Contents lists available at IdNNS

Indonesian Journal of Cyber-AI and Security Intelligence

Volume 1 ; Issue 1

journal homepage: https://journal.idnns.org/index.php/ijcasi

Deterministic Preprocessing Techniques in Deep Packet Inspection-Based Anomaly Detection: A Systematic Thematic Analysis

Adamu Abubakar Ibrahim[a], Shafana Muhammed Shareef[a,b]

[a]Kulliyyah of Information and Communication Technology, International Islamic University Malaysia, P.O. Box 10, Kuala Lumpur, 50728, Malaysia
[b]Department of Information and Communication Technology, South Eastern University of Sri Lanka, University Park, Oluvil, #32360, Sri Lanka

ARTICLE INFO

Keywords:
anomaly detection
cfmo framework
conditional filtering
deep packet inspection
dscp tagging
flow enrichment
preprocessing taxonomy
qos-aware inspection
theme-based review
time-series alignment
traditional dpi pipelines

ABSTRACT

Classical approaches to anomaly detection that utilize traditional Deep Packet Inspection (DPI) are inherently based on stringent preprocessing activities to ensure detection precision, scalable implementation, as well as resistance to advanced network threats. As this preprocessing layer contains heterogeneous, implementation-specific decisions, which are often not comparable in terms of their outcome counts alone, a more traditional systematic review, where such decisions are usually intended to contribute to aggregating headline results and weighing the quality of the study results in a fixed, reproducible fashion, would give only a partial picture of the evidence base. Our goal, however, is not merely to count how often packet reassembly or flow aggregation succeeds, but to trace the recurring patterns and new directions in deterministic pre-processing for traditional DPI-based anomaly detection. The paper is a synthesis and thematically analyzes 38 peer-reviewed articles published in 2020-2025, which involve deterministic preprocessing techniques incorporated in traditional DPI pipelines. Based on the CFMO(Context–Feature–Mechanism–Outcome) theoretical model, the review establishes eight fundamental preprocessing themes, including Packet Capture and Parsing, Stream Reassembly, Payload Fragment Handling, Packet Conditioning, Traffic Sampling and Aggregation, Quality of Service (QoS) Tagging, Session Reconstruction and Time-Series Alignment, and Flow Enrichment. Analysis reveals that Packet Capture and Parsing remains the predominant technique, present in 87 percent of the examined literature, with consistent usage over the period. In contrast, higher level mechanisms like flow enrichment and QoS based tagging have seen limited implementation. Notably, the focus has largely been placed on Session Reconstruction and Time-Series Alignment because it enables an organisation to organise data to be used in a behaviour-aware or flow-based detector. Stream reassembly and Fragment Handling are still essential to protocol integrity, but have visibly declined in use over the last few years. The paper uses thematic diagrams and trend charts to depict the hierarchical and interrelated nature of modern preprocessing mechanisms. This review confirms that DPI preprocessing is a dynamic sequence of intelligent modules, ranging from accurate packet acquisition to flow reconstruction with semantically rich contents, designed to handle such issues as network fragmentation, encryption, evasion, and high data rates. Moreover, it determines the main gaps in research, such as the lack of comparative evaluation of the sampling methods and underutilisation of DSCP/ToS fields in adaptive policy routing. The study consolidates existing practices and new directions in research and offers practical guidance to scholars and developers who want to improve DPI capabilities through next-generation preprocessing architectures.

1. Introduction

Deep-packet inspection (DPI) engines scan all the bytes of a packet to identify network anomalies. However, the raw traffic which enters these engines is frequently fragmented, mis-sequenced, or encoded in non-standard ways. Pre-processing serves as the interface between this "messy" reality and the deterministic signature logic that follows. By pre-processing traffic, anomaly-detection results depend on the threat itself, not on quirks of packet delivery or protocol variation.

1.1. Pre-processing in Traditional DPI-Based Anomaly Detection

Traditional DPI appliances first reassemble and normalise packets in a single, integrated pipeline, joining IP fragments, restoring TCP order, and translating headers, encodings, and URLs into a canonical form that strips out evasion tactics such as overlapping bytes or disguised paths (Arshad et al., 2023). Then, packets are aggregated into flows that share a five-tuple, with each packet enhanced with timing and size characteristics by the same pipeline, such as

*Corresponding author. Phone: +94 773626707.
adamu@iium.edu.my (A.A. Ibrahim); zainashareef@seu.ac.lk (S.M. Shareef)
ORCID(s):

inter-arrival gaps, total bytes, and session duration. Adaptive time-outs discard idle flows while retaining long, low-rate exchanges which typically carry stealth attacks (Radamski et al., 2024; Koumar et al., 2023).

According to vendor documentation on SonicWall IPS and Cisco Firepower, enabling this combined stage reduces the false-positive rates and maintains line rates up to 100 Gb/s. Lastly, publicly available corpora including the ICS-Flow repository release both unprocessed packets and sequences of pre-processed packets so that a researcher can use them to provide a ground-truth benchmark for validating cleaning routines before tuning anomaly signatures (Dehlaghi-Ghadim et al., 2023). All of these measures provide DPI detectors with dependable, context-rich data to deliver accurate real-time alerts.

1.2. Objective of the Thematic Review

This thematic review examines the usage pattern on how packet- and flow-level pre-processing has been applied in deterministic Deep Packet Inspection (DPI) systems for anomaly detection between 2020 and 2025. We define pre-processing as the deterministic treatment of raw traffic such as fragment reassembly, header normalisation, TCP-stream ordering, and flow aggregation carried out before any signature or grammar check (Arshad et al., 2023). To keep the scope clear, learning-based or adaptive pipelines are excluded so that attention remains on preprocessing with byte-accurate, Deterministic and rule-based DPI methods. The review pursues two linked goals: first, to record where, why, and by whom these deterministic steps are embedded in cloud, industrial, service-provider, and enterprise DPI stacks; and second, to identify the technique-level patterns such as fragment-reassembly variants, header-normalisation rules, and flow-timeout strategies, and trace how their use shifts across thirty-eight studies analysed. By coding these papers we map the research landscape for scholars who optimise deterministic pipelines, engineers who maintain DPI rule sets, and educators who teach byte-level inspection. Accordingly, the guiding question is:

What patterns and trends characterise deterministic packet- and flow-level pre-processing in traditional DPI-based anomaly detection, as reported between 2020 and 2025?

1.3. Theoretical Framework

We use a Realist CFMO(see Figure 1) lens grounded in CMO logic (Pawson et al., 2005) and the view that a mechanism comprises resources + reasoning (Dalkin et al., 2015); here we make the resource explicit as a Feature, akin to CIMO extensions (Denyer et al., 2008). Features are deterministic pre-processing steps that precede traditional DPI such as capture & parsing, TCP reassembly, payload/fragment handling, conditioning/normalization, sampling/aggregation, QoS tagging, session reconstruction, and flow enrichment. These Features enable Mechanisms such as canonical byte streams & header normalization, stateful ordering, fragment reconciliation, scale control, QoS-aware branching, session alignment, context enrichment, and integrity/consistency checks. Context includes micro-contexts explicitly reported (e.g., high-speed targets and offloads—SmartNIC/DPDK/NBA; deployment—cloud/ISP/enterprise/ICS; encryption/evasion pressure; toolchain/capture settings; flow-timeouts/LB/asymmetry; public corpora) and a macro-context (proposing country/region). Outcomes summarize, per Feature→Mechani usage patterns (counts, bundles, trends), advantages, and limitations. Because performance is rarely attributable to a single step, outcomes are synthesized qualitatively (Pawson et al., 2005).

