Detecting 802.11 Wireless Hosts from Remote Passive Observations

Based on paper by
Valeria Baiamonte, Konstantina Papagiannaki, and Gianluca Iannaccone

Presented by
Daniel Banky
Faculty of Informatics
Eötvös University, Budapest

2009

Detecting 802.11 Wireless Hosts from Remote Passive Observations

(1)Introduction

(2)Extracting a Wireless Signature

(3)Experimental Scenario

- Spectral Analysis
- Sample Entropy

(4)Detection Algorithm

(5)Evaluation

(6)Conclusion

Detecting 802.11 Wireless Hosts from Remote Passive Observations

(1)**Introduction**

(2)Extracting a Wireless Signature

(3)Experimental Scenario

- Spectral Analysis
- Sample Entropy

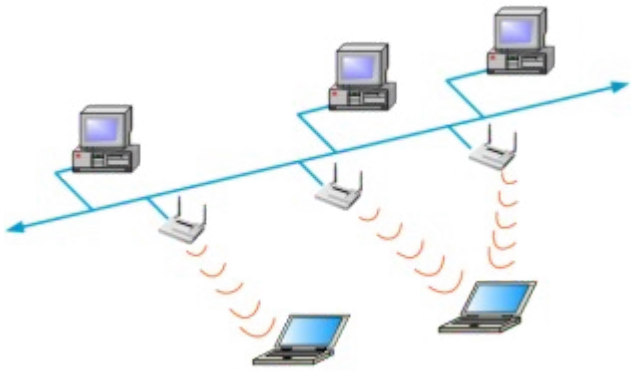
(4)Detection Algorithm

(5)Evaluation

(6)Conclusion

LANs in nowadays

- Hybrid local area networks
  - with wired and wireless clients
- Network access through 802.11
  - increased mobility
  - lower throughput, harder latency
  - unpredictable nature of the wireless medium



Chapter 1

Introduction

Aim of the research

- Characterizing and studying wireless traffic
- Highlighting essential differences and analogies with wire-line traffic
- Question:
  Are there any techniques to detect 802.11 wireless hosts?

Chapter 1

Introduction

Passive identification

- They looked into the problem of passive identification of wireless clients
- Observing at a location inside a network where a diverse set of client traffic is aggregated
- The observed clients use TCP/UDP protocols

Chapter 1

Introduction

What can we expect?

- “Uncertainty” of wireless traffic
  - unique signature
  - distinguish wireless form wired
- Depend on:
  - channel access mechanism
    - shared radio medium
  - channel conditions
  - link quality

Chapter 1

Introduction

Detecting 802.11 Wireless Hosts from Remote Passive Observations

(1)Introduction

(2)Extracting a Wireless Signature

(3)Experimental Scenario

- Spectral Analysis
- Sample Entropy

(4)Detection Algorithm

(5)Evaluation

(6)Conclusion

Chapter 2

Extracting a Wireless Signature

The aim

- Capturing the physical layer “signature” in a traffic flow through the inspection of the packet interarrival times
- We exploit two main features that a-priori differentiates wired from wireless access:
  - the unpredictability of the wireless medium
  - the impact of the 802.11 medium access control (MAC) mechanism

Conjecture

Chapter 2

Extracting a Wireless Signature

- Conjecture:
  These mechanisms are bound to random delays in the transmission of packets from a wireless host

- Shared radio medium
  - Packet loss, retransmit packet, delaying the packet delivery at the receiver
  - Client needs to contend for channel access

Measurement and problems

Chapter 2

Extracting a Wireless Signature

- Create constant bit rate (CBR) flows
  - using wired medium
  - using wireless medium
- Problems with the purely periodic data
  - highly loaded host
  - network congestion
  - multiplexing
- Not periodic traffic stream (like VoIP)
- They used
  - controlled experiments
  - periodic traffic

Detecting 802.11 Wireless Hosts from Remote Passive Observations

(1)Introduction

(2)Extracting a Wireless Signature

(3)Experimental Scenario

- **Spectral Analysis**
- **Sample Entropy**

(4)Detection Algorithm

(5)Evaluation

(6)Conclusion

Generating traffic

Chapter 3

Experimental Scenario

- Generating traffic flows from wired and wireless clients to destinations outside the LAN
- *tg* - Traffic Generator Tool
  - http://www.postel.org/tg/
- UDP/TCP traffic streams
  - UDP: CBR of 1.6 Mbps
    - 1000 byte packets sent every 5 ms
  - TCP: bulk transfer of large files
- Test uses ethernet interface and 802.11b wireless NIC
- The collection point is at the edge of the private LAN

Wireless experiments 1

Chapter 3

Experimental Scenario

- Good
  - 1 client
  - good location
  - no channel contention
  - signal level: -43 dBm
- G-Cong
  - 3 clients
  - good locations
  - contention for channel access
  - signal level: -43 dBm

Wireless experiments 2

Chapter 3

Experimental Scenario

- Bad
  - 1 client
  - bad location
  - no channel contention
  - signal level: -65 dBm
- Bad-Cong
  - 3 clients
  - bad locations
  - contention for channel access
  - signal level: -65 dBm

Spectral Analysis

Chapter 3

Experimental Scenario

- The purpose of this analysis is to investigate how the physical layer characteristics influence the explicit or implicit periodicity of a traffic flow
- Discrete Fourier Transform (DFT)
  - Fast Fourier Transform (FFT)
  - Processing of packet interarrival times
- Process of packet arrivals:

$$x[n]=\begin{cases}1, nT_0 \leqslant t_{arr} < (n+1)T_0 \\ 0, elsewhere\end{cases}$$

