Wireless Security
Device Identification

Boris Danev / Srdjan Capkun

Department of Computer Science

ETH Zurich

2010

Agenda

- Motivation Example
- Introduction
  - Definitions & Perspectives
  - Device Identification Basics
- Device Identification Techniques
  - Passive Device Identification
  - Active Device Identification
  - Summary
- Applications of Device Identification
- Attacks on Device Identification Systems
- Conclusion

Motivation Example (1/2)

- iPhone Self-Localization Feature
  - Uses known locations of Wi-Fi access points
  - Detects access points and looks for them in a database
  - Provided by Skyhook Inc. and contains information about access points around the world, collected by the company and partially provided by the users.
- More Information
  - http://www.syssec.ch/press/location-spoofing-attacks-on-the-iphone-and-ipod

- Do you see possible security issues?

Motivation Example (2/2)

- Results of spoofing the New York City location on an iPhone. (left) The iPhone has a GSM signal and GSM localization overrules the NY access points. The position is still in Zurich but much less accurate. (right) The result is in New York City if no GSM signal is available.

Introduction

- Device identification is the act
  - of identifying a component in a networked environment (e.g., an operating system, drivers, a physical device)
  - based on the analysis of the communication with that component with or without the component's knowledge
  - Also referred to as device **fingerprinting**!
- Device identification covers a
  - broad spectrum of technologies (wired, wireless) and
  - many different methods have been suggested
- Device identification presents *two perspectives*
  - The network authority perspective (defensive)
  - The attacker perspective (offensive)

Different Perspectives

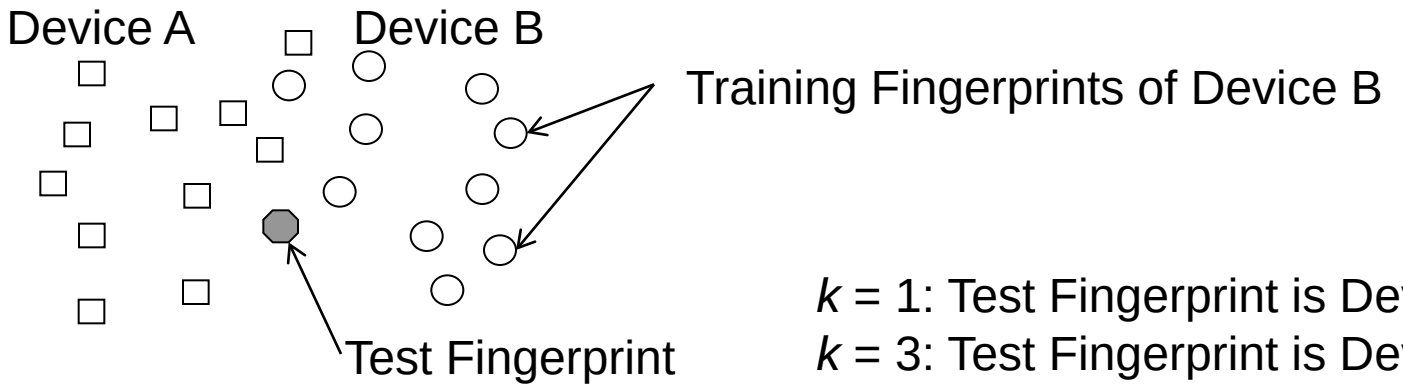
- The network authority perspective
  - Law enforcement agencies to maintain some measure of control and regulatory power (e.g., illegal transmitters)
  - Mobile operators to identify cloned cell phones
  - Administrator to identify and track problematic hosts
  - Prevention/protection against some types of attacks on the network operation (e.g., identity spoofing)
- The attacker perspective
  - Identify valuable targets (e.g., hosts) to break a network
  - Privacy violation (e.g., unauthorized tracking)
  - Protocol compromise (e.g., “Shake them up”)

Device Identification Basics (1/2)

- Device identification is also commonly referred to as *fingerprinting* (inspired from Biometrics)
- In a typical scenario
  - the *fingerprinter* observes traffic to and from a targeted device (*fingerprintee*) in order to find characteristics that (uniquely) distinguish the device or its components.
- Fingerprinting looks for characteristics in all layers:
  - Application/network layer
  - Link and physical layers
- Characteristics are usually differences in
  - software implementations of specs (e.g., 802.11)
  - hardware imperfections (e.g., clock skew)

Device Identification Basics (2/2)

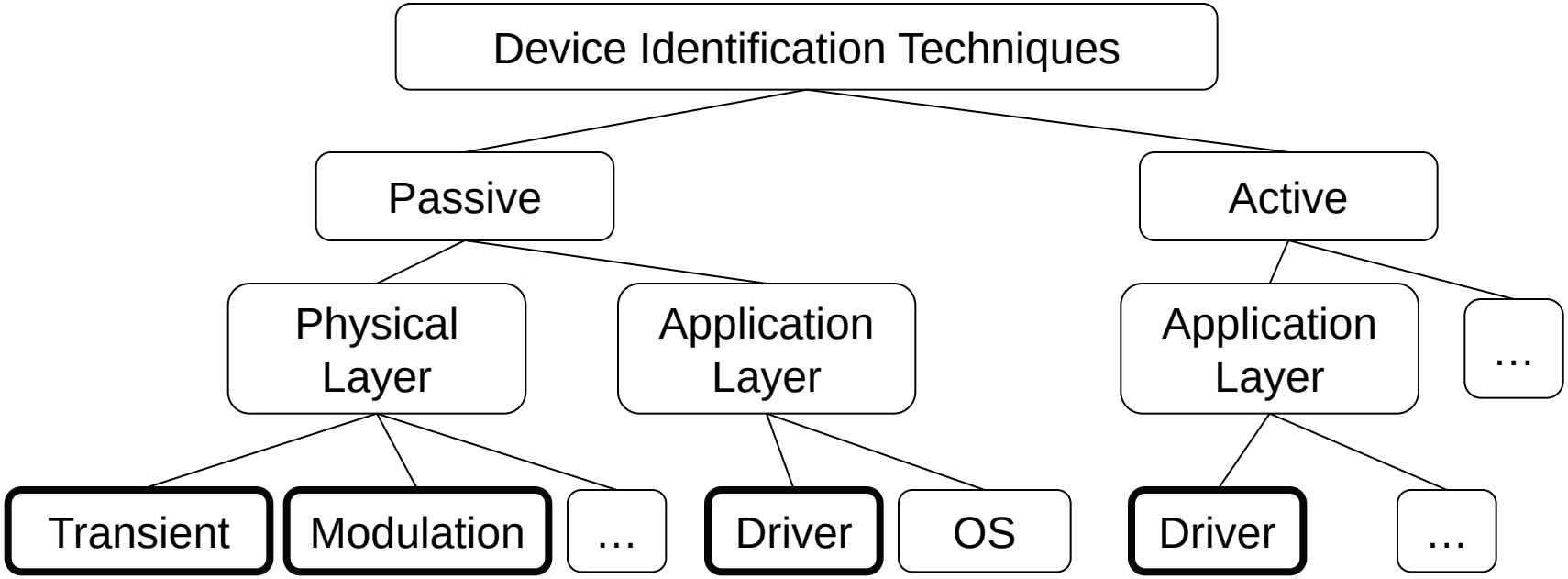
- Classification of the characteristics (fingerprints)
  - Typically by means of some standard classifier
  - When the number of devices is known in advance
- Nearest Neighbor Classifier
  - *k*-nearest neighbour rule



