Journal of Informatics and Computer Science Vol. 12, No. 1 Oktober 2026
Universitas Ubudiyah Indonesia
e-ISSN : 2615-5346

AN ADAPTIVE, PMF-AWARE MITIGATION FRAMEWORK FOR DEAUTHENTICATION ATTACKS IN IEEE 802.11 WIRELESS NETWORKS

KERANGKA MITIGASI ADAPTIF BERBASIS PROTECTED MANAGEMENT FRAMES (PMF) UNTUK SERANGAN DEAUTHENTICATION PADA JARINGAN NIRKABEL IEEE 802.11

Walaa H. Allawi [(1)], Saddam Madallah Obaid[(2)]
[(1)]Al-Mussaib Polytechnic College, Al-Furat Al-Awsat Technical University, Najaf, Iraq,
Walaa.allawi@atu.edu.iq
[(2)]Department of Public Health, College of Veterinary Medicine, Al-Qasim Green University, Babylon 51013, Iraq , saddam@vet.uoqasim.edu.iq

Abstrak— Penyalahgunaan autentikasi deauthentication (deauthentication abuse) merupakan permasalahan nyata yang dapat mengganggu ketersediaan layanan pada jaringan IEEE 802.11. Protected Management Frames (PMF) membawa perubahan terhadap cara lalu lintas manajemen diinterpretasikan dari perspektif keamanan, tetapi tidak serta-merta menghilangkan seluruh kemungkinan penyebab terputusnya koneksi. Dalam penelitian ini, kami mengusulkan *PMF-Aware Adaptive Response Framework* (PARF), yaitu kerangka kerja respons adaptif yang mudah dijelaskan (*explainable*) dan memiliki kompleksitas rendah. Kerangka kerja ini menggabungkan anomali volume frame, inkonsistensi asal frame yang telah dikonfirmasi melalui bukti pendukung, bukti konektivitas, frekuensi kejadian berulang, konteks operasional yang telah diverifikasi, serta status negosiasi PMF pada setiap stasiun. Evaluasi dilakukan melalui tiga laporan independen yang saling melengkapi, yaitu analisis baru terhadap total data tangkapan AWID3 yang telah dipublikasikan, mencakup 3.640.191 frame dan 114.076 frame berlabel serangan; audit analitis terhadap 48 kasus yang mencakup empat profil PMF; serta audit yang dilaporkan oleh peneliti menggunakan 160 jendela observasi *holdout* yang memenuhi kriteria kelayakan. Aturan berbasis penghitungan (*count-only rule*) berhasil mengidentifikasi 8 dari 24 kasus serangan dengan 16 positif palsu (*false positive*), sedangkan PARF berhasil mengidentifikasi 20 dari 24 kasus serangan dengan 4 positif palsu pada ambang batas $S \geq 4$. Pada perbandingan *holdout* yang dilaporkan secara terpisah, metode *count-only* mencapai sensitivitas dan presisi masing-masing sebesar 60,00%, sedangkan PARF mencapai sensitivitas sebesar 85,00% dan presisi sebesar 83,61%. Analisis *base rate* tidak menggunakan pengukuran aktual tingkat kesalahan detektor, melainkan mengkaji secara terpisah bagaimana tingkat kesalahan detektor yang diasumsikan berinteraksi dengan prevalensi serangan. Hasil *holdout* diperlakukan sebagai hasil yang dilaporkan oleh peneliti. Namun, identitas paket tangkapan, catatan pada tingkat jendela observasi, dan log verifikasi yang mendasarinya belum disertakan dalam naskah ini sehingga diperlukan untuk memungkinkan reproduksi hasil secara independen dalam penelitian selanjutnya.
PARF berfungsi sebagai sistem pendukung triase keamanan melalui analisis paket dan telemetri administratif, bukan sebagai mekanisme yang secara aktif melakukan pemblokiran atau penekanan frame melalui media nirkabel (*over-the-air frame suppression*).
Kata kunci: IEEE 802.11; Protected Management Frames; deauthentication; PMF; deteksi intrusi nirkabel; mitigasi berbasis aturan; AWID3; sistem pendukung keputusan yang dapat dijelaskan.

Abstract*-DE authentication abuse is indeed a realistic issue for availability in IEEE 802.11 networks, and Protected Management Frames (PMF) are making changes to how management traffic is interpreted with regards to security, but not every possible cause of disconnection is removed. In this paper, we propose an explainable low-complexity PMF-Aware Adaptive Response Framework (PARF) that combines frame-volume anomalies, corroborated-origin inconsistencies, connectivity evidence, recurrence, verified operational context and per-station PMF negotiation state. Evaluation consists of three independent reports made to each other: a new analysis of published AWID3 capture totals from 3,640,191 frames and 114,076 attack-labelled frames, a 48-case analytical audit that included four PMF profiles, and a researcher-reported audit using 160 eligible holdout observation windows. The count-only rule had 8 of 24 attack cases identified and 16 false positive and PARF had 20 of 24 attack cases identified with 4 false positive at $S \geq 4$. For the separately reported "holdout" comparison, count-only achieved sensitivities and precisions of 60.00% and 60.00%, respectively, while PARF achieved 85.00% and 83.61%, respectively. The base-rate analysis does not consider actual measures of detector error rates, but examines separately how detector error rates (assumed) interact with the prevalence of attacks. The holdout outcomes have been considered as the reported results of the researchers and underlying packet-capture identifiers, window-level records and verification logs have not been included in this manuscript and hence are necessary for independent research reproducing these results. PARF is supporting security triage via packet analysis and administrative telemetry, but not that PARF is some kind of active over-the-air frame suppression.*

Keywords: *IEEE 802.11; Protected Management Frames; deauthentication; PMF; wireless intrusion detection; rule-based mitigation; AWID3; explainable decision support.*

1. Introduction

1.1 Background and motivation

Wireless LAN (WLAN) networks are getting more and more to play a critical role with regards to the connectivity that needs to remain available even during mobility, during authentication changes as well as when devices are being churned out. Although the operations are allowed via IEEE 802.11 management exchanges, the behavior of the management plane can also be attacked for availability purposes. The disconnection messages are especially important because they play a part in the discontinuation of an authenticated relationship, a disconnection event could be confused with disconnection messages from an impersonated transmitter. Under the modern definition, only a wide MAC and PHY context are formed, though not necessarily every management action is cryptographically secured [1], [2, p. 34188].