Fourier transformed signal

Chapter 3

Experimental Scenario

- Discrete time interval: $[nT_0, (n+1)T_0]$
  - overlapping problem
- Packet arrivals are never spaced by less than 20μs
  - T_0 := 20μs
- Resolution of frequency domain:
  - $f_0 = 1/T_0$ (= 50 KHz)
- Observation window: NT_0
- Fourier transformed signal:

$$X[kf_0] = \frac{1}{N}\sum_{n=0}^{N-1} x[nT_0]e^{-j2\pi\frac{nk}{N}}$$

Cumulative Power Spectrum

Chapter 3

Experimental Scenario

- DFT power spectrum P[n]:

$$P[kf_0] = |X[kf_0]|^2$$

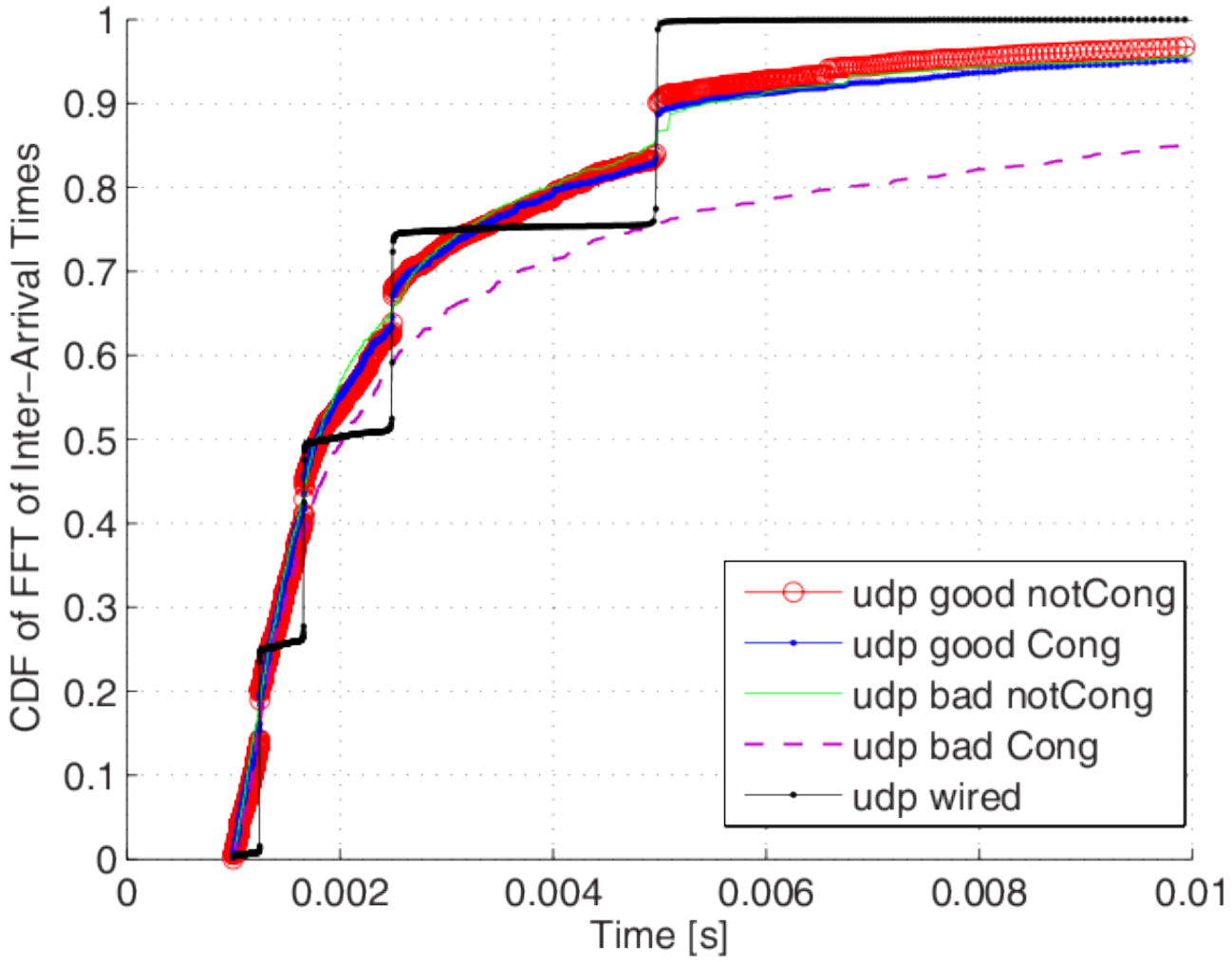
- Cumulative Power Spectrum (CPS) is use to compare different power spectra
- It is normalized to the total power of the signal as

$$C(n) = \frac{\sum_{k=0}^{n} P[kf_0]}{\sum_{k=0}^{N-1} |X[kf_0]|^2}, 0 \leqslant n < N$$

