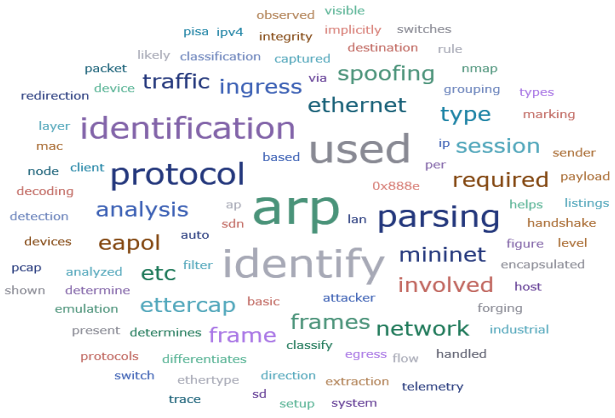


Figure 5: Word Cloud Generated from Articles

The visual summary of the field provided in Figure 5 is presented in the form of a word cloud that has been generated from the most significant terms that occur within the reviewed articles. This visualisation was created before formal coding in ATLAS.ti, and it became apparent that the terms “arp,” “identify,” “protocol,” “frames,” “spoofing,” “session,” and “ethernet” appeared frequently in the literature. The prevalence of the words “identification,” “parsing,” “ingress,” “analysis,” “used,” and “mininet” highlights the central focus on protocol processing, device identification, Address Resolution Protocol (ARP) analysis, and Ethernet frame parsing fundamental operations in the application of data link layer fields within traditional DPI for anomaly detection. The appearance of the terms “ettercap,” “session,” and “spoofing” further emphasises the significance of practical tools and attack scenarios discussed in this field of research.

## III. Results and Discussion

The main results of this thematic review can be found in two sections. The quantitative analysis summarises the number of studies meeting each identified theme, annual trends from 2020 to 2025, and maps publication activity by country and continent. Each theme is given a focused discussion through the qualitative analysis based on how the chosen studies implement the traditional DPI methods of parsing, protocol identification, frame analysis, and ARP spoofing detection based on data link layer header fields used as indicators of anomaly detection. Combined, these quantitative and qualitative views emphasise not only the scale and the geographic coverage of existing studies, but also the underlying analytic techniques and emerging trends that characterise data link layer field usage in conventional DPI to detect anomalies.

### *A. Quantitative Findings*

Eight papers were found to specifically discuss the application of data link layer fields in conventional DPI methods for anomaly detection within the six-year span of 2020 to 2025 (see Table 2). These papers appeared in an assortment of reputable journals and conference proceedings, among them ACM Conference, IEEE Conference, Sensors, Connection Science, and Electronics (Switzerland). The relatively small number of works to consider, spread over a number of locations and years, points to both the pioneering essence and the niche focus of this research subject. Among the eight chosen articles, five were published in peer-reviewed journals and three in conference proceedings, which reflects the current interest in the topic but also indicates the relatively limited scope of work specifically focused on data link layer field analysis in rule-based DPI anomaly detection during this period.

https://doi.org/10.31436/ijpcc.v12i2.711

TABLE 1
ARTICLES REVIEWED BASED ON TYPES OF JOURNALS

| Journal/Conference Name | 2020 | 2021 | 2022 | 2023 | 2025 |
|---|---|---|---|---|---|
| *ACM Conference* | | | 1 | | |
| *Connection Science* | | | 1 | | |
| *Electronics (Switzerland)* | | | | 1 | |
| *IEEE Conference* | | 1 | | | 1 |
| *IEEE/ACM Transactions on Networking* | | | | 1 | |
| *Sensors* | | 1 | | | |
| *Sensors (Switzerland)* | 1 | | | | |

As publication activity depicted in Figure 6, scholarly publication operations covering fields in data link layer in their application in conventional DPI-based anomaly detection were small over the years. In 2020, there was one relevant article, but in 2021, there were two articles, and from 2021 through 2023, the number remained stable at two per year. No study was found for 2024. A single article in 2025 suggests a minimal resurgence after the 2024 gap. In general, this trend indicates that the field of collecting and analysing data-link layer fields in the realm of classical DPI-based anomaly detection has received steady but low scholarly attention, and the lack of publications in 2024 may be either a momentary weakness or may positively indicate that other protocol layers are being addressed or alternative methods of detection are being pursued.

