Wavelet-Based IoT Device Fingerprinting for Network Security

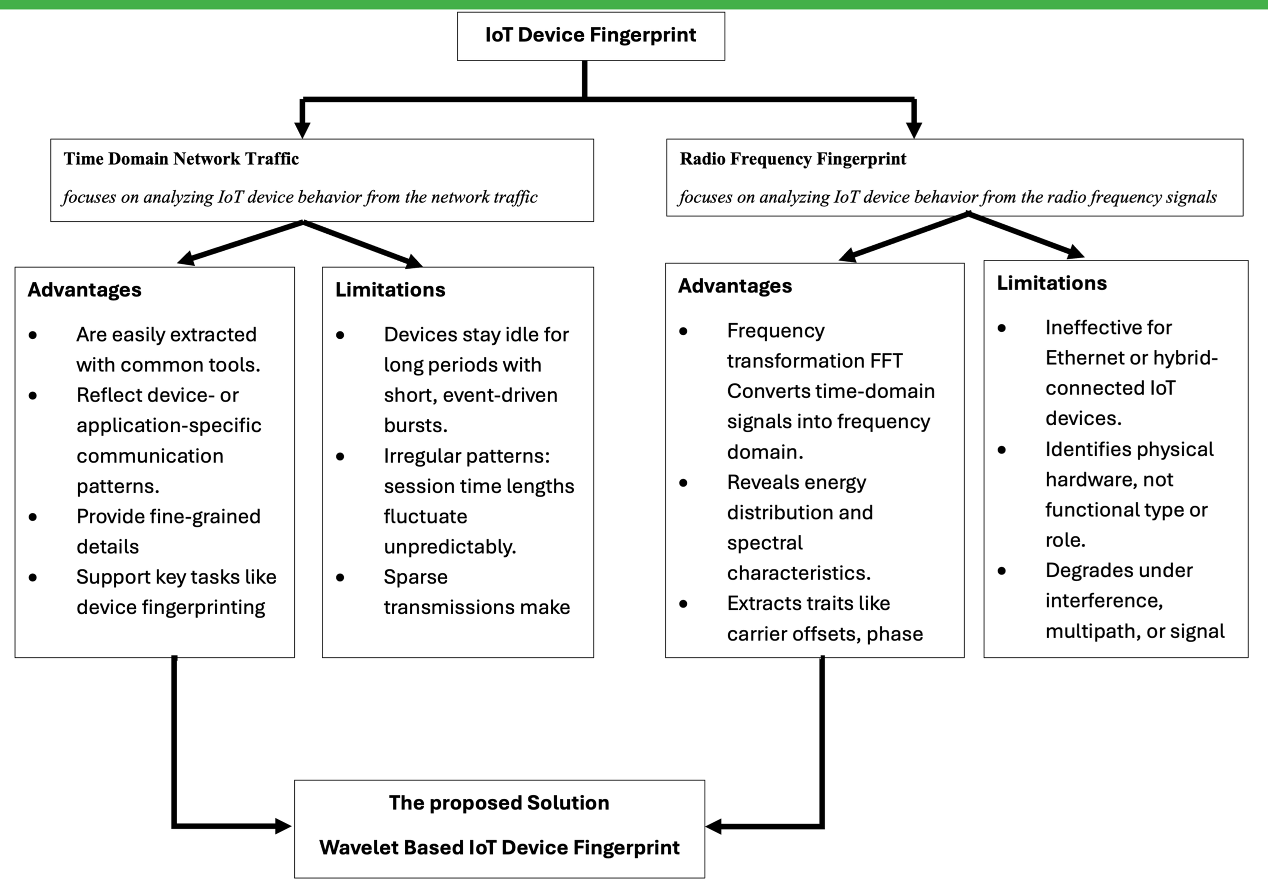
Cal Poly Pomona
College of Business Administration