CPS of UDP wired and wireless CBR flows are computed

Chapter 3

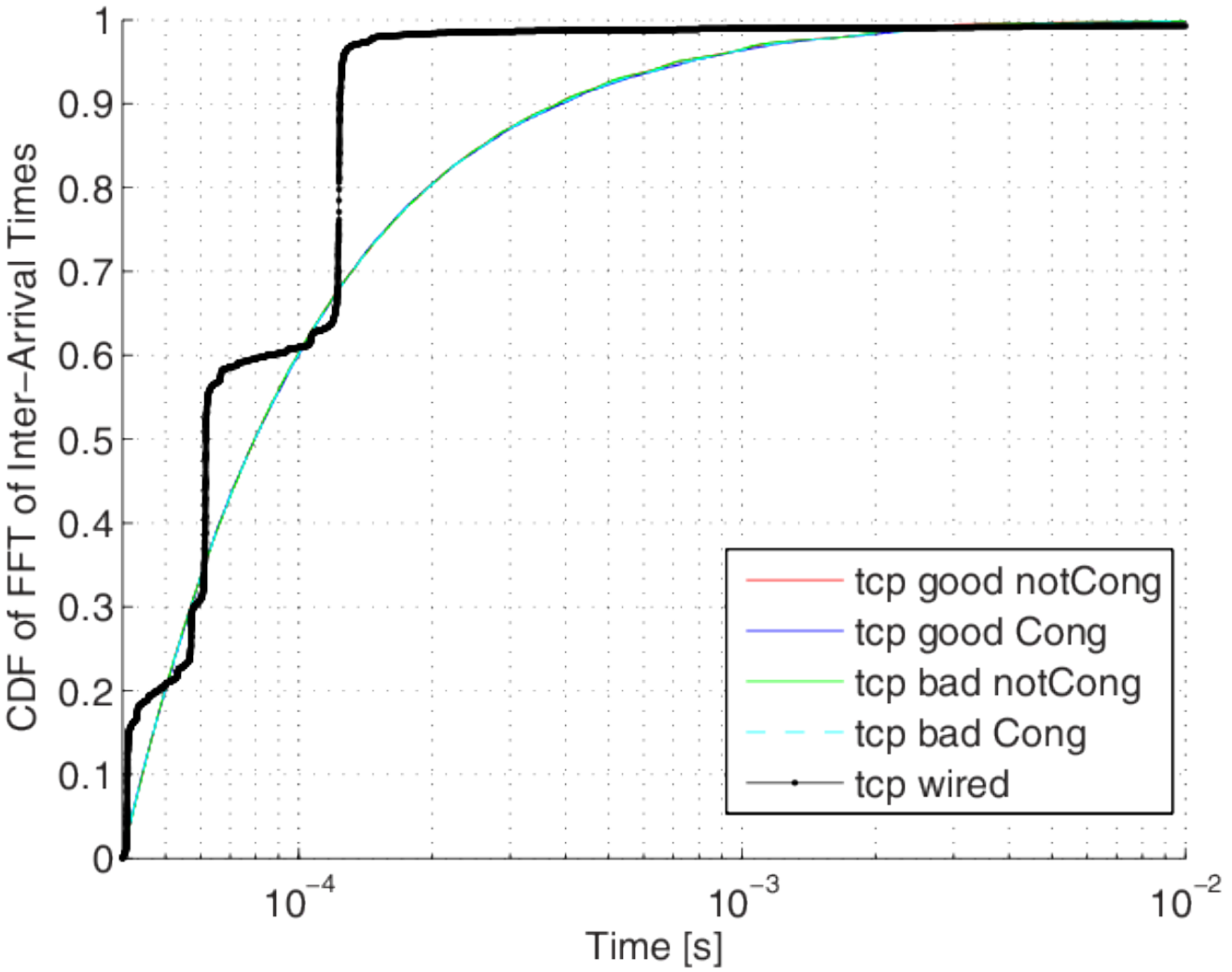
Experimental Scenario



CPS of TCP wired and wireless CBR flows are computed

Chapter 3

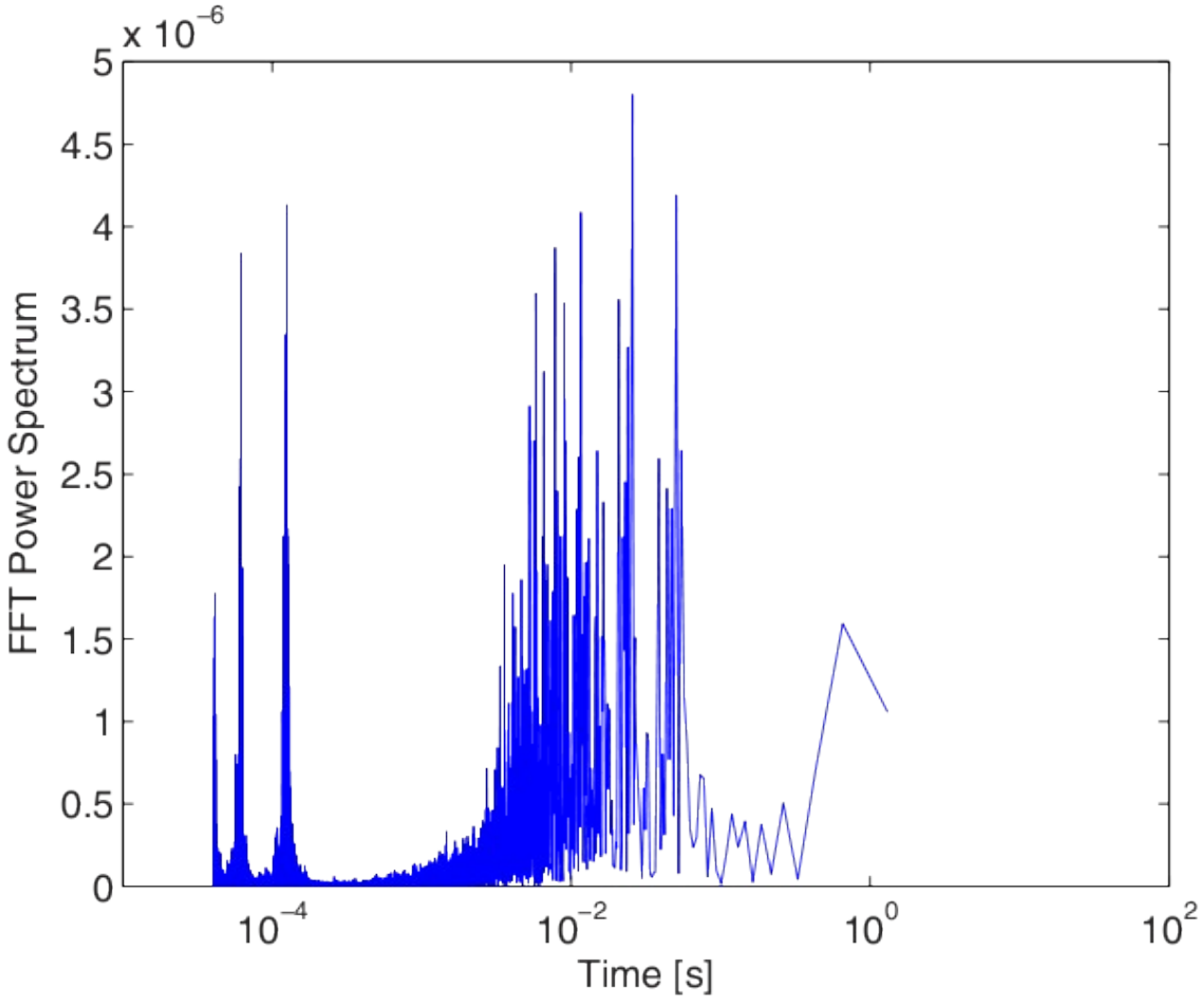
Experimental Scenario



DFT Power Spectrum of TCP (wired)