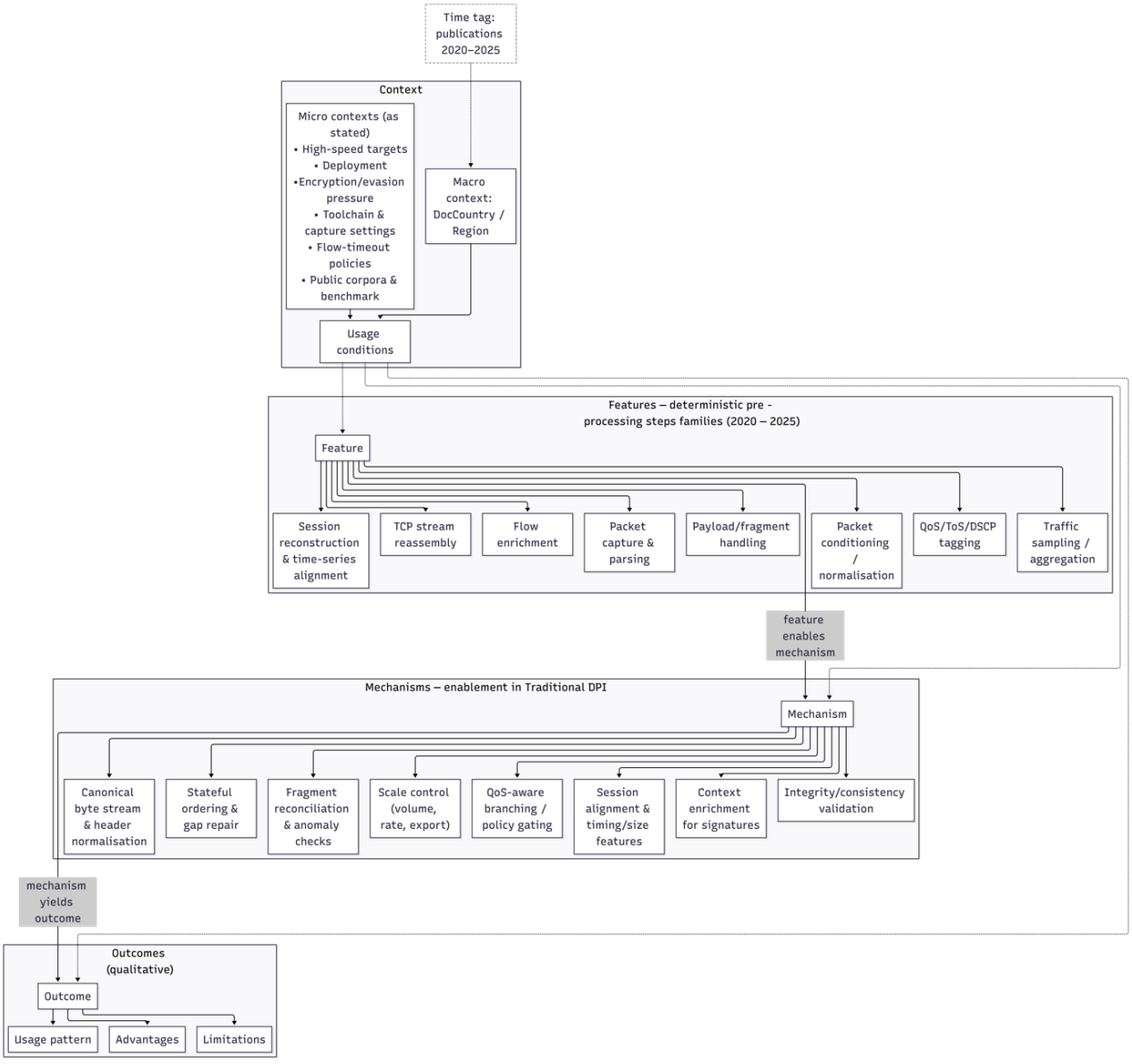

Figure 1: CFMO based Theoretical Framework of this Thematic analysis

2. Materials and Methods

The first author, Sh, whose current research concentrates on packet- and flow-level pre-processing within traditional deep-packet-inspection (DPI) pipelines. Accordingly, author is well versed in capture libraries, fragment reassembly, header normalisation, and flow aggregation. Other co-authors have likewise completed several studies on traffic conditioning for deterministic inspection, and their research interests cover SmartNIC off-load, flow-timeout tuning, and related pre-processing techniques. Thus, all of us understood what a review on DPI pre-processing would demand. Because Sh routinely refines pre-processing modules, and predisposed to regard these steps as crucial for accurate anomaly detection. To temper this inclination, the review was structured around strict inclusion criteria and a transparent coding protocol that admitted only peer-reviewed studies with explicit descriptions of deterministic pre-processing. Even so, some of the themes and conclusions that surfaced may still reflect our professional leaning toward the value of pre-processing in traditional DPI.

2.1. Rationale for Employing a Thematic Review

Systematic reviews are typically designed to aggregate headline results and evaluate study quality in a fixed, replicable manner (Noyes et al., 2008). Our goal, however, is not merely to count how often packet reassembly or flow aggregation succeeds, but to trace the recurring patterns and new directions in deterministic pre-processing for traditional DPI-based anomaly detection. Because the research problem touches on varied, practice-oriented topics such packet-capture libraries, fragment reassembly rules, header-normalisation choices, and flow-timeout policies, in this

Table 1
Search Strings from Eight Academic Databases

Database	Boolean query for Search String	Results
ScienceDirect	(“Deep Packet Inspection” OR “Traffic Inspection” OR “Packet Filtering” OR “Packet Inspection” OR “Payload Inspection” OR “Encryption Inspection”) AND “Anomaly Detection”.	317
Emerald Insight	(“Deep Packet Inspection” OR “Traffic Inspection” OR “Packet Filtering” OR “Packet Inspection” OR “Payload Inspection” OR “Encryption Inspection”) AND “Anomaly Detection”.	12
ProQuest	(“Deep Packet Inspection” OR “Traffic Inspection” OR “Packet Filtering” OR “Packet Inspection” OR “Payload Inspection” OR “Encryption Inspection”) AND “Anomaly Detection”.	764
Scopus	TITLE-ABS-KEY("Deep Packet Inspection" OR "Traffic Inspection" OR "Packet Filtering" OR "Packet Inspection" OR "Payload Inspection" OR "Encryption Inspection") AND "Anomaly Detection" AND PUBYEAR > 2019 AND PUBYEAR < 2026 AND PUBYEAR > 2019 AND PUBYEAR < 2026 AND (LIMIT-TO (SUBJAREA , "COMP")) AND (LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "re")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SRCTYPE , "j") OR LIMIT-TO (SRCTYPE , "p")) AND (LIMIT-TO (PUBSTAGE , "final"))	596
Taylor & Francis Online	[[All: "deep packet inspection"] OR [All: "traffic inspection"] OR [All: "packet filtering"] OR [All: "packet inspection"] OR [All: "payload inspection"] OR [All: "encryption inspection"]] AND [All: "anomaly detection"] AND [Publication Date: (01/01/2020 TO 31/12/2025)]	18
IEEE Xplore	("All Metadata":"Deep Packet Inspection" OR "All Metadata":"Traffic Inspection" OR "All Metadata":"Packet Filtering" OR "All Metadata":"Packet Inspection" OR "All Metadata":"Payload Inspection" OR "All Metadata":"Encryption Inspection") AND ("All Metadata":"Anomaly Detection")	110
ACM Digital Library	[[All: "deep packet inspection"] OR [All: "traffic inspection"] OR [All: "packet filtering"] OR [All: "packet inspection"] OR [All: "payload inspection"] OR [All: "encryption inspection"]] AND [All: "anomaly detection"] AND [E-Publication Date: Past 5 years]	405
Wiley Online Library	("Deep Packet Inspection" OR "Traffic Inspection" OR "Packet Filtering" OR "Packet Inspection" OR "Payload Inspection" OR "Encryption Inspection") AND "Anomaly Detection"" anywhere.	79

case a thematic review provides the most suitable lens. It lets us cluster diverse studies under common conceptual headings and highlight subtleties that a purely quantitative meta-analysis would likely miss.

2.2. Search String

The literature search aimed to capture recent work on deterministic pre-processing within traditional DPI pipelines that serve anomaly-detection tasks. Because authors rarely label a study “pre-processing” in isolation, we did not run a separate query for that term; doing so would have surfaced many articles on generic traffic grooming that never feed a DPI engine. Instead, we searched for publications that clearly position pre-processing as an early stage of a Deterministic and rule-driven DPI workflow. To cover the varied language used in the field, we built deliberately broad Boolean strings as presented in the following Table 1.

Table 1: Search Strings from Eight Academic Databases

and applied it uniformly across eight scholarly databases (Scopus, IEEE Xplore, ACM DL, ScienceDirect, ProQuest, Emerald Insight, Wiley Online Library, and Taylor & Francis Online). The query was limited to English-language, peer-reviewed journal and conference papers published between 2020 and 2025.

2.3. Inclusion and Exclusion Strategy

To keep methodological rigor and thematic focus, we applied a set of criteria. Studies were included if they (i) lay within the 2020-2025 window, (ii) were written in English, (iii) presented preprocessing with traditional DPI techniques for anomaly detection, and (iv) appeared in a peer-reviewed publication. Papers were excluded if they pre-dated to 2020, were non-English, or came from non-refereed outlets, and retracted.

The dataset was refined through a four-stage filtering process. First, a title screen removed records that were obviously unrelated to traditional DPI. Second, an abstract review eliminated papers whose scope or detail did not match the study objectives. Third, diagonal (skim) reading verified that each remaining article contained substantive content on traditional DPI rather than a passing reference. Studies that only discussed high-level monitoring or traffic summarisation without touching packet- or flow-level preprocessing as part of traditional DPI were excluded.