Viet Nguyen, Adam Cheung, Sarah Acosta, Thuy Linh Pham

Under the supervision of Professor Abdelfattah Amamra

Computer Science Department, Cal Poly Pomona

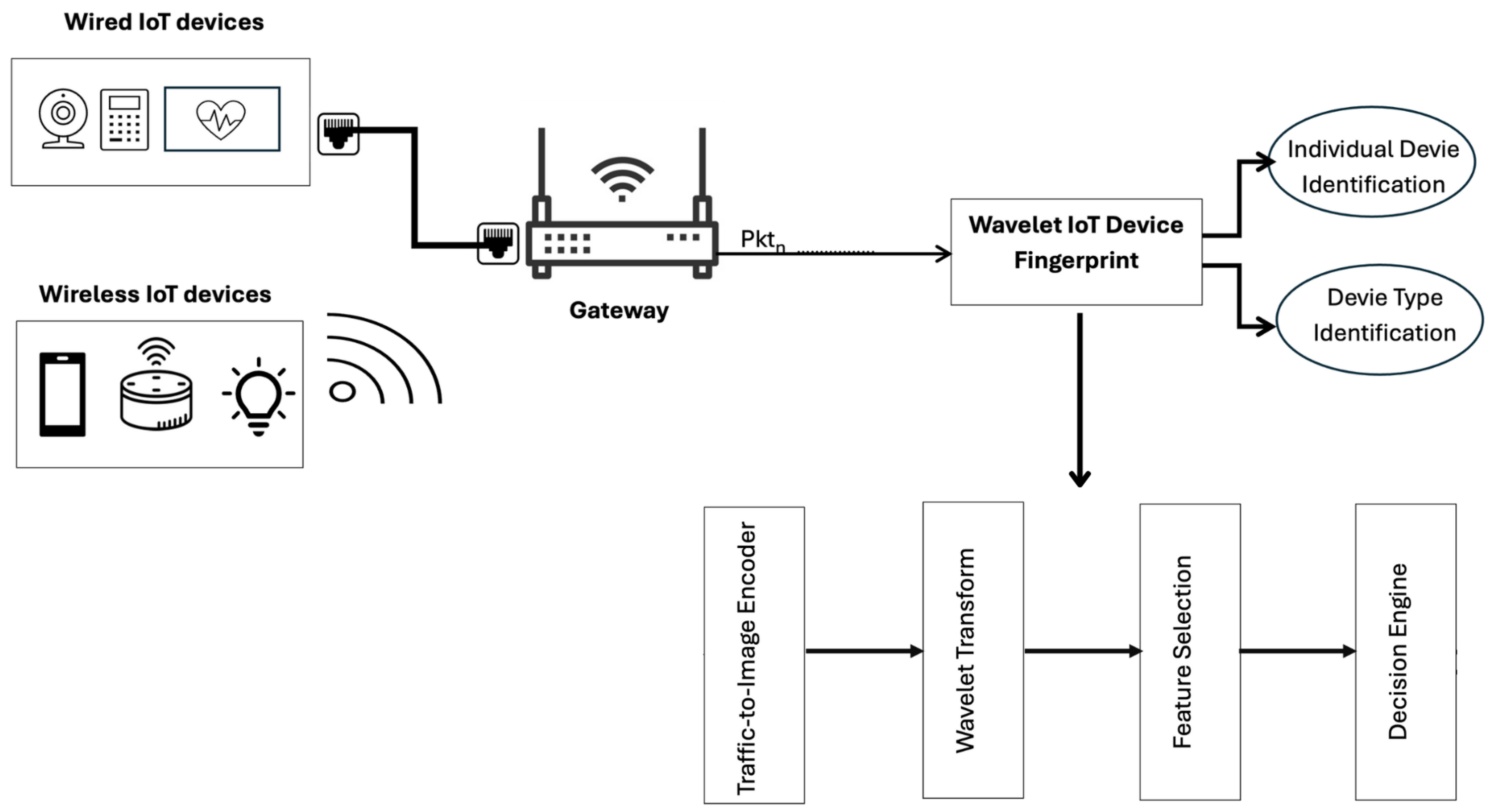
Problem statement



Proposed Solution

Key Features of Our Approach:

- **Overcomes RF Limitations:** Unlike Radio Frequency Fingerprinting which only works for wireless IoT devices, our solution enables fingerprinting of both wired and wireless devices through passive network traffic analysis
- **Wavelet-Based Analysis:** Applies DWT and WST to capture multi-scale behavioral patterns in network traffic, providing robust frequency-domain features that outperform traditional time-domain statistics
- **Network-Layer Collection:** Passively collects traffic at the gateway level, supporting heterogeneous IoT environments without requiring device modification or specialized hardware



Experimental Datasets

IoT Datasets Summary:

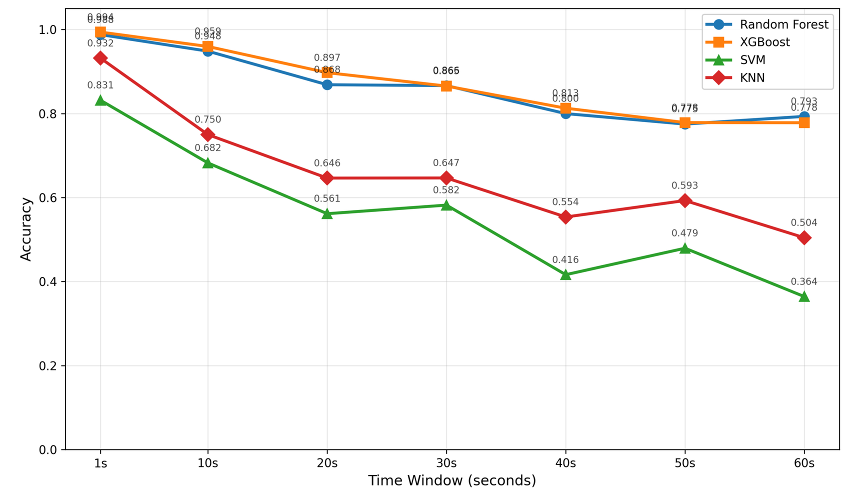
Evaluation Scope:

- **Dataset Diversity:** Three datasets (28, 60, 105 devices) assess scalability
- **Multi-Protocol Support:** WiFi, Zigbee, Z-Wave, MQTT, CoAP, Ethernet
- **Attack Scenarios:** Benign and DDoS attack traffic for robustness testing
- **Comprehensive Testing:** Device ID, type classification, sampling rates, feature reduction

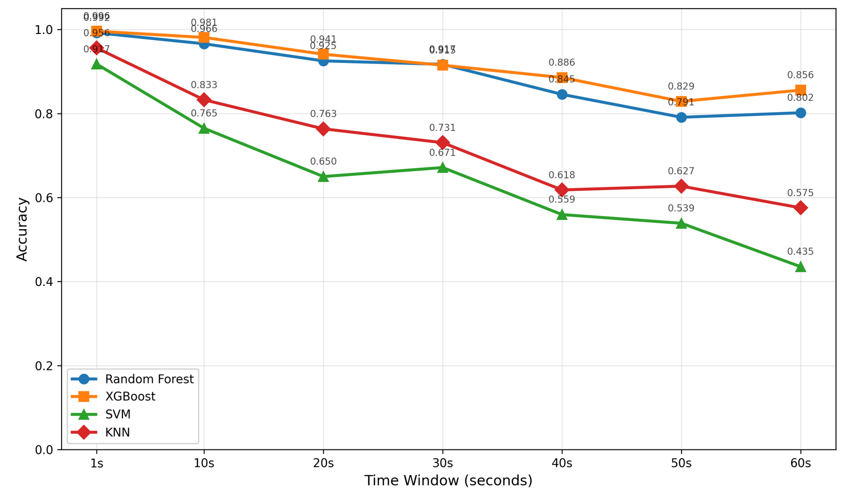
Dataset	Devices	Protocols	Traffic
UNSW IoT [13]	28	WiFi, MQTT, CoAP, Ethernet	Benign
CIC IoT 2022 [12]	60	WiFi, Zigbee, Z-Wave, Ethernet	Benign, Attack
CIC IoT 2023 [14]	105	WiFi, Zigbee, Z-Wave, MQTT, Ethernet	Benign, Attack

Sampling Rate Sensitivity Analysis

The sampling rate determines the temporal granularity of network traffic features. We evaluated classification performance across sampling intervals from 1 to 60 seconds, finding that a 1-second sampling rate achieves optimal accuracy by capturing fine-grained behavioral patterns while minimizing noise.