Once the suitable security association is created, PMF secures chose strong administration outlines. In fact it has quite a meaning, but you shouldn't replace the word without interruption with it. Experimental Security analysis has found standards level edge cases, implementation-disconnect paths and attacks against other protocol levels [3],[4],[5]. Therefore, in addition to determining the presence of the correct number of DE authentication frames, a security monitor can enquire as to which station initiates negotiation for a protection, be it prior to key establishment or subsequent thereto, and whether there is separate telemetry information available to aid service disturbance.

1.2 Problem statement and research gap

There are two common methodological gaps that tend to undermine the cost of cheap wireless defenses. This static DE authentication-count threshold will trigger excessive false alerts in benign scenarios such as roaming, or access point restarts, and in the opposite extreme scenario, a low-rate origin spoofing event may never hit the threshold. A second weakness is binary language regarding PMF — if an entire SSID is set as PMF, the difference between PMF-optional clients that negotiated protection and the old clients that had no idea what protection was may not be known. Systems already available span from advanced traffic classifiers up to dedicated low level defensive mechanisms and maintaining a portable decision policy is beneficial for the resource limited administration [6, p. 88], [7, p. 182796].

1.3 Objective, questions, and contribution

The ultimate goal is to develop and test a conservatively biased decision model for monitoring and response that considers the PMF. The research questions include: (RQ1) what are some observable cues that can be added to frame counts (RQ2) how does PMF state changing at the client level influence the recommended action? (RQ3) what are the dynamics of the framework for a set of normal and adversarial reference cases? and (RQ4) what is the impact of class imbalance on alert interpretation and the effectiveness of the holdout comparison? Contributions are limited to the PARF evidence model, a PMF specific response matrix, a duplicate researcher reported holdout comparison, and a prevalence sensitive precision calculation. The work does not attempt to pull out an active wireless intrusion prevention system.

2. Technical Background and Related Work

2.1 Deauthentication semantics and observability

The DE authentication and disassociation are separate management functions, and should not be mixed in terms of numbers of packets. Passive sensor can only see transmitted frames, not necessarily constitute evidence of a customer's transmitter being genuine, of the frame passing the PMF process at the receiving station or of the source of the receiver disconnecting. Observed rate can be affected by capture-channel selection, missed frames, re-transmissions, and monitor placement. There is research available that reports the consequences of availability in the context of management frame manipulation and reasons for correlating monitoring to actual end-point behavior [8], [9].

Attacks can also happen outside of the case of protected-DE authentication in WPA3. Dragon blood analyzed vulnerabilities in the SAE handshake and later, more study delved into denial-of-service attacks during SAE processing. The results underline the need for a different security dimension for each of the authentication stage and the implementation state and that not all outages can be blamed on forged DE authentication frames [10, p. 517–519], [4], [5].

2.2 PMF scope and limits

An access point that implements PMF is not supportive of an all-clients-protected deployment. Some of the associated stations negotiate PMF and some do not; required mode will not accept stations that do not have the capability. Even if strong management frames are secured after association, PMF does not validate all discovered frames or frames prior to association, or prevent radio frequency jamming and will not correct for incorrect client firmware. Weaknesses in implementation or processing defeat the expected DE authentication

resilience in some cases, while related frame-handling research illustrated weaknesses that are not shielded by management-frame protection in all cases [3], [11, p. 53–55], [12].
The above Wi-Fi attack overviews also categorically discuss multi-channel interception, association steering, rogue access points and denial-of-service. This universal blocking policy may thus lead to false reassurance. Instead, good recommendations should be split between configuration hardening, evidence verification, forensic investigation and firmware remediation [13], [14, p. 357–359], [15].

2.3 Detection research and the interpretability trade-off

Statistics, NN, feature selection and hybrid models have been recently used in the wireless IDS studies. While they are successful in showing high classification accuracy on certain datasets, their accuracy can be influenced by the features used, the nature of the attack, and the choice of validation set. Kolias et al. pointed out that improper preprocessing or "freebies" might lead to overestimation of the generalization properties [16]. Transferability of cross-layer information and transfers across capture conditions has also been studied in the AWID3 domain [16, p. 64761–64763], [4], [5], [17, p. 11248–11250].
The necessity of having to deal with the evidence of heterogeneous wireless environments is confirmed by both the results of surveys regarding intrusion detection in both wireless and converged networks and by the studies of whether any results can be transferred between various generations of AWID. Both these suggest the need to explain detection conditions and deployment assumptions in tandem with high test accuracy [18, p. 395–398], [19, p. 40950–40952], [20]. Explainability and low operational overhead thus are a hard design goal that can be argued and given that the policy is not sold as a cryptographic alternative to PMF, it can be met.

2.4 Datasets, contemporary defenses and research positioning

The AWID3 corpus serves as a valuable resource for contemporary 802.11 security scenarios and experimentation, including enterprise-oriented scenarios (e.g. DE authentication and disassociation). When performance figures need to be compared though, capture distribution, per frame labels and the evaluation unit must be documented. General focus studies on datasets highlight representativeness, data quality, different cross-sections, and reproducibility [2, p. 34188–34190], [21, p. 34–36], [22], [23].
Some of the recent options are a specially designed eBPF/XDP responder architecture, deep-learning classifiers for WPA3 traffic, feature-rich AWID detection studies, and PMF related IoT provisioning security analysis. Unlike PARF, their contribution has its own merits: this research makes the administrative decision logic explicit in a testable manner without the need to train a model or to modify a kernel. It does not claim superiority of any of these systems, nor does it directly contrast the percentages reported in these systems as these have not undertaken the same operations or the same desired results [24], [25], [26, p. 34–35], [27, p. 1–2].

3. Proposed PMF-Aware Adaptive Response Framework



Figure 1. The original PARF workflow and its five auditable decision stages.

3.1 Design principles and scope

PARF is a determinate security decision-support policy. It takes as its input a monitoring window that refers to a single AP–station pair or to a known set of AP–station pairs. This model aims to be really specific – it is about suspicious events and proportionate response. It doesn't proof cryptographic authenticity with only knowing the MAC address, nor does it prevent any frames from being sent on the air, nor does it indicate if a spoofing attack has succeeded based on a passive capture. The three design principles that underlie the architecture are: protect by channel per station, not per SSID by using a shorthand technique; perform independent coronation of abnormal activities per station; and treat benign operational activities with caution.

3.2 Evidence model

Variable D for each 1 min observation window counts the number of DE authentication events observed or, for a normalized equivalent, and a variable M is defined to be one if an origin or sequence inconsistency was observed and substantiated by other related AP/client log information rather than just assumed from a purported sender address, and the variable Z is used to record an independent connectivity anomaly, and the variable R is used to indicate repetition across adjacent observation windows, and the variable L is used to

indicate a verified maintenance, roaming, or administrative reason. U = 1 if the station not in PMF, Q = 1 if PMF required deployment has inconsistent management-frame observation. But U and Q are context signals and not indicators of maliciousness.