Figure 2 (PRISMA diagram) traces the selection flow. From 2301 initial records, 186 duplicates and 62 entries flagged by automation were removed. Manual screening of the remaining 2053 records, followed by the staged filtering above, yielded 38 primary studies for qualitative synthesis and thematic analysis of pre-processing trends in traditional DPI-based anomaly-detection pipelines.

2.4. The Thematic Review Process

Screening and coding followed a multistep workflow modelled on established qualitative-review practice. First, all 2301 search hits were imported into Rayyan web tool, whose duplicate-detection and blind-screen functions accelerated title-and-abstract triage (Ouzzani et al., 2016). After applying the inclusion and exclusion rules, 38 papers remained and were exported to Mendeley for final metadata cleaning and DOI verification. Full texts were then read in detail, and a brief analytic memo capturing each study's core pre-processing contributions was drafted. These memos, along with the PDFs, were loaded into ATLAS.ti 25, following the project-setup guidelines of a study (Musfira et al., 2022; Zairul, 2020). Inside ATLAS.ti(see Figure 3), documents were first grouped by publication year (2020 – 2025) and then by the high-level pre-processing categories defined earlier. Initial codes captured country, venue, and the specific pre-processing step reported. Later, similar codes were merged and elevated into broader themes. Iterative refinement produced a compact set of themes that reveal how deterministic pre-processing for DPI-based anomaly detection has evolved year by year, pinpointing the design patterns that dominate the 2020-2025 corpus.

Based on the codes generated in ATLAS.ti, the thirty-eight papers were organised into the hierarchical scheme illustrated in Figure 4 to answer our guiding question, Because any taxonomy reflects both the material reviewed and the researchers' interpretive lens, we adopted the smallest set of layers that still captured the distinct engineering choices visible across the corpus. As first-level, all codes fall under two parent themes that mirror the natural order of a DPI pipeline: Packet-Level Pre-processing and Flow-Level Pre-processing. Second-level themes splits into several technical strands such as Packet Capture and Parsing, TCP Stream Reassembly, Payload Fragment Handling, Packet Conditioning, Traffic Sampling / Aggregation, QoS Tagging, Session Reconstruction and Time-Series Alignment, Flow Enrichment. This layered structure preserves fine-grained methodological detail, down to specific capture libraries or reconstruction rules, while still exposing the broader patterns that link related practices across studies. Theme names were taken from the dominant terminology in each code cluster and align with the architecture sketched in contemporary DPI researches, ensuring consistency with established vocabulary while reflecting the empirical trends of the 2020-2025 literature window.

Figure 5 visually summarises the field with a word cloud created from key terms extracted across all included studies. The recurring use of such terms as "packet," "flow," "parsing," "libpcap," "session," "timing," "aggregation," and "alignment", even before the formal coding, highlighted the importance of both packet- and flow-level pre-processing in traditional DPI-based anomaly detection. The use of high-frequency words like "capture," "reconstructed," "feature," "tcp," and "timing" further corroborated the key thematic strands, emphasising how practical work in the field consistently revolves around handling raw packets, rebuilding flows, and extracting session-level context prior to anomaly detection. This early word-cloud analysis supported the coding practice and thematic taxonomy that was developed in later stages of the review

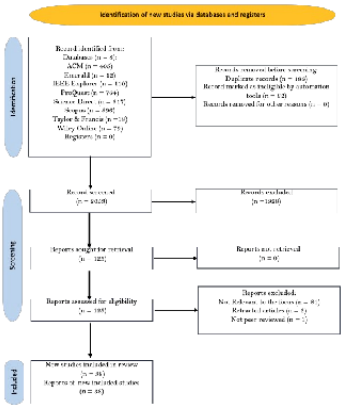

Figure 2:

PRISMA Flow Diagram for Article Selection Process

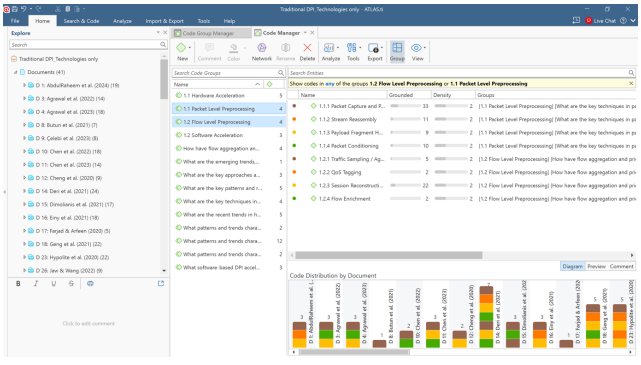

Figure 3: Established Code Group in ATLAS.ti

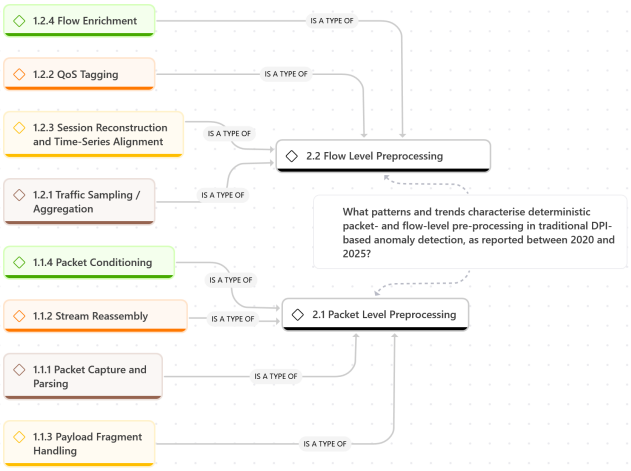

Figure 4: Thematic Code layer Taxonomy

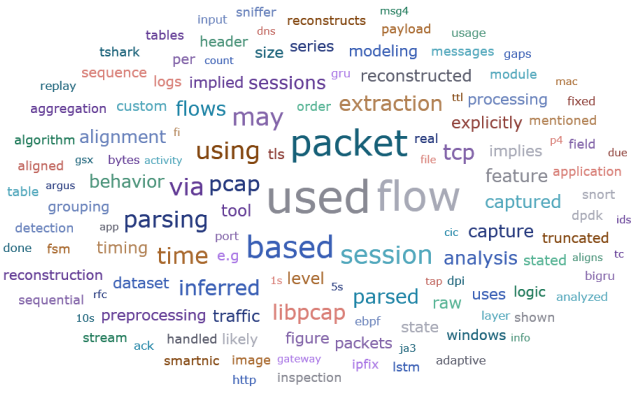

Figure 5: Word Cloud Generated from Articles

Table 2
Articles reviewed based on types of journals and conferences

Journal/Conference Name	2020	2021	2022	2023	2024	2025
ACM Conference	1	3	2	1	1	
Applied Sciences (Switzerland)				1		
Computer Networks					1	
Computer Systems Science and Engineering		1				
Connection Science			1			
Electronics (Switzerland)				1	1	
IEEE (CSR Conference)		1				
IEEE (S&P Conference)					1	
IEEE Access		1			1	
IEEE Conference	1	2	1	1		1
IEEE Conference (CCWC)			1			
IEEE Conference (TCSET)	1					
IEEE/ACM Transactions on Networking		1		1		
IEEE Trans. on Dependable and Secure Computing						1
International Journal of Information Technology (Singapore)					1	
International Journal of Information Technology and Web Eng.			1			
Journal of Theoretical and Applied Information Technology				1		
Mathematical Problems in Engineering		1				
PeerJ Computer Science			1			
Scientific Reports					1	
Security and Communication Networks	1					
Sensors		1				
Sensors (Switzerland)	1					

3. Results and Discussion

The main results of this thematic review are given in two parts. The quantitative analysis provides the count of studies that fit into each pre-processing theme, making use of the taxonomy of packet-level and flow-level procedures. It also monitors annual publication patterns since 2020 through to 2025, and it also shows the regional publishing activity by country and continent. The outlet profile created on the basis of the databases reveals that there are 15 journal articles and 23 conference papers(Table 2), yielding a curated corpus of 38 articles for this thematic review. The qualitative analysis proceeds to provide a theme-by-theme analysis, describing how each of the studies incorporates deterministic pre-processing into a more traditional, deterministic DPI pipeline for anomaly detection. Together, these analyses not only provide the magnitude and geographic distribution of the recent work, but also the practical methods, recurring engineering patterns, and evolving trends that shape current pre-processing strategies in rule-based DPI pipelines.

Table 2: Articles reviewed based on types of Journals

Figure 6 shows that in the year 2021, more scholarly articles focused on pre-processing as a part of DPI anomaly detection appeared than in any previous year (11). Even though the number dropped to seven in 2022 and six in 2023, it was still relatively steady in 2024 with seven publications. This trend shows an increased and relatively stable interest in pre-processing methods. This decline to two articles in 2025 may be due to more integrated or machine-oriented methods of inspection, or a temporary change in the research focus. Taken together, all these yearly changes highlight the changing research environment in which pre-processing remains a relevant aspect, as network traffic becomes increasingly complex and encrypted and solutions to DPI-based risks are becoming increasingly complex.

