Paper Title:
A Study of MAC Address Randomization in Mobile Devices and When it Fails

DOI:
10.1515/popets-2017-0054

Problem Identified:
MAC address randomization is intended to protect user privacy by preventing tracking of devices through WiFi probe requests. However, the adoption of this privacy measure is inconsistent across device manufacturers and operating systems, and many implementations contain critical flaws. These flaws allow attackers to bypass randomization and identify or track devices using their real MAC addresses.

Proposed Solution / Contribution:
The paper presents the first large-scale empirical study of MAC address randomization in the wild. It:

- Analyzes over 600 GB of 802.11 traffic from 2.8 million devices.
- Breaks down how different manufacturers and OS versions implement MAC randomization.
- Demonstrates new passive and active attacks that can defeat these privacy measures, including:
  - Improvements to UUID-E reversal for Android devices.
  - Sequence number correlation.
  - A novel active attack using RTS/CTS control frames that reveals the real MAC address of any tested device.
- Suggests best practices for designing secure MAC randomization policies.

Limitations in Findings:

- Focuses mainly on devices in an unassociated WiFi state.
- Device model identification is sometimes based on inference rather than ground truth.
- iOS devices are more resistant to passive attacks, but still vulnerable to some active techniques.

- Many flaws are due to low-level chipset behavior, which cannot be fixed through software alone.

Future Work Addressed:

- Encourage universal adoption of MAC address randomization with:
    - Full 46-bit randomness.
    - Fixed or removed sequence numbers in probe requests.
    - Anonymized or standardized 802.11 Information Elements.
- Further study on countermeasures against RTS-based attacks.
- Cooperation between OS vendors and chipset manufacturers to mitigate hardware-level leaks.
- Continuous evaluation as mobile hardware and firmware evolve.

Paper Title:
Self-Supervised Association of Wi-Fi Probe Requests Under MAC Address Randomization

DOI:
10.1109/TMC.2022.3205924

Problem Identified in This Paper:
Modern smartphones use MAC address randomization when sending Wi-Fi probe requests to protect user privacy. However, this disrupts the ability to associate probe requests from the same device, which negatively impacts applications like people counting, crowd flow estimation, and trajectory tracking—where identifying device continuity is crucial without violating user privacy.

Proposed Solution:
The paper introduces **Cappuccino**, a privacy-preserving framework that associates probe requests under MAC address randomization.
Key contributions include:

- A **multi-modal self-supervised frame correlation estimator** using information elements (IEs), sequence numbers, and received signal strength (RSS).

- A **neural network with contrastive learning** to estimate association probabilities without labeled data.
- A **multi-frame association algorithm** formulated as a minimum-cost network flow optimization to find optimal associations across probe requests.
- Cappuccino runs on standard Wi-Fi infrastructure without requiring specialized hardware or external calibration.

Limitations in Findings:

- Ground truth for evaluating randomized MAC address association is hard to obtain; validation uses a limited number of controlled devices or non-randomized MACs.
- It focuses only on **local probe request association**, not on tracking across locations or over longer timeframes.
- The approach may need tuning to adapt to other environments or different wireless device behaviors.

Future Work Addressed in This Paper:

- Adapt Cappuccino for **online, real-time environments** using mini-batch processing.
- Extend to **cross-location probe association** for broader anonymized tracking use cases.
- Explore more **privacy-preserving analytics** using associated frame sequences.
- Investigate generalization of this framework to other wireless or IoT sensing systems.

Paper Title:
Over-the-Air Runtime Wi-Fi MAC Address Re-randomization

DOI:
arXiv:2405.15747v1 [cs.NI] – May 2024

Problem Identified in This Paper:
Current MAC address randomization methods in Wi-Fi networks only change the MAC address when a device disconnects from an access point. While this provides some privacy protection, it allows ongoing sessions to be fully linkable, meaning that any eavesdropper can still track a device for as long as it remains connected, undermining user privacy.