Individual Device ID



Device Type ID

Wavelet Transform Techniques

Why Wavelets Are Powerful:

Wavelet transforms are versatile signal processing techniques that analyze signals in both time and frequency domains simultaneously. They excel at capturing transient features, localized patterns, and multi-scale structures that traditional time-domain methods cannot detect.

Discrete Wavelet Transform (DWT):

- Uses low-pass and high-pass filters followed by downsampling to decompose signals into coefficients at different scales
- Produces **approximation coefficients** (low-frequency, smooth components) and **detail coefficients** (high-frequency, rapidly changing components)
- Effectively captures low-frequency, steady-state patterns in IoT traffic useful for identifying baseline device behaviors
- Multi-resolution analysis with linear time complexity O(N)

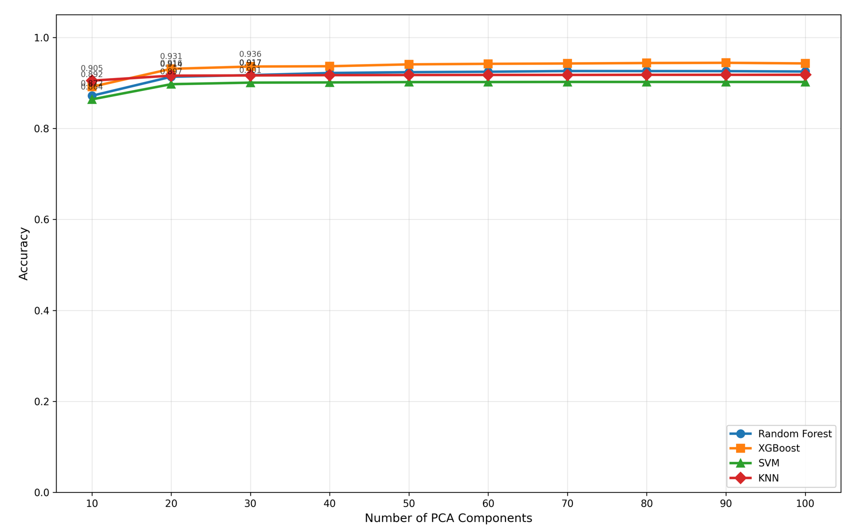
Wavelet Scattering Transform (WST):

- Deeper, hierarchical representation producing stable, translation-invariant features across multiple layers
- Cascades wavelet filters, modulus operations, and local averaging to build translation invariance
- Three key properties: translation invariance (robust to timing shifts), stability to small deformations (packet timing variations), and rich multi-scale detail
- Captures both coarse trends (periodic telemetry) and fine-grained quirks (millisecond-level jitter)

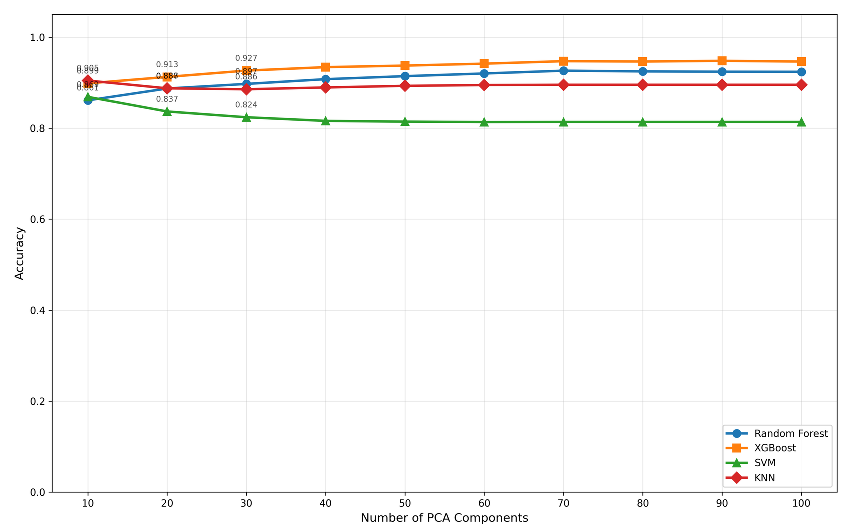
Both techniques leverage the informative richness of network traffic alongside the robustness of frequency-domain representations, significantly outperforming time-domain baselines across all scenarios.