Table 1. PARF input dictionary and audit controls

Field	Meaning	Recording rule
D	Deauth count per window	Count only applicable observed 802.11 subtype 0x0c frames; document channel and duration.
M	Corroborated origin anomaly	1 only with independent AP/client evidence of inconsistency; else 0.
Z	Connectivity correlation	1 when independent telemetry documents coincident link impact.
R	Repeated activity	1 when a comparable suspicious pattern recurs in adjacent windows.
L	Known benign operational context	1 only for confirmed restart, roaming or maintenance.
U	Client unprotected under PMF	1 for Disabled or Optional/unprotected station; else 0.
Q	Inconsistency under PMF Required	1 if PMF Required and M=1; else 0.

3.3 Explicit scoring and safeguards

The evaluated policy uses a count threshold of 12 events per observation window as a declared design parameter. Let $H = 1$ if $D \geq 12$ and 0 otherwise. The integer risk score is $S = 2H + 2M + 2Z + R + U + Q - 2L$. If $L = 1$ and the pair (M,Z) is not simultaneously (1,1), S is capped at 2. An event enters the investigation queue when $S \geq 4$; lower scores remain recorded for trend inspection. The count-only comparator flags a window when $D \geq 12$. The same stated parameters underpin the 48-case audit and the separately reported holdout comparison. Site-specific deployment may require subsequent validation and recalibration.

In an administrative sense, the policy responds adaptively with acceptance of both U (the level of protection that each station negotiates) and Q (the level of security that each station verifies) and changing of the H threshold following nonincidental baseline assessment. The same set of fixed parameters have been used for the evaluation reported here. No claim was made for machine-learning weights, automatic calibration or active packet suppression. This delineation helps avoid misidentifying policy adaptation as online learning.

3.4 Risk-to-response matrix

Table 2. PMF-aware mitigation policy (proposed, not deployed)

PMF state	Evidence level	Recommended administrative action
Disabled	Low / moderate	Record, verify compatibility; plan secure upgrade path with PMF where feasible.
Disabled	High	Correlate AP/client logs, confirm spoofing evidence, isolate event; enable stronger protections after compatibility review.
Optional / PMF not negotiated	Any confirmed anomaly	Treat client as unprotected; prioritize compatibility audit and migration.
Optional / PMF negotiated	Persistent anomaly	Investigate adjacent stages and implementation; do not infer successful forged deauth.
Required	Repeated disruption	Investigate firmware, authentication-stage abuse, RF interference and capture errors.
All	Verified maintenance	Lower immediate escalation; retain log and check post-maintenance stability.

The suggested answer is designed to be reversible. Strictly speaking, a score is not a reason to deny access to the network and operator should not disable PMF to reduce groups due to a score. High-Risk Output is triggered for investigation and verification of configuration – policy rollout should include firmware updates, endpoint compatibility tests and change control. The statements made in the recommendations never say that failure of a PMF check shows that the client is disconnected.

4. Methodology and Reproducible Analytical Design

4.1 Research design and boundaries

Three sets of analytical components are reported by the study. A secondary quantitative re-analysis, first, re-calculates published AWID3 capture total for attack labelled shares. Second, under 4 rules an attempted decision audit of 12 predefined event profiles is performed in a deterministic fashion; it's the testing of rule behavior in its declared reference cases. Third, the study reports aggregate results for the comparison provided by the researcher to the holdout data (160-window samples) with the rule (count-only) and PARF. Due to different sampling units of published frame totals, designs of audit cases, and holdout windows, the three sources of evidence are interpreted separately. For the aggregate re-analysis, raw AWID3 PCAP files were not processed. The holdout summaries are provided as study results, but the corresponding capture identifiers or even the packet level records or evidence for these events presented as labels in the displayed AP/client log excerpts are not provided in the manuscript to make external verification as extensive as possible.

4.2 Secondary quantitative dataset and derived quantities

The chosen publication contains two different AWID3 capture files, one with 1,626,472 total frames (including 38,945 attack labelled frames), and another with 2,013,719 total frames (including 75,131 attack labelled frames). Importantly, these totals represent frames contained in attack-oriented capture files and not the number of DE authentication subtype per minute and also not a representative prevalence of attacks detected from real networks. No packets were found with a label check here. In the present paper, detector assumptions as disclosed in the abstract are applied and the attack labelled proportions, differences between the files and conditional positive predictive values were re-computed..

Table 3. Published capture totals and original calculations (secondary analysis)

Capture	All frames	Attack-labelled	Non-attack	Attack share
Deauth capture	1,626,472	38,945	1,587,527	2.394%
Disassociation capture	2,013,719	75,131	1,938,588	3.731%
Combined (descriptive)	3,640,191	114,076	3,526,115	3.134%

Source: numerical inputs transcribed from the AWID3 capture summary in Winiarski and Natkaniec (2026), Table 2; all shares and the combined row were recalculated for this paper. Capture files are not independent, uniformly sampled network epochs; the combined proportion is descriptive, not an estimate of deployment-wide attack prevalence.

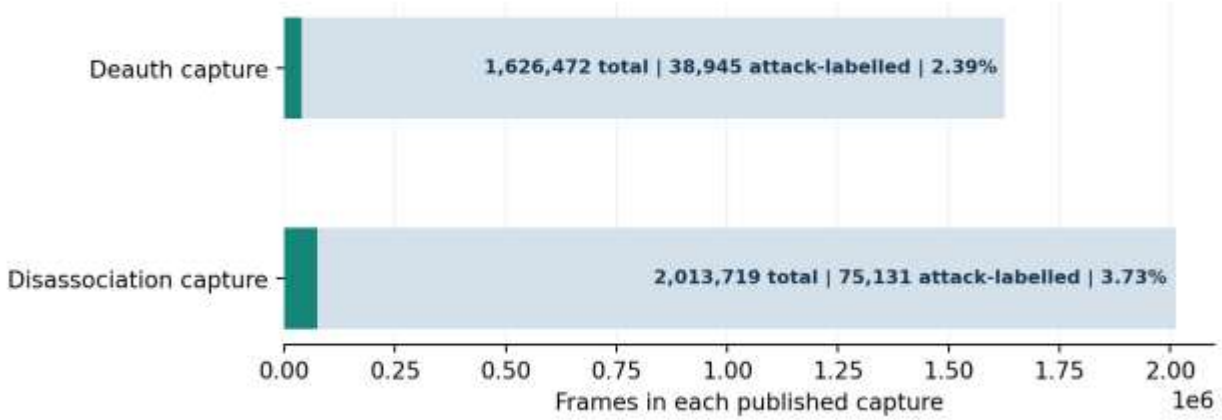


Figure 2. Re-expressed published AWID3 capture totals; percentages are independently recalculated.