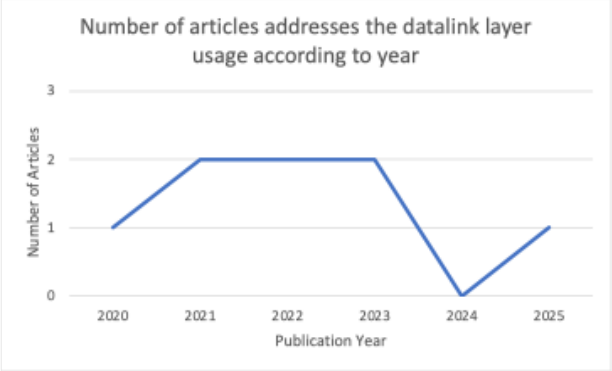


Figure 6: Breakdown of published articles according to the Year of Publication

As Table 2 illustrates, there is broad, though uneven, participation in the international research on the usage of a data-link layer field in traditional DPI-based anomaly detection. China shows sustained participation, with the additions of publications in 2020, 2022, and 2023, which indicates a long-term interest in conducting research in this field. Malaysia is also active with subsequent issues in 2021 and 2022 which mark a trend in the region. India contributes in 2022, reflecting continued interest. A number of other countries such as Poland, South Korea, Sweden, Turkey and Ukraine each contributed a single study within the same time period, which represents less habitual but significant use of this specialised topic. Notably, South Korea's publication in 2023 and Uzbekistan's entry in 2025, along with scattered contributions from Europe and Eastern Europe, signal the gradually diversifying geographic reach of research efforts.

TABLE II
COUNTRY AFFILIATIONS REPRESENTED IN THE EIGHT INCLUDED STUDIES BY YEAR

| Country | 2020 | 2021 | 2022 | 2023 | 2025 |
|---|---|---|---|---|---|
| *India* | | | 1 | | |
| *Sweden* | | 1 | | | |
| *Turkey* | | 1 | | | |
| *China* | 1 | | 1 | 1 | |
| *Ukraine* | 1 | | | | |
| *Republic of Korea* | | | | 1 | |
| *Malaysia* | | 1 | | | |
| *Poland* | 1 | | | | |
| *Uzbekistan* | | | | | 1 |

The yearly change is noticeable and the fact that there are no new articles as of 2024 indicates that there is either a transient gap or change of focus in the research. Altogether, although no single country can be considered a leader in the data-link layer sphere in the traditional DPI research, the involvement of many countries supports the idea of the growing international scope of this domain.

As Figure 7 and the table of country summative results indicate, the data-link layer field utilisation in conventional DPI-based anomaly detection is spread amongst the few national communities. China emerges as the leading contributor, with three publications over the review period, while Malaysia produced two studies. India, along with Poland, South Korea, Sweden, Turkey, and Ukraine, each contributed a single publication. This spread indicates that there is an established research presence in East and Southeast Asia, with meaningful yet less frequent engagement from European and Eurasian countries.


