Uncovering Threats Before Connection: A Physical-Layer Approach for Intrusion Detection Using Wi-Fi CSI

Felipe Silveira de Almeida[1], **Lourenço Alves Pereira Júnior**[1]

[1]Divisão de Ciência da Computação - Instituto Tecnológico de Aeronáutica (ITA)
São José dos Campos, SP - Brasil

{felipefsa,ljr}@ita.br

Abstract. *This paper presents a proactive intrusion detection approach that uses Wi-Fi Channel State Information (CSI) to detect suspicious behavior at the physical and data link layers before access point association. The proposal was validated through two experimental studies using low-cost ESP32 devices: the first analyzed 70,513 instances in two environments and achieved 94.24% accuracy with Random Forest; the second, named Spider-Sense, analyzed approximately 800,000 instances in three environments and reached 99.95% accuracy and an F1-score of 1.000 with Decision Tree. The main contribution is a passive, affordable, and application-agnostic early-warning layer that complements traditional NIDS before network-layer traffic is available.*

1. Introduction

Wireless networks are now part of the critical attack surface of residential, corporate, academic, healthcare, and defense infrastructures. This dependence increases the need for intrusion detection mechanisms that do not rely exclusively on traffic already generated inside a protected network. Traditional Intrusion Detection Systems (IDS) and Network Intrusion Detection Systems (NIDS) usually inspect packets, flows, headers, payloads, or traffic statistics at the network, transport, or application layers [Mirsky et al. 2018, De Carvalho Bertoli et al. 2021, Aleesa et al. 2020]. This is effective for many attack classes, but it also means that detection often starts only when the suspicious device is already associated with the network or when traffic evidence is already observable.

This work investigates a complementary direction: using the wireless channel itself as a security sensor. Channel State Information (CSI) describes how a Wi-Fi signal propagates between transmitter and receiver across Orthogonal Frequency Division Multiplexing (OFDM) subcarriers, within the broader IEEE 802.11 physical-layer mechanisms [IEEE 2021]. Its amplitude and phase components are affected by reflection, attenuation, diffraction, and multipath, making CSI a fine-grained indicator of physical changes in the environment [Ma et al. 2019, Liu et al. 2020]. The central research question was: *can unauthorized Wi-Fi connection attempts be detected from CSI variations before conventional network-layer evidence becomes available?*

1.1. Motivations

Wireless attacks may start before a device becomes a legitimate network participant. Scenarios such as rogue devices, evil twin preparation, unauthorized association attempts, spoofing, handshake capture, and brute-force attempts create a relevant pre-association stage in which traffic-based NIDS has limited visibility. At the same time, Wi-Fi

sensing has matured as a research field, demonstrating that commodity radio signals can reveal motion, presence, gestures, and device-related changes [Tan et al. 2022, Cominelli et al. 2023]. This work therefore explores whether part of cybersecurity monitoring can be shifted to this earlier physical-layer stage.

This direction is also motivated by cost and deployability. The experiments used ESP32-based devices, which are inexpensive, compact, and compatible with IoT-style deployment. The goal was not to replace current NIDS, but to add a defense-in-depth layer able to raise an alert before the attacker effectively uses the victim network.

1.2. CSI-based Early Intrusion Detection

The proposed approach treats CSI as an Indicator of Compromise (IoC) generated by physical and link-layer behavior. A transmitter emits Wi-Fi frames; a receiver collects CSI; the raw CSI data are transformed into amplitude and phase features; and machine learning models classify the observed patterns as benign/neutral or malicious/offensive. Figure 1 summarizes the CSI-based early-detection concept and the Spider-Sense workflow.

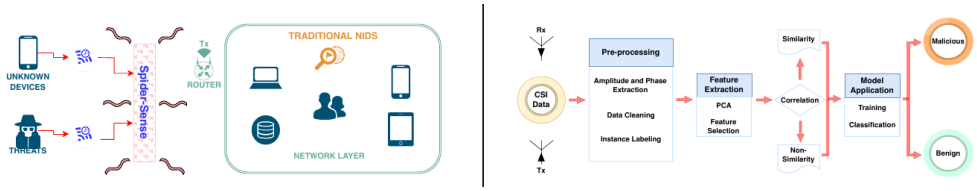


Figure 1. CSI-based early detection concept and Spider-Sense workflow.