The total number of publications by country was calculated by adding all author affiliations of an article and counting each country once per publication, ensuring that multinational collaborations were represented accurately. Figure 7 indicates that China has led pre-processing research in the background of conventional DPI anomaly detection, with two or three articles every year and reaching its peak of three publications in 2022. The next ones are the United States and India with more fluctuating outputs. In the case of the U.S., there was consistently one to two papers each year before a dip to zero in 2023, after which the numbers recovered slightly with one article in 2024. India was somewhat active as it published one or two articles per year during 2021–2023. The Republic of Korea, Malaysia, and a few European countries (Poland and Switzerland) were active across multiple years. On the other hand, Turkey, Pakistan, and Italy were intermittent contributors, having been observed only in certain years but not consistently. The long tail includes the countries of Australia, Nigeria, Kuwait, Uzbekistan, Netherlands, and The Gambia, each providing one paper, reflecting the expanding global interest in DPI preprocessing research. It is noteworthy that Uzbekistan and the Netherlands appear in 2025, which indicates that this area is becoming increasingly popular with the developing research communities. Even though there was a decline in output of certain core nations in 2023, the sustained activity in China and newcomers in 2024 and 2025 suggest a vibrant, globalised research practice, which is the result of the maturation of established research centres and the diverse international perspectives shaping the development of pre-processing methodologies to identify DPI anomalies.

As indicated in Figure 8, the literature covering preprocessing in traditional DPI anomaly detection spans over 25 countries, accounting for a total of nearly 38 publications. China has topped the list with 12 contributions, followed by the United States (6). India (4), and notable multi-year

contributions from Malaysia, the Republic of Korea, and Pakistan. This consistent national effort indicates the development of their growing cybersecurity research capacity and support for network-defense innovation. The existence of Finland, The Gambia, Uzbekistan, and Kuwait among other countries only adds to the global diversity in contributions. This worldwide allocation consequently confirms the trend and growing importance of pre-processing studies in traditional DPI anomaly detection pipelines, which indicates the involvement of old and new research groups in the development of the research area.

3.1. Qualitative Findings

The qualitative results based on the 38 papers, summarised in Figures 9 and 10, suggest that Packet Capture and Parsing is easily the most commonly used pre-processing operation in traditional DPI anomaly detection pipelines, occurring in 33 out of 38 articles (≈87%). It is the only methodology that has been applied in the entire six-year implementation period (2020–2025), and it is of paramount value in almost all implementations. The second most commonly used methods are the Session Reconstruction and Time-Series Alignment that are used in 22 articles (≈58%). This approach was also used in all years of the review, especially heavily in 2020–2023, thus confirming that it is a means of allowing the detection of anomalies based on flow.

Other functions, such as Stream Reassembly, Payload Fragment Handling and Packet Conditioning were not as widely adopted but still have a role to play in the handling of protocol fragmentation and layered payload issues. Stream Reassembly was used in 11 studies (29%), primarily between 2020 and 2024, but not in 2023 or 2025. Payload Fragment Handling was present in 9 studies (24%), showing steady use from 2020 through 2024 before disappearing entirely in 2025. Packet Conditioning was used in 10 studies (≈26%). Its adoption gained traction starting in 2021 and maintained a consistent presence through 2024, but is not found in 2020 or 2025, possibly due to the influence of normalisation and filtering trends at the time. Traffic Sampling/Aggregation appeared in five studies (13%), with a very large spike in 2021 when the technique was used in four studies, indicating sporadic but context-sensitive adoption for resource optimization. The least used features are QoS Tagging and Flow Enrichment, which are represented in only two works (≈5%). QoS Tagging was used only in 2020 and 2025, suggesting isolated exploration of DSCP-based techniques at the beginning and end of the review period. Flow Enrichment appeared in 2021 and 2022 but was not observed in later years, indicating that higher-layer semantic enrichment is still an underutilised feature in traditional DPI pre-processing chains. Overall, even though core capabilities such as packet capture and session reconstruction dominate consistently, the limited uptake of tagging and enrichment functions reveals a promising direction for future innovation.

3.2. Qualitative Findings

This part summarises the qualitative trends in the preprocessing mechanisms of classic DPI anomaly detection. All of the articles were systematically coded against the functional taxonomy presented in Figure 4. The primary mechanisms, optimisation strategies, deployment scenarios, and reported or inferred limitations were distilled for each category, and how later studies refined, extended, or validated previous designs was also tracked. Figure 9 visualizes the comprehensive mapping of preprocessing functions across the reviewed studies, providing a clear depiction of functional coverage and overlap.

The large volume and interrelationship of the codes make a single, comprehensive network (polyline) view visually impractical. To improve clarity and interpretive value, the polyline is further segmented into sub-figures, each addressing a key pre-processing task and the inter-article relationships. Such staged visualisation allows for fine-grained inspection of methodological clusters while preserving the integrity of the review's global structure. It should be noted that the page numbers referenced in our quotation identifiers are page numbers in our internally prepared summary, not the original published articles' page numbers.

3.3. Theme 1: Packet Capture and Parsing

The use of packet capture and parsing became a core functionality in virtually all the systems studied(Figure 11) that use DPI to detect anomalies. This type of functionality is backed by an extensive collection of implementations, including open-source software libraries as well as custom capture modules within more complex frameworks, as illustrated in the diagram. Many of these studies (e.g., (Chen et al., 2022; Gulomov et al., 2025; AbdulRaheem et al., 2024)) make use of commonly available tools such as libpcap, WinSock, or DPDK, often in conjunction with packet analyzers like Wireshark, Snort, or Suricata. In other systems, the process of parsing is performed through platform-specific tools; for instance, Song et al. (2020) utilize WinSock 2.2, while Li et al. (2023) apply frame parsing from PROFIBUS relay traffic.

Various papers highlight the refinement or extension of fundamental capture routines. Park et al. (2021), for example, benchmark capture performance using DPDK, whereas SmartNIC logic is added to PF_RING, Netcope, or Napatech. Rawat et al. (2022) and Pratomo et al. (2021) highlight the use of custom-designed or scripted capture modules, including how the developers adapt toolchains to suit DPI-specific workloads. In addition to mere acquisition, parsing modules usually identify structured fields (TCP/UDP headers, timestamps, and session identifiers). This parsed data forms the foundation of further processes: Moon et al. (2024) extract finite state machine transitions from raw traces; Khan and Abdul Rahman (2023) describe system flowcharts that integrate parsing logic into custom classifiers; and Agrawal et al. (2023) apply Python-based sniffers to support lightweight feature extraction from Wi-Fi traffic.

Packet capture, which has been embedded as a subroutine in several studies, also appears within larger preprocessing or filtering pipelines. Dimolianis et al. (2021) and Kim et al. (2021) incorporate capture logic into the XDP

and OpenSSL stacks, respectively, optimising for speed and low-level parsing. Conversely, Lin et al. (2024) and Deri et al. (2021) suggest capture is implicitly embedded within broader PCAP or NetFlow exports, especially in environments involving extension-header parsing or passive probes. Taken as a whole, the literature reviewed represents a broad range of capture techniques, ranging from low-level kernel-based methods to high-performance SmartNIC offloads. However, few studies provide comparative benchmarks on packet loss, interface bottlenecks, or timestamp fidelity, and most omit details on how capture accuracy was validated. In future, more transparent reporting of these metrics, along with shared testbeds, would improve repeatability and inform best practice in the DPI research community.

3.4. Theme 2: Stream Reassembly

Reconstruction of message streams is also a critical element in traditional DPI pipelines, since it allows message fragments of TCP or UDP packets to be reassembled into coherent session streams for deep inspection. Several studies as you can see in Figure 12 have explicitly referenced this process. As an example, Liu et al. (2022) refer to TCP data stream capture, while Song et al. (2020) show that both TCP and UDP traffic are reassembled using session tables. Similarly, Einy et al. (2021) provides native support for stream reassembly in Suricata, and Deri and Fusco (2021) utilizes reconstructed flows for advanced analysis tasks such as TLS handshake parsing and HTTP inspection.

Some of the scholars integrate or assume stream reassembly as part of broader preprocessing stages. In Geng et al. (2021), bidirectional request-response sessions are reconstructed, whereas Liang and Kim (2022) infer flow-level reconstruction within application gateway architectures. AbdulRaheem et al. (2024) suggest that Snort-type preprocessors may handle TCP reconstruction implicitly, while Rawat et al. (2022) adopts a flow-based reconciliation approach. At the same time, Pratomo et al. (2020) reconstructs application-layer messages from lower TCP layers prior to protocol inspection.