- Classification Error Rate
  - The percentage of misclassified test fingerprints over the total number of test fingerprints

Device Identification Techniques

- Two main approaches
  - *Passive* identification: Only observes the communication traffic of the targeted device
  - *Active* identification: Generates purpose-built traffic with the device and then observes the device's behavior



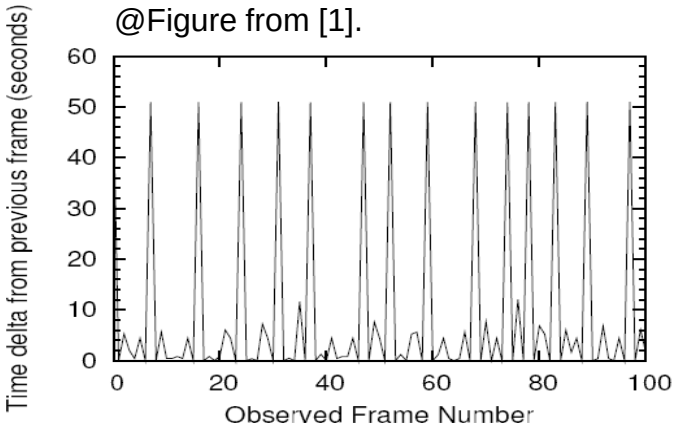
Agenda

- Motivation Example
- Introduction
  - Definitions & Perspectives
  - Device Identification Basics
- Device Identification Techniques
  - Passive Device Identification
  - Active Device Identification
  - Summary
- Applications of Device Identification
- Attacks on Device Identification Systems
- Conclusion

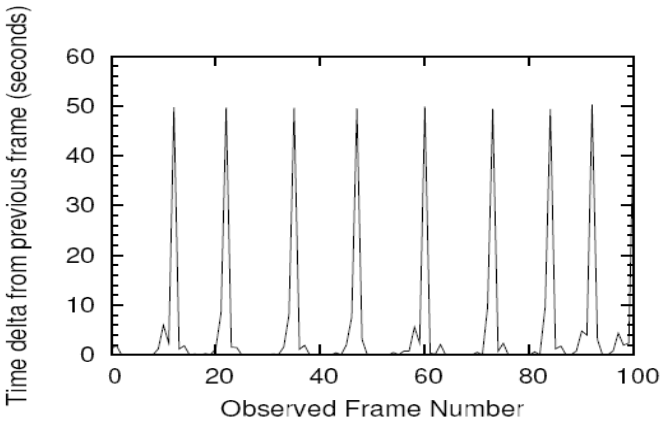
Passive Identification - Application Layer (1/3)

- Franklin et al. “Passive Data Link 802.11 Wireless Device Driver Fingerprinting” [USENIX Security’06]
- Overview of technique
  - A station sends *probe request* frames when it needs to discover access points in a wireless network. This process is known as *active scanning*.

@Figure from [1].



D-Link driver
D-Link DWL-G520 PCI Wireless NIC

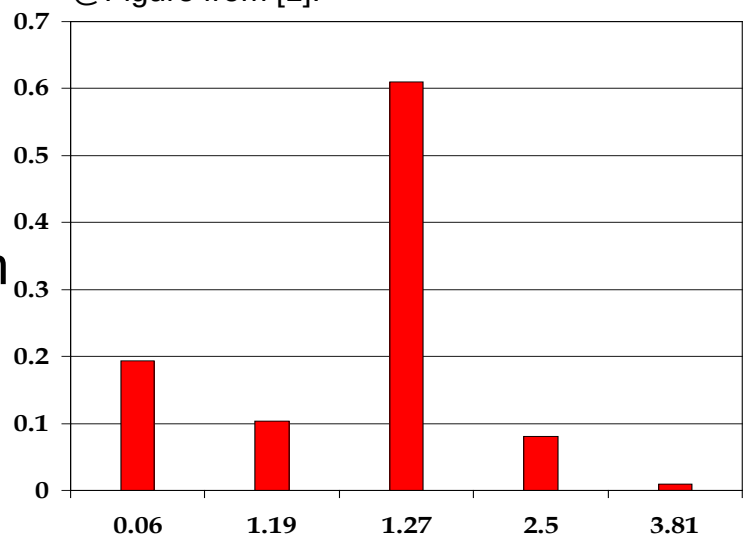


Cisco driver
AIR-CB21AG-A-K9 PCI Wireless NIC

Passive Identification - Application Layer (2/3)

- Driver signatures
  - Based on the *delta arrival time* between probe requests.
  - Signatures are obtained via binning with an empirically tuned and fixed bin width.
    1. Record the percentage of probe requests placed in each bin
    2. Record the average, for each bin, of all actual (non-rounded) delta arrival time values in that bin
    3. Generate a vector initialized with these parameters as the signature for that driver

@Figure from [1].