1.3. Objectives

The work pursued the following objectives:

- identify whether Wi-Fi CSI can expose distinguishable patterns associated with unauthorized connection attempts before network association;
- design an experimental methodology using low-cost ESP32 devices to collect and label CSI data in controlled environments;
- evaluate classical machine learning algorithms for classifying benign/neutral and malicious/offensive behaviors;
- compare the first feasibility study with an extended Spider-Sense evaluation, showing the evolution from a proof-of-concept dataset to a larger, more robust validation;
- consolidate contributions, datasets, preprocessing pipelines, limitations, and future directions for CSI-based cybersecurity.

2. Related Work and Research Gap

CSI-based Wi-Fi sensing has been widely explored for activity recognition, localization, gesture recognition, vital-sign monitoring, and authentication [Ma et al. 2019, Liu et al. 2020, Wang et al. 2021]. These applications use the fact that the propagation channel changes when people, objects, or devices affect the electromagnetic environment. CSI is richer than Received Signal Strength Indicator

(RSSI) because it provides fine-grained subcarrier-level information instead of a single aggregate power measurement [Yang et al. 2013]. Research access to CSI has also been enabled by practical extraction toolchains, from early 802.11n CSI collection with the Intel 5300 NIC to more recent Nexmon-based extraction platforms [Halperin et al. 2011, Gringoli et al. 2019].

Security-oriented CSI research has also shown that wireless sensing creates both opportunities and risks. Prior work has investigated CSI capabilities and limitations in modern Wi-Fi sensing [Cominelli et al. 2023], the security and privacy consequences of commodity Wi-Fi sensing [Tan et al. 2022], and the possibility of revealing sensitive behavior such as mobile passwords through Wi-Fi signals [Meng et al. 2020]. Nevertheless, the literature mapping conducted in this work found that the direct use of Wi-Fi CSI as the central evidence for pre-association NIDS remained underexplored.

Traditional NIDS research, in turn, has advanced through machine learning, deep learning, autoencoders, online learning, and deployment-oriented frameworks [Mirsky et al. 2018, De Carvalho Bertoli et al. 2021, Aleesa et al. 2020]. These systems, however, generally depend on traffic features observed at higher OSI layers. Table 1 summarizes the gap addressed by this work.

Table 1. Research Gap and Positioning.

Area	Evidence	Objective	Pre-assoc. NIDS?
Traffic NIDS	Packets/flows	Intrusion detection from network traffic	No
Wi-Fi sensing	CSI/RSSI	Presence, activity, gesture, or environment recognition	No
CSI authentication	CSI profiles	User or device authentication	No
Our work	CSI amplitude/phase	Early IDS at layers 1 and 2	Yes

The originality is therefore not the isolated use of CSI or machine learning, but their combination in a cybersecurity architecture that detects suspicious pre-association behavior using physical and link-layer features collected by low-cost hardware. This makes the proposal application-agnostic: it does not depend on the payload, service, or application being attacked.

Accordingly, the novelty is application-driven and experimentally grounded: CSI-based classification, commonly explored for sensing and authentication, is used here as evidence for early intrusion detection before association with the target Wi-Fi network.

3. Physical-Layer IDS with Wi-Fi CSI

3.1. Methodology

The research followed an experimental pipeline composed of passive CSI capture, preprocessing, labeling, feature extraction, feature selection, and supervised classification. The ESP32 receiver collected CSI from Wi-Fi transmissions using the CSI support documented for ESP-IDF [Espressif Systems 2018] and stored records in CSV files. Repeated or null attributes were discarded, while the raw CSI field was transformed into amplitude and phase information across 52 subcarriers. During

exploratory analysis, the first extracted subcarrier position presented constant values across all capture files and environments. Since it carried no observable variance for distinguishing the monitored behaviors, it was treated as a non-informative feature and disregarded; the analysis therefore concentrated on the remaining 51 subcarriers.

Several preprocessing alternatives were evaluated, including Hampel filtering, phase calibration, low-pass filtering, Discrete Hilbert Transform, Discrete Wavelet Transform (DWT), interpolation, Principal Component Analysis (PCA), and outlier analysis. Local Outlier Factor (LOF) was used to study temporal deviations and support the interpretation of connection-attempt patterns. Feature selection with WrapperSubsetEval, J48/Decision Tree, and GreedyStepwise helped identify the subcarriers most affected in each environment.

3.2. First Experimental Evaluation

The first study, published at SBRC 2024, evaluated the feasibility of using CSI for intrusion detection with physical-layer features. The experiments were conducted in two controlled environments: a semi-anechoic chamber and Lab-C2DC at ITA. The setup used an ESP32 Wi-Fi board as receiver, a TP-Link Archer C60 router as transmitter, and a DELL Inspiron 15 Gaming 7567 notebook for data collection and processing. The protocol considered two users, four devices, and situations such as static presence, movement, absence of users/devices, and connection attempts with an incorrect password.