Chapter 3

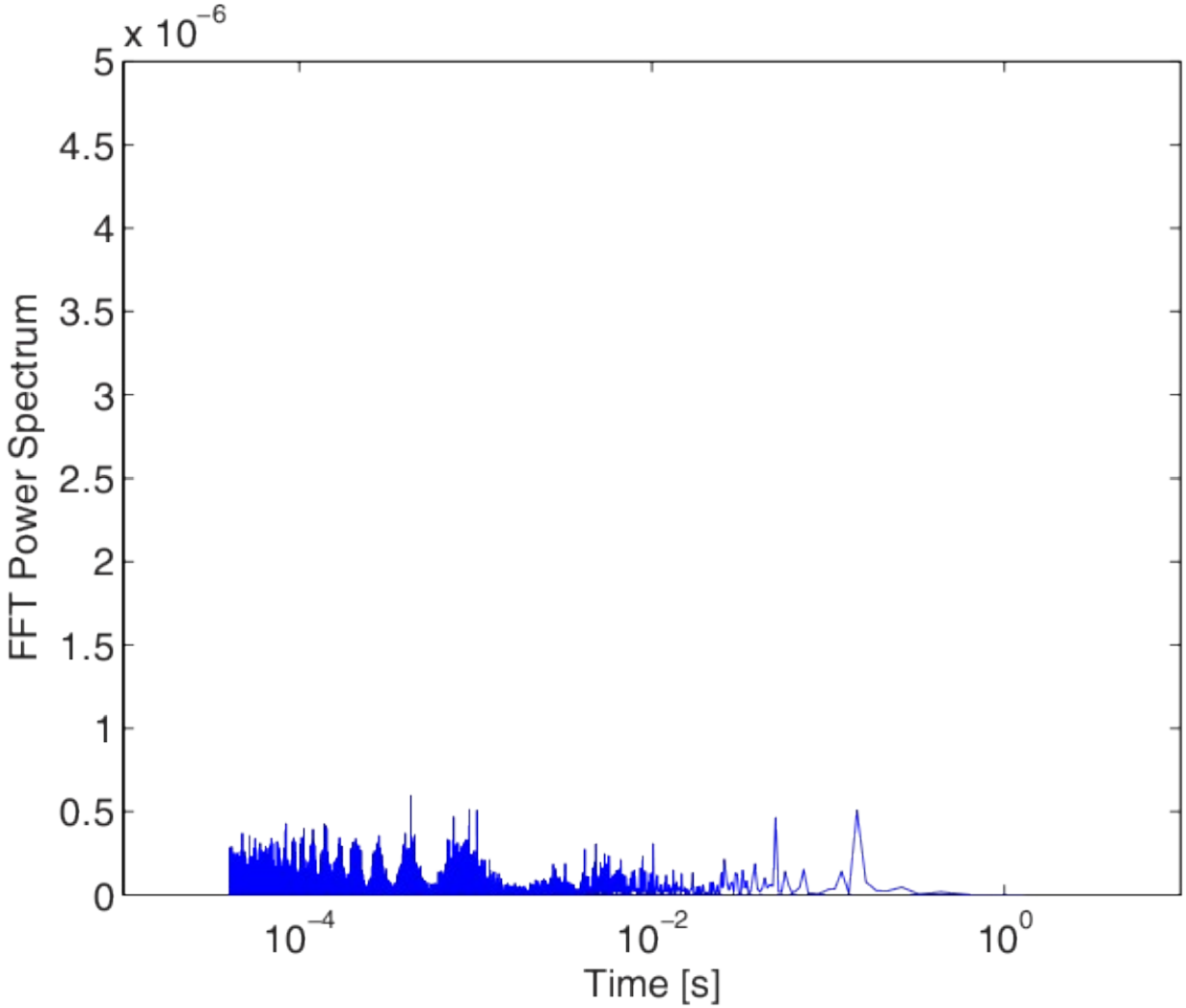
Experimental Scenario



DFT Power Spectrum of TCP (wireless)

Chapter 3

Experimental Scenario



Behaviour of TCP traffic

Chapter 3

Experimental Scenario

- Original traffic stream is not periodic
  - periodicity distortion
- Analysing those parts of the traffic streams where packets are sent in a back to back fashion
  - network effects are more visible
- Dropping all the big interarrival times exceeding a certain threshold, T_{RTT}.
- Choosing a big T_{RTT} := 10 ms

Sample entropy

Chapter 3

Experimental Scenario

- DFT methodology
  - computationally expensive
  - difficult to be incorporated within an automated algorithm
- Empirical entropy
  - because of empirical distributions
- Empirical entropy of interarrival time distribution:

$$H(x) := -\sum_{x_i \in X} p(x_i) \log_2 p(x_i),\ p(x_i) = \frac{m_i}{m}$$

Empirical entropy

Chapter 3

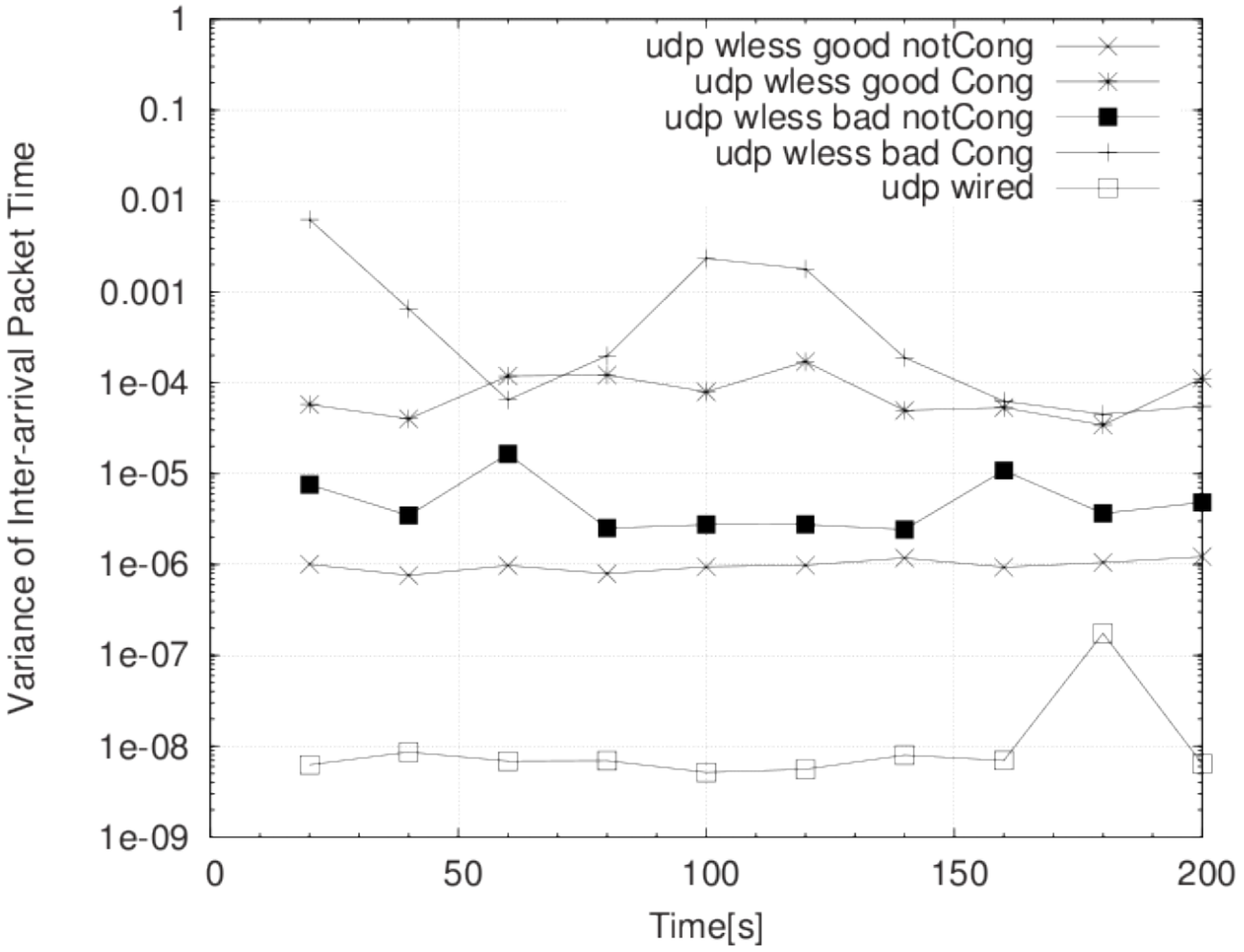
Experimental Scenario

- Empirical entropy values computed on the same distribution can be different
  - depend on time scales to discretize the sample
- Bin size to compute Probability Mass Function (PMF): 100μs
- Computing the entropy of the PMF of packet interarrivals across time

Variance of interarrival packet time (UDP)

