Analyzing and Protecting Communication Metadata

Ludovic Barman

Laboratory for Data Security & Decentralized and Distributed Systems Laboratory

PhD Public Defense, 17.09.2021

Analyzing and Protecting Communication Metadata

Ludovic Barman

Laboratory for Data Security & Decentralized and Distributed Systems Laboratory

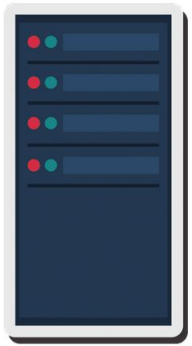
PhD Public Defense, 17.09.2021

WARNING

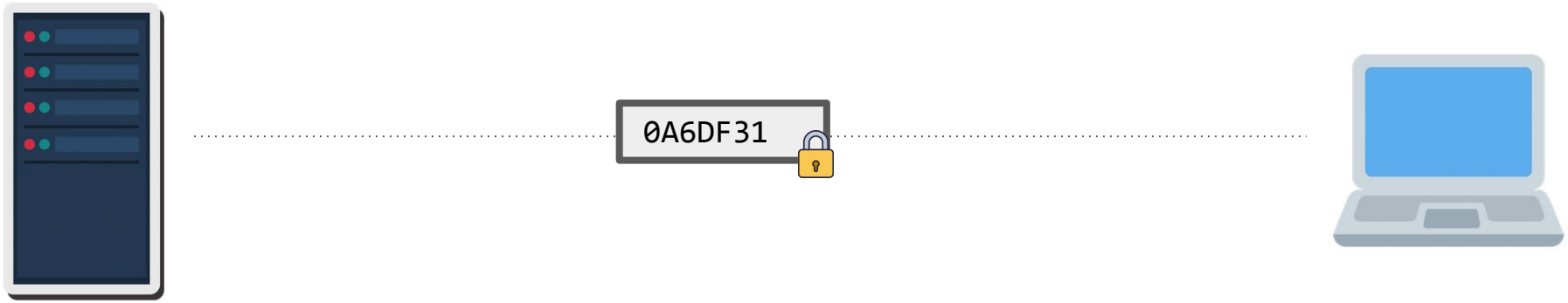
Simplified presentation. Some technical details have been simplified for the sake of clarity.

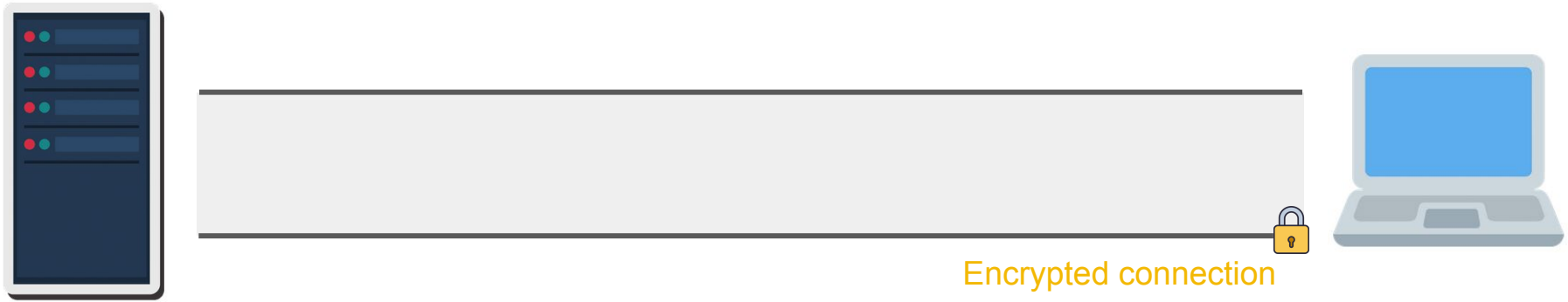
Please check the last slide for the proper references.


I wonder what she ordered...



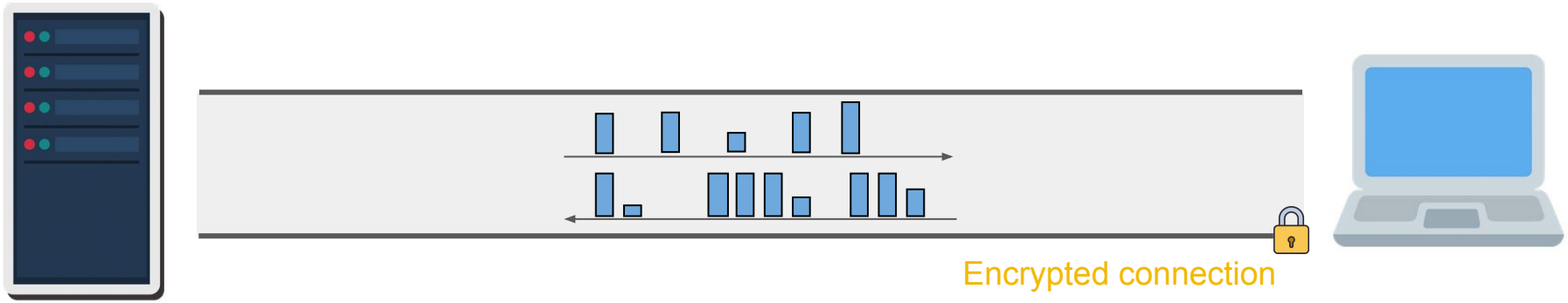
Hello !


0A6DF31



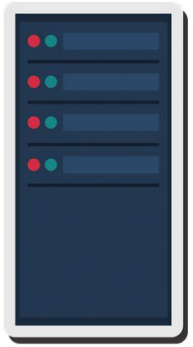
Encrypted connection

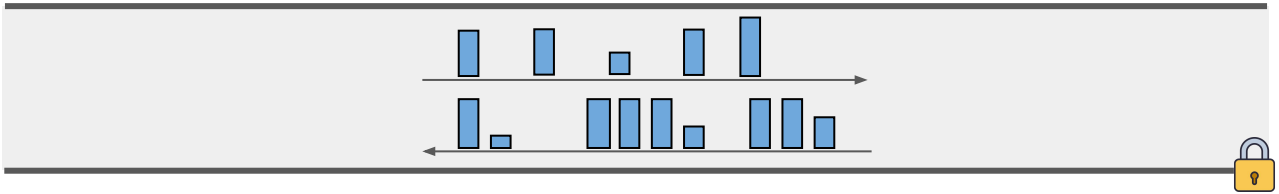
implies confidentiality from 3rd parties



Encrypted connection

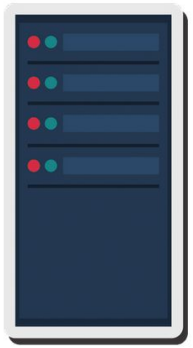
implies confidentiality from 3rd parties

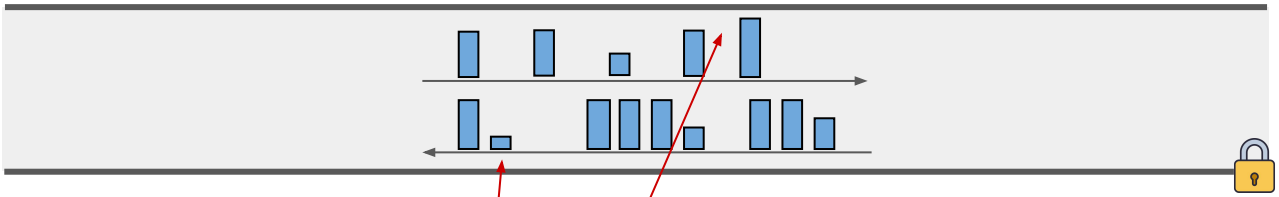




Encrypted connection

implies confidentiality from 3rd parties

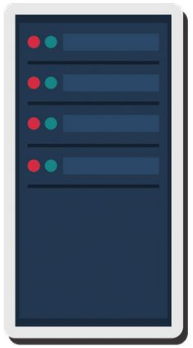


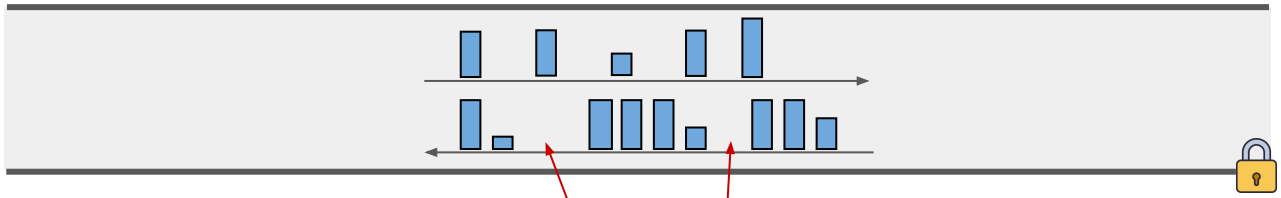


Encrypted connection

implies confidentiality from 3rd parties

Packet sizes



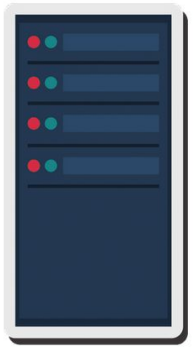


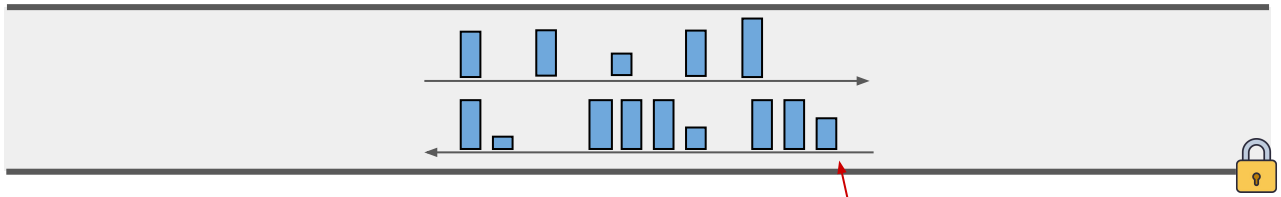
Encrypted connection

implies confidentiality from 3rd parties

Packet sizes

Pauses





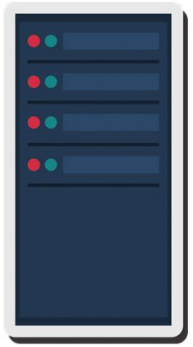
Encrypted connection

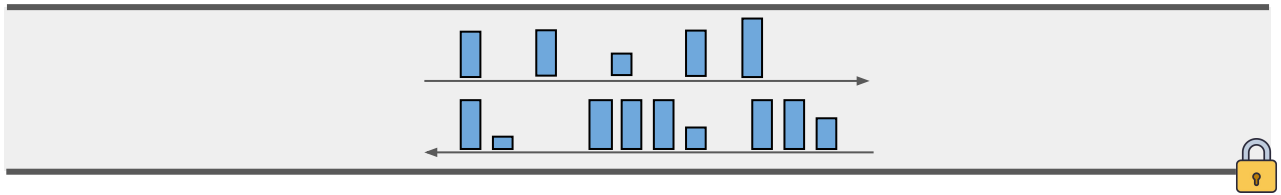
implies confidentiality from 3rd parties

Packet sizes

Total duration

Pauses





Encrypted connection

implies confidentiality from 3rd parties

Packet sizes

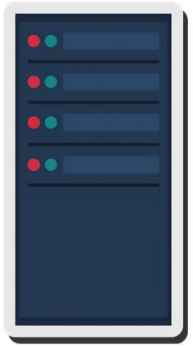
Total duration

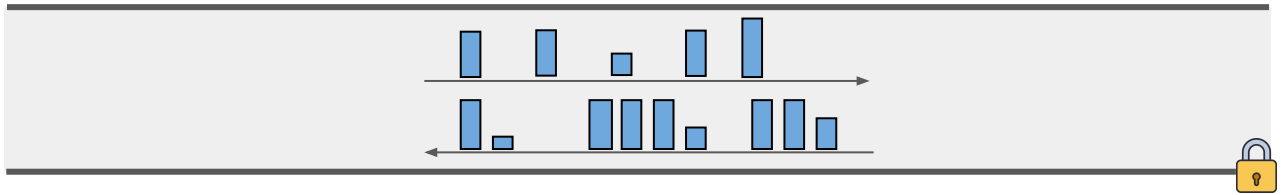
Pauses

Packet counts

Ordering

etc.