4.3 Base-rate and alert-quality equations

For sensitivity t, false-positive rate f, and an attack-labelled proportion p, Bayes' rule gives the positive predictive value PPV = tp / [tp + f(1−p)]. The sensitivity analysis sets t = 0.95 as an explicitly assumed operating point and varies f = {0.25, 0.50, 1, 2, 3, 5}% against the two published capture shares. This sweep is a quantitative analysis of how change in base rate affects the expected accuracy of alerts. It is a mathematical analysis rather than measured-performance values as provided in the separately reported “Holdout comparison” analysis in Section 5.5.

4.4 Constructed decision-audit cases

There are twelve predefined audit profiles, six of which relate to valid or slightly ambiguous operational use contexts (B1 through B6) and six origins of attack scenarios (A1 through A6). All four configurations of PMF

are tested against each, Disabled, Optional with unprotected station, Optional with PMF protected station, and required. Each of the 48 cases in the matrix has 24 cases in each of the reference categories. The case labels are based on the predefined event descriptions, not on PARF's score. This is a controlled decision-audit input, and is separate from the holdout observations reported in Section 5.5. Input statistics to replicate the audit computations are shown in Table 4.

Table 4. Twelve original case archetypes, evaluated across four PMF contexts

ID	Reference event	D	M	Z	R	L	Truth
B1	Quiet traffic	2	0	0	0	0	Benign
B2	Ordinary roaming	14	0	1	0	1	Benign
B3	Planned AP restart	24	0	1	1	1	Benign
B4	Single client sleep	3	0	1	0	0	Benign
B5	Dense benign churn	16	0	1	0	0	Benign
B6	Capture artifact	13	0	0	0	0	Benign
A1	Low-rate spoofing	6	1	1	1	0	Attack
A2	High-rate spoofing	22	1	1	1	0	Attack
A3	Pre-association probe	3	1	0	1	0	Attack
A4	Distributed probes	15	1	0	1	0	Attack
A5	Attack amid maintenance	11	1	1	1	1	Attack
A6	Residual DoS mechanism	7	0	1	1	0	Attack

The reference label does not depend on the active PMF protection state. For example, in the case of pre-association spoofing (attempt), it is still considered an attempt, not an established-client disconnection, if using PMF Required. In order to be "deliberate" A6 inexplicably does not provide an anomaly of origin confirmation, creating the possibility for false negatives due to a remaining implementation issue. B5 is intentionally a dense but legitimate churn, no maintenance explanation, no other way to see the risk of false alarms other than this churn. These "adversarial" cases demonstrate failure modes but are not necessarily a good indicator of scores.

4.5 Metrics and statistical reporting policy

The decision audit reports TP, FP, TN, FN, sensitivity TP/(TP+FN), specificity TN/(TN+FP), precision TP/(TP+FP), and balanced accuracy for the two rules. The equal class counts are a result of the 48-case test design and are not interpreted as estimates of occurrence of attacks in the live network. On the unchanged matrix the sensitivity to decision thresholds (S = 3, 4, 5 and 6) is studied. These are pre-defined cases that are not taken through the population-level significance tests. Each comparison using holdout is a report of confusion matrix measures, F1 scores, and alert rates for the 160-window result set.

4.6 Separate packet-level holdout evaluation: methodology and reporting

4.6.1 Independent material and study design

The holdout comparison examines the fixed PARF rule for the case of the reported set of 160 eligible 60-second AP-station observation windows. The study summary contains six capture sessions, and compares the number of animals captured by PARF against the count-only threshold on these matching eligible windows. The reproducible evidence standard is a packet level PCAP/PCAPNG file that is tied to an out-of-band Attack/Benign label, an established observation period, negotiated PMF status, and documented access point/client telemetry for fields M, Z and L. All the numerical results of 5.5 are the work of the researcher for this manuscript. The capture identifiers and source-file hashes, along with their respective per-window records and log references are not included here. As such, the window labels and features are not easily retrievable from the manuscript; only the metric arithmetic in the published publication can be checked.

Each comparison unit is a 60-second window for a specified AP–station pair. There are 180 windows reported observed, 20 windows excluded, and 160 windows eligible for the reported dataset, which equates to 88.9% window eligibility recorded. From the eligible windows 60 are labelled as Attack and 100 as Benign. There are 44 Disabled, 38 Optional/unprotected, 39 Optional/protected, and 39 Required windows in the PMF state totals. The authors use the same denominator of eligible at the two rules. Adjacent-exposing windows of a capture session may be correlated; these overall comparisons should not be considered as independent experimental replications or necessarily be considered to carry window-level significance. To enable external replication, session identifiers should be part of the archived source records along with exclusion reasons, and provenance of class annotations.

4.6.2 Reference labeling, feature extraction and PMF verification

The reproducible labeling procedure uses an independent incident timeline, authorized testbed record, or validated dataset annotation to designate Attack or Benign status; classification must not be derived from D or the PARF score. In Wireshark, subtype 0x0c is isolated using wlan.fc.type_subtype == 0x0c, with the relevant

BSSID/station addresses and 60-second time bounds documented. Disassociation subtype 0x0a is tabulated separately. The RSN fields wlan.rsn.capabilities.mfpc and wlan.rsn.capabilities.mfpr identify advertised PMF capabilities, while station-level PMF negotiation requires association evidence and/or access-point logs. An MFPC beacon by itself does not establish the PMF status of an Optional-mode client [1], [2]. This extraction specification makes the comparison reproducible when the source captures and logs are available.
For each window, the feature record contains D = deauthentication count; M = corroborated origin/sequence inconsistency based on AP/client evidence; Z = independent connectivity disruption; R = comparable activity in an adjacent window; L = verified roaming/maintenance context; U = 1 when the station is unprotected under Disabled or Optional mode; and Q = 1 only when PMF Required and M = 1. Spoofing cannot be inferred from an unverified MAC header, and reception of a frame does not establish a successful disconnection. The exclusions (20) from full-feature relate to the number of windows for which evidence was missing and are therefore included in denominator for coverage. This copy of the manuscript does not contain supporting window-level exclusion reasons.