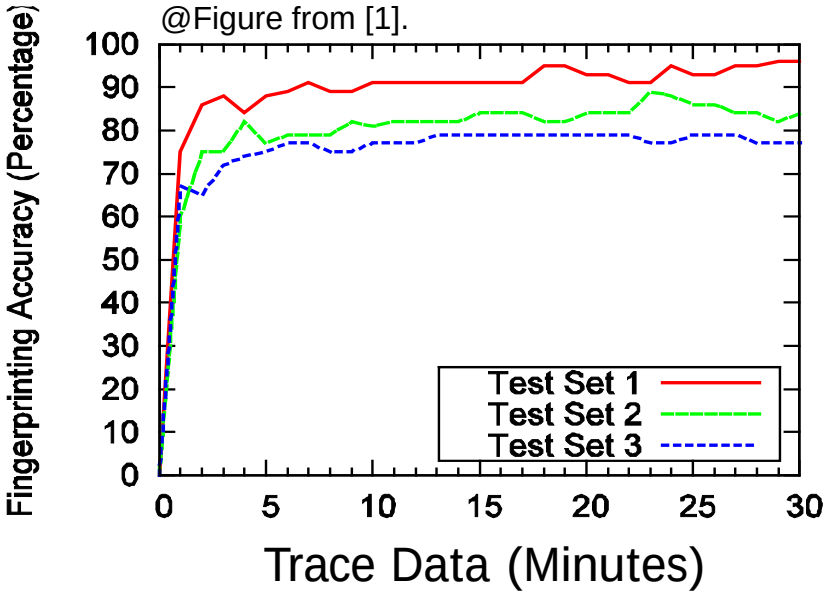


Windows Engenius driver signature.

Passive Identification - Application Layer (3/3)

- Results
  - Test set 1: (Lab) No background traffic, no obstructions
  - Test set 2: (Home Network) No background traffic, wall between fingerprinter and victim
  - Test set 3: (Coffee House) Background wireless traffic, miscellaneous objects between fingerprinter and victim

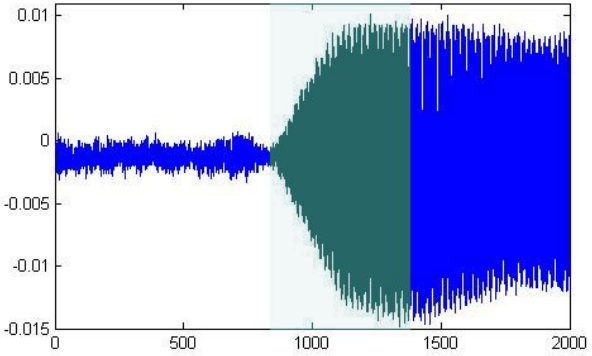
@Figure from [1].

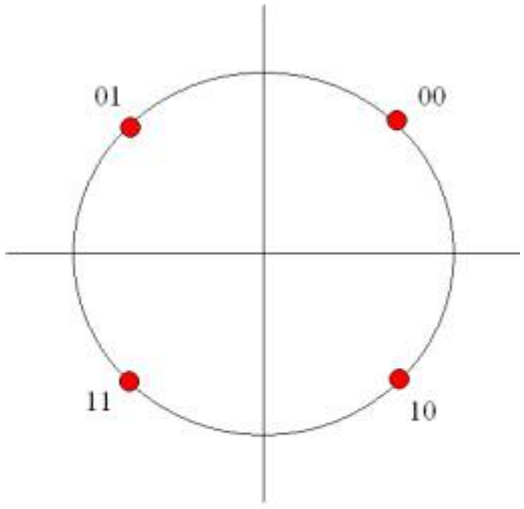


Test Set	Successful	Total	Accuracy
1	55	57	96%
2	48	57	84%
3	44	57	77%

Passive Identification – Physical Layer

- Transient-based Identification
  - The signal ramps up from channel noise to full power before a new transmission
  - The time between the start of the ramping and reaching full power is the *transient signal*
- Modulation-based Identification
  - Demodulation of a QPSK-modulated signal generates a number of errors
  - Exploiting these errors is in the essence of the approach