Encrypted connection

implies confidentiality from 3rd parties

Unprotected *Metadata* :

Packet sizes

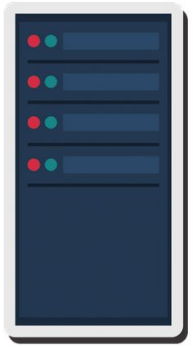
Total duration

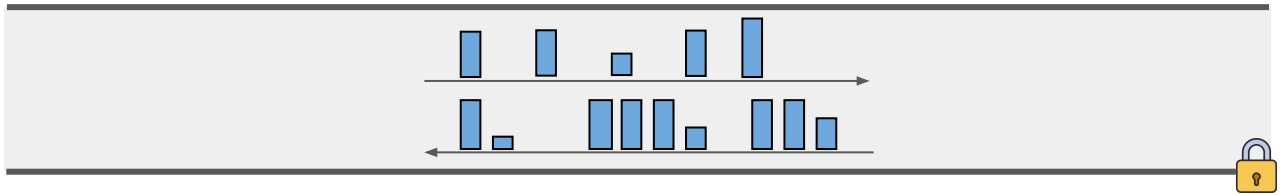
Pauses

Packet counts

Ordering

etc.





Encrypted connection

implies confidentiality from 3rd parties

Unprotected *Metadata* :

Packet sizes

Total duration

Pauses

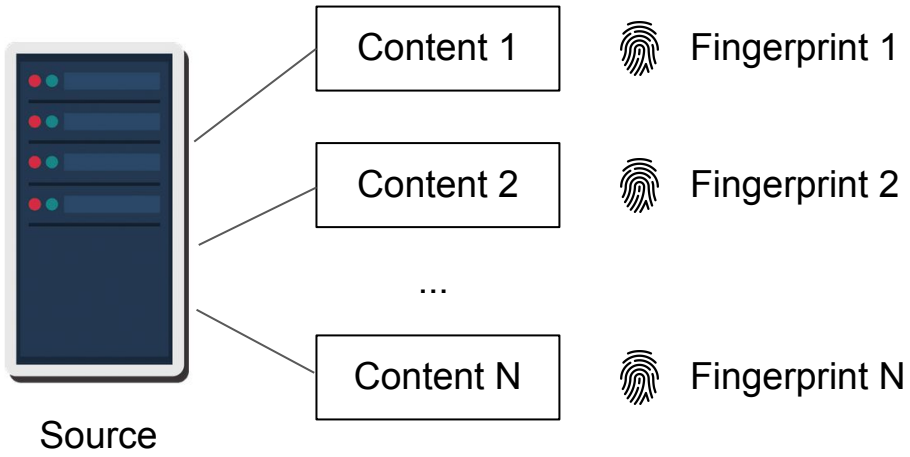
Packet counts

Ordering

etc.

also called a *Fingerprint*

Often, the source is known



My thesis in one slide

1. Assess the threat, raise awareness

Fingerprint → Something sensitive

2. Present solutions

Reduce, remove or perturb the Fingerprint (at a reasonable cost)

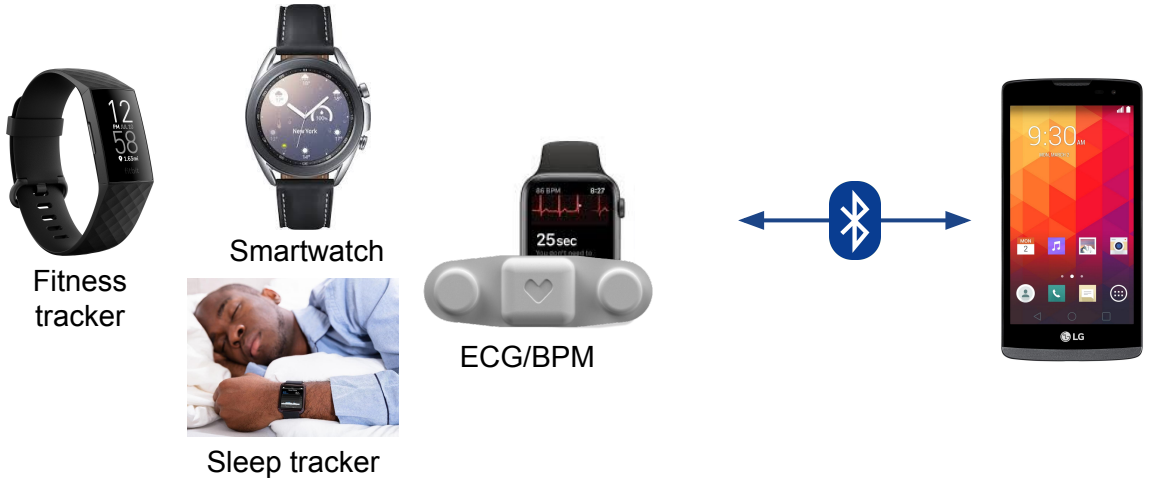
Brief Highlight

Every Byte Matters:
Traffic Analysis of Wearable Devices

joint work with Alexandre, Apostolos and Jean-Pierre

Setting

Wearable devices communicate with a smartphone over Bluetooth



The data exchanged is personal and sensitive

Consumer devices → Medical devices

Smartwatch

Fitness tracker

Sleep tracker

ECG/BPM

The data exchanged is personal and sensitive

Consumer devices → Medical devices

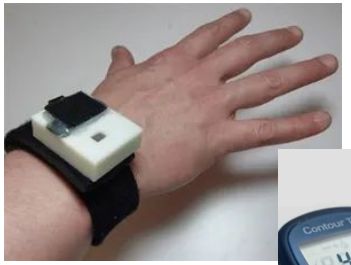
Smartwatch

Fitness tracker

Sleep tracker

ECG/BPM

...



Asthma monitor

Blood sugar monitor

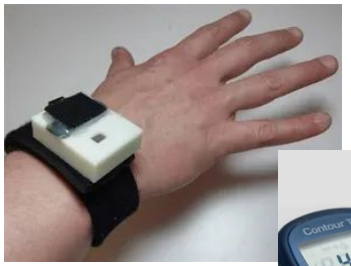
The data exchanged is personal and sensitive

Consumer devices → Medical devices

Smartwatch

Fitness tracker

...



Asthma monitor

Blood sugar monitor

- List of apps (app to stop smoking, medication reminder)
- User activities (e.g., recording an insulin injection)

Goal

Do wearable devices produce stable fingerprints,
and what can be inferred then ?

Step 1: buy some devices

Bluetooth Classic

Vendor	Model	OS
Samsung	Galaxy Watch	Tizen
Fossil	Explorist HR	Wear OS 2
Apple	Watch 4	watchOS 5
Huawei	Watch 2	Wear OS 2
Fitbit	Versa 2	Fitbit OS 4
Sony	MDR-XB9	-
Apple	AirPods	-

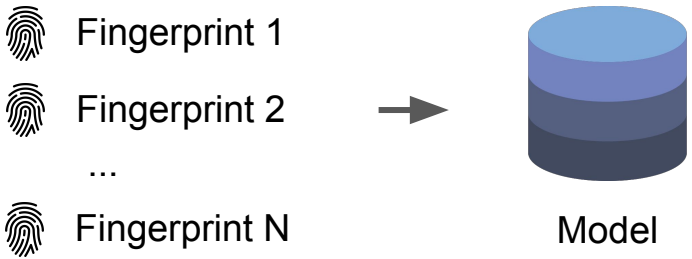
Bluetooth LE

Vendor	Model
Apple	Watch 4
Fitbit	Charge 2
Fitbit	Charge 3
Huawei	Band 3e
Mi	Band 2
Mi	Band 3
Mi	Band 4

smartwatches
headphones
step counters & fitness trackers

Step 2: collect a dataset

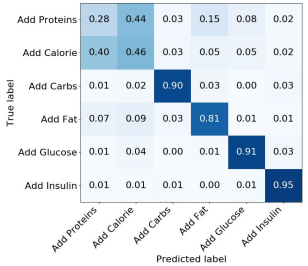
Step 3: design & extract fingerprints, train a model



Step 4: identify what can be learned

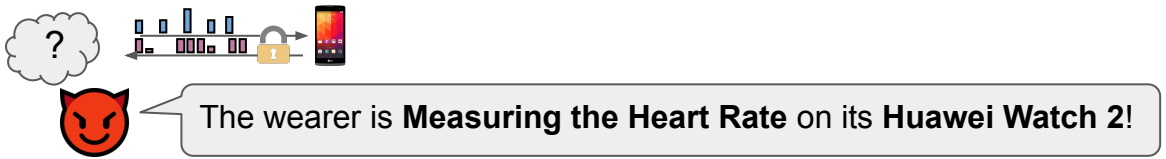
unknown fingerprint

Classify



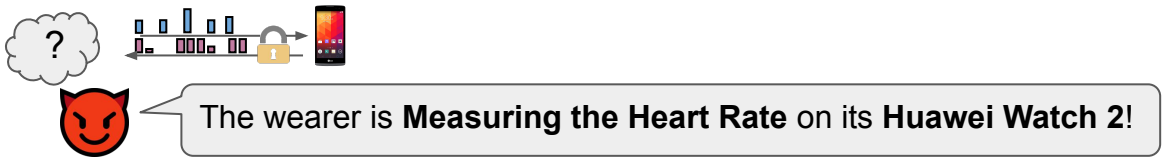
(Some) successful attacks:

Device/Action Identification:

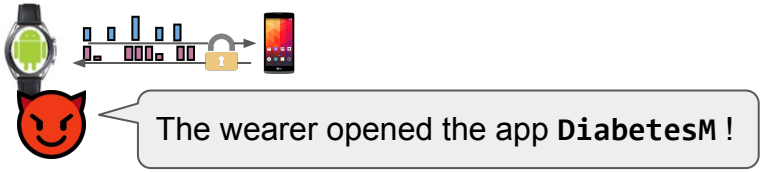


(Some) successful attacks:

Device/Action Identification:

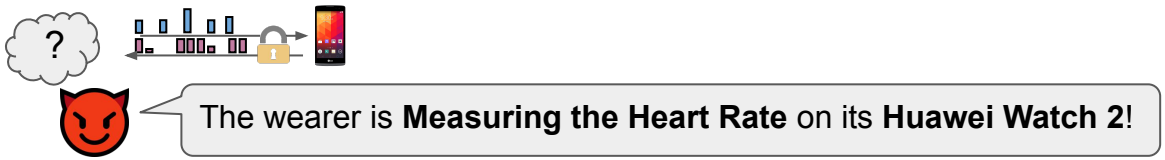


WearOS Apps Identification:

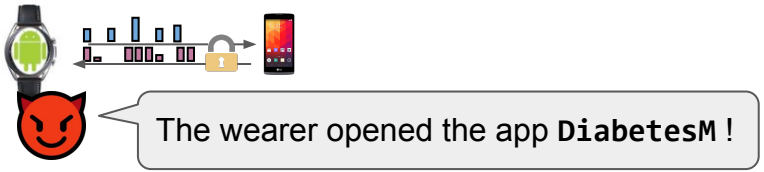


(Some) successful attacks:

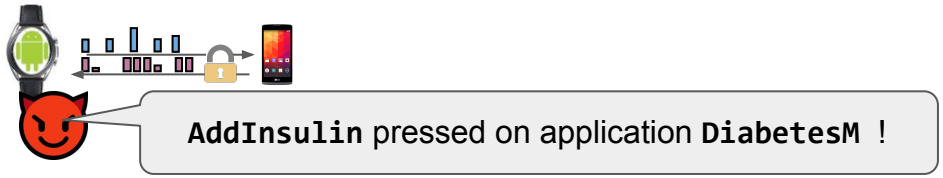
Device/Action Identification:



WearOS Apps Identification:



In-App Action Identification:



Defenses

- We evaluate 3 standard defenses (pad, delay, and group packets)
- They are expensive and mildly effective (e.g., 200x data sent, -20% attack accuracy)
- No easy fix to the problem, probably no overarching solution
- In *specific scenarios*, defending might be easier:
    - Wear OS app openings: packet **sizes** matter
    - In-app action fingerprinting: packet **timings** matter
- Other defense strategies: Data minimization, Bulk transfers

Project 2:

Focus on just one packet. What can we do ?