Figure 7: Breakdown of Geographical map of authors

https://doi.org/10.31436/ijpcc.v12i2.711

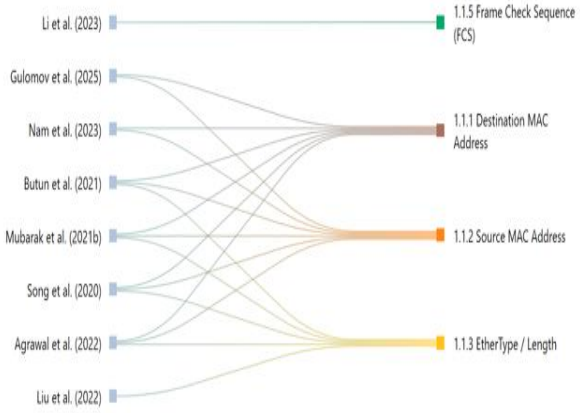


Figure 8: Thematic Mapping of the Publications

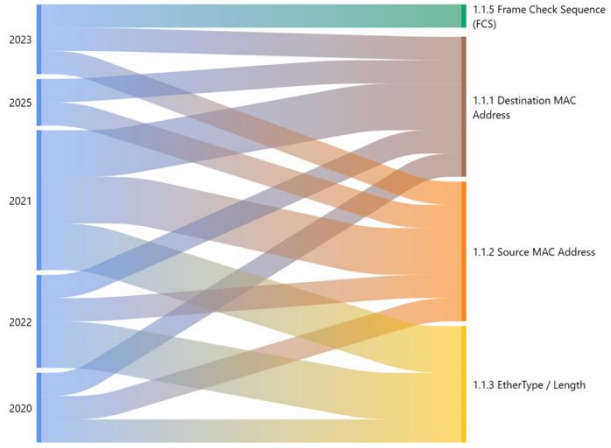


Figure 9: Sankey diagram of the Themes According to Year.

Note: The author–year labels in this ATLAS.ti network visualisations are software-generated document identifiers used for coding traceability. They are not formal in-text citations. The corresponding works are cited in the manuscript text and listed in the reference list according to IEEE style

As illustrated in Figure 8 and 9, the review of traditional DPI-based anomaly detection (n = 8 articles, 2020–2025) indicates that the studies related to data-link layer fields are centred around a limited number of fundamental attributes of Ethernet. Destination and Source MAC Address were addressed in six articles each, and EtherType/Length analysis was found in five research works. The Frame Check Sequence (FCS) was given very little focus because it was examined in only one article, meaning that integrity-check fields are rarely incorporated into DPI-based anomaly detection pipelines.

The trends in the number of publications annually suggest a steady but modest scholarly output, with publication activity peaking in 2021 and gradually declining in subsequent years. In the literature considered, it is shown that the field of data-link layer undergoes extensive inspection, and the majority of works use the combination of one or several types of fields [7], [5], [16], which indicates the further relevance of multi-attribute parsing strategies. Such focused reliance on common Ethernet frame fields highlights the technical consistency of the domain, and the lack of more varied or probabilistic data-link layer mechanisms indicates that there is a prevailing methodological preference for deterministic, header-driven inspection in traditional DPI-based anomaly detection.

## B. *Qualitative Findings*

In this part, the qualitative findings of the traditional DPI techniques based on the data link layer field usage are synthesised. Every article was coded as per the thematic taxonomy, including the main areas, which are field-level parsing, device identification, and protocol differentiation. Each theme has been summarised in terms of core methods, deployment contexts, evaluation strategies, and limitations and how later studies build on earlier work. Figure 10 shows the entire polyline representation of all thematic linkages of the data link layer corpus, which assists in unambiguously visualising the clusters of techniques and their relationships. The numbers in the quote identifiers are those used as page numbers of our internal summary documents and not the original articles.

### 1) *Polyline View of Data Link Layer Themes*

Figure 10 shows a polyline thematic map of the utilisation of data-link layer header fields in conventional DPI-based anomaly detection (2020–2025). This mapping shows how the major Ethernet frame fields, namely, Destination MAC Address, Source MAC Address, EtherType/Length, Frame Check Sequence (FCS), are linked to the research studies implementing the said fields in their practical anomaly detection systems. Both similarities and functional overlaps in roles are depicted by documented use cases across the reviewed literature representing each of the themes. It is noteworthy that the VLAN Tag (802.1Q) field is presented as a stand-alone node that lacks corresponding studies, which indicates a lack of adoption or a research gap regarding the use of VLAN-based classification in the conventional DPI pipelines. The map emphasises a methodological preference for deterministic, field-based inspection, while highlighting the constant emphasis on standard Layer 2 characteristics.