More refined techniques appear in studies like Moon et al. (2024), who model sequence/ack manipulation to track connection state transitions, and Hypolite et al. (2020), who address out-of-order queuing and stateful stream logic. These studies emphasize robustness, particularly in high-throughput or lossy settings. Overall, many of the papers rely on built-in libraries or intrusion detection system (IDS) platforms to reassemble the streams, while a few pursue custom logic to ensure correct session assembly in complex traffic conditions—thus making it a vital enabler of accurate, protocol-aware DPI analysis.

3.5. Theme 3: Payload Fragment Handling

Payload fragmentation is a salient preprocessing step in DPI systems that ensures that packet-level fragments are reassembled or normalised before deeper inspection. This function has explicit implementations in several studies as depicted in the above Network View of Figure 13. After decompressing encapsulated payloads, Song et al. (2020) reconstruct flows, and Geng et al. (2021) reconstruct entity bodies on the session level to support application layer analysis. Liu et al. (2022) also performs payload feature extraction on the reconstructed application data, highlighting its usefulness in preparing the content for downstream classifiers. In Deri and Fusco (2021), fragmentation management is implicitly included in the logic of TCP stream reconstruction; Hypolite et al. (2020) addresses fragmented and out-of-order segments using DMA-assisted techniques.

Other works adopt specialized or indirect strategies. Pratomo et al. (2020) truncate large messages, meaning that fragment management may occur earlier in the pipeline, while Liang and Kim (2022) imply fragment processing within their scanning architecture, though details are not fully described. Some approaches aim to normalise rather than reassemble, something that Chen et al. (2023) achieve through fixed-length padding to ensure payload consistency, or Lin et al. (2024) achieve with an in-depth study of the headers of fragments, implementing rule-based processing to ensure uniformity before analysis.

Taken together, these methodologies indicate that payload fragment handling is implemented through a mix of explicit reassembly, out-of-order correction, size normalization, and header-based inspection. On the one hand, detailed mechanisms have not always been documented, but on the other, the literature reflects a common objective: that fragmented or irregular payloads do not undermine the completeness, accuracy, or reliability of DPI-based anomaly detection.

3.6. Theme 4: Packet Conditioning

The use of packet conditioning encompasses a diverse set of preprocessing operations aimed at normalizing, filtering, and structuring packet data before deep inspection. The reviewed articles(see Figure 14) implement several strategies to improve data consistency and reduce noise. Shlingbaum et al. (2024) and Chen et al. (2023) focus on direct payload access and removal of hidden or irrelevant content prior to vectorization—a critical step for machine learning pipelines. Kang et al. (2024) addresses input redundancy by implementing packet deduplication and source filtering, while Agrawal et al. (2022, 2023) apply state-machine logic and group packets by EAPOL sessions, maintaining sequence integrity and reducing the risk of replay attacks.

Other studies enhance formatting or apply inline processing. Geng et al. (2021) employ ASCII normalization and image formatting for content transformation, and Deri and Fusco (2021) introduces inline classification that parses traffic in real time. At a more advanced level, Nam et al. (2023) demonstrates eBPF-based redirection for inline packet filtering, and Dimolianis et al. (2021) filters DNS flows to focus on relevant features. Lin et al. (2024) handles complex protocol structures using a finite state machine-driven adaptive parsing approach to efficiently process chained extension headers. These methods show that packet conditioning combines normalization, classification, filtering, and adaptation

strategies—each tailored to meet the inspection objectives of modern DPI systems.

3.7. Theme 5: Traffic Sampling / Aggregation

The preprocessing functions of traffic sampling and aggregation enable reduced data volumes and improved scalability by summarizing flow-level behaviors in DPI systems. As of Figure 15, several studies propose distinct methods for condensing network traffic and managing flow visibility. Panda et al. (2021) introduce a hybrid approach combining FlowCache and sketch-based logic to aggregate flow statistics such as microburst detection and cardinality estimation. Ma et al. (2021) infers aggregation behavior through flow hash tables, using sketch techniques to emulate real-time summary views. Gulomov et al. (2025) shifts aggregation to the controller plane, where an SDN controller collects flow records centrally, enhancing overall visibility. Another subject of focus is time-based sampling. Dimolianis et al. (2021) evaluates mitigation efficacy by varying sampling windows across 1 s, 5 s, and 10 s, whereas Deri and Fusco (2021) supports standard NetFlow/IPFIX formats to enable broader flow monitoring and external integration. Together, these studies demonstrate the strategic use of sampling and aggregation to reduce computational overhead, balance inspection depth and coverage, and ensure scalable flow analytics. Though techniques vary from sketching and caching to time-window sampling, they collectively aim to improve DPI responsiveness and visibility in high-speed networks.

3.8. Theme 6: QoS Tagging

Quality-of-Service (QoS) tagging remains underutilized but important in conventional DPI preprocessing pipelines. This capability is addressed in only a few studies evidenced by Figure 16, reflecting its limited adoption in mainstream workflows. Hypolite et al. (2020) demonstrates how DPI operations can influence the Type of Service (ToS) field for routing behavior, highlighting the potential of inline classification to redirect traffic paths. Gulomov et al. (2025) explicitly incorporates Differentiated Services Code Point (DSCP) values into its detection logic, achieving approximately 91% detection success, as shown in their evaluation.

Although the corpus is limited, these findings reflect a growing interest in leveraging ToS, DSCP, and QoS labels not only for traffic prioritization but also as features for intelligent flow classification and policy enforcement. Future DPI systems could benefit from tighter integration of QoS-based tagging, especially in latency-sensitive or encrypted environments where conventional inspection techniques face challenges.

3.9. Theme 7: Session Reconstruction and Time-Series Alignment

Session reconstruction and time-series alignment are two fundamental processes used to convert raw packets into meaningful flow-level or behavioral timelines in DPI pipelines. As in Figure 17, across the literature, these techniques appear in diverse forms, ranging from rule-based state tracking to learning-driven sequence modeling. Agrawal et al. (2023) employ a state machine where gaps between protocol messages signal key transitions, while Pratomo et al. (2020) preserve message boundaries or rely on SSH retry/timeout behaviors to align flows accurately. Rawat et al. (2022) derive session-level features such as packet counts and durations, and Moon et al. (2024) apply full connection state timing to support symbolic attack derivation.

Other works emphasize temporal modeling. Liu et al. (2022) apply sequential flow modeling through an STCN architecture, whereas Xiang et al. (2024) use sliding window segmentation to structure time-aligned flow data. Protocol-driven alignment also appears in Lee et al. (2022), where TCP ports and timestamps guide direction inference. At a finer granular level, Lin et al. (2023) track message transitions using a finite-state machine, and Nam et al. (2023) reconstruct directional sessions to support policy matching. Deri and Fusco (2021) implies real-time flow alignment through monitoring tools, and Einy et al. (2021) infer session boundaries from logged or observed states.

Together, these studies reveal a rich spectrum of alignment strategies — from timing gap reconstruction and behavioral markers to window-based segmentation and state-machine modeling — highlighting the critical role that session reconstruction and time-series alignment play in enabling context-aware, behaviorally meaningful DPI analysis.

3.10. Theme 8: Flow Enrichment

Flow enrichment appears as a specialised yet valuable preprocessing function in traditional DPI pipelines, though only a small number of studies(see Figure 18) implement it explicitly. In Chen et al. (2022), a flow rate classification is introduced using a dual-stack memory structure, enriching each flow with rate-based characteristics that support more granular traffic differentiation. Similarly, Deri and Fusco (2021) enhances DPI by incorporating bidirectional flow context, using features such as per-session JA3 fingerprints to characterize encrypted traffic more precisely. Together, these studies demonstrate that while flow enrichment remains one of the least commonly applied preprocessing techniques, it offers meaningful improvements by providing session-aware contextual information that downstream DPI and anomaly detection modules can leverage for higher accuracy and robustness.

3.11. Findings and Discussion

The following findings were observed from the eight themes identified in the thematic analysis of preprocessing methods within traditional DPI-based anomaly detection studies.

3.12. Preprocessing as the Central Role of DPI Pipelines

Across all analysed papers, researchers consistently highlight the importance of preprocessing operations, including packet capture, stream reassembly, fragment handling, and session alignment, in preparing network traffic

for effective DPI. These preprocessing modules extract, normalize, and structure raw packet data, enabling downstream DPI engines to operate with greater speed and precision. By segmenting traffic, reconstructing sessions, and standardizing payloads, preprocessing ensures that anomaly detection algorithms work on complete and relevant data, rather than fragmented or noisy input.