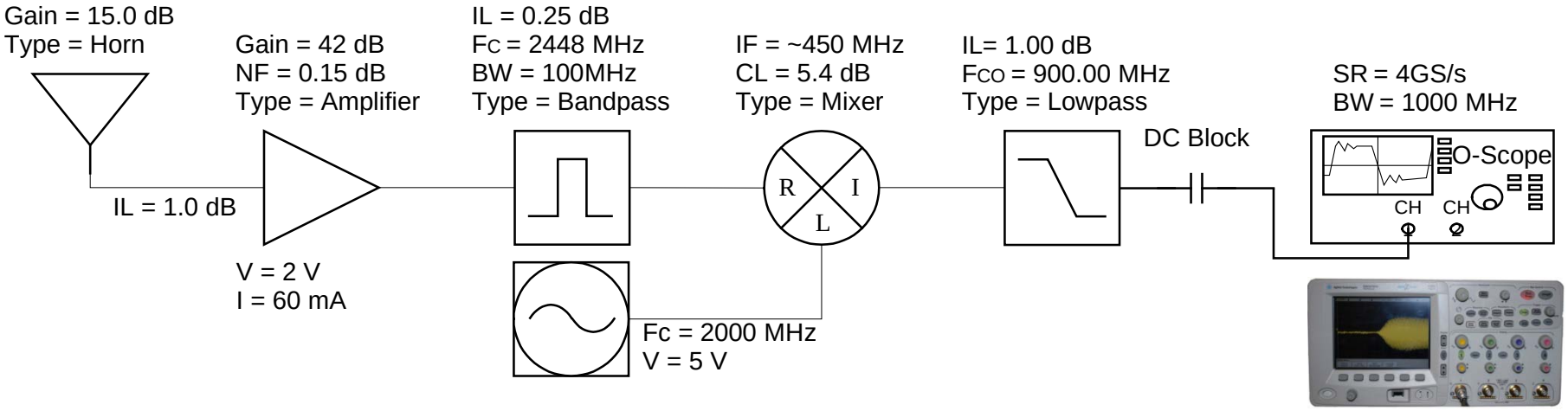


QPSK Signal Constellation

@Wikipedia

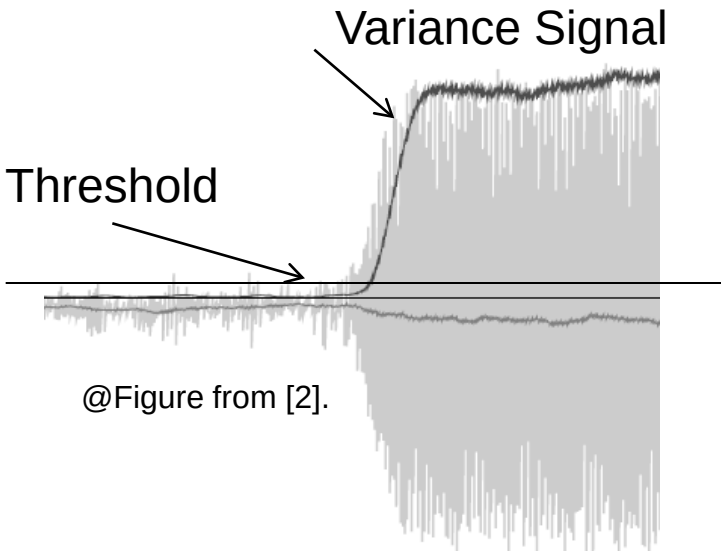
Transient-based Identification (1/4)

- Danev and Capkun., “Transient-based Idenfication of Wireless Sensor Nodes” [IPSN’09]
  - Improved method to identifying identical (same manufacturer and model) sensor node
- Signal Capturing Process
  - Hardware Components

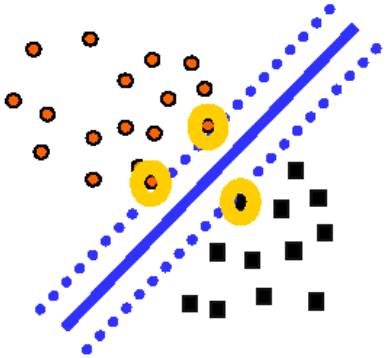


Transient-based Identification (2/4)

- Fingerprinting Approach Details
  1. Extract the transient part
     - Threshold-based algorithm
  2. Extract features from the transient signal (fingerprints)
     - Different spectral features
     - + statistical analysis (LDA)
  3. Equal Error Rate (EER) in verification
     - Compute False accept and reject rates
     - EER is the rate of FAR = FRR



@Figure from [2].

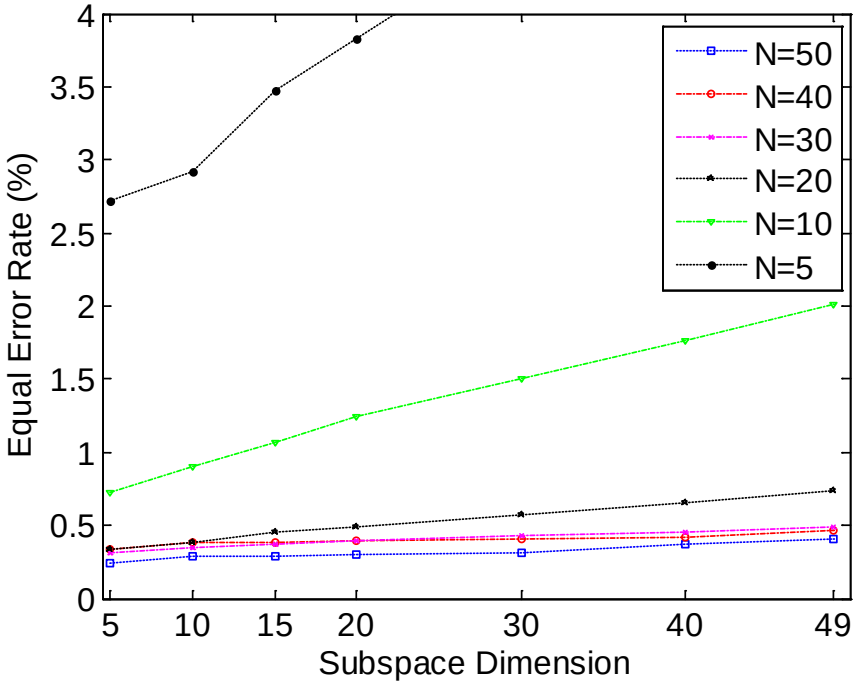


@Internet

Transient-based Identification (3/4)

- Experimental Conditions
  - 50 CC2420 identical Tmote Sky nodes
  - Distance, location, voltage, temperature
- Results on fixed distance, 10 meters
  - EER = 0.0024 (0.24%)

@Tmote Sky Datasheet.



FAR	FRR	100 % - FRR
0.01%	0.72%	99.28%
0.1%	0.65%	99.35%
1%	0%	100%
>1%	0%	100%

Table 1: FAR/FRR accuracy

Transient-based Identification (4/4)

Pros

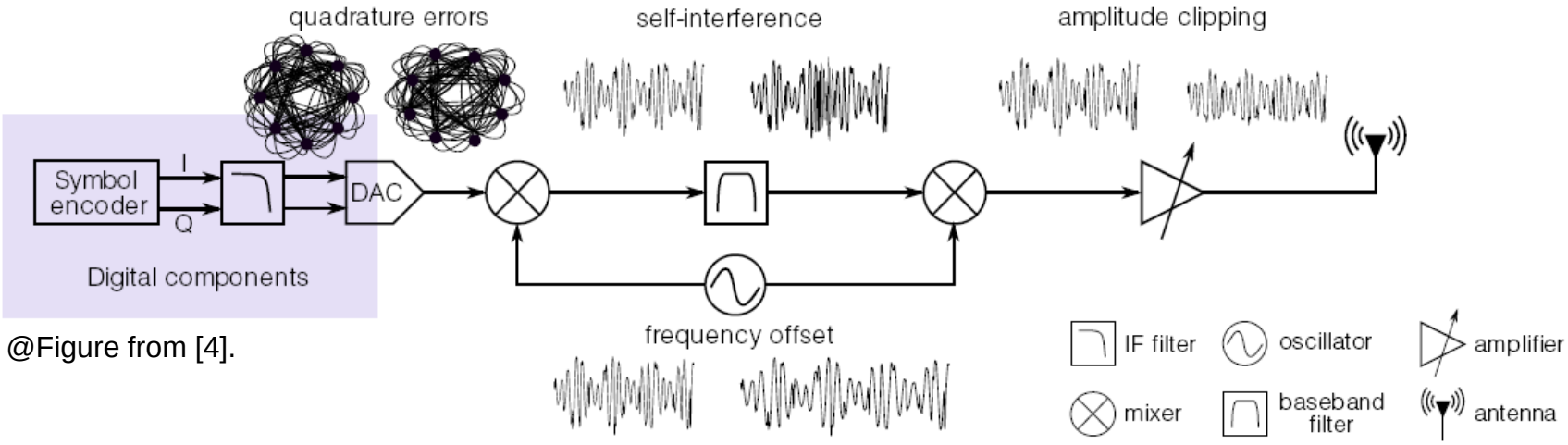
- Fixed Distance
  - EER (10 meters) ≈ EER (40 meters)
- Voltage and Temperature
  - EER (2.4 V) ≈ EER (3 V)
  - ~ 5°C had no effect.
- Static scenario is ok