Chapter 3

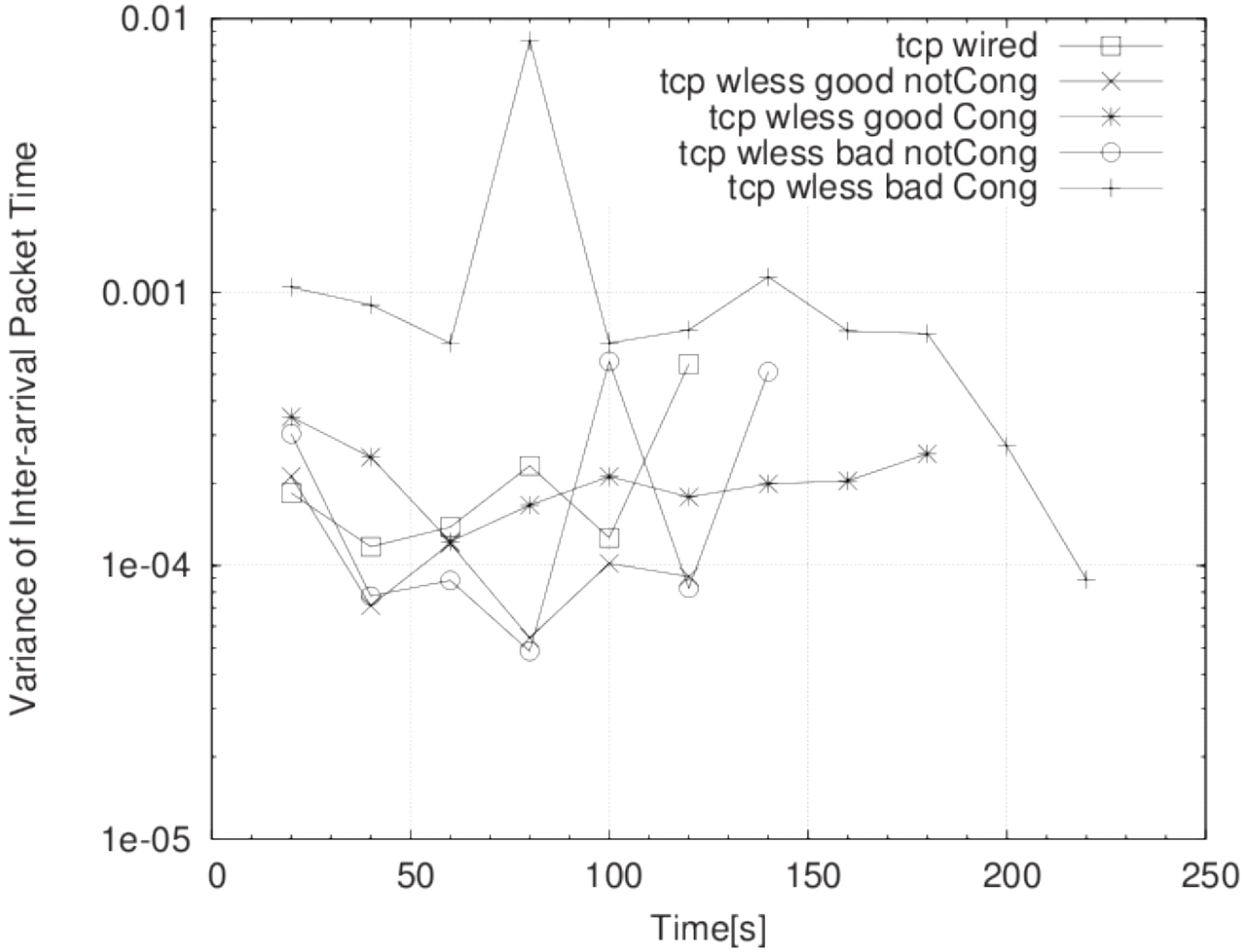
Experimental Scenario



Variance of interarrival packet time (TCP)

Chapter 3

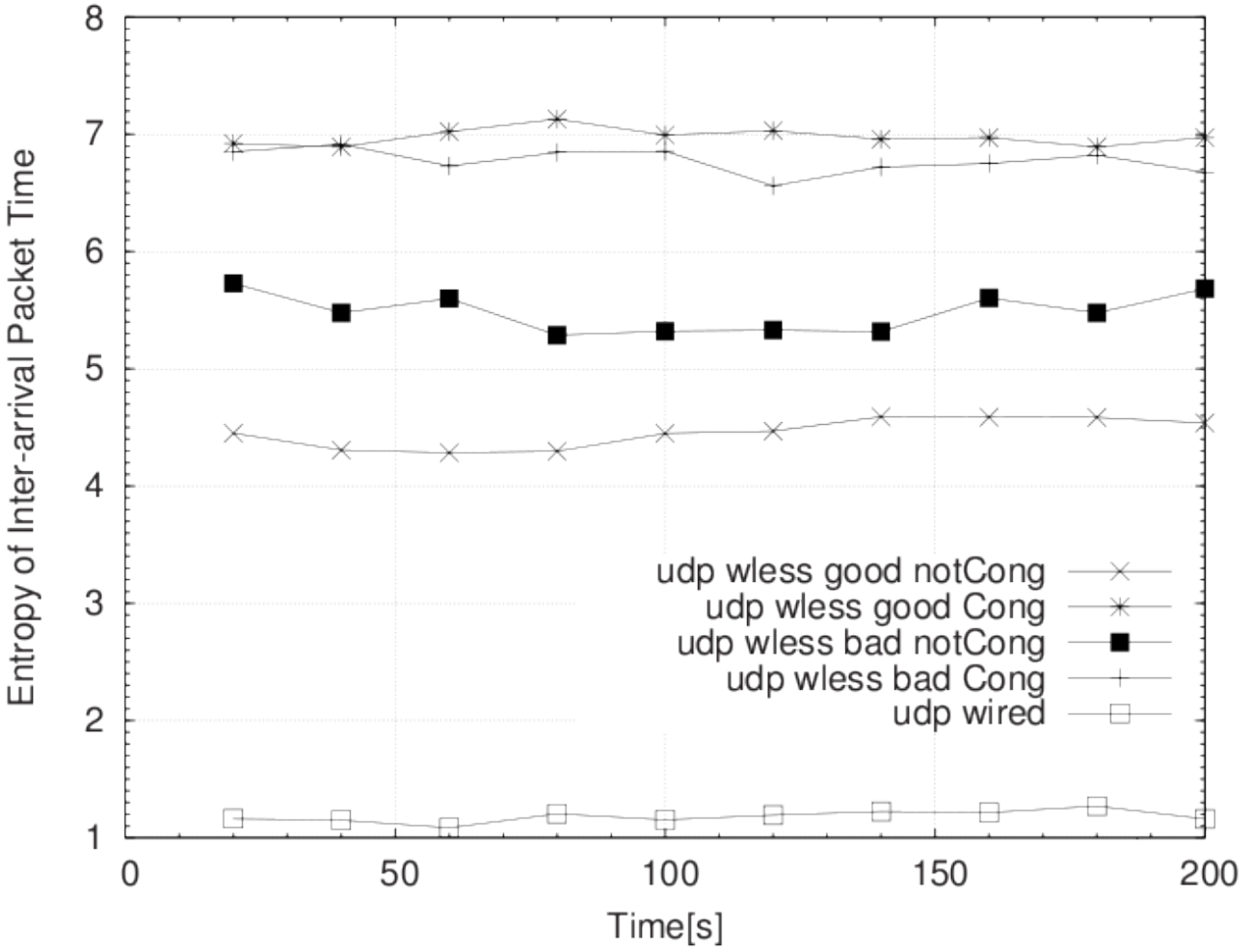
Experimental Scenario



Entropy of interarrival packet time (UDP)