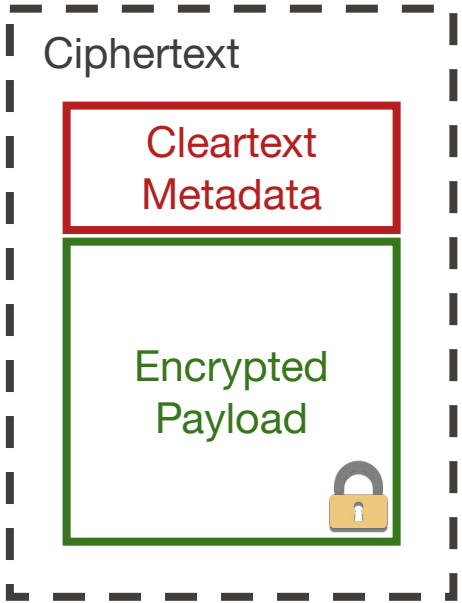
Brief Highlight

Reducing Metadata Leakage from Static Files

Padmé

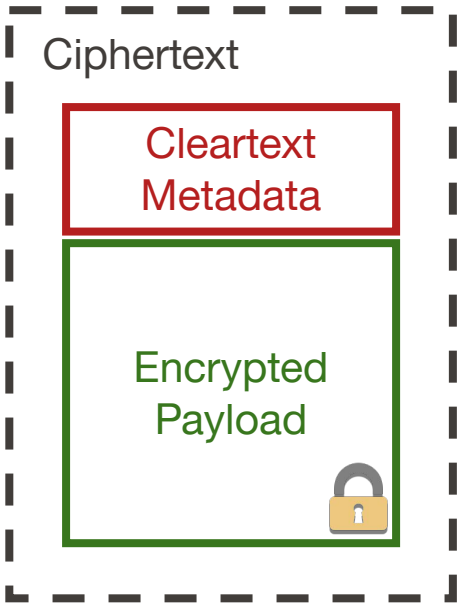
joint work with Kirill, Wouter, Bryan and Jean-Pierre

Ciphertexts often expose metadata in clear



(Contextual Metadata)

Ciphertexts often expose metadata in clear

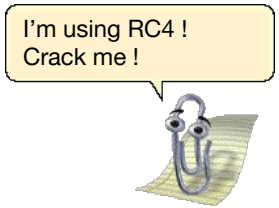


This Metadata may reveal:

What application produced the ciphertext

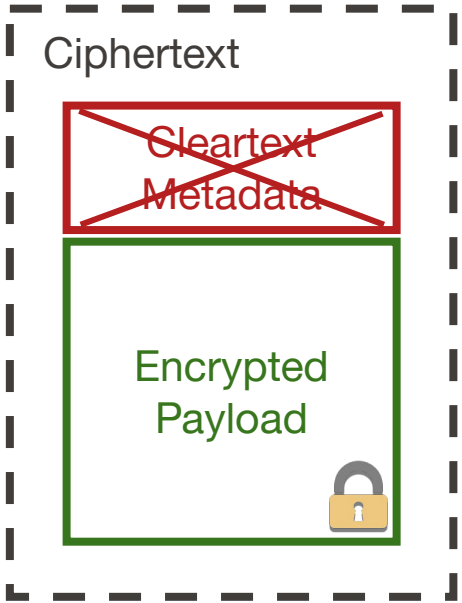
- it tells the context (could be incriminating in itself)
- it enables searching for implementation errors

The cipher suites used

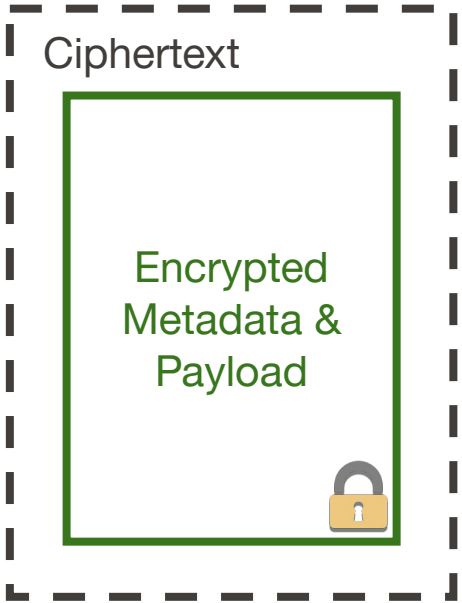


The list of the recipients (PGP)
The length of the contents

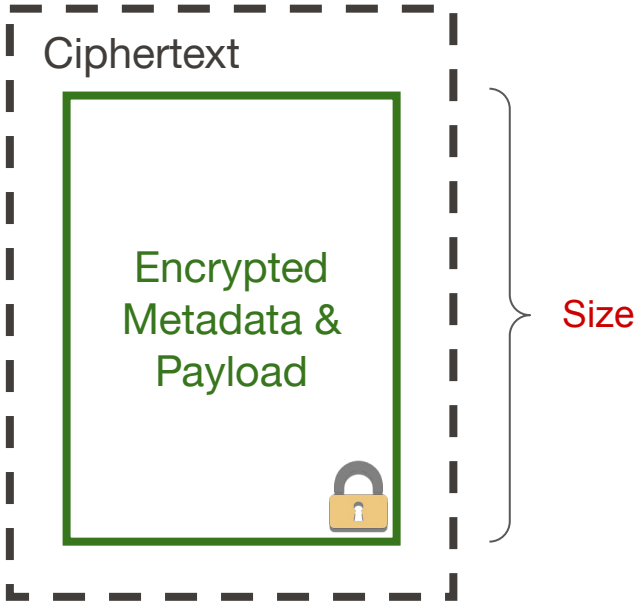
Ciphertexts often expose metadata in clear



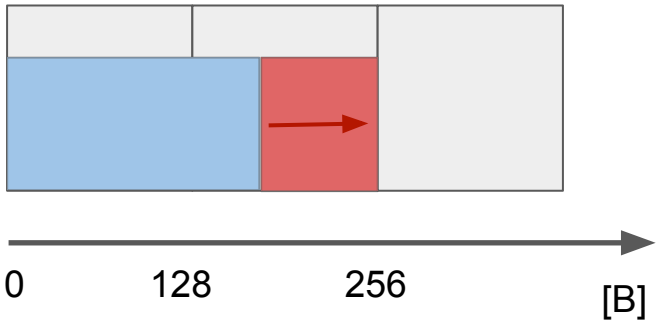
Ciphertexts often expose metadata in clear