https://doi.org/10.31436/ijpcc.v12i2.711

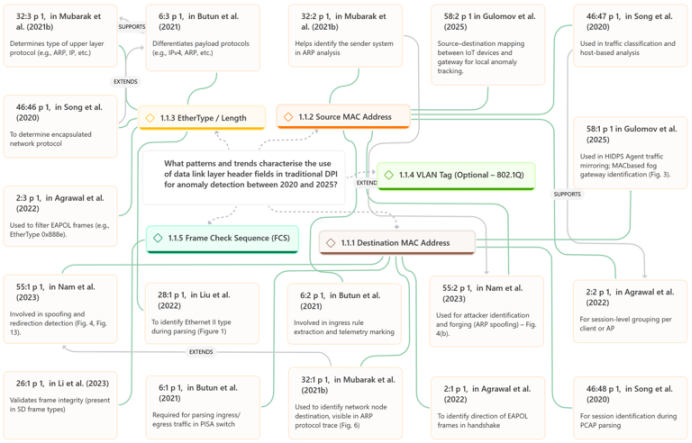


Figure 10: Polyline view of the Themes

### 2) *Theme 1: Destination MAC Address*

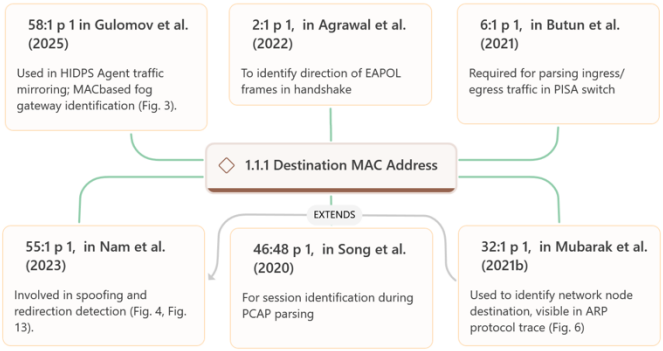


Figure 11: Network View of the research articles on the "Destination MAC Address" Theme

All of them, together, in Figure 11 demonstrate that the Destination MAC Address is a fundamental field for data link layer parsing and anomaly detection in the traditional DPI pipelines. Across recent studies, this area is actively used for key network operations, including identifying sessions, parsing ingress/egress traffic and detecting spoofing as well as device-level classification.

Agrawal et al. [7] show how the Destination MAC field could be used to identify the directionality of Extensible Authentication Protocol over LAN (EAPOL) frames during handshake-based authentication, and thus enabling fine-grained session tracking. According to [5], it is essential in parsing incoming and outgoing traffic in the Protocol-Independent Switch Architecture (PISA) switch environment, which facilitates detailed telemetry. Nam et al. in [17] use Destination MAC parsing to identify spoofing and redirection attacks, especially in particularly within hostile or redirected traffic scenarios. Mubarak et al. in [16] also use the field to determine the destinations of network nodes that can be viewed in ARP traces, which helps to attribute low-level traffic. Song et al. in [8] apply the field to session identification tasks when parsing packet capture (PCAP) sessions, grounding behavioural analysis on MAC-based session granularity. The authors of [18] capitalise on the Destination MAC in host-based intrusion detection and prevention system (HIDPS) agent configurations for fog-based gateway identification via traffic mirroring techniques.