Cons

- Fast ADC required
- Varying Distance
  - Cross matching (10 vs. 40 meters), EER = 38%
- Antenna Polarization
  - 2 polarizations distant 45°, EER = 30%
  - Only 6 out of 10 were correctly recognized
- Channel fingerprint inside

Modulation-based Identification (1/3)

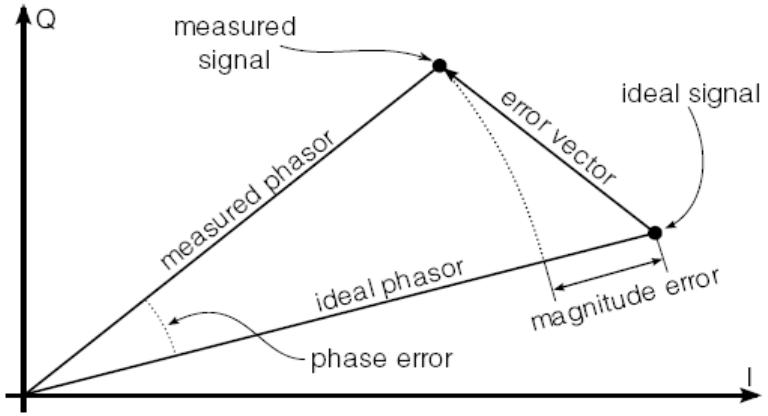
- Brik et al, “Wireless Device Identification with Radiometric Signatures” [Mobicom 08]
  - Explores the variance of demodulation of QPSK signals
  - Uses 1-NN and SVM classifiers to classify the fingerprints
- Signal Capturing Process
  - Hardware setup



@Figure from [4].

Modulation-based Identification (2/3)

- Fingerprinting Approach Details
  1. Capture the signals in the frequency domain
  2. Extract the following errors due to QPSK modulation
     - Frequency offset
     - SYNC correlation
     - I/Q origin offset
     - Magnitude/Phase offset

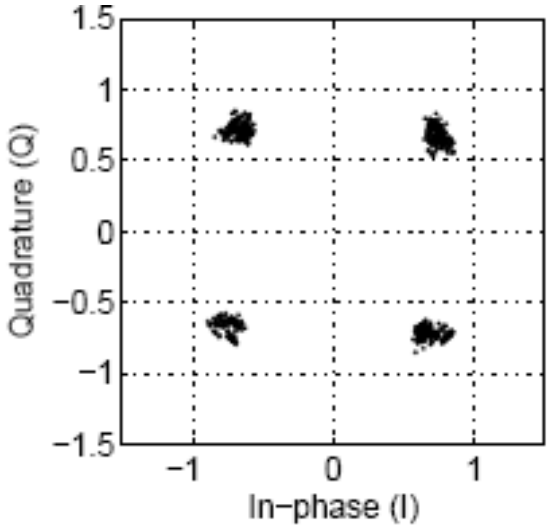


@Figure from [4].

  3. Fingerprints are represented by a vector of the above three errors
  4. Compute the classification error rate

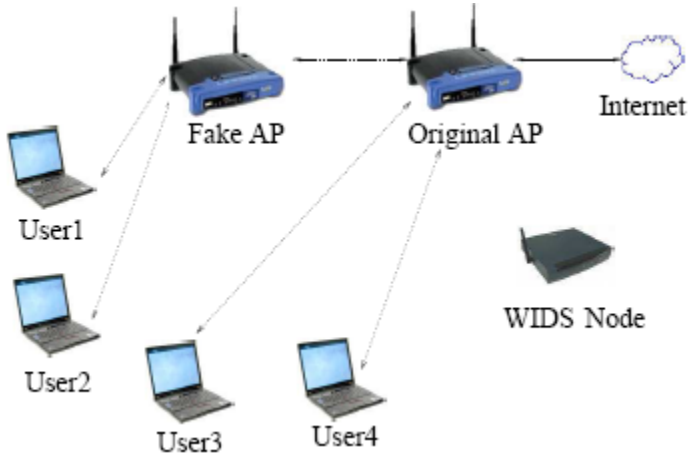
Modulation-based Identification (3/3)

- Experimental Conditions
  - 100 identical 802.11 NICs (same manufacturer and model)
  - distance and location
- Results
  - Classification error rate
    - k-NN Classifier - ~3%
    - SVM Classifier - ~0.34%
- Open issues
  - Feature stability
  - Resilience to attacks



Passive Identification – Cross Layer (1/2)

- Jana and Kasera, “On Fast and Accurate Detection of Unauthorized Wireless Access Points Using Clock Skews” [Mobicom 08]
  - Computes the clock skew of WiFi access points
  - Often reasonable to assume that a clock’s skew is constant
- Signal capturing setup
  - Capture AP probe response frames or
  - beacon probe frame broadcasts
  - Extract the timestamps
  - Compute the clock’s skew