Feature Reduction Analysis

Wavelet coefficients produce high-dimensional features that may contain redundancy. Using PCA, we identified that 30 features capture the most discriminative information, achieving optimal accuracy while reducing computational complexity and preventing overfitting.



DWT Feature Reduction



WST Feature Reduction

Experimental Results

Individual Device Identification Results:

Dataset	Algorithm	Baseline Accuracy	DWT Accuracy	WST Accuracy
CIC 2022	XGBoost	72%	**99%**	98%
	Random Forest	69%	**99%**	91%
	SVM	64%	81%	**90%**
	KNN	54%	90%	**92%**
CIC 2023	XGBoost	68%	**99%**	96%
	Random Forest	64%	**97%**	90%
	SVM	41%	74%	**86%**
	KNN	43%	83%	**86%**
UNSW IoT	XGBoost	77%	99%	**100%**
	Random Forest	68%	**99%**	**99%**
	SVM	61%	95%	**99%**
	KNN	61%	94%	**99%**

Key Findings:

- DWT shows dramatic improvement over baseline (up to +45%)
- WST achieves perfect 100% accuracy on UNSW dataset
- Both wavelet methods significantly outperform traditional approaches

Device Type Classification Results:

Dataset	Algorithm	Baseline Accuracy	DWT Accuracy	WST Accuracy
CIC 2022	XGBoost	72%	**100%**	**99%**
	Random Forest	69%	**99%**	**97%**
	SVM	64%	85%	**94%**
	KNN	54%	93%	**94%**
CIC 2023	XGBoost	68%	**99%**	**96%**
	Random Forest	64%	**98%**	93%
	SVM	41%	75%	**88%**
	KNN	43%	84%	87%
UNSW IoT	XGBoost	77%	**100%**	**100%**
	Random Forest	68%	**99%**	**99%**
	SVM	61%	**99%**	**99%**
	KNN	61%	**99%**	**99%**

Key Findings:

- DWT achieves perfect 100% on CIC2022 & UNSW
- WST reaches 100% on UNSW dataset
- Device type ID easier than individual device ID
- Both wavelets show dramatic improvements

Conclusion and Future Work

Key Contributions:

- **Robust Framework:** Introduced Wavelet IoT Device Fingerprint framework that overcomes limitations of traditional time-domain and RF-based methods through wavelet-based network traffic analysis
- **Multi-Scale Pattern Capture:** Leveraged DWT and WST to capture multi-scale behavioral patterns, enabling accurate device identification and type classification in heterogeneous IoT environments
- **Superior Performance:** Demonstrated that wavelet-based features consistently outperform time-domain baselines across three real-world datasets (CICIoT2022, CICIoT2023, UNSW)
- **Optimal Model Integration:** Achieved highest performance with XGBoost ensemble models, particularly when integrated with WST features, showing strong potential for federated learning applications
- **Deployment-Ready Solution:** Validated framework effectiveness for secure, scalable IoT environments through passive network-layer traffic collection supporting both wired and wireless devices

Future Research Directions:

- Real-time application deployment and optimization
- Federated learning architectures for privacy-preserving fingerprinting
- Advanced deep learning model integration
- Large-scale distributed IoT environment testing

This work has been submitted to the NDSS 2026 Conference and is currently under review. 2025