Duplicate detection methodology for IP network traffic analysis

Iñaki Ucar[1], Daniel Morato[2], Eduardo Magaña[2], Mikel Izal[2]

inaki.ucar@uc3m.es, {daniel.morato|eduardo.magana|mikel.izal}@unavarra.es

[1]Department of Telematic Engineering
University Carlos III of Madrid

[2]Department of Automatics and Computer Science
Public University of Navarre

upna Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

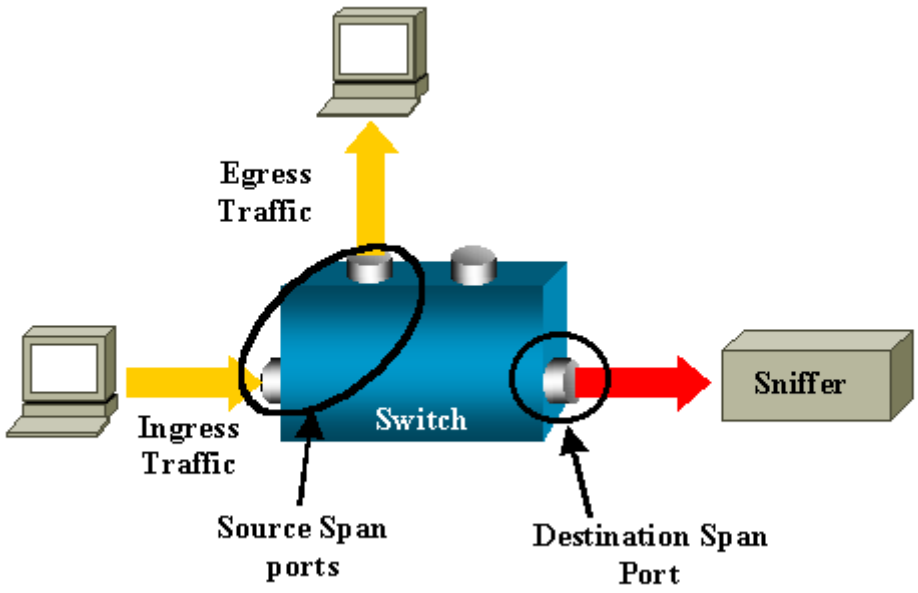
Contents

1. Introduction
2. Theoretical analysis
3. Duplicate detection methodology
4. Efficiency aspects
5. Conclusions

1. Introduction

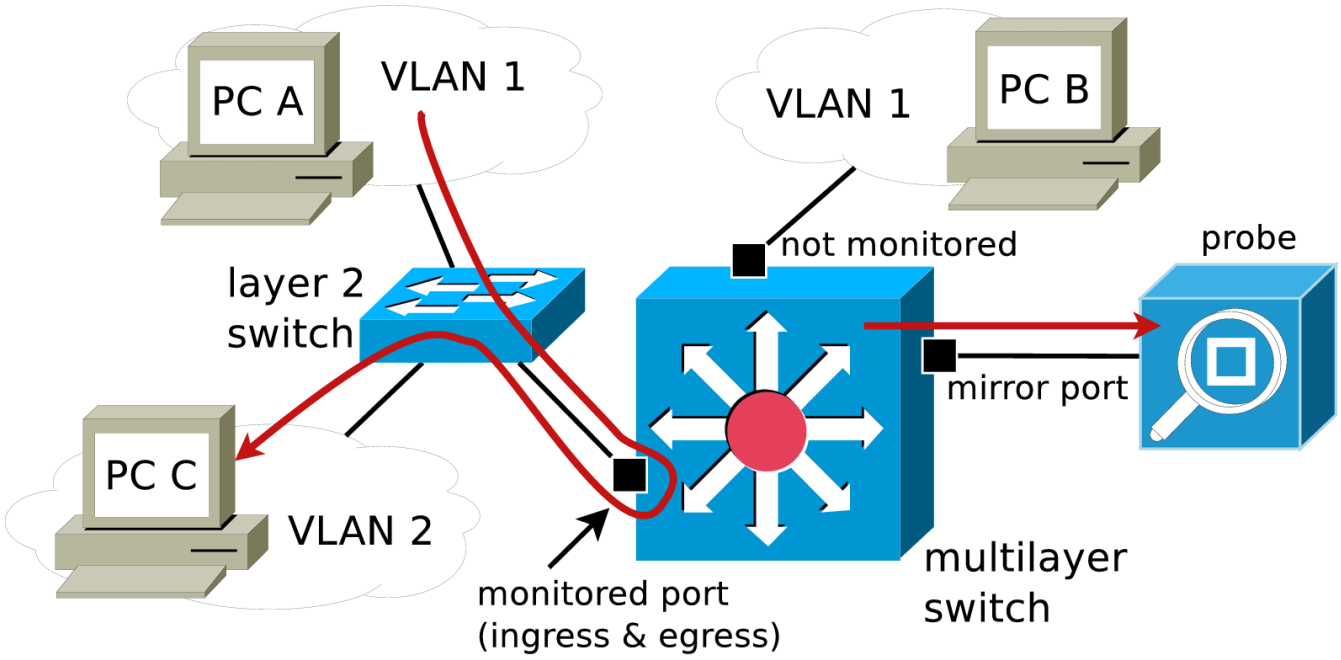
1. Introduction

- Traffic monitoring in Ethernet-based packet-switched networks
  - Port mirroring (Cisco´s SPAN)