Ciphertexts often expose metadata in clear



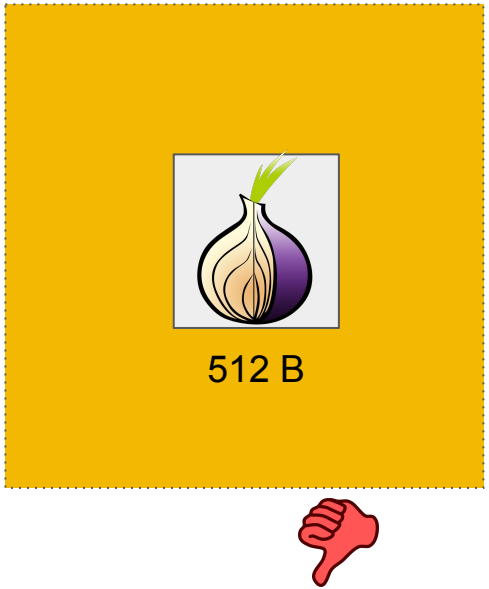
Constant block-size padding



Constant block-size padding

Problem: no good value for block size

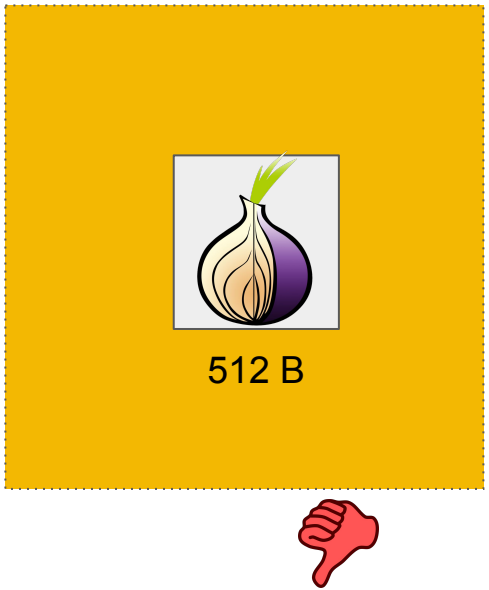
Example: b = 1 MB

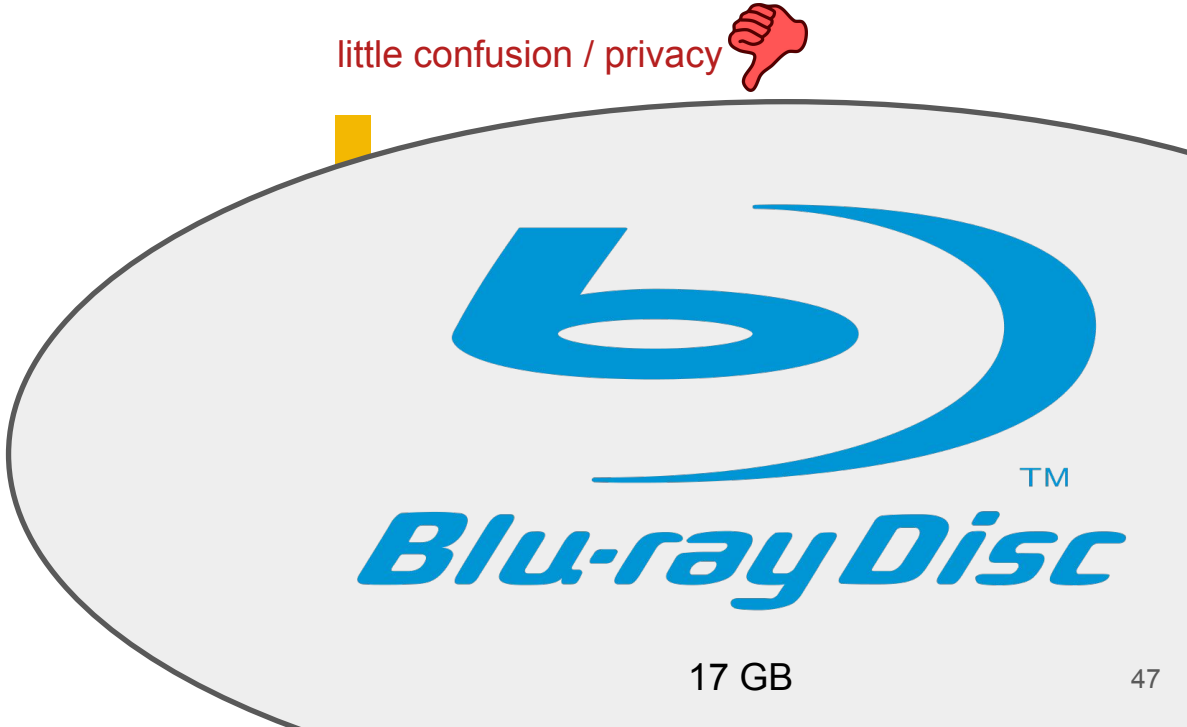


Constant block-size padding

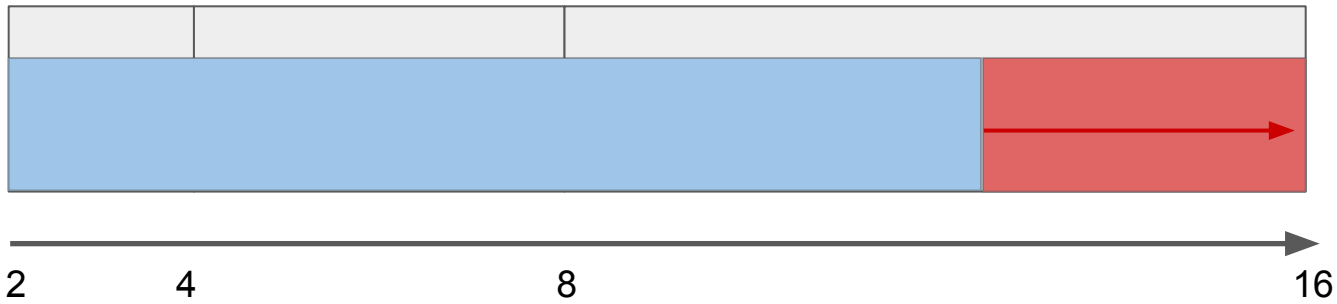
Problem: no good value for block size

Example: b = 1 MB





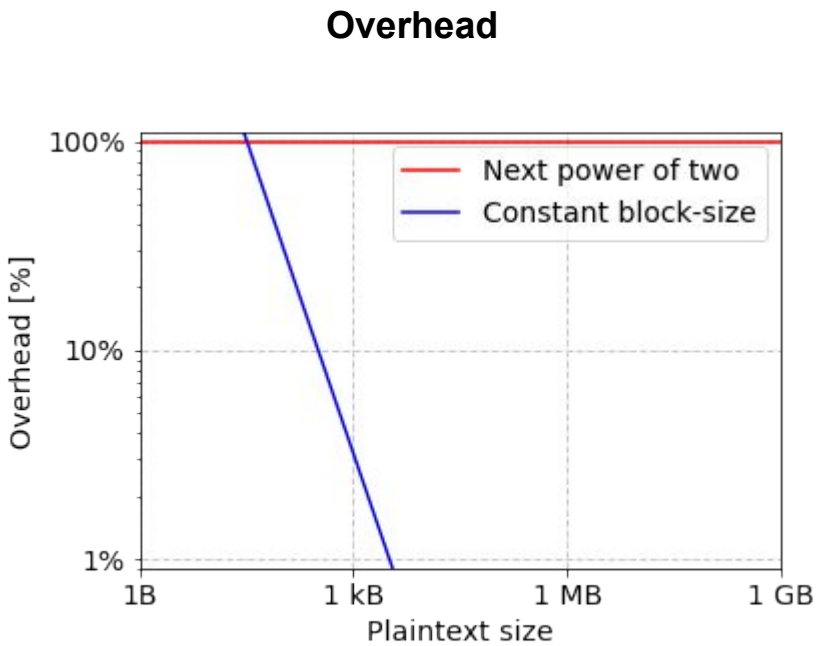
Next power of 2



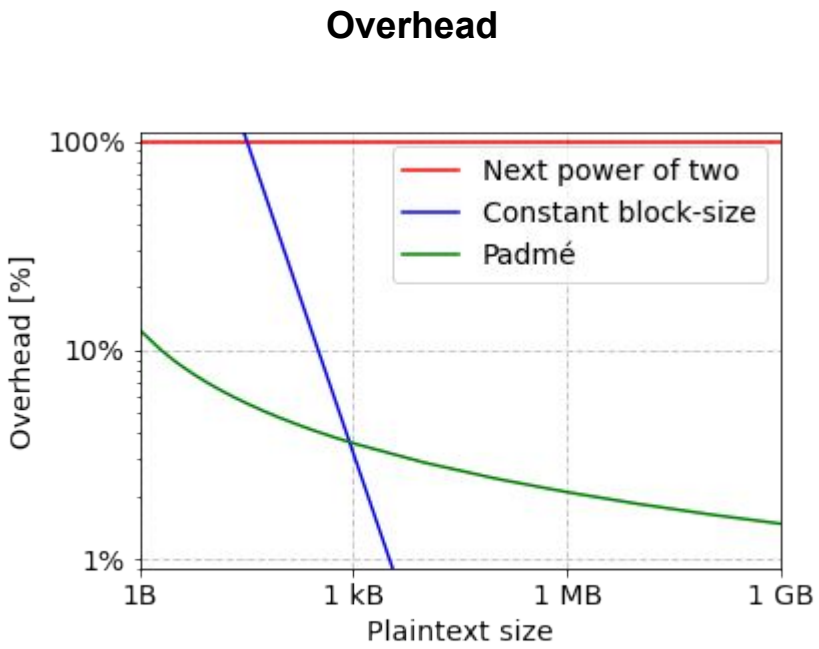
better hiding properties

large overhead (17 GB -> 32 GB)

Padmé

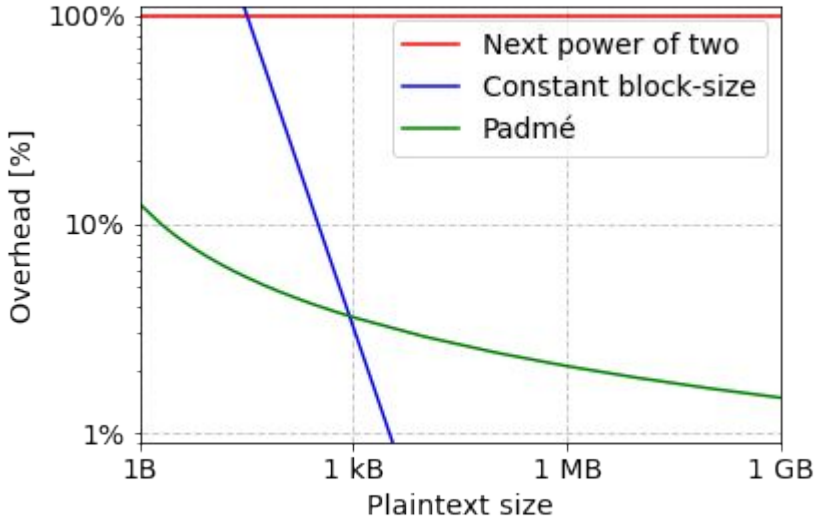

Overhead
100%
10%
1%
Overhead [%]
Next power of two
Constant block-size
1B
1 kB
1 MB
1 GB
Plaintext size

Padmé


Overhead
100%
10%
1%
Overhead [%]
1B
1 kB
1 MB
1 GB
Plaintext size
Next power of two
Constant block-size
Padmé