### 3) *Theme 2: Source MAC Address*

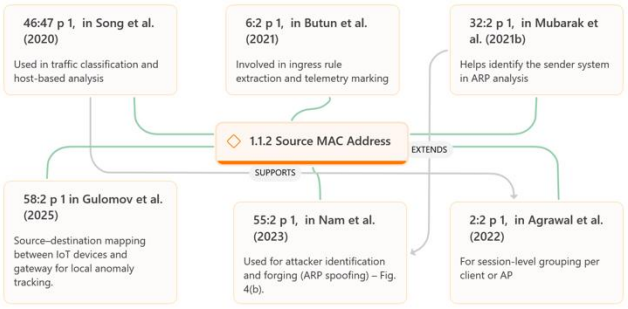


Figure 12: Network View of the research articles on the "Source MAC Address" Theme

The Source MAC Address is essential in tracing the identity of the sender and enforcing security in traditional DPI-based anomaly detection. Recent research uses this field for a broad range of functions, such as ingress rule extraction, telemetry marking, attacker identification, session grouping, host-based analysis, and ARP spoofing as shown in Figure 12.

Butun et al. [5] utilise the Source MAC Address for ingress rule extraction and telemetry marking in programmable network switches. Mubarak et al. [16] use it in the analysis of ARP to detect sender systems that are caught in cyber attack traces. Nam et al. [17] show how it can be applied to attacker identification and MAC address forging detection, particularly to identify ARP spoofing attacks. Song et al. in [8] use it in the field of host-based traffic classification. Agrawal et al. in [7] adopt the use of Source MAC for session-level tracking of clients or access points. In the Internet of Things (IoT) context, [18] apply it to source–destination mapping to support local anomaly tracking.

Taken together, these studies prove the versatility and security relevance of Source MAC Address inspection in different DPI applications. Nevertheless, certain obstacles still remain inconsistent traceability of spoofing, reliance on tool-related results, and insufficient attention to computational efficiency. Improving these limitations will be important in standardising Source MAC Address analysis as a foundational practice in DPI-based anomaly detection systems.

https://doi.org/10.31436/ijpcc.v12i2.711

*4) Theme 3: EtherType / Length*

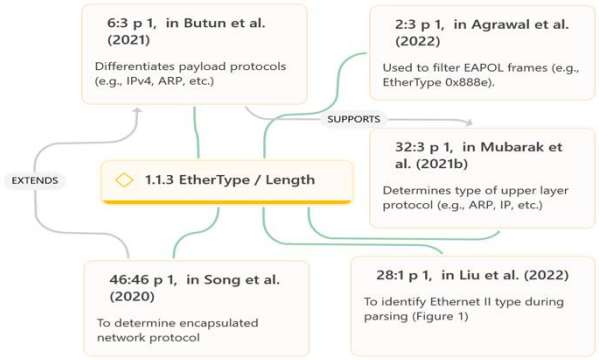


Figure 13: Network View of the research articles on the "EtherType/Length" Theme

The EtherType/Length field is a critical discriminator in traditional DPI-based anomaly detection, enabling precise protocol identification and targeted packet filtering at the data link layer. Recent work employs this field to distinguish between encapsulated payload types (as shown in Figure 13) like IPv4, ARP, and EAPOL, support upper-layer protocol recognition and apply selective filtering for downstream analysis.

Butun et al. [5] use EtherType to differentiate various payload protocols during parsing, ensuring DPI elements are fed with protocol-specific data. Authors of [7] use EtherType filtering to isolate EAPOL and better reflect the authentication traffic segmentation. Liu et al. [19] emphasise that the EtherType is used to determine the types of Ethernet II frames during detailed packet parsing. Mubarak et al. [16] use this area to identify the upper-layer protocol nature in both ARP as well as IP-based flows, contributing to effective protocol dissection in ARP and IP-based flows. EtherType analysis is also used by authors of [8] to detect encapsulated protocols for flow labelling and session classification. Through these contributions, it is clear that EtherType/Length is a pivotal element in facilitating proper differentiation of protocols and interpretation of layers of traffic through DPI pipelines. However, to fully harness its potential, the way forward is to ensure its integration with dynamic protocol stacks and the definition of consistent evaluation criteria.