The resulting dataset contained 32 captures and 70,513 labeled instances, with 61,713 neutral instances and 8,800 offensive instances. The evaluation used ten-fold cross-validation: the dataset was divided into ten parts, with nine used for training and one for testing in each of ten rounds. SVM, Random Forest, KNN, Random Tree, and Naive Bayes were compared. Although SVM achieved the highest accuracy and F1-score, with 94.43% and 0.969, respectively, its construction time was 360.79 seconds. Random Forest was therefore selected as the best trade-off for the feasibility study, reaching 94.24% accuracy and 0.968 F1-score with a construction time of 28.67 seconds.

This first evaluation established the feasibility of distinguishing neutral and offensive behavior from CSI captured with low-cost hardware. It also showed that the affected subcarriers varied by environment, reinforcing the importance of environment-aware calibration.

4. Spider-Sense: Extended Evaluation

4.1. Experimental Setup

The second study, published at IEEE GLOBECOM 2024, extended the initial methodology and introduced Spider-Sense. The evaluation expanded the dataset, labels, device diversity, and environmental diversity. CSI was captured in three controlled environments: a semi-anechoic chamber, an electronic studies laboratory, and a computational research laboratory. The third environment included more furniture, reflective objects, electronic equipment, and noise, making the scenario closer to realistic multipath conditions.

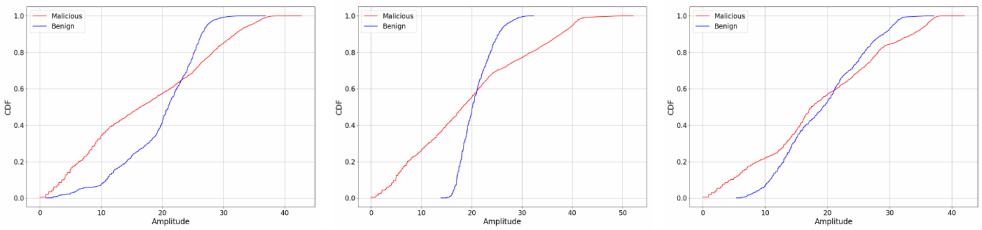
The experimental setup used a Heltec Automation WiFi LoRa 32(V2) module based on ESP32 + SX127x as receiver, a TP-Link Archer C60 router as transmitter, and

a DELL Inspiron 15 Gaming 7567 laptop. Two individuals and ten portable devices were observed, including laptops, iPhones, and Android devices. Each capture lasted 60 seconds, with line-of-sight positioning and calibration before acquisition.

4.2. Results and Analysis

The Spider-Sense dataset comprised approximately 800,000 instances: 400,000 benign instances from successful connection attempts and 400,000 malicious instances from brute-force connection attempts. Using ten-fold cross-validation, the study evaluated SVM, Random Forest, KNN, Decision Tree, and Naive Bayes across the three environments. The Decision Tree algorithm emerged as the most appropriate model because it preserved high accuracy while remaining compact and computationally suitable for constrained devices. In Lab 2, the most challenging environment, Decision Tree achieved 99.95% accuracy and an F1-score of 1.000.

Figure 2 illustrates the Cumulative Distribution Function (CDF) analysis of CSI amplitudes in the Spider-Sense dataset. Benign samples represented successful connection attempts, whereas malicious samples represented brute-force attempts. Therefore, the separation between the curves provides evidence that brute-force connection attempts produce CSI amplitude patterns distinguishable from successful connection attempts, rather than merely reflecting device presence.



(a) Laptop. (b) iPhone operating system. (c) Android operating system.

Figure 2. CSI amplitude CDF for successful and brute-force connection attempts.

Table 2 summarizes the progression from feasibility to expanded Spider-Sense validation.

Table 2. Dataset Evolution and Best Results.

Study	Dataset	Best result
SBRC'24	2 environments; 32 captures; 70,513 instances (61,713 neutral / 8,800 offensive)	Random Forest: Acc. 94.24%; F1 0.968; build time 28.67 s
GLOBECOM'24	3 environments; 120 captures; ≈800,000 instances (400,000 benign / 400,000 malicious)	Decision Tree: Acc. 99.95%; F1 1.000; build time 1.29 s

The results demonstrate that CSI can provide early evidence for intrusion detection before conventional traffic-based NIDS receives meaningful input. This work does not claim that CSI alone is sufficient for all deployment conditions; instead, it validates CSI as an early-warning source that can be correlated with network-layer evidence in a defense-in-depth architecture.