The enduring value of preprocessing in modern DPI pipelines can be attributed to two key factors: (i) advances in capture and parsing technologies supported by scalable, open-source tools and efficient hardware interfaces now enable high-throughput, low-latency data handling at the packet level; and (ii) the integration of adaptive mechanisms such as traffic sampling, session grouping, and selective enrichment, which filter and prioritize traffic before deep inspection. As a result, every theme demonstrates that robust preprocessing is essential for enabling DPI systems to maintain high detection accuracy and real-time responsiveness, even in complex or high-speed network environments.

3.13. Discussion

The utilisation patterns of preprocessing functions over the six-year review period indicate that basic operations of the method's core workflow, including packet capture and session reconstruction, are widely used, highlighting the importance and central role of these operations in DPI pipelines. Middle-level operations, such reassembly and conditioning, showed a concentration of usage between 2021–2025, likely due to changing demand on data normalisation and protocol processing. There were more advanced additions, on the other hand, such as QoS tagging and flow enrichment, which were observed only sporadically, which implies limited but emerging interest in the additions that would involve context-aware preprocessing.

The literature on classical DPI based anomaly detection emphasises that preprocessing is a stratified sequence of complex and interlocking modules and not a separate process; raw packet streams are fed through an elaborate preprocessing sequence before undergoing deep packet inspection. The present review is given a thematic format that consists of eight major themes that explain the dynamic aspect of preprocessing from basic handling of packets to intelligent, context-aware conditioning. In all pipelines, packet capture and parsing is always the first phase using tools like libpcap, tcpdump, and hardware assisted engines to ensure lossless, high throughput acquisition. The data taken are later reconstructed through data stream reassembly, and hence DPI systems are able to examine full TCP streams or session layers despite out-of-order delivery or fragmentation.

At the same time, payload fragment operations include how partial packets are dealt with by normalisation and reassembly of fragmented payloads, thus ensuring integrity of data in high-speed environments. Meanwhile input consistency is enforced through advanced packet conditioning mechanisms, including deduplication, windowing, and state-machine-driven filtering, reduce noise. These operations provide the foundation of sampling and aggregation of traffic in a manner that is adaptive, such like reducing large quantities of traffic into flows or sketches using algorithms such as NetFlow or Tranalyzer2, to allow behavioural monitoring at both temporal and flow scales. QoS tagging is their less popular counterpart that can be used to give higher priority to traffic classes in the DPI pipeline by repurposing ToS or DSCP fields to align with quality-of-service semantics, now pointing toward the eventual integration of quality-of-service semantics and security processes.

Lastly, flow enrichment enhances the behavioural context of such sessions via adopting JA3 fingerprints, correlation features, and memory-based identifiers are shifting DPI beyond highly syntactic analysis over to semantic interpretation. Along with the themes outlined above, the findings confirm that preprocessing is essential in modern DPI design, is also embarks a fact that not only must it be fast and accurate, but it also needs to be able to handle encrypted, evasive, and high-throughput threats.

The invisible foundation of effective DPI pipelines, as shown in the layered architecture in the supplemental diagrams, is the layered preprocessing, beginning with the capture of data and ending with the enrichment. Starting with data capture, then moving to different phases until it reaches the enrichment stage, preprocessing is the basis of all the aspects of the inspection process. Such results affirm that the DPI performance cannot exist without preprocessing quality and state that further development of such modules is required to meet the multifaceted and challenging nature of modern network traffic.

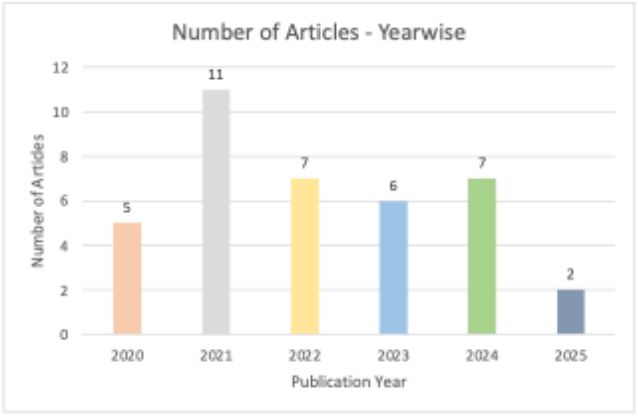

Figure 6: Breakdown of published articles according to the Year of Publication

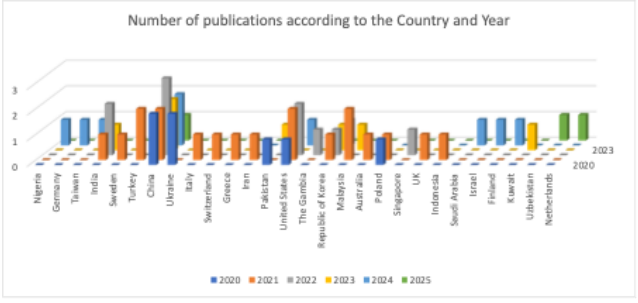

Figure 7: Breakdown of Publications According to Country

Figure 8: Breakdown of Geographical map of authors

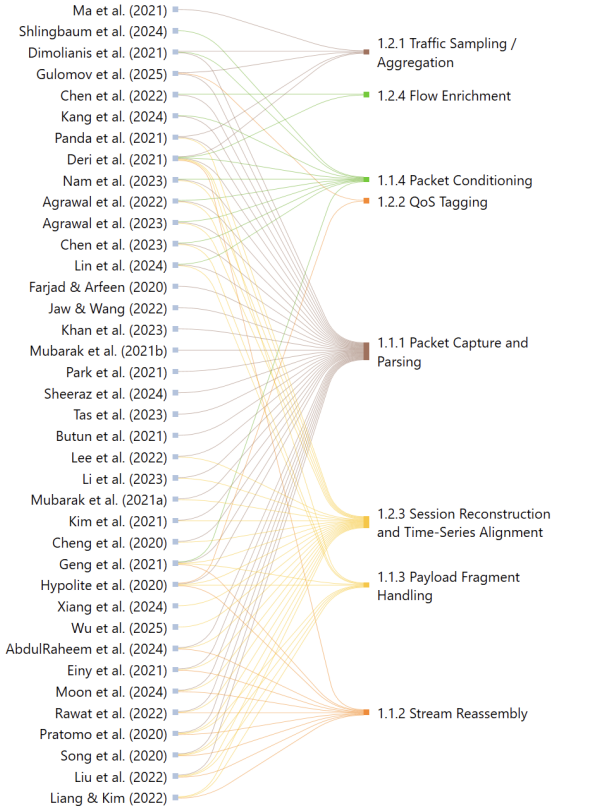

Figure 9: Sankey Thematic Mapping of the Publications

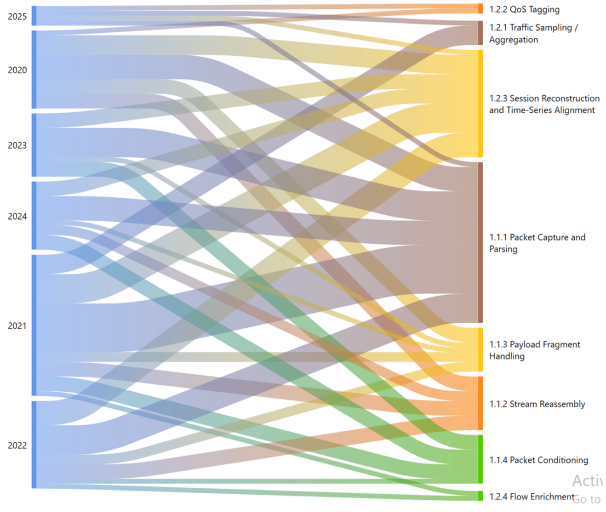

Figure 10: Sankey diagram of the Themes According to Year

Figure 11: Network View of the research articles on the "Packet Capture and Parsing" Theme

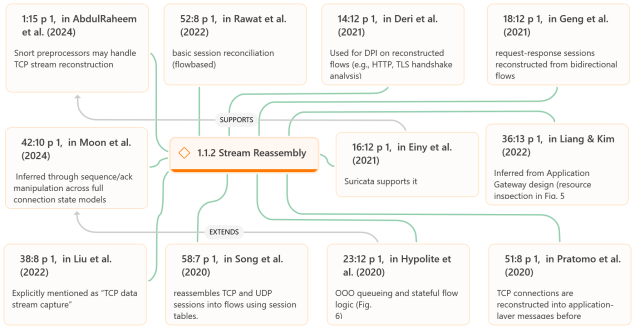

Figure 12: Network View of the research articles on the "Stream Reassembly" Theme