*5) Theme 5: Frame Check Sequence (FCS)*

The Frame Check Sequence (FCS) field is a specialised field with a critical role in the analysis of the data link layer in traditional DPI-based anomaly detection, especially in contexts that demand robust frame integrity verification. As Figure 14 indicates, the application of its role is not common in the reviewed literature but remains vital where deployed. The authors of [20] use the FCS field to validate the integrity of SD (Service Data) frame types, making sure that only valid, uncorrupted frames are sent to the further inspection layer. This check is vital to prevent false anomalies due to frame-level transmission errors or tampering at the physical or link layer.

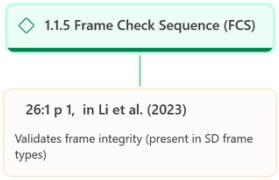


Figure 14: Network View of the research articles on the "Frame Check Sequence (FCS)" Theme

This method strengthens the integrity of subsequent DPI modules by eliminating corrupted frames early in the processing pipeline. Even though the FCS field appears in only one study within the review period, its practical use proves a strategic application of low-level error detection to enhance the accuracy of DPI. With DPI systems also contending with multi-layered threats, such as low-level evasion techniques and manipulation of protocols, a helpful way of providing such defence is to integrate integrity checks such as FCS.

## C. Discussion

The subsequent findings are the summary of the principal themes identified in the thematic analysis of the use of data link layer fields in traditional DPI anomaly detection literature from 2020 to 2025.

In all the papers under review, scholars have always pointed out the enduring relevance of data link layer header fields in conventional DPI pipelines in detecting anomalies. Through the fundamental fields of Destination and Source MAC Address, EtherType/Length, and Frame Check Sequence (FCS), these studies can provide detailed traffic visibility and enable device identification, protocol differentiation, and integrity verification, which do not fully scale to flow-based and machine-learning-based methods.

This value is motivated by two factors: (i) the direct interpretability and determinism of header-level parsing, enabling security mechanisms to detect anomalies at line rate quickly and accurately, and (ii) the interplay between this set of fields with effective pre-filtering or session classification strategies, which provide yet further optimisation to DPI performance. Taken together, the evidence demonstrates that data link layer inspection remains a fundamental, low-latency foundation for high-fidelity anomaly detection, particularly in networks where context-aware packet-level scrutiny is essential.

The present paper discusses the application trend and thematic patterns of data link layer header fields in traditional DPI-based anomaly detection from 2020 to 2025. Code-to-document analysis in ATLAS.ti was able to identify recurring technical themes, however, one of the most notable findings is quantitative: of the six years represented in the ATLAS.ti data, just eight works had explicitly utilized data link layer fields in traditional DPI-based anomaly detection pipelines, and a portion of those also included the inspection of payloads. Such a low rate of adoption is in sharp contrast to the increased utilisation of network, transport, and application layer fields seen in the larger DPI literature. Although core data link layer fields such Destination MAC Address, Source MAC Address, EtherType/Length, and Frame Check Sequence (FCS) demonstrate practical utility for functions such as session identification, protocol parsing, and integrity validation, it can be argued that these fields remain underexplored or are perceived as less effective than more upper-layer elements.

The review claims that link-layer inspection is still applicable in specific cases, such as spoofing detection, device fingerprinting, and frame integrity verification. A relative lack of use may reflect technical constraints, such as reduced visibility in encrypted or tunnelled packets, reduced reliability due to MAC spoofing, or lack of data link parsing tools in most DPI frameworks. In addition, the fact that the VLAN Tag (802.1Q) has not been mentioned in any of the reviewed studies underscores the marginal role currently played by optional data link extensions. It is unclear whether this omission reflects oversight, irrelevance, or implementation difficulty, but it leaves open the potential for further research. Future studies that consider fields of the data link layer as part of multi-layer DPI pipelines and document their relative efficacy can help surface underutilised capabilities. Regardless of the existing constraints, data link field analysis can contribute to better anomaly detection, especially when paired with protocol-specific understanding and hardware-level visibility.