4.3. Operational Perspective

In an operational scenario, the proposed layer would serve as a wireless perimeter-security layer before conventional traffic-based inspection. In principle, a low-cost receiver could observe the wireless channel around a protected access point and extract CSI variations from regular Wi-Fi transmissions. When the observed pattern deviates from the calibrated baseline or resembles connection-attempt behavior, the system could generate an early alert to be correlated with authentication logs, access point telemetry, and traditional NIDS sensors. This correlation is important because CSI should be interpreted as contextual evidence rather than as an isolated final verdict.

This perspective clarifies the role of the proposed approach in a layered defense architecture. Traffic-based NIDS remains essential after association, when packet headers, flows, and payload-related metadata become available. The CSI-based layer, however, explores an earlier moment in the attack timeline: the physical and link-layer interaction between devices and the wireless medium. Its value is therefore not limited to classification accuracy, but also to reducing detection latency and increasing situational awareness before the attacker becomes an active network participant.

5. Conclusions

The results show that Wi-Fi CSI can be used as a physical-layer source of evidence for proactive intrusion detection. Its main scientific contribution is shifting part of the detection process to OSI layers 1 and 2, before the suspicious device becomes associated with the target Wi-Fi network. This approach fills a gap between traffic-based NIDS and Wi-Fi sensing research: it uses sensing not only to recognize activity or presence, but to support cybersecurity decisions.

Together, the two studies validate a progression from a 70,513-instance feasibility study to an approximately 800,000-instance Spider-Sense evaluation, reaching 99.95% accuracy and an F1-score of 1.000 with Decision Tree.

5.1. Contributions

1. a CSI-based methodology for early intrusion detection at the physical and data link layers, independent of payload inspection and network-layer traffic;
2. two labeled datasets and experimental protocols based on low-cost ESP32 devices;
3. evidence that brute-force and unauthorized connection attempts produce distinguishable CSI amplitude patterns;
4. comparative evaluation of classical machine learning algorithms for CSI-based intrusion detection;
5. a practical research direction for low-cost, passive, application-agnostic Wi-Fi security monitoring;
6. discussion of limitations, including environment dependency, calibration needs, supervised learning dependence, and embedded-processing constraints.

5.2. Scientific Production

- Almeida, F. S.; Trindade, E. F. G.; Pettersson, M. I.; Machado, R.; Pereira Júnior, L. A. *Spider-Sense: Wi-Fi CSI as a Sixth Sense for Early Detection in Network Intrusion Detection Systems*. IEEE GLOBECOM 2024.

- Almeida, F. S.; Trindade, E. F. G.; Pereira Júnior, L. A. *Detecção de pessoas e dispositivos utilizando Channel State Information: IDS com features de camada física*. SBRC 2024.
- Almeida, F. S.; Trindade, E.; Pereira Júnior, L. P. *Detecção Proativa de Intrusos em Redes Wi-Fi Utilizando CSI e Machine Learning*. SIGE 2024.
- Almeida, F. S.; Freitas, O.; Pereira Júnior, L. P. *Defesa Cibernética com IA: Aplicações do Wi-Fi CSI para Proteção, Exploração e Ataque*. SIA-EMPN 2024.
- Trindade, E. F. G.; Almeida, F. S.; Pereira Júnior, L. A. *Além do Sinal: Autenticação Biofísica com Wi-Fi CSI e Raspberry Pi*. SBSeg 2024.
- Trindade, E.; Almeida, F. S.; Pereira Júnior, L. P. *Autenticação de Usuários com Wi-Fi CSI: Uma Fonte Estratégica para Inteligência*. SIGE 2024.

5.3. Limitations and Future Work

The experiments were conducted in controlled environments, which was necessary to isolate the effect of connection attempts on CSI and to label the dataset consistently. The semi-anechoic chamber reduced internal reflections and served as a cleaner baseline for observing signal variations. The laboratory environments, in turn, introduced furniture, electronic equipment, reflective objects, and multipath effects, bringing the evaluation closer to realistic indoor deployments. Even so, real networks may include moving furniture, multiple access points, simultaneous users, channel changes, interference from neighboring networks, and adversarial attempts to mask or mimic signal patterns.