Chapter 3

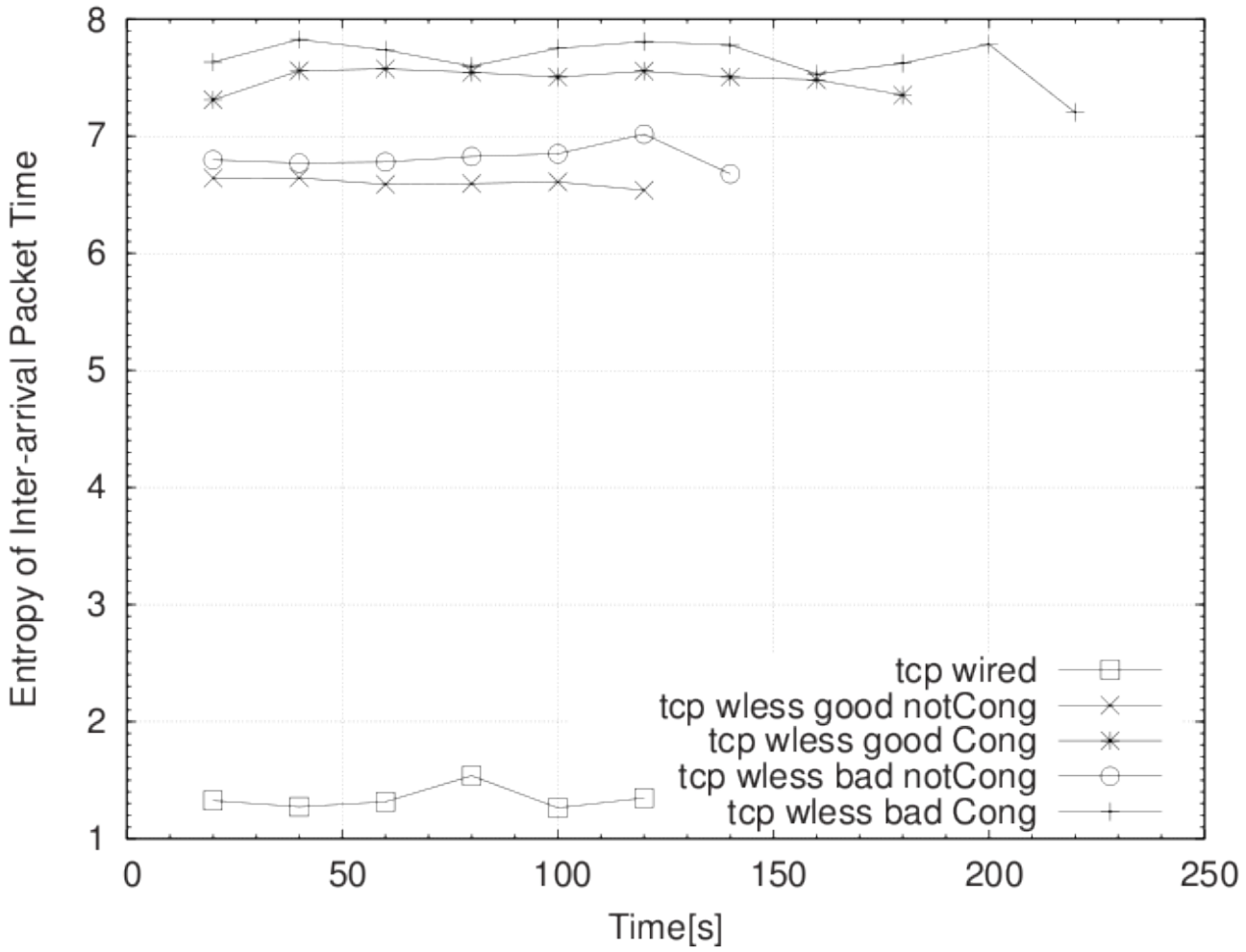
Experimental Scenario



Entropy of interarrival packet time (TCP)

Chapter 3

Experimental Scenario



Detecting 802.11 Wireless Hosts from Remote Passive Observations

(1)Introduction

(2)Extracting a Wireless Signature

(3)Experimental Scenario

- Spectral Analysis
- Sample Entropy

(4)Detection Algorithm

(5)Evaluation

(6)Conclusion

Algorithm to classify

Chapter 4

Detection Algorithm

- Within each IP-source set, 5-tuple flows are then isolated
- Interarrival times between consecutive packets are computed and then filtered: only those, falling below T_{RTT}, are kept
- Computing:
  - the empirical entropy $H_{IP,}$ evaluated on the whole IP-source aggregated traces
  - the empirical entropy of the largest 5-tuple flow of each IP-source trace, $H_{IP,5}$
- Variation of Entropy: $\Delta H = H_{IP} - H_{IP,5}$

Pseudo code of the algorithm

Chapter 4

Detection Algorithm

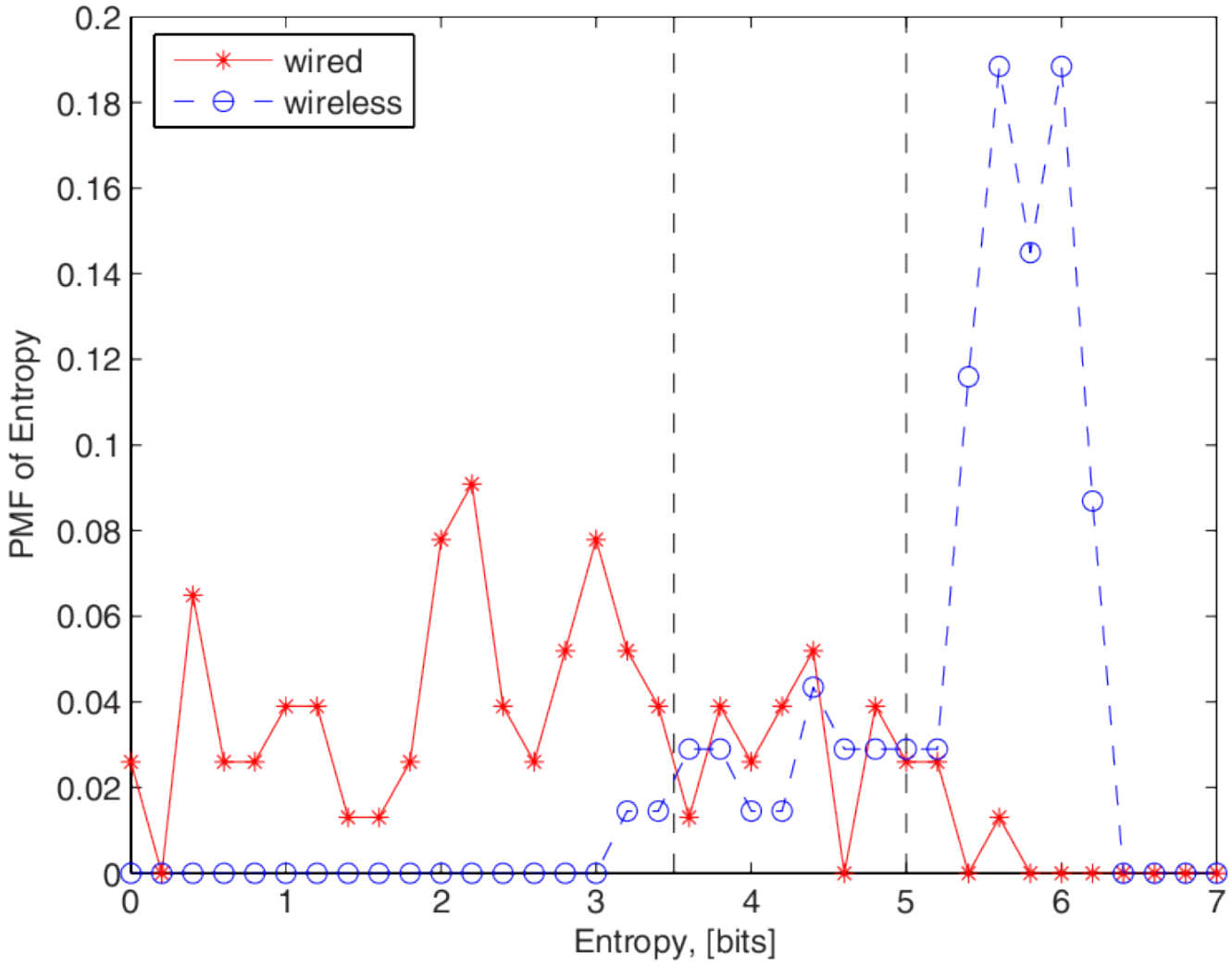
```
If H_IP ≤ H_lower
        then the host is wired
 else if H_IP ≥ H_upper
        then the host is wireless
 else if H_lower < HIP < H_upper
        if ΔH ≥ ΔH_THR
                then the host is wired
        else if ΔH ≤ ΔH_THR
                then the host is wireless
```