## IV. Implications and Paths for Future Research

According to the survey, data-link layer fields have the potential to support the identification of spoofing, device identification and frame-integrity checks, however, their role in the traditional DPI-based anomaly detection has not been fully explored as only eight related studies have been reported in the six-year period. Moreover, the literature also does not provide abundant reporting on detection-error values, standardised performance baselines or real-world deployment evidence, which would facilitate a reliable appraisal of the practical viability of Layer 2 features. More importantly, the performance improvements reported cannot be attributed to data-link fields only, as other confounding variables such as characteristics of datasets, feature-engineering choices, classifier parameters, traffic mix, hardware acceleration and data pipeline optimisation can equally promote performance gains. In turn, the cross-study performance comparisons are inherently difficult and, in most cases, methodologically unreliable. Future investigations are hence necessary to include more rigorous and transparent evaluation designs, including ablation experiments that disaggregate the incremental worth of data-link fields, cross-dataset validation, and experimentation under a variety of operational circumstances.

The research gap is highly pronounced, as there is no usage of VLAN Tag (802.1Q) at all, despite the widespread use of segmented, virtualised, and multi-tenant networks. This gap highlights opportunities for future research on VLAN-aware anomaly detection strategies. Also, the analysis of data link layer may be combined with multi-layer or machine-learning–assisted DPI to detect encrypted or evasive traffic. Overall, to enhance the utility of data link layer fields in DPI-based security, better reporting practices, extended experimental validation, and hybrid inspection strategies are justified.

## V. Contributions and Benefits of the Study

The present paper is a systematic analysis of the use of data link layer fields in traditional DPI-based anomaly detection between 2020 and 2025. It reports common patterns of application, identifies dominant patterns involving Destination and Source MAC addresses, EtherType/Length, and frame-integrity checks, and highlights significant gaps, particularly limited performance reporting and the complete absence of VLAN Tag analysis. The results provide researchers with concrete directions for future investigation and guide practitioners in understanding which security functions the reviewed evidence associates with each Layer 2 field, including device and session identification, spoofing detection, protocol discrimination, and frame-integrity validation. However, the available evidence is insufficient to determine the independent efficiency contribution of individual Layer 2 fields within traditional DPI pipelines.

## ACKNOWLEDGMENT

The authors would like to thank colleagues who provided feedback on the study
design and manuscript.

## CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

## AUTHOR(S) CONTRIBUTION STATEMENT

All authors contributed equally to the study design, methodology, software development, data analysis, and manuscript preparation. All authors reviewed and approved the final manuscript.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ETHICS STATEMENT

Ethical approval is not required for the publication of the paper.

## REFERENCES

[1] M. Çelebi, A. Özbilen, and U. Yavanoğlu, "A comprehensive survey on deep packet inspection for network security," NOHU Journal of Engineering Sciences, vol. 12, no. 1, pp. 1–29, 2023.

[2] F. A. Khan and A. A. Ibrahim, "Network traffic classification analysis on differentiated services code point using deep learning models for efficient deep packet inspection," International Journal of Innovative Computing, vol. 14, no. 2, pp. 15–24, 2024, doi: 10.11113/ijic.v14n2.438.

[3] Snort Project, "What are community rules?" Snort.org FAQ, 2025. [Online]. Available: https://www.snort.org

[4] M. Çelebi and U. Yavanoğlu, "Accelerating pattern matching using a novel multi-pattern-matching algorithm on GPU," Applied Sciences, vol. 13, no. 14, Art. no. 8104, 2023, doi: 10.3390/app13148104.

[5] I. Butun, Y. K. Tuncel, and K. Oztoprak, "Application layer packet processing using PISA switches," Sensors, vol. 21, no. 23, Art. no. 8010, 2021, doi: 10.3390/s21238010.

[6] S. Mubarak, M. H. Habaebi, M. R. Islam, F. D. Abdul Rahman, and M. Tahir, "Anomaly detection in ICS datasets with machine learning algorithms," Computer Systems Science & Engineering, vol. 37, no. 1, pp. 33–46, 2021, doi: 10.32604/csse.2021.014384.