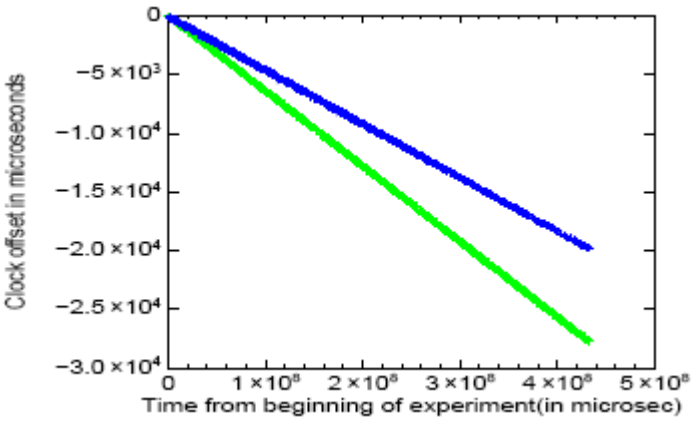


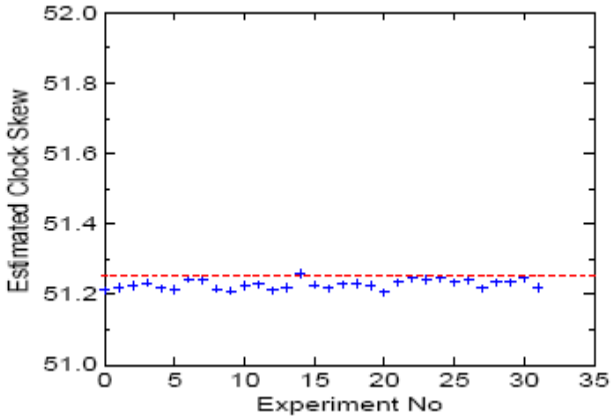
@Figure from [5].

Passive Identification – Cross Layer (1/2)

- Clock(C) offset and skew
  - off[C] = R[C](t) – t
  - skew[C] = d(R[C](t) – t)/dt
- Clock's skew estimation
  - Linear programming
  - Least square fitting (LSF)
- Experimental Conditions
  - 5 WiFi access points
  - Classification error rate?
- Included issues
  - Skew stability to internal T and NTP synchronization
  - Spoofing attacks?

@Figures from [5].





Agenda

- Motivation Example
- Introduction
  - Definitions & Perspectives
  - Device Identification Basics
- Device Identification Techniques
  - Passive Device Identification
  - Active Device Identification
  - Summary
- Applications of Device Identification
- Attacks on Device Identification Systems
- Conclusion

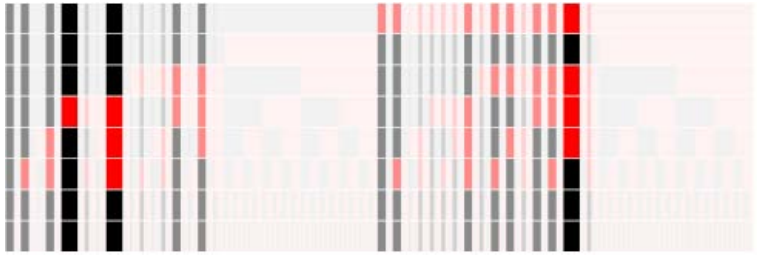
Active Identification – Application Layer (1/2)

- Bratus et al. “Active Behavioral Fingerprinting of Wireless Devices” [ACM WiSec 08]
  - Test driver implementations on non-standard or mal-formed packets
- Hardware Setup
  - No need of specialized software
  - Only a generator and injector of link-layer frames
- Fingerprinting Approach Details
  1. Send “stimulus” non-standard and/or malformed frame to the device
  2. Record the device’s response (any type of frame, specific frame, lack of response)

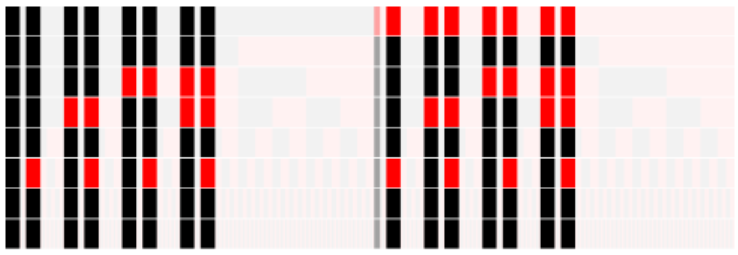
Active Identification – Application Layer (2/2)

- Fingerprinting an Access Point/Client
  - Build different scenarios
  - “Probe Requests and Authentication Requests are not expected to be fragmented. The AP reaction to a fragmented request may differ.”
- Results
  - Patterns maps of different driver implementations

@Figure from [5].



Linksys WRT54g AP



Madwifi-ng soft AP

Passive vs. Active Identification Summary

- Passive approach
  - Application layer fingerprinting can be circumvented by changing the application parameters
  - Physical-layer fingerprinting is more difficult to compromise due inherent physical properties (imperfections)
  - The accuracy depends on the features extracted and good features are very difficult to find

- Active approach
  - More flexible and allows to explore different scenarios
  - Usually can distinguish only different manufacturers
  - Easier to detect from the authority