4.6.3 Frozen comparison, calculations and reporting

The reported holdout comparison applies the same fixed rules to eligible windows: count-only alert = 1 if $D \geq 12$; PARF uses $H = 1$ if $D \geq 12$ and $S = 2H + 2M + 2Z + R + U + Q - 2L$, capped at 2 when $L = 1$ unless $M = Z = 1$, with an alert at $S \geq 4$. Evaluation measures include TP, FP, TN, FN, sensitivity, specificity, precision, F1 score, balanced accuracy, coverage, and false alerts per 60 eligible AP–station pair-windows. Both methods use the same class totals (60 Attack and 100 Benign). These measures assess alert classification, not active frame blocking or measured link-uptime improvement.
The per-window analysis sheet needs to be able to hold capture ID, the time of the capture, the AP-station key, reference label, PMF state, D, M, Z, R, L, evidence link, exclusion reason, and calculated classifications. The details for aggregation (section 5.5) can be used to check the calculations of the cell values in the confusion matrix and percentages derived, but no confusion matrix event annotations and window-level worksheet is not provided or attached. The session-level analysis still plays a major role as there are devices and root causes that can be shared between adjacent windows. If such underlying records were shared, these confidence intervals should take this clustering into account.

Table 8. Holdout data documentation and eligibility criteria

Item	Required record / inclusion check	Source
Capture provenance	Distinct capture/session ID; date and recording duration; file hash	PCAP/PCAPNG metadata
Independent truth	Attack or Benign, with separate incident-label evidence	Dataset labels or authorized lab log
Packet-level unit	60-s window, AP–station key, channel and subtype 0x0c count	Wireshark packet list
PMF and telemetry	AP mode; negotiated client PMF; logs for M, Z and L	Association evidence + AP/client logs
Missing-data control	Eligibility, exclusions and reasons recorded before scoring	Window-level audit log
Frozen evaluation	D >= 12; PARF S >= 4; identical eligible windows	Independent Excel worksheet

5. Results and Original Analysis

5.1 Empirical source totals and prevalence implications

The two published files have a total of 3,640,191 frames of which 114,076 frames are attack labelled, equating to a recalculated within-file attack share of 2.394% and 3.731% respectively. So the disassociation capture enjoys a 1.56-fold greater labelled attack share in comparison to the DE authentication capture. This is because the gap between the numbers suggests detection-performance percentage cannot be directly translated across differently composed captures because you must take into account how the labels are distributed. The pooled prevalence of 3.134% is a descriptive aggregation of prevalence of the two files and not an estimate of the prevalence in the live network.

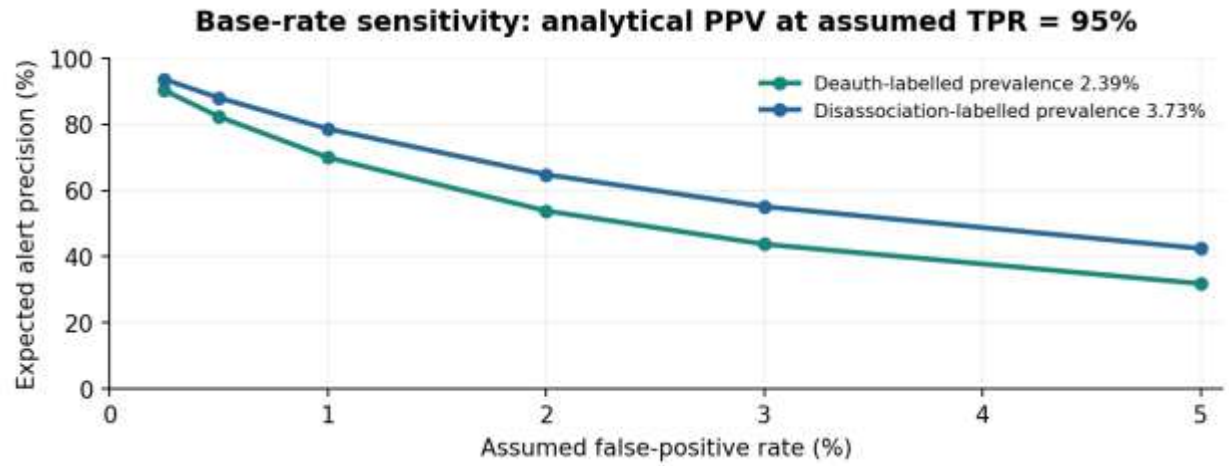


Figure 3. Bayesian sensitivity analysis using published attack-labelled shares and explicitly assumed detector error rates.

Table 5. Expected alert precision calculated under an assumed TPR of 95%

Assumed FPR	Deauth share = 2.394%	Disass share = 3.731%
0.25%	90.31%	93.64%
0.50%	82.34%	88.04%
1.00%	69.97%	78.64%
2.00%	53.82%	64.80%
3.00%	43.72%	55.10%
5.00%	31.79%	42.41%

The overall precision of the deauthentication capture and disassociation capture, at the same stated sensitivity parameter and an assumed 1% FPR, are approximately 70.0% and 78.6%, respectively, and 31.8% and 42.4%, respectively, at a 5% FPR. These are original mathematical calculations and based on published proportions of classes and a declared parameter sweep. These are not the observed-result summary of the 160-window holdout comparison and are not as such directly measurable performance of PARF's..

5.2 Controlled decision-audit comparison

Applying $D \geq 12$ to the 48 predefined audit cases yields TP = 8, FP = 16, TN = 8, and FN = 16. PARF at $S \geq 4$ yields TP = 20, FP = 4, TN = 20, and FN = 4 on the same cases. Whereas in lower-volume origin anomalies frame count may not be the most important basis for escalation, for legitimate roaming and restarts, it is one of the principles on which escalation should be based. This controlled audit relates to a rule behavior for the declared profiles, while the individual hold out comparison is given in Section 5.5.