[7] A. Agrawal, U. Chatterjee, and R. R. Maiti, "kTRACKER: Passively tracking KRACK using ML model," in Proc. 12th ACM Conf. Data and Application Security and Privacy (CODASPY '22), 2022, pp. 1–12, doi: 10.1145/3508398.3519360.

[8] W. Song, M. Beshley, K. Przystupa, H. Beshley, O. Kochan, A. Pryslupskyi, D. Pieniak, and J. Su, "A software deep packet inspection system for network traffic analysis and anomaly detection," Sensors, vol. 20, no. 6, Art. no. 1637, 2020, doi: 10.3390/s20061637.

[9] R. Pawson, T. Greenhalgh, G. Harvey, and K. Walshe, "Realist review—A new method of systematic review designed for complex policy interventions," Journal of Health Services Research & Policy, vol. 10, suppl. 1, pp. 21–34, 2005, doi: 10.1258/1355819054308530.

[10] S. M. Dalkin, J. Greenhalgh, D. Jones, B. Cunningham, and M. Lhussier, "What's in a mechanism? Development of a key concept in realist evaluation," Implementation Science, vol. 10, Art. no. 49, 2015, doi: 10.1186/s13012-015-0237-x.

[11] D. Denyer, D. Tranfield, and J. E. Van Aken, "Developing design propositions through research synthesis," Organization Studies, vol. 29, no. 3, pp. 393–413, 2008, doi: 10.1177/0170840607088020.

[12] J. Noyes, J. Popay, A. Pearson, K. Hannes, and A. Booth, "Qualitative research and Cochrane reviews," in Cochrane Handbook for Systematic Reviews of Interventions, vol. 5.0.1, J. Higgins and S. Green, Eds. Wiley, 2008, pp. 1–18.

[13] M. Ouzzani, H. Hammady, Z. Fedorowicz, and A. Elmagarmid, "Rayyan—A web and mobile app for systematic reviews," Systematic Reviews, vol. 5, Art. no. 210, 2016, doi: 10.1186/s13643-016-0384-4.

[14] A. F. Musfira, N. Ibrahim, and H. Harun, "A thematic review on digital storytelling (DST) in social media," The Qualitative Report, vol. 27, no. 8, pp. 1590–1620, 2022, doi: 10.46743/2160-3715/2022.5383.

[15] M. Zairul, "A thematic review on student-centred learning in the studio education," Journal of Critical Reviews, vol. 7, no. 2, 2020, doi: 10.31838/jcr.07.02.95.

[16] S. Mubarak, M. H. Habaebi, M. R. Islam, and S. Khan, "ICS cyber-attack detection with ensemble machine learning and DPI using cyber-kit datasets," in Proc. IEEE Conf., 2021, pp. 1–6, ISBN: 978-1-7281-1065-3.

[17] J. Nam, S. Lee, P. Porras, V. Yegneswaran, and S. Shin, "Secure inter-container communications using XDP/eBPF," IEEE/ACM Transactions on Networking, vol. 31, no. 2, pp. 672–685, 2023, doi: 10.1109/TNET.2022.3206781.

[18] S. R. Gulomov, T. R. Khudayberganov, T. T. Turdiyev, A. B. Xasanov, and R. R. Raximov, "Hybrid traffic filtering and anomaly detection model for next-generation networks," in Proc. 2025 IEEE Ural-Siberian Conf. Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT), 2025, doi: 10.1109/USBEREIT65494.2025.11054242.

[19] X. Liu, Z. Liu, Y. Zhang, W. Zhang, D. Lv, and Q. Zhou, "TCN enhanced novel malicious traffic detection for IoT devices," Connection Science, vol. 34, no. 1, pp. 1322–1341, 2022, doi: 10.1080/09540091.2022.2067124.

[20] Z. Li, Q. Wei, R. Ma, Y. Geng, Y. Yang, and Z. Lv, "DpGuard: A lightweight attack detection method for an industrial bus network," Electronics, vol. 12, no. 5, Art. no. 1121, 2023, doi: 10.3390/electronics12051121.