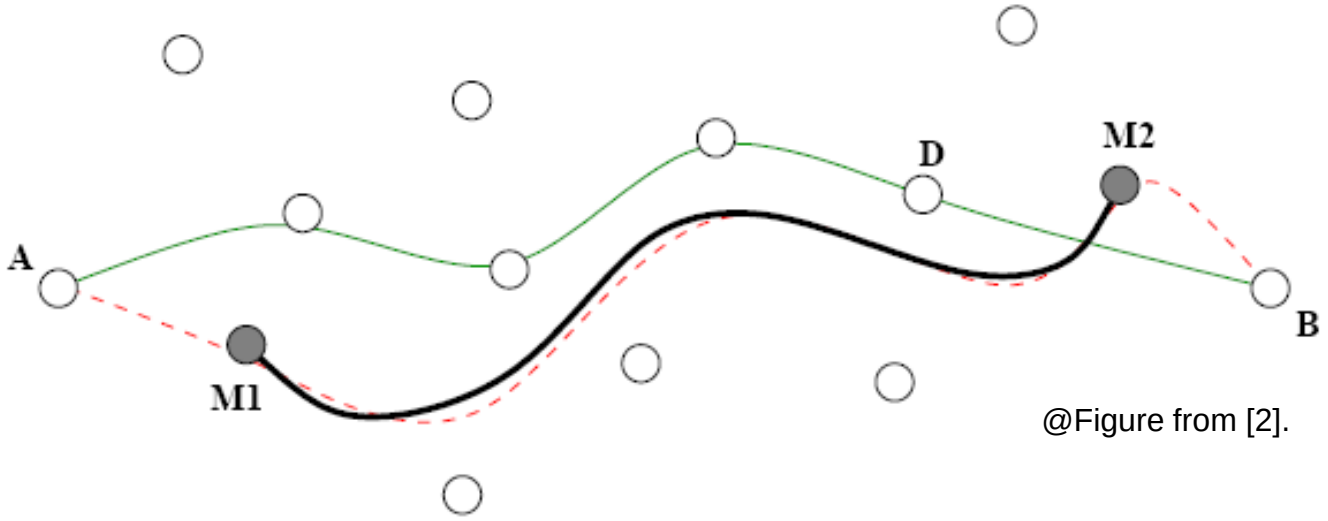
Agenda

- Motivation Example
- Introduction
  - Definitions & Perspectives
  - Device Identification Basics
- Device Identification Techniques
  - Passive Device Identification
  - Active Device Identification
  - Summary
- Applications of Device Identification
- Attacks on Device Identification Systems
- Conclusion

Applications of Device Identification (1/3)

- Defensive Use
  - Intrusion detection in all-wireless multi-hop networks
    - Wormhole attack
    - Sybil and Replication attacks

- Wormhole attack case



@Figure from [2].

Applications of Device Identification (2/3)

- Defensive Use
  - Second layer of security in access control in current wireless networks
  - Can compliment existing solutions such MAC authentication or cryptographic schemes

- Problems with existing solutions
  - MAC authentication is easy to spoof and therefore not secure
  - Cryptographic schemes require key distribution and
  - In particular, such schemes are not robust to detecting and revoking compromised keys

Applications of Device Identification (3/3)

- Offensive Use
  - Compromise a communication protocol

- “Shake Them Up” key establishment
  - A movement-based pairing protocol for CPU-constrained devices
  - Allows two wireless devices to establish a shared key without public key crypto, out-of-band channels
  - Alice and Bob broadcast bits
  - An attacker cannot retrieve the secret bit since it cannot figure out whether the packet was actually sent by Alice or Bob

[Solution] iPhone Location Spoofing

- Passive Wi-Fi Driver Fingerprinting
  - Does neither require hardware modification of the LN device nor changes on the scanned access points
  - But relies on characteristic behavior of different AP models
  - Does not fully prevent location-spoofing attacks, it makes them more difficult since the attacker needs to know the exact model. This would require his prior physical presence at the access point whose location is to be spoofed

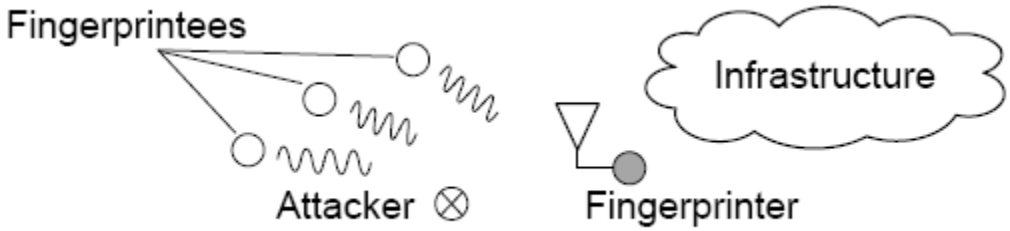
- Physical-layer Fingerprinting
  - Potentially more robust solution
  - But requires hardware modifications

Attacks on Device Identification Systems

- Impersonation attacks
  - Involves recreating the device fingerprint in order to impersonate a targeted device
  - E.g., faked transient signal concatenated with data
- Replay attacks
  - Involves recording the physical representation of the signal and replaying it
  - E.g., record the modulated signal and replay it
- Denial-of-Service attacks
  - Involves preventing a device identification procedure from correctly recognizing the devices
  - E.g., jamming only the transient signal
- Others

Impersonation Attacks on Modul.-based ID (1/3)

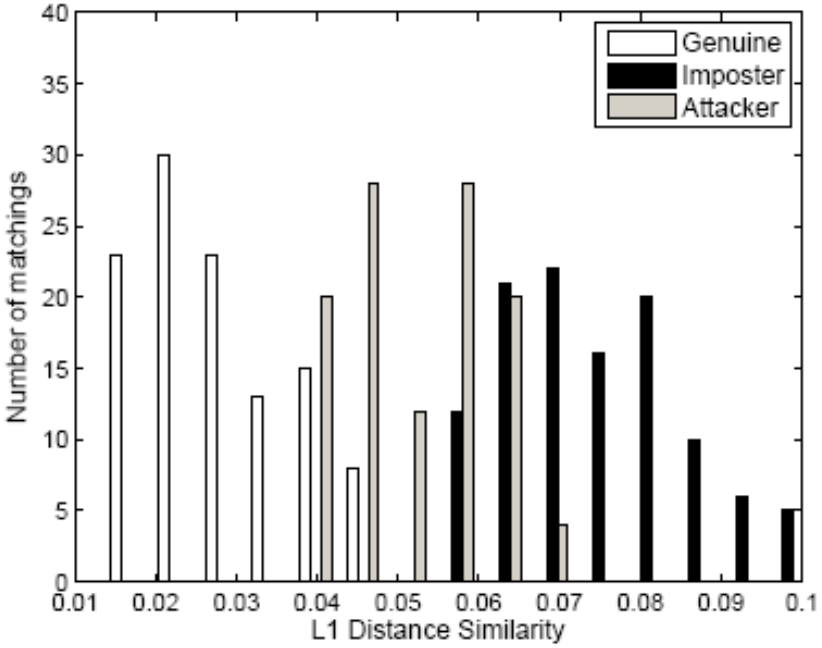
- System overview
  - 3 genuine devices
  - 1 attacker device
  - USRP and AWG