Proposed Solution:
The authors propose a novel scheme for **runtime MAC re-randomization** that:

- **Re-randomizes the MAC address on the fly**, before each transmission, without requiring disconnection.

- Keeps the original MAC address hidden over the air while maintaining communication integrity.
- Ensures **synchronized re-randomization** across all connected devices to enlarge the anonymity set.
- Includes reset of sequence numbers and WPA2/3 nonces to prevent linking frames based on metadata.
- Implements the system using off-the-shelf Wi-Fi hardware and open-source Linux drivers.

Limitations in Findings:

- Experiments were conducted on **legacy hardware (Atheros AR5414)** that supports only Wi-Fi a/b/g, limiting insights on modern Wi-Fi technologies (n/ac/ax).
- The evaluation was **small-scale**, with only a few devices, in a controlled indoor environment.
- **Time synchronization is critical** for unlinkability but challenging in practice.
- The `T` (re-randomization interval) is hard-coded and not dynamically communicated from the AP.
- The scheme does **not enhance privacy against the access point itself**, only external eavesdroppers.

Future Work Addressed in This Paper:

- **Port the scheme to modern Wi-Fi chipsets** with non-free firmware.
- **Enable dynamic re-randomization scheduling**, broadcasted by the AP to connected devices.
- Develop solutions for **loosely synchronized devices** to maintain unlinkability.
- Extend the scheme to support **Wi-Fi ad hoc networks**.
- Conduct **large-scale evaluations** to assess real-world performance and scalability.

Paper Title:
Defeating MAC Address Randomization Through Timing Attacks

DOI:
10.1145/2939918.2939930

Problem Identified in This Paper:
MAC address randomization is intended to protect user privacy by preventing tracking of devices through probe requests. However, the paper identifies that this countermeasure is insufficient, as timing information in probe requests can still be exploited to track devices, even when randomized MAC addresses are used.

Proposed Solution:
The authors propose a timing-based attack that uses inter-frame arrival time (IFAT) patterns to group probe requests and identify frames originating from the same device. They introduce a signature based on timing, several timing-based distance metrics (D1, D2, D3), and an incremental learning algorithm to cluster probe requests that likely come from the same device despite using different MAC addresses. Their method achieves up to 77.2% accuracy.

Limitations in Findings:

- The method's accuracy decreases when fewer bursts are available, making near real-time tracking less reliable.
- The approach relies on statistical consistency, which could lead to false positives or negatives in less controlled environments.
- Results vary depending on device behavior and operating system implementation of MAC randomization.
- The offline algorithm and tuning parameters significantly affect performance, and overuse of prior knowledge may introduce classification errors.

Future Work Addressed in This Paper:
The authors suggest exploring countermeasures such as:

- More frequent MAC address changes (per burst or per frame).
- Introducing random delays between probe frames and bursts to disrupt the timing patterns used for fingerprinting.
- Combining timing-based attacks with other sensor data (e.g., location from multiple sensors) to improve accuracy.

Paper Title:
Efficient Association of Wi-Fi Probe Requests under MAC Address Randomization

DOI:
10.1109/INFOCOM42981.2021.9488769

Problem Identified in This Paper:
Modern Wi-Fi devices use MAC address randomization to protect user privacy. While this helps prevent device tracking, it also disrupts important data analysis tasks like people counting, trajectory inference, and crowd flow estimation because probe requests from the

same device appear disconnected. This results in fragmented data and limits the utility of Wi-Fi sensing systems.

Proposed Solution:

The authors propose **Espresso**, a novel, efficient method to associate Wi-Fi probe requests even when MAC addresses are randomized. Espresso models the association problem as a **minimum-cost flow network**, estimating correlation between frames using **multimodal features**:

- Information Elements (IEs)
- Sequence numbers
- Received Signal Strength (RSS)

Espresso does not require external localization, manual labeling, or specialized hardware. It uses a probabilistic model for frame correlation and solves the global association using network flow optimization. A mini-batch mechanism is introduced for scalability.