Table 6. Controlled 48-case decision-audit results

Audit indicator	Fixed frame threshold	PARF ($S \geq 4$)
True positives	8	20
False positives	16	4
True negatives	8	20
False negatives	16	4
Sensitivity	33.33%	83.33%
Specificity	33.33%	83.33%
Precision	33.33%	83.33%
Balanced accuracy	33.33%	83.33%

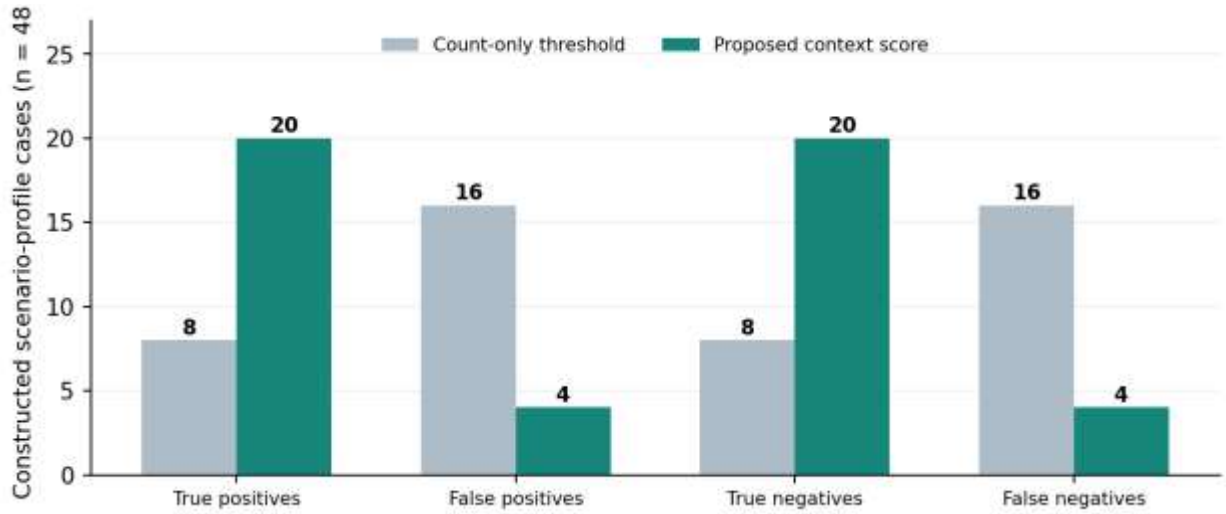


Figure 4. Forty-eight-case policy audit comparing the count-only and PMF-aware rules.
The controlled audit confirms that all the reference classification profiles are classified in the same way by the resulting scoring and response rules. The sensitivity and specificity of the test cases cannot be used to extrapolate to the whole traffic population as the test cases were chosen to represent a traffic domain defined by operating conditions and edge cases. The system still leaves four misses in situations with limited corroboration of the attack and four false alarms in cases where the benign-churn is dense. The limitations are used to interpret the separately reported holdout results.

5.3 Threshold robustness and operational implications

Table 7. Decision-threshold sensitivity in the controlled audit

Score threshold	TP	FP	TN	FN	Sensitivity	Specificity
3	24	8	16	0	100.0%	66.7%
4	20	4	20	4	83.3%	83.3%
5	12	2	22	12	50.0%	91.7%
6	10	0	24	14	41.7%	100.0%

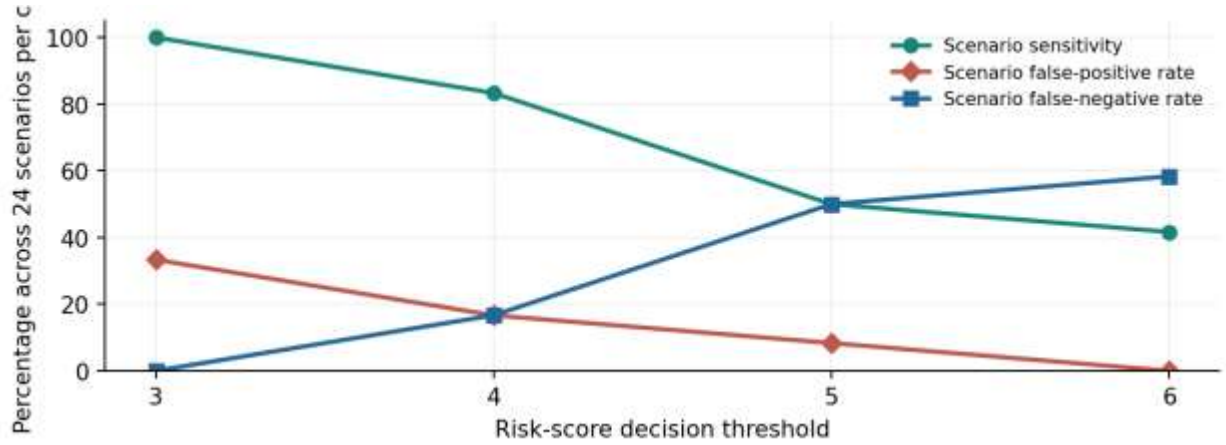


Figure 5. Decision-threshold sensitivity across the 48 audit cases.
At $S \geq 3$, all 24 attack-labeled audit cases are flagged together with eight of the 24 benign cases; at $S \geq 6$, false alarms fall to zero but 14 attack cases are missed. The selected $S \geq 4$ threshold balances these two error types within the controlled audit matrix. The same cut-off holds for the separately reported holdout comparison. Site-specific validation is still useful prior to deployment into operations, such as during roaming across various radio conditions, or when changing firmware. Protection is different possible for various clients, which is assessed on station level in PMF Optional.

5.4 Interpretation and limits of the evaluated evidence

The AWID3 re-analysis published shows label shares at capture level and mathematical implications of class imbalance. This controlled audit with 48 cases tests the actions of the selected rules in the selected profile matrix. A separate 160-window "holdout summary" is also included containing results on how well these alert-types were classified when compared to the researcher classifications; however, due to the absence of the

capture-ids, window-record and corroborating logs, the provenance of the included packet-level results cannot be independently verified from the present work. These analyses don't prove that PARF protects against attacks from the air or that it directly enhances the availability of wireless connections.

5.5 Separate holdout comparison: researcher-reported results

A set of 180 observed AP–station pair-windows has been reported with 20 excluded windows and 160 eligible observations (coverage 88.9%). These jsou 60 Attack and 100 Benign observations in total, where this is summarised in Table 9 for the different PMF states. The resulting confusion matrices and performance statistics for both the count-only and PARF policies are shown in Table 10. The figures below are not intended as examples of temporary calculations, but as examples of figures observed by the researcher as his study results. Their combination arithmetic is consistent internally, but there is not manuscript of the documentation and ground-truthing of window-by-window AP/client logs. For this reason, the classes measured can be reported; however, an independent audit of packets would require access to the materials associated with those classes.