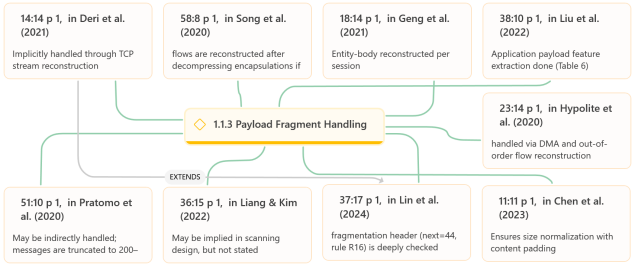

Figure 13: Network View of the research articles on the "Payload Fragment Handling" Theme

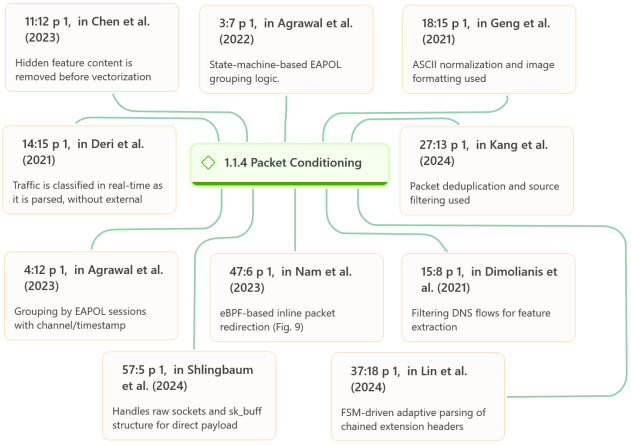

Figure 14: Network View of the research articles on the "Packet Conditioning" Theme

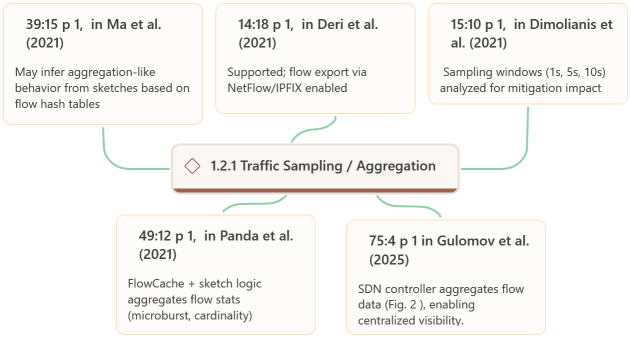

Figure 15: Network View of the research articles on the "Traffic Sampling / Aggregation" Theme

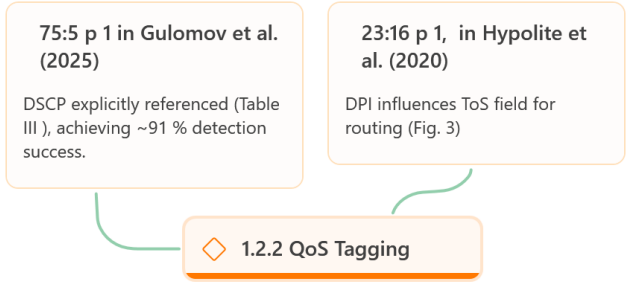

Figure 16: Network View of the research articles on the "QoS Tagging" Theme

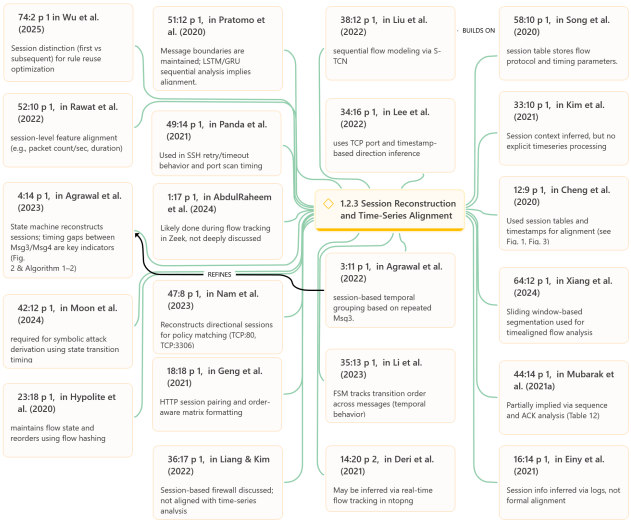

Figure 17: Network View of the research articles on the "Session Reconstruction and Time-Series Alignment" Theme

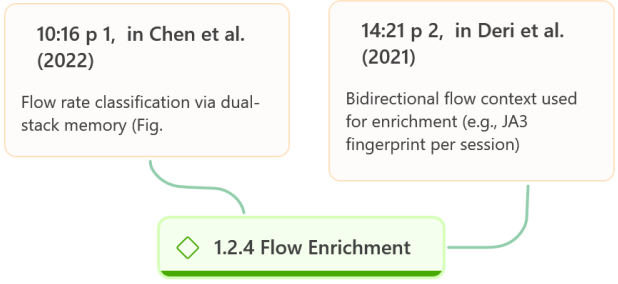

Figure 18: Network View of the research articles on the "Flow Enrichment" Theme

4. Conclusion and Future Research Directions

Although the analysed body of literature on the problem of DPI preprocessing indicates significant steps in the right direction, such as effective capture models, adaptive session alignment, and more effective ways of aggregation, a number of gaps warrant further investigation. To begin with, despite the frequent praise of flow aggregation and sampling strategies, including sketching algorithms, and NetFlow-like grouping, there is little research conducted that provides direct, side-by-side comparisons of the strategies using consistent data sets and measures. This lack of comparati ve benchmarking prevents the evaluation of the optimal method of balancing memory occupancy, latency, as well as detection accuracy in comparable traffic scenarios. Rigorous, head-to-head assessments using a variety of traffic characteristics and attack models would offer invaluable insights on the best aggregation logic of new-generation DPI pipelines.

Additionally, there are relatively few studies which use QoS/ToS or DSCP fields to make preprocessing decisions or enable conditional inspection paths, representing a significant untapped research opportunity in the future. As network-layer markers, these fields may be used to support real-time policy pathing, selective enrichment, or fast-path filtering, and thereby enhance both efficiency and inspection granularity. To improve responsiveness and scalability in DPI systems, implementing preprocessing modules that dynamically adapt based on DSCP or ToS values could help balance system resource utilisation with inspection depth.

4.1. Advantages and Merits of the Research

This review systematically maps eight key preprocessing techniques used in deterministic DPI-based anomaly detection from 2020 to 2025. It emphasises the preprocessing strategies that are most widely adopted, identifies key gaps that warrant further research, including the absence of real-world error reporting and the limited use of adaptive network-layer signals, and outlines clear directions for future research. The results equip researchers with specific open questions to pursue and help practitioners select and refine preprocessing techniques by emphasising the need to evaluate accuracy, efficiency, and adaptability under realistic conditions.

References

AbdulRaheem, M., Oladipo, I.D., Imoize, A.L., Awotunde, J.B., Lee, C.C., Balogun, G.B., Adeoti, J.O., 2024. Machine learning assisted snort and zeek in detecting ddos attacks in software-defined networking. International Journal of Information Technology 16, 1627–1643. doi:10.1007/s41870-023-01469-3.

Agrawal, A., Chatterjee, U., Maiti, R.R., 2023. Checkshake: Passively detecting anomaly in wi-fi security handshake using gradient boosting based ensemble learning. IEEE Transactions on Dependable and Secure Computing doi:10.1109/TDSC. iEEE Transactions on Dependable and Secure Computing. Advance online publication.

Arshad, S., Zanib, R., Akram, A., Saeed, T., 2023. A short review on faster and more reliable tcp reassembly for high-speed networks in deep packet inspection, in: Proceedings of the 1st International Conference on Advanced Innovations in Smart Cities (ICAISC 2023), pp. 1–6. doi:10.1109/ICAISC56366. in Proceedings of the 1st International Conference on Advanced Innovations in Smart Cities (ICAISC 2023) (pp. 1–6). IEEE.

Chen, J., Zhang, X., Wang, T., Zhang, Y., Chen, T., Chen, J., Xie, M., Liu, Q., 2022. Fidas: Fortifying the cloud via comprehensive fpga-based offloading for intrusion detection, in: Proceedings of the 49th Annual International Symposium on Computer Architecture (ISCA 2022), pp. 1–12. doi:10.1145/3470496. in Proceedings of the 49th Annual International Symposium on Computer Architecture (ISCA 2022) (pp. 1–12). ACM.

Chen, X., Lu, B., Sun, R., Jiang, M., 2023. Honeypot detection method based on anomalous requests response differences, in: Proceedings of the 6th International Conference on Electronics, Communications and Control Engineering (ICECC 2023), pp. 1–9. doi:10.1145/3592307. in Proceedings of the 6th International Conference on Electronics, Communications and Control Engineering (ICECC 2023) (pp. 1–9). ACM.