Limitations in Findings:

- Espresso assumes that some probe requests with real MAC addresses are available for model training.
- Its performance might degrade in extreme environments with very high device density or limited RSS data.
- The assumption of independence between modalities might not hold perfectly in all real-world settings.
- There's potential sensitivity to parameter tuning like number of RSS partitions or sequence modes.

Future Work Addressed in This Paper:

- Enhancing Espresso’s adaptability to different environments and device behaviors.
- Applying the system in more dynamic or sparser settings to improve generalizability.
- Exploring privacy-preserving extensions that ensure robustness without sacrificing user anonymity.

- Investigating real-time online deployment scenarios to enhance system responsiveness.

Paper Title:
Three Years Later: A Study of MAC Address Randomization In Mobile Devices And When It Succeeds

DOI:
10.2478/popets-2021-0042

Problem Identified in This Paper:
MAC address randomization was introduced to protect mobile users' privacy by preventing device tracking. However, early implementations were flawed, allowing adversaries to bypass randomization. Although newer devices claim better privacy protections, the current effectiveness of MAC randomization—especially with evolving standards and devices—remained unclear and underexplored.

Proposed Solution:
The authors conducted a large-scale empirical study of 160 mobile phone models across 18 manufacturers to analyze the current state of MAC address randomization. They examined when devices randomize addresses, how consistently randomization is applied, and whether known vulnerabilities (e.g., sequence numbers, WPS UUIDs, RTS/CTS attacks) are still present. Their study included both controlled laboratory experiments and wild data collection (20 million probe requests) to validate real-world behavior.

Limitations in Findings:

- The study's device sample, though large, may not fully represent the global market or real-world usage patterns.
- Wild data collection lacks definitive ground truth due to MAC randomization itself and absence of persistent identifiers.
- Some behavior variations (e.g., chipset-specific or OS-level effects) could not be fully explained.
- The analysis focuses on passive observations and selected active attacks, leaving out other privacy vectors like inter-frame timing or app-level data leakage.

Future Work Addressed in This Paper:

- Exploring inter-frame timing and other side-channel leaks not covered in this study.
- Investigating how hardware components (e.g., chipsets) influence privacy behavior more deeply.
- Standardizing post-association MAC randomization, which is still inconsistently deployed.
- Expanding fingerprinting resistance by encouraging manufacturers to create more generic device signatures and remove persistent identifiers like sequence numbers and UUID-E fields.

Paper Title:
McMatcher: A Symbolic Representation for Matching Random BLE MAC Addresses

DOI:
https://doi.org/10.1109/icce59016.2024.10444395

Problem Identified in This Paper:
Bluetooth Low Energy (BLE) devices frequently randomize their MAC addresses to protect user privacy, making it difficult to track or associate different MAC addresses with the same physical device. This creates challenges in applications such as contact tracing, mobility tracking, or device fingerprinting where consistent identification is needed without violating privacy.

Proposed Solution:
The authors introduce **McMatcher**, a novel, privacy-preserving method that uses **symbolic representations (SAX)** of RSSI (Received Signal Strength Indicator) time-series data to generate characterizing vectors for BLE signals. It employs **cosine similarity** to match signals and detect if they originate from the same device. Unlike other approaches, McMatcher does not require machine learning model training, works in real-time, and solely relies on RSSI data.

Limitations in Findings:

- The recall rate, while generally high, was lower (93%) compared to the precision (99%), showing that the method may miss some valid matches.
- Most false negatives occurred when smartphones had similar RSSI distributions or were farther from the sniffer.

- The performance degrades when using lower sampling rates or suboptimal parameter configurations.

- The methodology has not been fully tested in dynamic, high-mobility environments with frequent environmental changes.

Future Work Addressed in This Paper:
The authors plan to:

- Evaluate McMatcher in **dynamic real-world scenarios**, such as moving vehicles or people in motion, where signal variability is higher.

- Apply McMatcher to the **start and end segments of MAC time-series** (head and tail) to improve robustness in dynamic environments.

- Conduct more comprehensive testing beyond indoor, static experimental setups.

- Optimize parameters further for different environments and device types.