Padmé

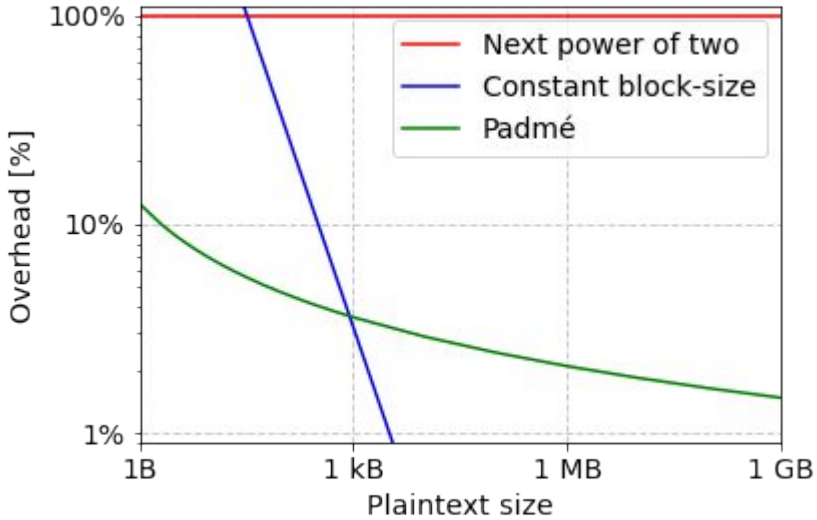
Overhead



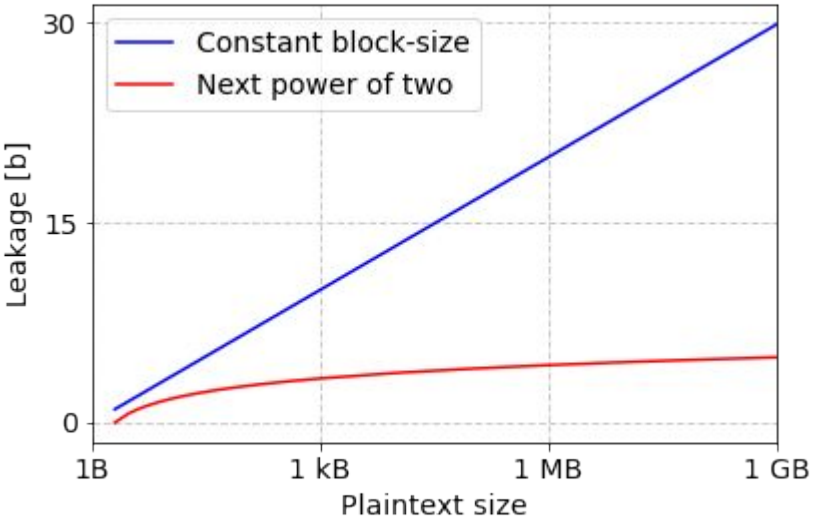
practical overhead

Padmé

Overhead



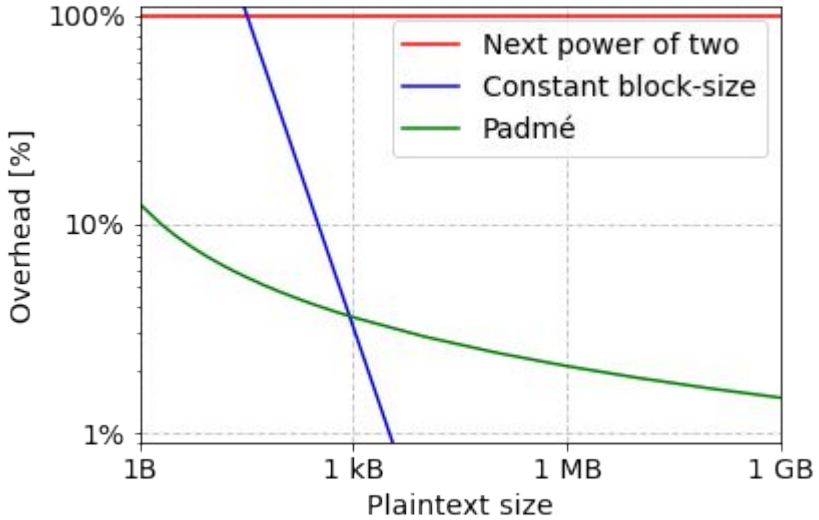
Leakage



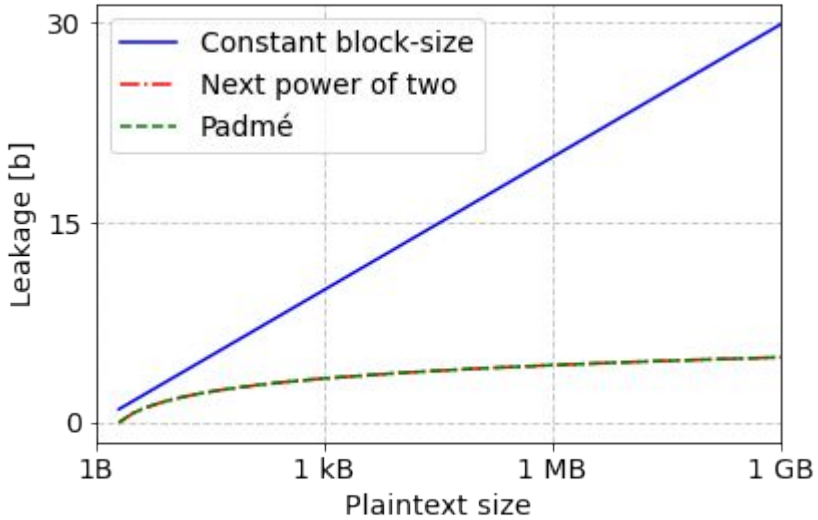
practical overhead

Padmé

Overhead



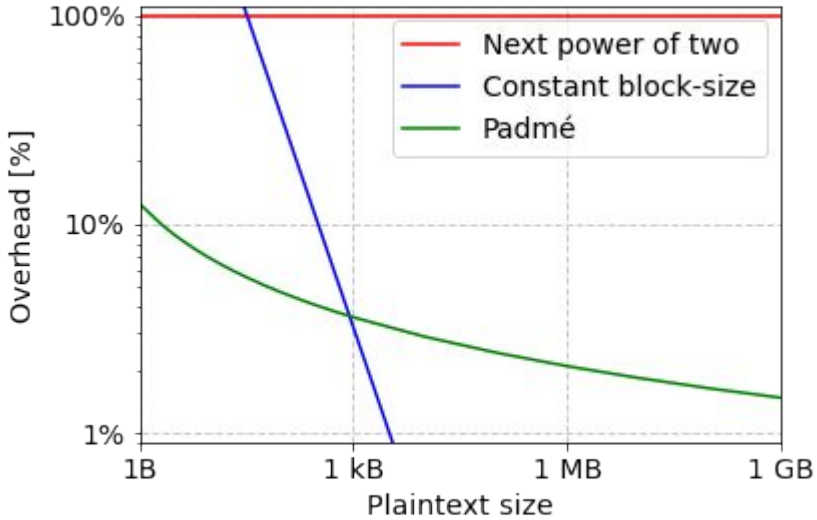
Leakage



practical overhead

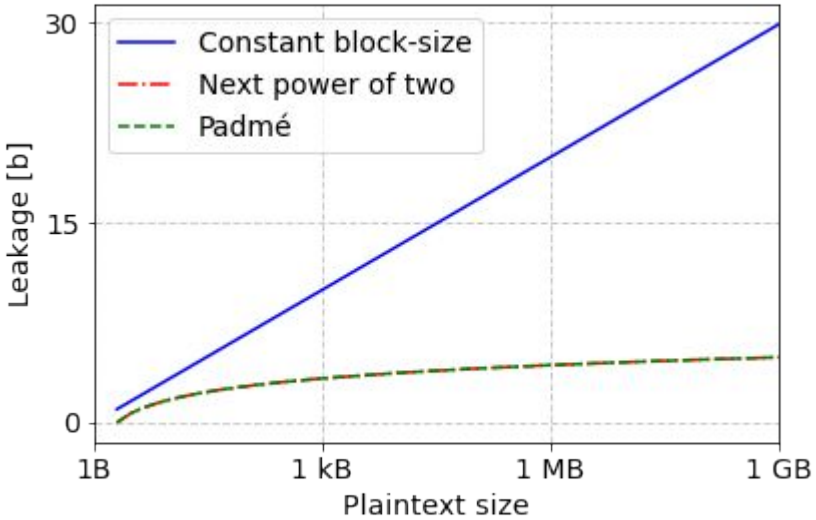
Padmé

Overhead



practical overhead

Leakage



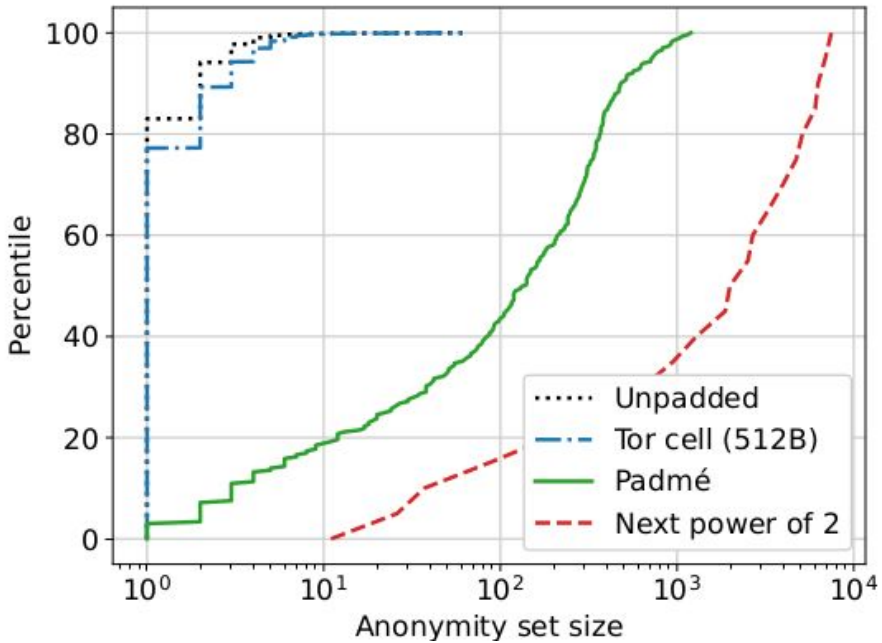
low leakage

Padmé

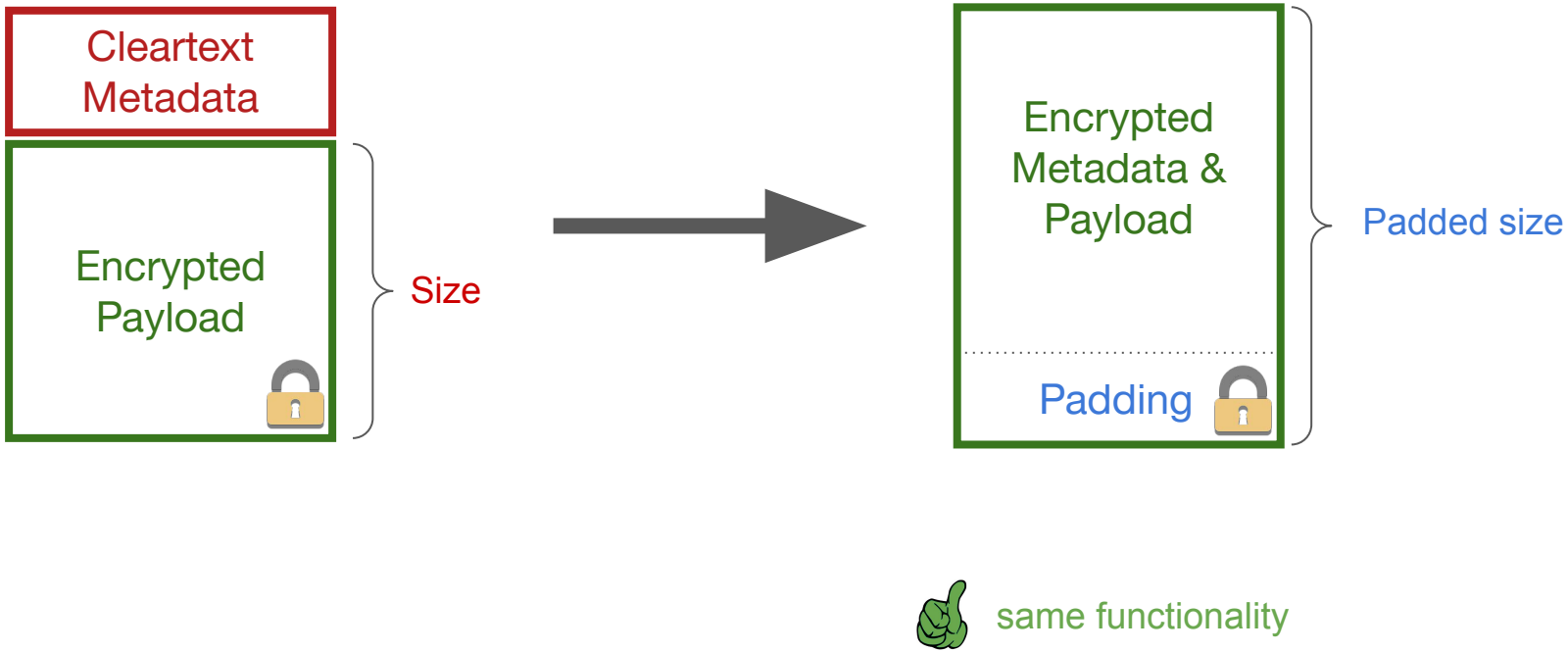
Overhead:

- max +12% ∀L
- max +6% ∀L > 1 MB
- max +3% ∀L > 1 GB

Ubuntu Packages

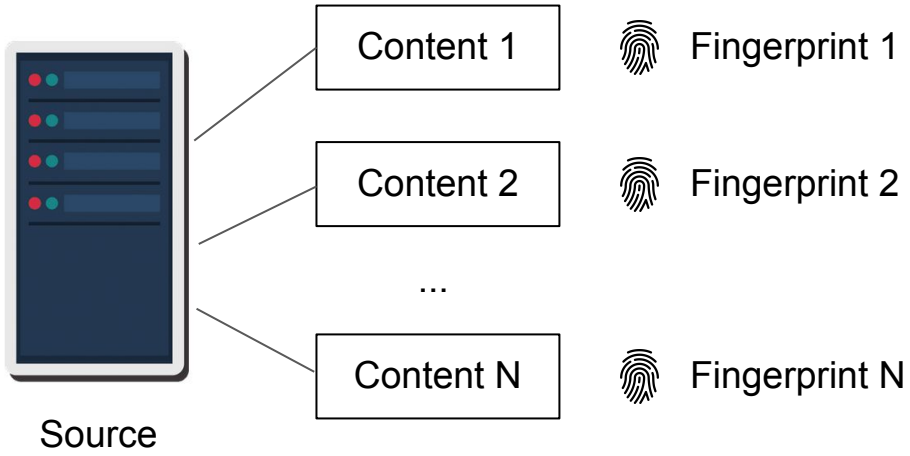


PURBs expose almost no metadata

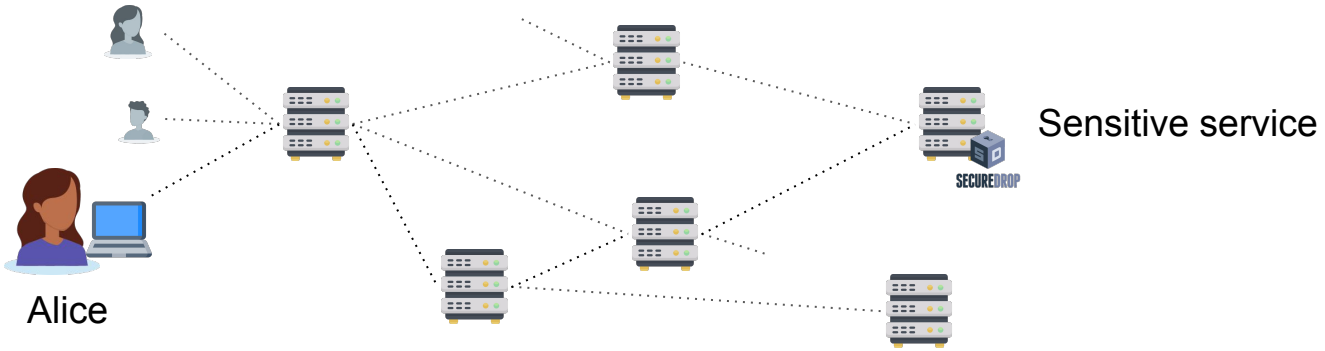


Projects 3 & 4: Anonymous Communication Networks

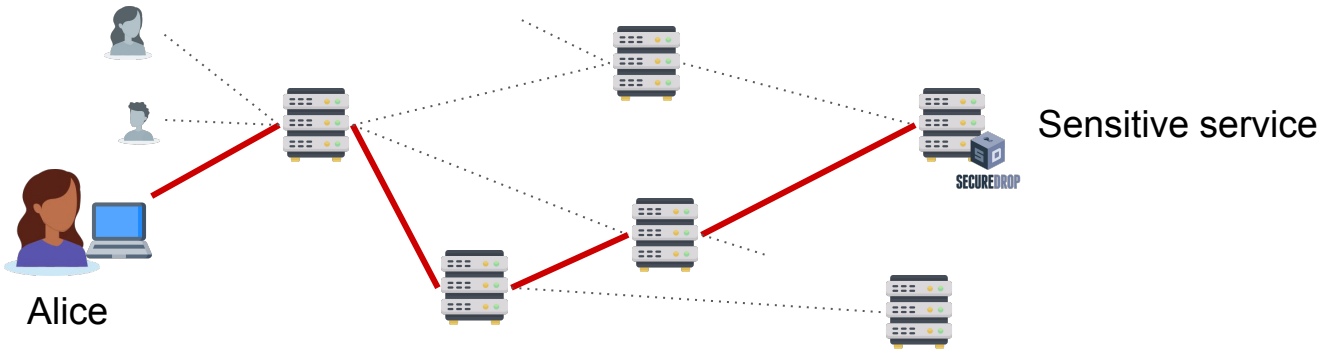
Knowing the source helps a lot



The source / destination can be a sensitive metadata

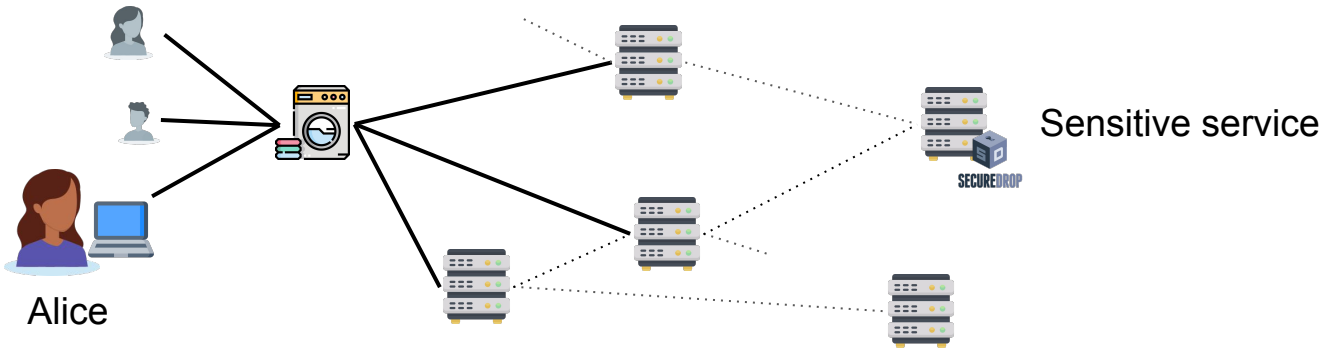


The source / destination can be a sensitive metadata



Alice is contacting SecureDrop !

The source / destination can be a sensitive metadata



Brief Highlight

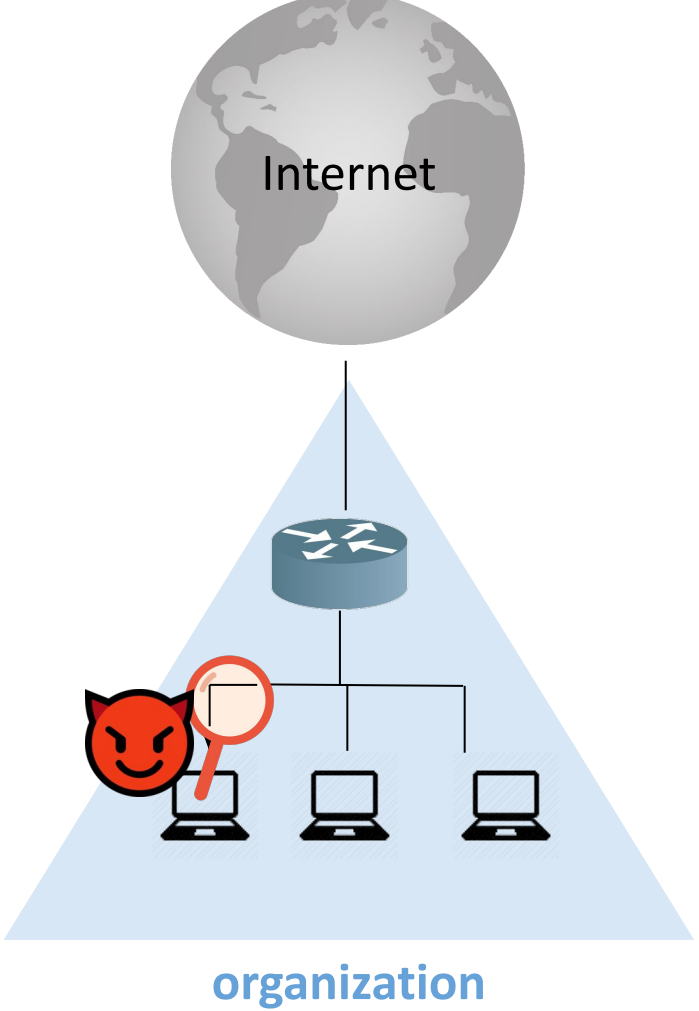
PriFi

A low-latency ACN for LANs and WLANs

joint work with Italo, Apostolos, Bryan and Jean-Pierre

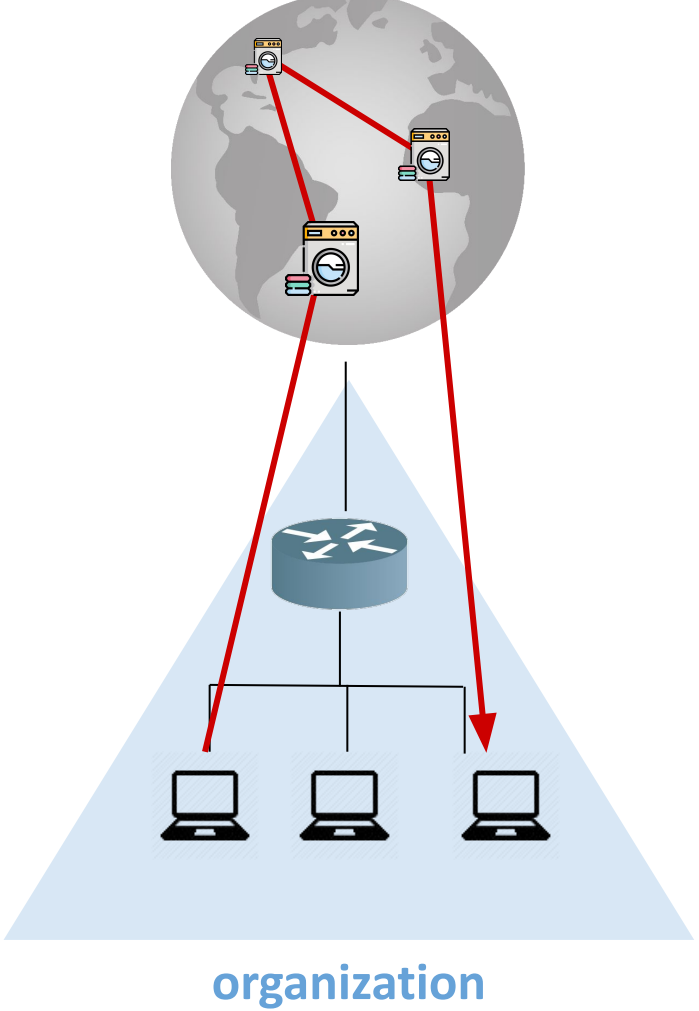
Anonymity for Local-Area Networks

- Problem:
  - Risk of targeted attacks in loosely trusted, sensitive WLANs (e.g., NGOs)
- Goal:
  - Hide source/recipient
  - “Blend in” the traffic of key individuals



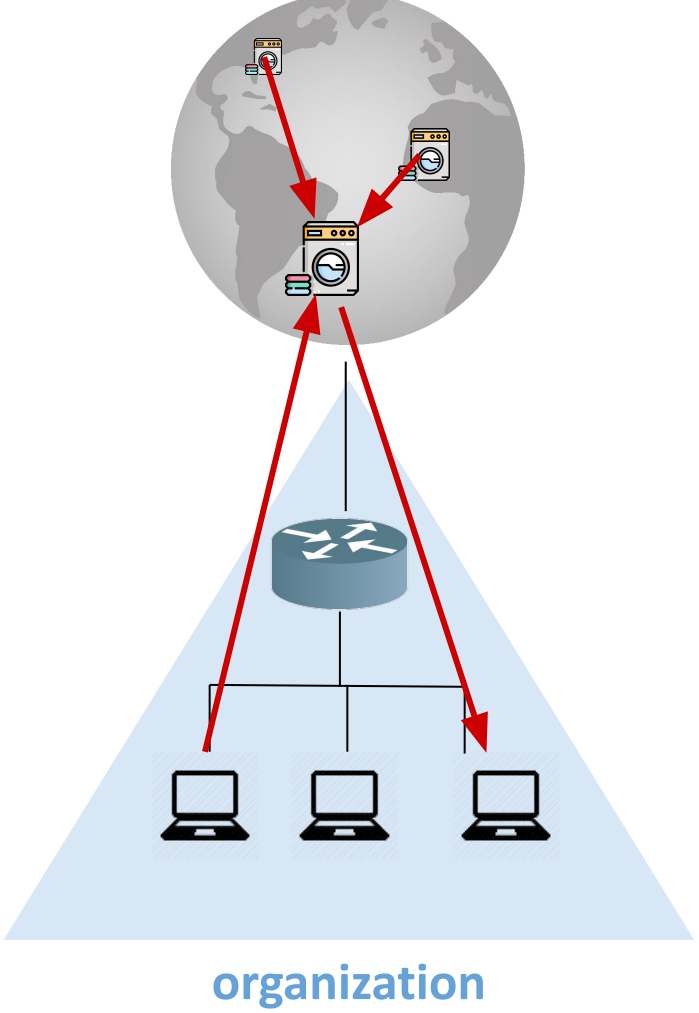
ACNs are poorly suited to LANs

- Tor / Mixnets add extra hops = extra latency
- Traffic leaves the organization



ACNs are poorly suited to LANs

- DC-nets can avoid this
- In practice, they don't [10]
- At each round, "chatty" protocol with the servers [10]