- H_{lower} = 3.5 bits
- H_{upper} = 5 bits
- H_{THR} = 0.5 bit

Probability Mass Function of Entropy

Chapter 4

Detection Algorithm



Detecting 802.11 Wireless Hosts from Remote Passive Observations

(1)Introduction

(2)Extracting a Wireless Signature

(3)Experimental Scenario

- Spectral Analysis
- Sample Entropy

(4)Detection Algorithm

(5)Evaluation

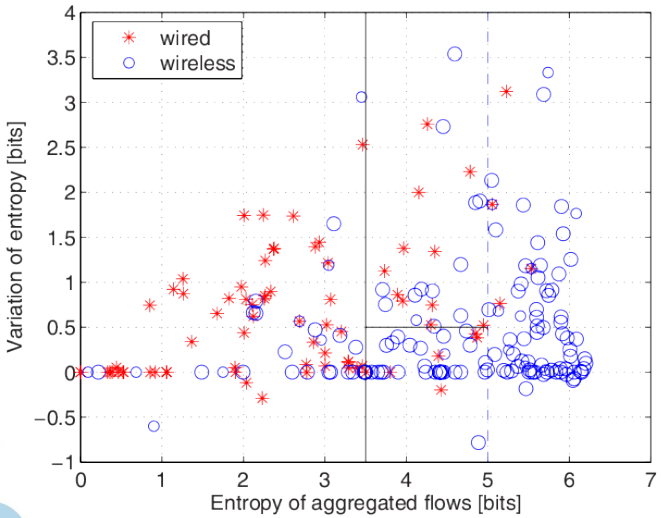
(6)Conclusion

Evaluation

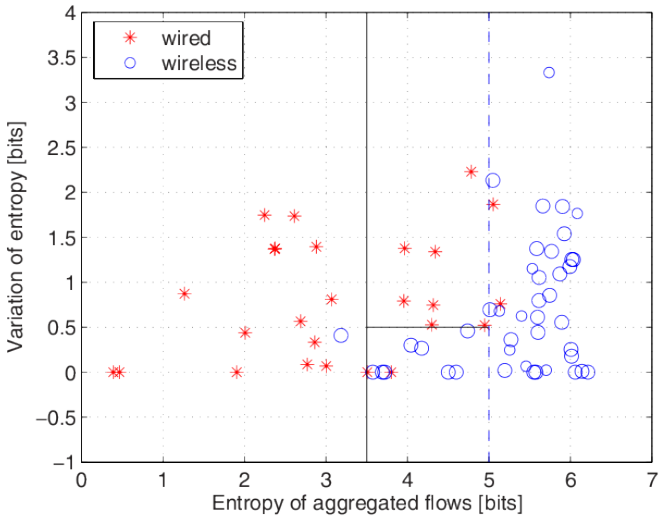
- Measured on 93 hosts in a laboratory
  - 27 WLAN hosts

Chapter 5

Evaluation



(a) Full datasets



(b) $Interarr > 200$

Detecting 802.11 Wireless Hosts from Remote Passive Observations

(1)Introduction

(2)Extracting a Wireless Signature

(3)Experimental Scenario

- Spectral Analysis
- Sample Entropy

(4)Detection Algorithm

(5)Evaluation

(6)Conclusion

Related works

- Bayesian inference algorithm
  - estimating and classifying whether TCP flows have traversed an 802.11 WLAN
  - time intervals occurs between TCP-ACK pairs and infers the connection type of a user
  - identification is possible only for clients that are employing the TCP protocol
  - making an inefficient use of traces

Chapter 6

Conclusion

Bayesian Inference Algorithm vs. Presented Algorithm

- Considers inter-packet pairs
  - so it is faster
- UDP and TCP
  - with time distance below 10ms

Chapter 6

Conclusion

Conclusion

- I have presented a method to classify wired and wireless hosts only based on passive traffic observations at a remote location
- The accuracy error reaches 1%-6%
- In the future this methodology will be tested on larger datasets

Chapter 6

Conclusion

Detecting 802.11 Wireless Hosts from Remote Passive Observations

Based on paper by
Valeria Baiamonte, Konstantina Papagiannaki, and Gianluca Iannaccone

Presented by
Daniel Banky
Faculty of Informatics
Eötvös University, Budapest

2009