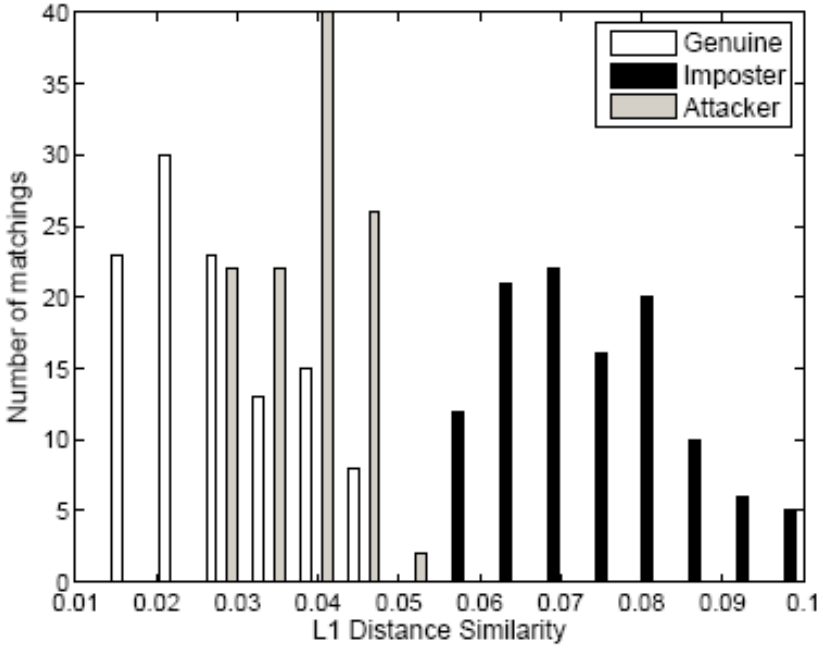


- Impersonation attack by Feature Replay
  - modify the signal characteristics to closely match the features used to identify a targeted genuine device
  - we assume knowledge of features and feature extraction
- Impersonation attack by Signal Replay
  - Record signals and retransmit at RF with high-end AWG
  - we do not assume any knowledge of the features

Impersonation Attacks on Modul.-based ID (2/3)

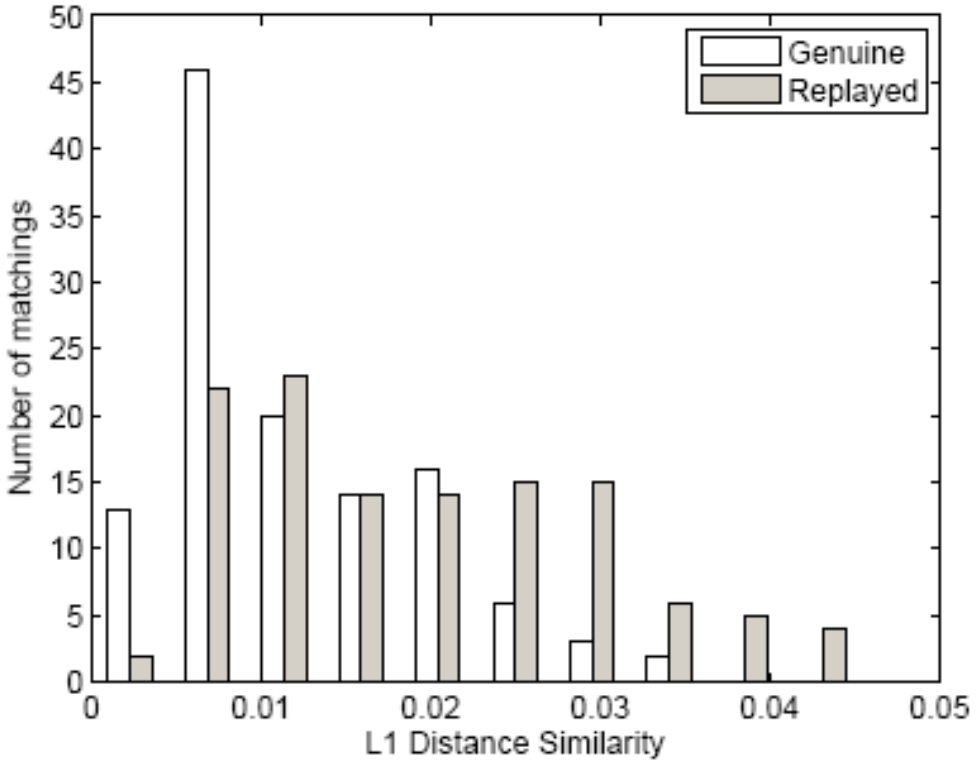
- Modulation-based features [Brik et. al, 2008]
  - F1: Frame frequency offset
  - F2: Frame SYNC correlation
  - F3: Frame I/Q Origin Offset
  - F4: Frame magnitude error
  - F5: Frame phase error
- Accuracy of impersonation by Feature Replay
  - USRP attacker (F1, F3 vs. F1, F3 and F5)





Impersonation Attacks on Modul.-based ID (3/3)

- Accuracy of impersonation by Signal Replay
  - All features
  - 20GS/s Arbitrary Waveform Generator (AWG) attacker



Conclusion

- Techniques for device identification span all network layers for a variety of objectives (defensive or offensive)
- The accuracy of identification depends on many factors (e.g., hardware, experimental conditions, features and classification procedures)
- There is little work on the resilience of identification with respect to the above and other attacks
- Research should not only focus on new identification approaches, but also carefully analyze the implications of using them under security threats

References

1. J. Franklin et al. “Passive Data Link Layer 802.11 Wireless Device Driver Fingerprinting” [USENIX SEC’06]
2. Rasmussen et al. “Implications of Radio Fingerprinting on the Security of Sensor Networks” [SECURECOMM’07]
3. Danev et al. “Transient-based Identification of Wireless Sensor Nodes” [IPSN’09]
4. Brik et al. “Wireless Device Identification with Radiometric Signatures” [MOBICOM’08]
5. Jana et al. “On Fast and Accurate Detection of Unauthorized Wireless Access Points Using Clock Skews” [MOBICOM’08]
6. Bratus et al. “Active Behavioral Fingerprinting of Wireless Devices” [WISEC’08]
7. Castelluccia et al. “Shake Them Up” [MOBISYS’05]