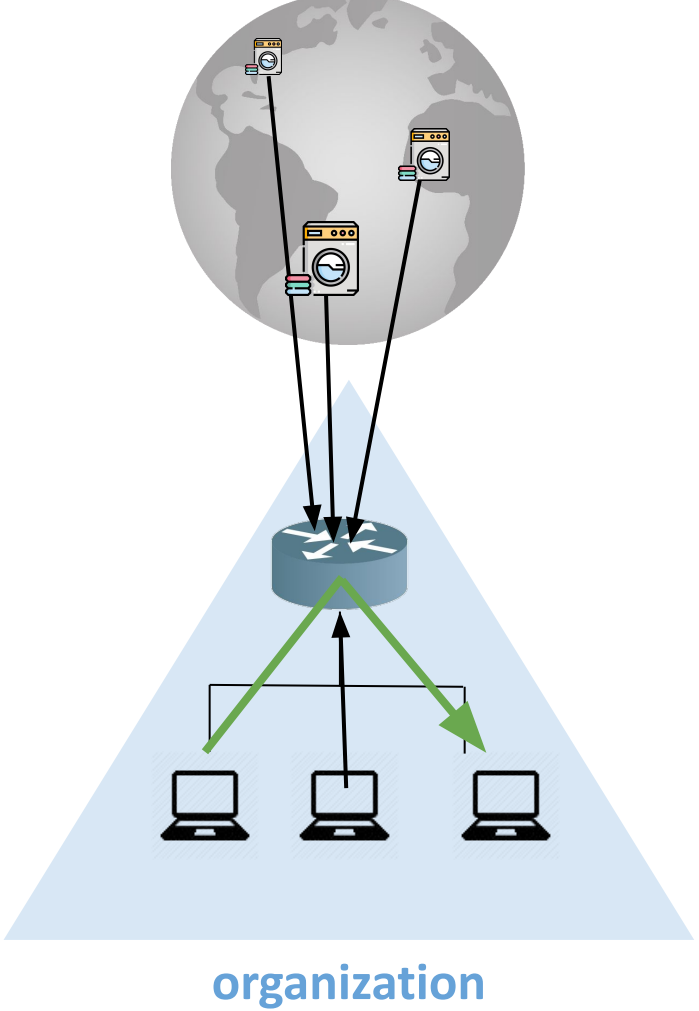
[10] D. Wolinsky, H. Corrigan-Gibbs, B. Ford. **Dissent in Numbers: Making strong anonymity scale.** OSDI 2012

PriFi

- New topology for DC-nets

- Redesign of the protocols
  - servers contributions are sent in advance
  - avoid server-to-server messages

=> Latency to the servers is not important



PriFi

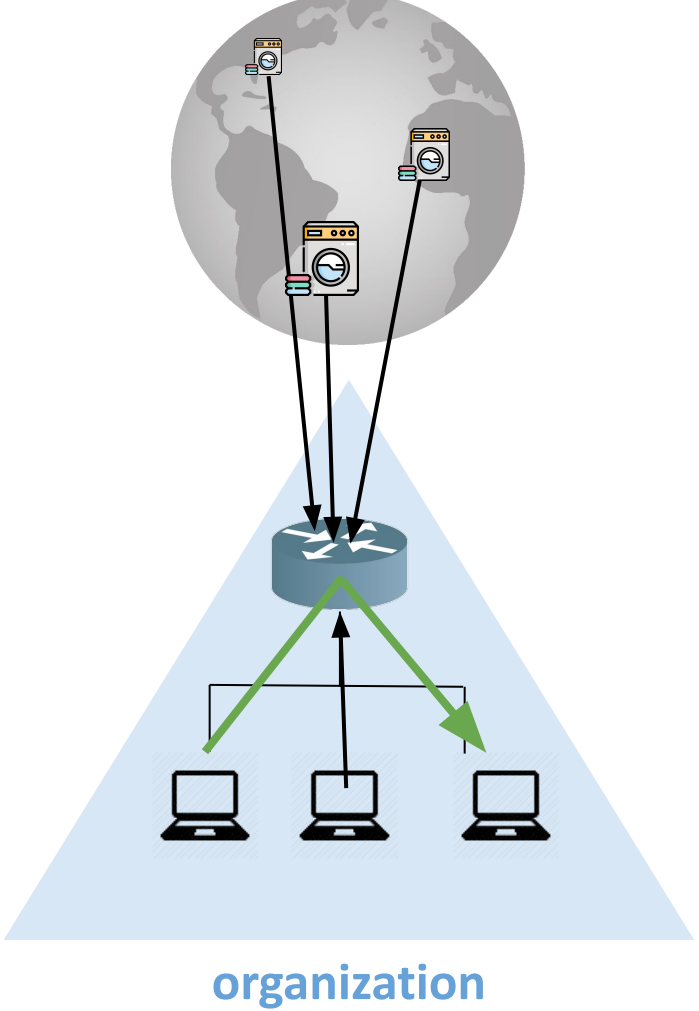
- New topology for DC-nets

- Redesign of the protocols
  - servers contributions are sent in advance
  - avoid server-to-server messages

=> Latency to the servers is not important

=> “on-path” anonymity

=> cheap broadcast in WLANs



PriFi

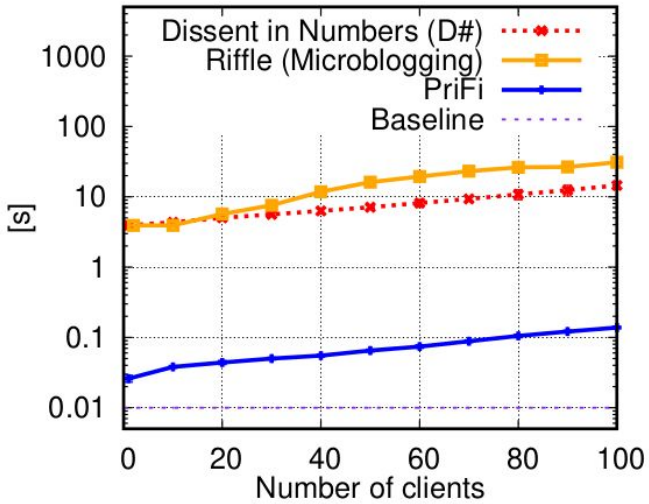
- New topology for DC-nets

- Redesign of the protocols
  - servers contributions are sent in advance
  - avoid server-to-server messages

=> Latency to the servers is not important

=> “on-path” anonymity

=> cheap broadcast in WLANs



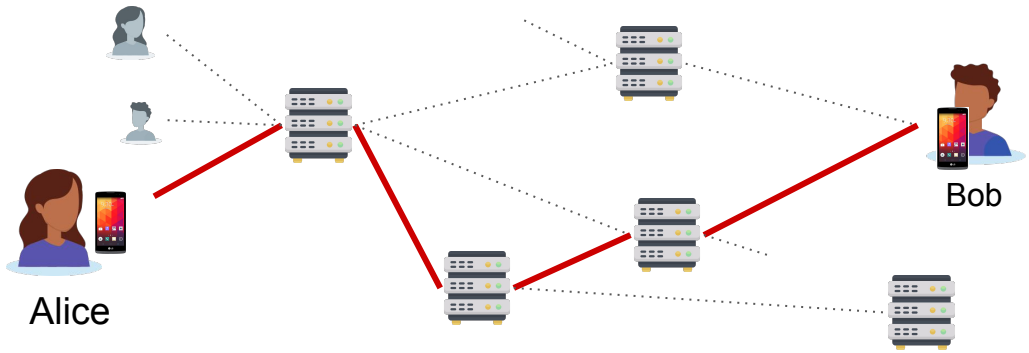
Brief Highlight

Rubato: Metadata-Private Communications for Mobile Devices

joint work with Moshe, David, Yossi, Nickolai

System for text communication on phones

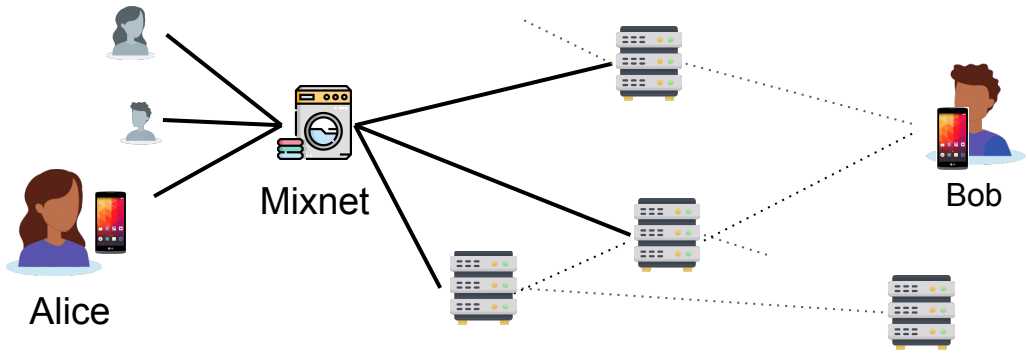
Hide relationships



Alice and Bob know each other !

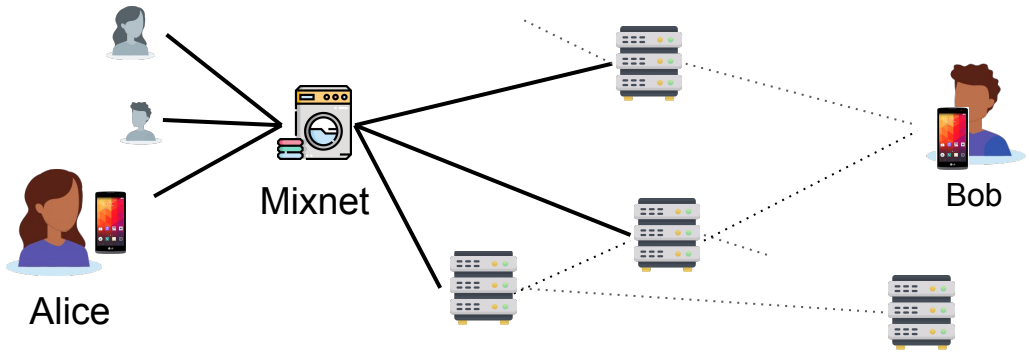
Model: Global Active Adversary, Honest Sender/Recipient

Hide relationships with Mixnets



Model: Global Active Adversary, Honest Sender/Recipient

Hide relationships with Mixnets

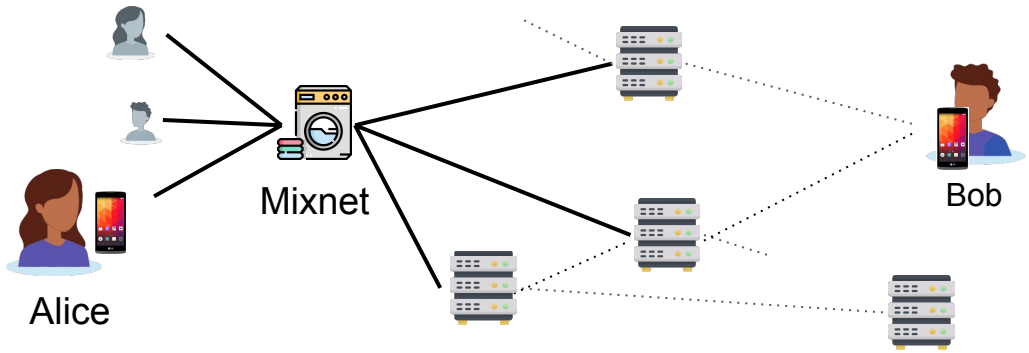


 safe

 very impractical for users:

- Mixnets supports only one device / user
- it must be always online
- every minute, it must send a message