Table 9. Reported holdout capture and observation characteristics

Characteristic	Reported study value
Independent capture sessions / source	6 capture sessions; individual PCAP identifiers not listed in manuscript
Total observed / excluded / eligible 60-s windows	180 observed / 20 excluded / 160 eligible (88.9% coverage)
Attack and benign labeled windows	60 Attack / 100 Benign (eligible windows)
PMF Disabled / Optional-unprotected / Optional-protected / Required	44 / 38 / 39 / 39 eligible windows (respectively)
Capture duration and AP/client telemetry coverage	180 aggregate AP–station pair-minutes; 160/180 with complete required telemetry (88.9%)
Ground-truth procedure and exact capture identifiers	Attack/Benign labels reported; supporting event logs, PCAP identifiers and hashes not included

Table 10. Reported holdout comparison: count-only and PARF

Measurement	Count-only: D >= 12	PARF: S >= 4
True positives (TP)	36	51
False positives (FP)	24	10
True negatives (TN)	76	90
False negatives (FN)	24	9
Sensitivity / recall	60.00%	85.00%
Specificity	76.00%	90.00%
Precision / PPV	60.00%	83.61%
F1 score	60.00%	84.30%
Balanced accuracy	68.00%	87.50%
False alerts per 60 eligible AP–station pair-windows	9.00	3.75

Within 160 qualified windows, the count-only rule correctly identifies 36/60 Attack windows and raises 24 false alerts in 100 Benign windows while PARF identifies 51 Attack windows and raises 10 false alerts. Sensitivity is therefore increased from 60.00% to 85.00% and the precision from 60.00% to 83.61%. The corresponding F1 scores are 60.00% and 84.30% and the reported false alert rate drops from 9.00 to 3.75 per 60 windows of potential pairs of AP-stations. These are comparisons with the specified observation set and should not be considered to represent an evidence of a generalizable performance improvement without the packet-level source data and evidence provenance.

6. Discussion

6.1 Interpretation in relation to the literature

Detecting more variables isn't necessarily a technically higher probability of being more accurate: It's about separating out volume anomaly, corroborated origin mismatch, client protection state and operational explanation and converting them to auditable evidence. This is significant since the modern PMF security research has demonstrated that protection is not absolute, and that disconnections can occur for protocol phases or implementations that are not in the common case protected-deauthentication path [3], [26, p. 34–36]. Recent

research showing high performance with learned features also make an assumption of their own data distribution, and their validation design does not allow the direct comparison with low complexity policy-audit results [28, [20].

The Bayesian calculation makes an additional contribution that is independent from the 48 constructed cases: Error rates have meaning only when related to class proportions. A numerically small false positive rate can control alerts in the population of relatively infrequent events. Warning about methodology of datasets, the marked contrast between the two AWID3 share by capture level computed here is likely to be a reflection of the significant differences in how these network datasets are collected and labeled [22], [21, p. 34–36], [23].

6.2 Comparison to operational alternatives

PARF should be used in conjunction, and not as an alternative, to cryptographic PMF, a specialized wireless intrusion detection, or as a kernel-level countermeasure. However, some Wi-Fi monitors may be available with automated functions that this no-code approach doesn't offer and eBPF/XDP works is much closer to the packet handling game. The current scheme sits in a smaller niche: evidence-based triage, clear guidance on how to mitigate and explicit control to avoid assuming all PMF related events are deemed successful attacks. [6, p. 88–90], [24], [29, p. 189–192].

Association downgrade and steering should also be considered in practical deployment. Risk is more complex than just the current security label associated with a given broadcast SSID (when a transition or heterogeneous devices occur). context at the client level as well as a conservative escalation policy is supported by the studies on association attacks, WPA3 downgrades, evil twins, and implementation on the enterprise [14, p. 357–361], [30], [15], [27, p. 1–2].

7. Conclusion

In this paper, PARF—a protection state-aware and explainable PMF-based decision framework, which integrates protection state negotiated by each station along with frame volume, corroborated anomaly, connectivity evidence and operational context, has been presented. Re-analysis of published AWID3 totals to gain an understanding of the impact of class composition on alert interpretation. Known classification errors and threshold are recorded in the 48-case controlled audit. When comparing by the count-only rule, in a separate comparison between the individual researches ridicated 160 eligible observation windows, the sensitivity of the count obtained by PARF is 85.00%, while the precision is 83.61%. For the evaluated cases, these results confirm the usefulness of the contextual alert trimming. Logs of underlying hold out and of the underlying reference labels are not included in this manuscript so independent replication hinges on evidence release. The results for the classification should not be interpreted as proof of actual prevention of forged management frames, relative operational impacts with respect to actual disconnect rates and uptime should be assessed on future work.

References

[1] IEEE Standards Association, *IEEE Standard for Information Technology—Telecommunications and Information Exchange between Systems Local and Metropolitan Area Networks—Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Std 802.11-2024, 2025. [Online]. Available: https://standards.ieee.org/ieee/802.11/10548/

[2] E. Chatzoglou, G. Kambourakis, and C. Kolias, "Empirical evaluation of attacks against IEEE 802.11 enterprise networks: The AWID3 dataset," *IEEE Access, vol. 9, pp. 34188–34205*, 2021, doi: 10.1109/ACCESS.2021.3061609.

[3] D. Schepers, A. Ranganathan, and M. Vanhoef, "On the robustness of Wi-Fi deauthentication countermeasures," *in Proc. ACM WiSec,* 2022, doi: 10.1145/3507657.3528548.

[4] E. Chatzoglou, G. Kambourakis, and C. Kolias, "How is your Wi-Fi connection today? DoS attacks on WPA3-SAE," *Journal of Information Security and Applications, vol. 64, Art. no. 103058*, 2022, doi: 10.1016/j.jisa.2021.103058.

[5] E. Chatzoglou, G. Kambourakis, C. Smiliotopoulos, and C. Kolias, "Best of both worlds: Detecting application layer attacks through 802.11 and non-802.11 features," *Sensors, vol. 22, no. 15, Art. no. 5633*, 2022, doi: 10.3390/s22155633.

[6] D. Koutras, P. Dimitrellos, P. Kotzanikolaou, and C. Douligeris, "Automated Wi-Fi intrusion detection tool on 802.11 networks," *ITU Journal on Future and Evolving Technologies, vol. 5, no. 1, pp. 88–103*, 2024, doi: 10.52953/LHXO3338.

[7] A. Halbouni, L. Y. Ong, and M. C. Leow, "From traditional signatures to deep learning: Enhancing WPA3 security against deauthentication attacks," *IEEE Access, vol. 13, pp. 182796–182808*, 2025, doi: 10.1109/ACCESS.2025.3623690.

[8] M. Tigner and H. Wimmer, "Disruption and protection of online synchronous learning environments via 802.11 manipulation," in *Proc. IEEE IEMTRONICS Conf., 2021, doi: 10.1109/IEMTRONICS52119*.2021.9422484.