Dalkin, S.M., Greenhalgh, J., Jones, D., Cunningham, B., Lhussier, M., 2015. What's in a mechanism? development of a key concept in realist evaluation implementation science, 10, 49. https://doi , 13012–015doi:10.1186/s13012-015-0237-x.

Dehlaghi-Ghadim, A., Helali Moghadam, M., Balador, A., Hansson, H., 2023. Ics-flow: An anomaly detection dataset for industrial control systems (arxiv no. 2305 URL: https://arxiv.org/abs/2305.09678.

Denyer, D., Tranfield, D., Van Aken, J.E., 2008. Developing design propositions through research synthesis. Organization Studies 29, 393–413. doi:10.1177/0170840607088020.

Deri, L., Fusco, F., 2021. Using deep packet inspection in cyber traffic analysis, in: 2021 IEEE International Conference on Cyber Security and Resilience (CSR), pp. 1–6. doi:10.1109/CSR51186. in 2021 IEEE International Conference on Cyber Security and Resilience (CSR) (pp. 1–6). IEEE.

Dimolianis, M., Pavlidis, A., Maglaris, V., 2021. Signature-based traffic classification and mitigation for ddos attacks using programmable network data planes. IEEE Access, 9, 111130–111145 , 111130–111145doi:10.1109/ACCESS.

Einy, S., Oz, C., Navaei, Y.D., 2021. The anomaly- and signature-based ids for network security using hybrid inference systems. Mathematical Problems in Engineering, 2021, Article 6639714 doi:10.1155/2021/6639714. article 6639714.

Geng, J., Li, S., Zhang, Y., Liu, Z., Cheng, Z., 2021. Lifh: Learning interactive features from http payload using image reconstruction, in: 2021 IEEE International Conference on Communications (ICC), pp. 1–6. doi:10.1109/ICC42927. in 2021 IEEE International Conference on Communications (ICC) (pp. 1–6). IEEE.

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

Hypolite, J., Sonchack, J., Hershkop, S., Dautenhahn, N., DeHon, A., Smith, J.M., 2020. Deepmatch: Practical deep packet inspection in the data plane using network processors, in: Proceedings of the 16th International Conference on Emerging Networking Experiments and Technologies (CoNEXT '20), pp. 1–15. doi:10.1145/3386367. in Proceedings of the 16th International Conference on Emerging Networking Experiments and Technologies (CoNEXT '20) (pp. 1–15). ACM.

Kang, H.S., Kim, K., Kim, S.R., 2024. A new mitigation method against dr-dos attacks using a snort udp module in low-specification fog computing environments. Electronics 13. doi:10.3390/electronics13152919.

Khan, R.A.S., Abdul Rahman, M.N., 2023. Efficiency of surveillance of tcp packet in iot in reducing the risk of ransomware attacks. Journal of Theoretical and Applied Information Technology 101, 1126–1132.

Kim, J., Camtepe, S., Baek, J., Susilo, W., Pieprzyk, J., Nepal, S., 2021. P2dpi: Practical and privacy-preserving deep packet inspection, in: Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security (ASIA CCS '21), pp. 1–12. doi:10.1145/3433210.

in Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security (ASIA CCS '21) (pp. 1–12). ACM.

Koumar, J., Hynek, K., Pešek, J., Čejka, T., 2023. Nettisa: Extended ip flow with time-series features for universal bandwidth-constrained high-speed network traffic classification. Computer Networks, 236, 110147 doi:10.1016/j.

Lee, J., Kang, M.S., Divakaran, D.M., Thet, P.M., Singhai, V., You, J.S., 2022. A step towards on-path security function outsourcing, in: Proceedings of the International Conference on Distributed Computing and Networking (ICDCN'22). In Proceedings of the International Conference on Distributed Computing and Networking (ICDCN'22) (13 pages). ACM.

Li, Z., Wei, Q., Ma, R., Geng, Y., Yang, Y., Lv, Z., 2023. Dpguard: A lightweight attack detection method for an industrial bus network. Electronics 12. doi:10.3390/electronics12051121.

Lin, B., Zhang, L., Zhang, H., Guo, Y., Ge, S., Fang, Y., Ren, M., 2024. An adaptive detection model for ipv6 extension header threats based on deterministic decision automaton. Scientific Reports, 14, Article 9534 , 41598–024doi:10.1038/s41598-024-59913-8. article 9534.

Liu, X., Liu, Z., Zhang, Y., Zhang, W., Lv, D., Zhou, Q., 2022. Tcn enhanced novel malicious traffic detection for iot devices. Connection Science 34, 1322–1341. doi:10.1080/09540091.

Ma, C., Chen, S., Zhang, Y., Xiao, Q., Odegbile, O.O., 2021. Super spreader identification using geometric-min filter. IEEE/ACM Transactions on Networking 30, 112–126. doi:10.1109/TNET.

Moon, S.J., Srivastava, M., Bieri, Y., Martins, R., Sekar, V., 2024. Pryde: A modular generalizable workflow for uncovering evasion attacks against stateful firewall deployments, in: Proceedings of the 2024 IEEE Symposium on Security and Privacy (SP), pp. 1–12. doi:10.1109/SP54263. in Proceedings of the 2024 IEEE Symposium on Security and Privacy (SP) (pp. 1–12). IEEE.

Musfira, A.F., Ibrahim, N., Harun, H., 2022. A thematic review on digital storytelling (dst) in social media. The Qualitative Report 27, 1590–1620. doi:10.46743/2160-3715/2022.

Nam, J., Lee, S., Porras, P., Yegneswaran, V., Shin, S., 2023. Secure inter-container communications using xdp/ebpf. IEEE/ACM Transactions on Networking 31, 672–685. doi:10.1109/TNET.

Noyes, J., Popay, J., Pearson, A., Hannes, K., Booth, A., 2008. Qualitative research and cochrane reviews, in: J. Higgins & S. Greens (Eds.), Cochrane handbook for systematic review of interventions (Vol. 5.0.1, pp. 1-18). Wiley.

Ouzzani, M., Hammady, H., Fedorowicz, Z., Elmagarmid, A., 2016. Rayyan—a web and mobile app for systematic reviews. Systematic Reviews, 5, 210 , 13643–016doi:10.1186/s13643-016-0384-4.

Panda, S., Feng, Y., Kulkarni, S.G., Ramakrishnan, K.K., Duffield, N., Bhuyan, L.N., 2021. Smartwatch: Accurate traffic analysis and flow-state tracking for intrusion prevention using smartnics, in: Proceedings of the 17th International Conference on Emerging Networking Experiments and Technologies (CoNEXT '21), pp. 1–12. doi:10.1145/3485983. in Proceedings of the 17th International Conference on Emerging Networking Experiments and Technologies (CoNEXT '21) (pp. 1–12). ACM.

Park, T., Nam, J., Na, S.H., Chung, J., Shin, S., 2021. Reinhardt: Real-time reconfigurable hardware architecture for regular expression matching in dpi, in: ACSAC '21: Annual Computer Security Applications Conference, pp. 1–14. doi:10.1145/3485832. in ACSAC '21: Annual Computer Security Applications Conference (pp. 1–14). ACM.

Pawson, R., Greenhalgh, T., Harvey, G., Walshe, K., 2005. Realist review—a new method of systematic review designed for complex policy interventions. Journal of Health Services Research & Policy , 21–34doi:10.1258/1355819054308530.

Pratomo, B.A., Burnap, P., Theodorakopoulos, G., 2020. Blatta: Early exploit detection on network traffic with recurrent neural networks. Security and Communication Networks, 2020, Article 8826038 doi:10.1155/2020/8826038. article 8826038.

Radamski, K., Ząbek, W., Domżał, J., Wójcik, R., 2024. Adaptive flow timeout management in software-defined optical networks. Photonics 11. doi:10.3390/photonics11070595.

Rawat, R., Garg, B., Pachlasiya, K., Mahor, V., Telang, S., Chouhan, M., Shukla, S.K., Mishra, R., 2022. Scnta: Monitoring of network availability and activity for identification of anomalies using machine learning approaches. International Journal of Information Technology and Web Engineering 17, 1–15. doi:10.4018/IJITWE.

Shlingbaum, E., Ben Yehuda, R., Kiperberg, M., Zaidenberg, N.J., 2024. Virtualized network packet inspection. Computer Networks, 251, 110619 doi:10.1016/j.

Song, W., Beshley, M., Przystupa, K., Beshley, H., Kochan, O., Pryslupskyi, A., Pieniak, D., Su, J., 2020. A software deep packet inspection system for network traffic analysis and anomaly detection. Sensors 20. doi:10.3390/s20061637.

Zairul, M., 2020. A thematic review on student-centred learning in the studio education. Journal of critical reviews .