Hide relationships with Mixnets

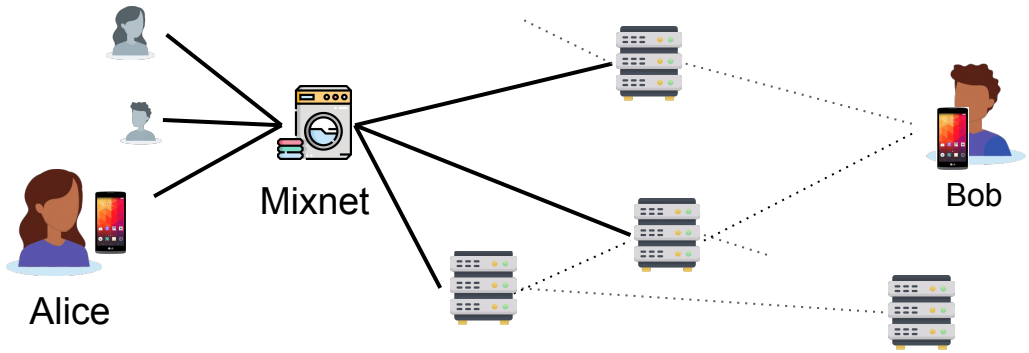


safe

very impractical for users:

- Mixnets supports only one device / user
- it must be always online
- every minute, it must send a message

Hide relationships with Mixnets

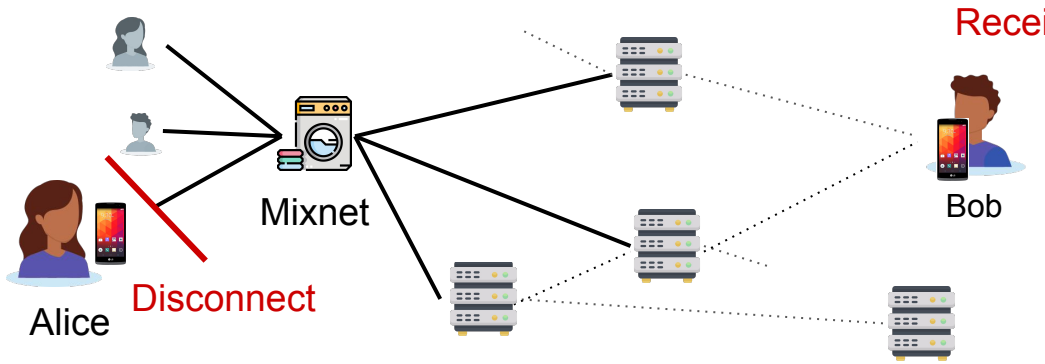


safe

very impractical for users:

- Mixnets supports only one device / user
- it must be always online
- every minute, it must send a message

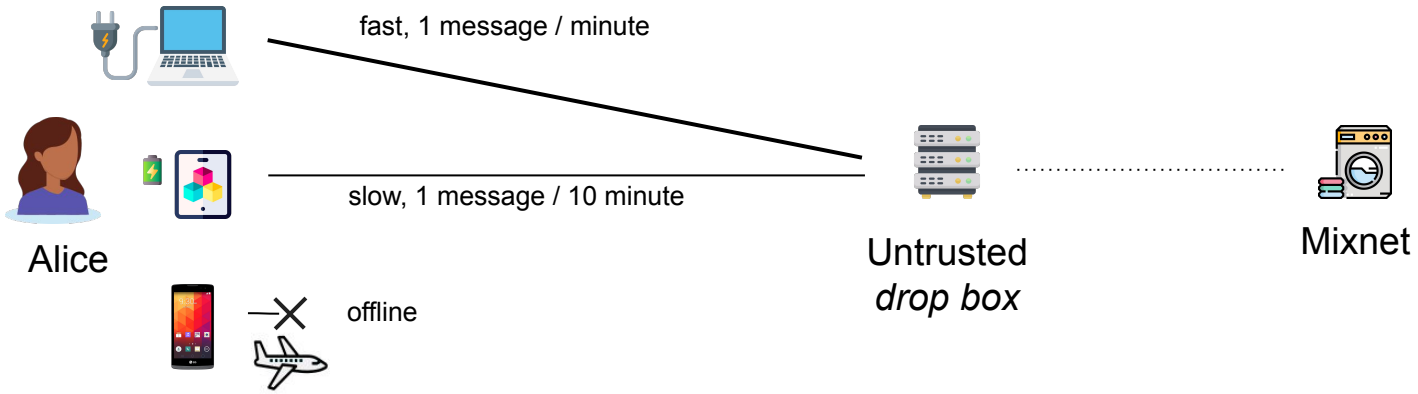
Cost of large-scale mixnets



Global *Active* Adversary

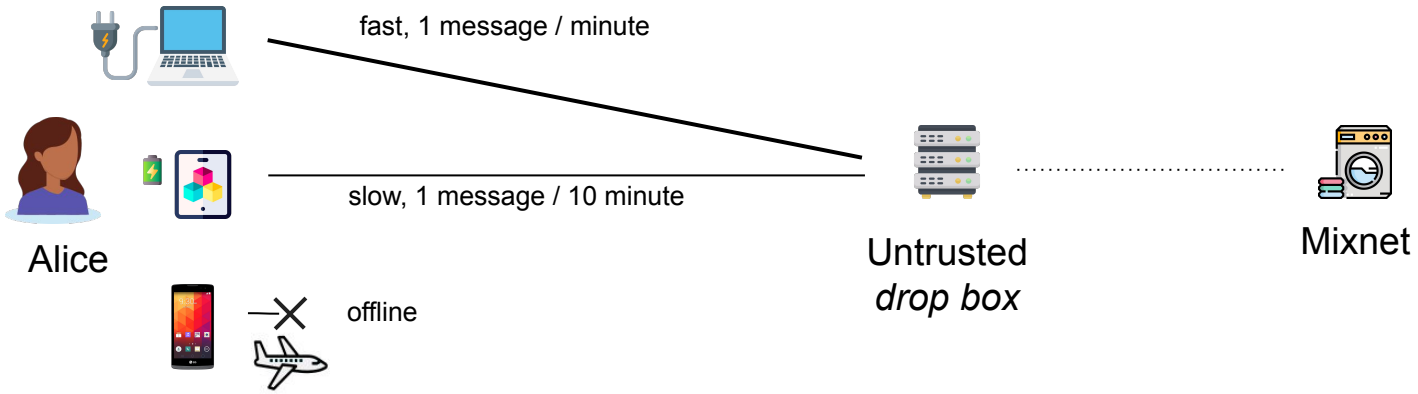
- Sees all the network
- Can delete / delay / inject packets

Rubato



- user can switch between devices
- each device can disconnect
- each device can choose its *schedule*

Rubato



steps Mixnets closer to standard messaging apps

Signal

Client Costs

Bandwidth:

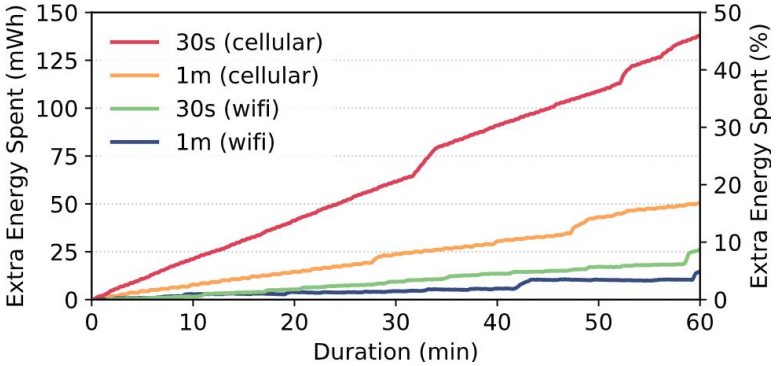
Setup: 110 KB/epoch = 100 MB/month

Messaging: 105 MB/month

Latency: between 32s and 64s

(with a 1-min schedule)

Energy usage:



With a 5-min schedule, after 1h:
≈ +5% energy usage

Conclusion

Contributions of the thesis

- Every Byte Matters: Traffic Analysis of Bluetooth Wearable Devices
  - First broad analysis of the communication metadata of wearable devices
  - We reveal a general susceptibility to traffic-analysis attacks, which can allow:
    - identifying devices, applications, user actions
    - tracking and profiling users
  - If we want to protect such information, we need defense strategies

Contributions of the thesis

- Every Byte Matters: Traffic Analysis of Bluetooth Wearable Devices
  - First broad analysis of the communication metadata of wearable devices
  - We reveal a general susceptibility to traffic-analysis attacks, which can allow:
    - identifying devices, applications, user actions
    - tracking and profiling users
  - If we want to protect such information, we need defense strategies

- Padmé
  - Padding function with low costs (<12%) that outperforms classic approaches asymptotically
  - In practice, we show that it has good hiding properties

Contributions of the thesis (cont')

- PriFi
  - Low-latency, traffic-agnostic anonymity for a small set of users (VoIP support)
  - The latency does not depend on the latency to the anytrust servers
  - "On-path" anonymization that provides low latency

Contributions of the thesis (cont')

- PriFi
  - Low-latency, traffic-agnostic anonymity for a small set of users (VoIP support)
  - The latency does not depend on the latency to the anytrust servers
  - "On-path" anonymization that provides low latency

- Rubato
  - First large-scale ACN with multi-device, asynchronous clients (Global Active Adversary setting)
  - Each device can choose its communication frequency & costs
  - It enables mobile devices to participate at a reasonable cost

Impact outside of research

- Every Byte Matters: Traffic Analysis of Bluetooth Wearable Devices
  - Contacted ~100 vendors and manufacturers, ~10 follow-ups by email, 2 follow-up meetings with large device manufacturers
  - Received a bug bounty
- Padmé
  - Maintainers of SequoiaPGP implemented Padmé
- PriFi
  - Demos at the Red Cross (ICRC) headquarters and at EPFL (one awarded a prize)
  - Patent

Next steps for metadata privacy ?

It is still an open problem

- No one-size-fits-all defense

  => Per domain, iteratively evaluate risks

- Compared to non-metadata-private alternatives, solutions are costly

  => Increase visibility of the attacks to justify the costs

Building safer apps

- Could we have (automated?) guidelines for app developers ?

- Could we have “defense strategies” provided by the OS ?

  This could be an opportunity for designing the defenses iteratively

Analyzing and Protecting Communication Metadata

Experimental

Every Byte Matters [4]
Traffic-analysis attack of wearable devices.

Theoretical

Padmé [2]
A padding function that efficiently hides sizes.

Attacks ⟷ Defenses

Systems: Anonymous Communication Networks (ACN)

PriFi [1,3]
Traffic-agnostic, low-latency ACN for local-area networks.

Rubato [5]
Large-scale ACN for text messaging on mobile devices.

Thesis & technical presentations soon on https://lbarman.ch

[1] L. Barman, M. Zamani, I. Dacosta, J. Feigenbaum, B. Ford, J.-P. Hubaux, D. Wolinsky. **PriFi: A Low-latency [...] Protocol for Local-Area Anonymous [...]**. WPES 2016.

[2] K. Nikitin*, L. Barman*, W. Lueks, M. Underwood, J.-P. Hubaux, B. Ford. **Reducing Metadata Leakage from Encrypted Files and Communication with PURBs**. PETS 2019

[3] L. Barman, I. Dacosta, M. Zamani, E. Zhai, A. Pyrgelis, B. Ford, J. Feigenbaum, J.-P. Hubaux. **PriFi: Low-latency Anonymity for Organizational Networks**. PETS 2020

[4] L. Barman, A. Dumur, A. Pyrgelis, J.-P. Hubaux. **Every Byte Matters: Traffic Analysis of Bluetooth Wearable Devices**. UbiComp 2021.

[5] L. Barman, M. Kol, D. Lazar, Y. Gilad, N. Zeldovich. **Rubato: Metadata-Private Messaging for Mobile Devices**. Under submission.