A practical deployment would require an initial calibration phase. During calibration, the receivers would observe the normal behavior of each monitored site, identify the most stable and most informative subcarriers, and update the detection model as the environment changes. This is important because CSI is strongly dependent on geometry, antenna placement, hardware characteristics, and propagation conditions. Therefore, the learned patterns should not be interpreted as universal signatures, but as site-aware evidence extracted from the monitored environment.

The current approach also depends on supervised learning and offline processing. This makes it suitable for validating the feasibility of CSI-based intrusion detection, but further work is required before operational deployment. Future versions should investigate online learning, unsupervised anomaly detection, transfer learning across environments, and lightweight inference on constrained hardware. These directions would reduce the cost of labeling new environments and improve adaptation to long-term variations in the wireless channel.

Future work also includes real-time detection, integration with traditional IDS, multiple ESP32 receivers for collaborative sensing, embedded inference with TinyML, multi-band analysis at 2.4 GHz, 5 GHz, and 6 GHz, and validation in offices, campuses, hospitals, public spaces, and defense facilities. The long-term vision is not a standalone replacement for existing NIDS, but a new sensing layer that detects earlier and enriches the security context available to traditional monitoring systems.

Thesis Availability

The full text of this thesis is available online at the ITA Digital Library:

`http://www.bdita.bibl.ita.br/tesesdigitais/80502.pdf`.

References

Aleesa, A. M., Zaidan, B. B., Zaidan, A. A., and Sahar, N. M. (2020). Review of intrusion detection systems based on deep learning techniques: coherent taxonomy, challenges, motivations, recommendations, substantial analysis and future directions. *Neural Computing and Applications*, 32(14):9827–9858.

Cominelli, M., Gringoli, F., and Restuccia, F. (2023). Exposing the csi: A systematic investigation of csi-based wi-fi sensing capabilities and limitations. In *2023 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 81–90.

De Carvalho Bertoli, G., Pereira Júnior, L. A., Saotome, O., Dos Santos, A. L., Verri, F. A. N., Marcondes, C. A. C., Barbieri, S., Rodrigues, M. S., and Parente De Oliveira, J. M. (2021). An end-to-end framework for machine learning-based network intrusion detection system. *IEEE Access*, 9:106790–106805.

Espressif Systems (2018). Csi documentation – esp-idf programming guide. `https://docs.espressif.com/projects/esp-idf/en/latest/esp32/api-guides/wifi.html#wi-fi-csi`. Accessed: 2024-01-20.

Gringoli, F., Schulz, M., Link, J., and Hollick, M. (2019). Free your csi: A channel state information extraction platform for modern wi-fi chipsets. In *Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, WiNTECH '19, pages 21–28, New York, NY, USA. Association for Computing Machinery.

Halperin, D., Hu, W., Sheth, A., and Wetherall, D. (2011). Tool release: gathering 802.11n traces with channel state information. *SIGCOMM Comput. Commun. Rev.*, 41(1):53.

IEEE (2021). 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 802.11-2020 (Revision of IEEE Std 802.11-2016)*, pages 1–4379.

Liu, J., Liu, H., Chen, Y., Wang, Y., and Wang, C. (2020). Wireless sensing for human activity: A survey. *IEEE Communications Surveys & Tutorials*, 22(3):1629–1645.

Ma, Y., Zhou, G., and Wang, S. (2019). Wifi sensing with channel state information: A survey. *ACM Comput. Surv.*, 52(3):46.

Meng, Y., Li, J., Zhu, H., Liang, X., Liu, Y., and Ruan, N. (2020). Revealing your mobile password via wifi signals: Attacks and countermeasures. *IEEE Transactions on Mobile Computing*, 19(2):432–449.

Mirsky, Y., Doitshman, T., Elovici, Y., and Shabtai, A. (2018). Kitsune: An ensemble of autoencoders for online network intrusion detection. *Network and Distributed System Security Symposium (NDSS)*.

Tan, S., Ren, Y., Yang, J., and Chen, Y. (2022). Commodity wifi sensing in ten years: Status, challenges, and opportunities. *IEEE Internet of Things Journal*, 9(18):17832–17843.

Wang, Z., Dou, W., Ma, M., Feng, X., Huang, Z., Zhang, C., Guo, Y., and Chen, D. (2021). A survey of user authentication based on channel state information. *Wireless Communications and Mobile Computing*, 2021(1):6636665.

Yang, Z., Zhou, Z., and Liu, Y. (2013). From rssi to csi: Indoor localization via channel response. *ACM Comput. Surv.*, 46(2).