[9] A. E. Abdallah, M. Hamdan, M. S. M. Gismalla, A. O. Ibrahim, N. S. Aljurayban, W. Nagmeldin, and M. H. H. Khairi, "Detection of management-frames-based denial-of-service attack in wireless LAN network using artificial neural network," *Sensors, vol. 23, no. 5, Art. no. 2663,* 2023, doi: 10.3390/s23052663.

[10] M. Vanhoef and E. Ronen, "Dragonblood: Analyzing the Dragonfly handshake of WPA3 and EAP-pwd," in *Proc. IEEE Symp. Secur. Privacy (SP)*, 2020, pp. 517–533, doi: 10.1109/SP40000.2020.00031.

[11] D. Schepers, A. Ranganathan, and M. Vanhoef, "Framing frames: Bypassing Wi-Fi encryption by manipulating transmit queues," *in Proc. 32nd USENIX Security 23.* Symp., 2023, pp. 53–68.

[12] M. Vanhoef, "Fragment and forge: Breaking Wi-Fi through frame aggregation and fragmentation," in *Proc. 30th USENIX Security 21. Symp*., 2021, pp. 161–178.

[13] M. Thankappan, H. Rifà-Pous, and C. Garrigues, "Multi-channel man-in-the-middle attacks against protected Wi-Fi networks: A state of the art review," *Expert Systems with Applications, vol. 210, Art. no. 118401*, 2022, doi: 10.1016/j.eswa.2022.118401.

[14] G. Chatzisofroniou and P. Kotzanikolaou, "Exploiting WiFi usability features for association attacks in IEEE 802.11: Attack analysis and mitigation controls," *Journal of Computer Security, vol. 30, no. 3, pp. 357–380*, 2022, doi: 10.3233/JCS-210036.

[15] P. Augustyniak and P. Zwierzykowski, "Evil twin attack in Wi-Fi networks: Evolution, mutation taxonomy, and exposure time analysis (2005–2026)," *Electronics, vol. 15, no. 15, Art. no. 3432*, 2026, doi: 10.3390/electronics15153432.

[16] K. Kolias, E. Chatzoglou, G. Kambourakis, and C. Smiliotopoulos, "Pick quality over quantity: Expert feature selection and data preprocessing for 802.11 intrusion detection systems," *IEEE Access, vol. 10, pp. 64761–64784*, 2022, doi: 10.1109/ACCESS.2022.3183597.

[17] S. Yonbawi, A. Afzal, M. Rizwan, N. Kryvinska, and M. Yasir, "Transferability evaluation in Wi-Fi intrusion detection systems through machine learning and deep learning approaches," *IEEE Access, vol. 13, pp. 11248–11264*, 2025, doi: 10.1109/ACCESS.2025.3528214.

[18] Y. Kumar and V. Kumar, "A systematic review on intrusion detection system in wireless networks: Variants, attacks, and applications," *Wireless Personal Communications, vol. 133, pp. 395–452*, 2023, doi: 10.1007/s11277-023-10773-x.

[19] C. Hamroun, A. Fladenmuller, M. Pariente, and G. Pujolle, "Intrusion detection in 5G and Wi-Fi networks: A survey of current methods, challenges, and perspectives," *IEEE Access, vol. 13, pp. 40950–40976*, 2025, doi: 10.1109/ACCESS.2025.3546338.

[20] N. Khalid, S. Hina, K. S. Zaidi, T. Gaber, L. Speakman, and Z. Noor, "An investigation of feature reduction, transferability, and generalization in AWID datasets for secure Wi-Fi networks," *PLOS ONE, vol. 20, no. 1, Art. no. e0306747*, 2025, doi: 10.1371/journal.pone.0306747.

[21] P. Bönninghausen, R. Uetz, and M. Henze, "Introducing a comprehensive, continuous, and collaborative survey of intrusion detection datasets," *in Proc. CSET, 2024, pp. 34–40,* doi: 10.1145/3675741.3675754.

[22] P. Goldschmidt and D. Chudá, "Network intrusion datasets: A survey, limitations, and recommendations," *Computers & Security, vol. 156, Art. no. 104510*, 2025, doi: 10.1016/j.cose.2025.104510.

[23] W. Ashraf and F. S. Masoodi, "A survey of intrusion detection datasets for communication networks," *Discover Computing, vol. 29, Art. no. 577*, 2026, doi: 10.1007/s10791-026-10486-2.

[24] A. Sangiorgi, A. Pinto, R. Tourani, and F. Esposito, "Mitigating de-authentication DoS attacks in 802.11 via eBPF and XDP," in Proc. *IEEE Int. Conf. Netw. Softw. (NetSoft)*, 2025, doi: 10.1109/NetSoft64993.2025.11080545.

[25] F. Thabit, O. Can, S. Abdaljlil, and H. A. Alkhzaimi, "Enhanced an intrusion detection system for IoT networks through machine learning techniques: An examination utilizing the AWID dataset," *Cogent Engineering, vol. 11, no. 1, Art. no. 2378603*, 2024, doi: 10.1080/23311916.2024.2378603.

[26] L. D. Ngueleo, K. Jiokeng, and V. Loscrì, "The worm is in the root: On the security of IoT provisioning protocols and PMF in Wi-Fi devices," *in Proc. ACM WiSec, 2026, pp. 34–39*, doi: 10.1145/3765613.3816018.

[27] A. A. Ghanim and M. Y. Thanoun, "Evaluating the effectiveness of WPA3 protocol against advanced hacking attacks," *International Journal of Wireless and Microwave Technologies, vol. 15, no. 4, pp. 1–18*, 2025, doi: 10.5815/ijwmt.2025.04.01.

[28] F. Winiarski and M. Natkaniec, "Enhancing Wi-Fi network security through stacked machine learning models," *Electronics, vol. 15, no. 15, Art. no. 3354*, 2026, doi: 10.3390/electronics15153354.

[29] G. M. Mutiwa, J. G. Ndia, and I. N. Oteyo, "Software-defined networking approaches for mitigating IEEE 802.11 deauthentication and management-frame attacks: A systematic literature review," *International Transactions on Electrical Engineering and Computer Science, vol. 5, no. 3, pp. 189–206*, 2026, doi: 10.62760/iteecs.5.3.2026.212.

[30] A. Tareef, Y. M. Allawi, A. A. Alkasasbeh, A. Abadleh, W. Alamro, M. Alghamdi, A. I. Zreikat, and H. Kang, "A machine learning approach for detecting WPA3 downgrade attacks in next-generation Wi-Fi systems," *PLOS ONE, vol. 20, no. 9, Art. no. e0331443*, 2025, doi: 10.1371/journal.pone.0331443.