1. Introduction

- A simple example



1. Introduction

- Impact of duplicate packets
  - Throughput duplication (some streams may be affected, others may not)
    - SLA planning
    - Threshold-based alerting
    - Traffic matrix characterization
    - Heavy hitters
    - Packet size distributions
    - …
  - Tracking of stateful connections
    - A duplicated TCP sequence can be mistaken for a valid retransmission

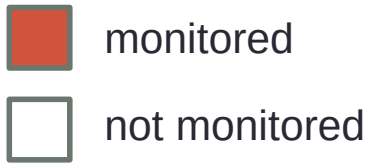
2. Theoretical analysis

Duplication mechanisms | Types of duplicates

2. Theoretical analysis

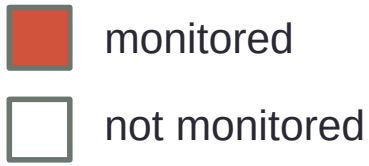
- Referred but not limited to a switched Ethernet environment
- IPv4 as layer 3 (IPv6 case is analogous)

2. Theoretical analysis

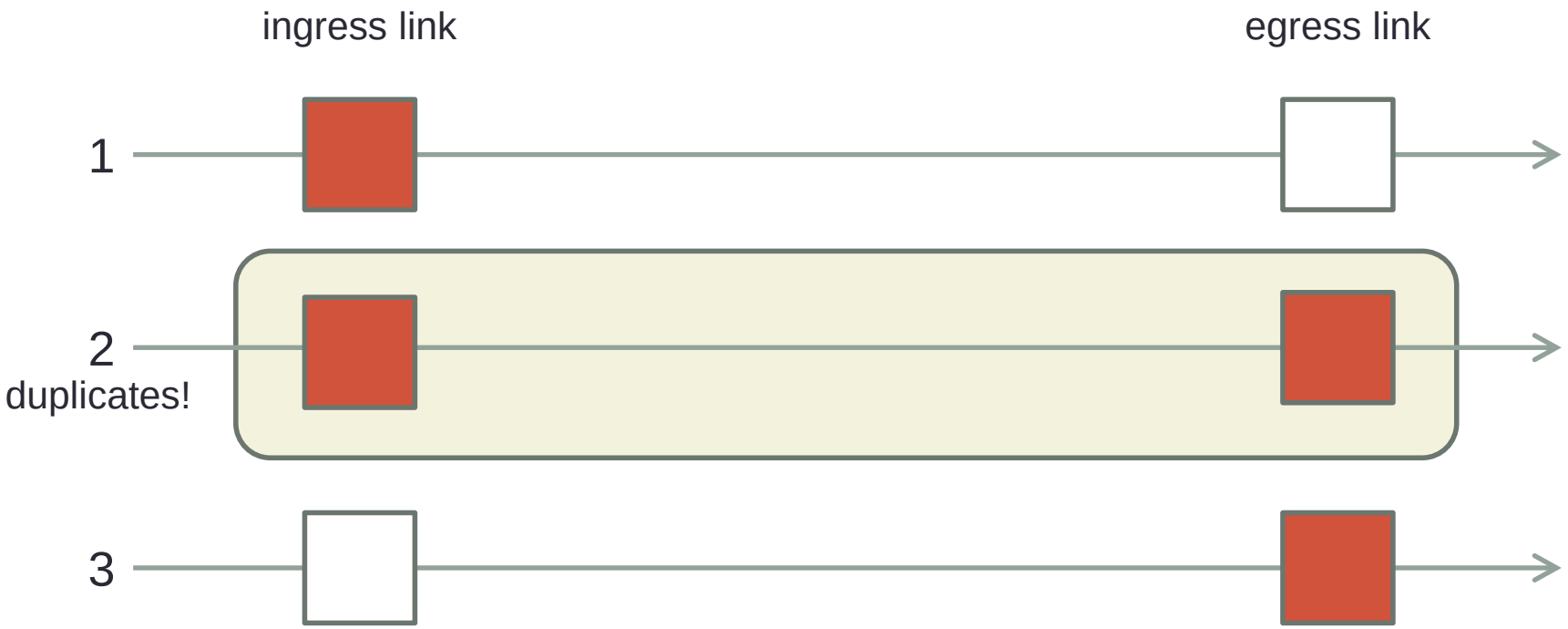


- Referred but not limited to a switched Ethernet environment
- IPv4 as layer 3 (IPv6 case is analogous)
- Packet traversing a monitored device, 3 possibilities…

2. Theoretical analysis



- Referred but not limited to a switched Ethernet environment
- IPv4 as layer 3 (IPv6 case is analogous)
- Packet traversing a monitored device, 3 possibilities…

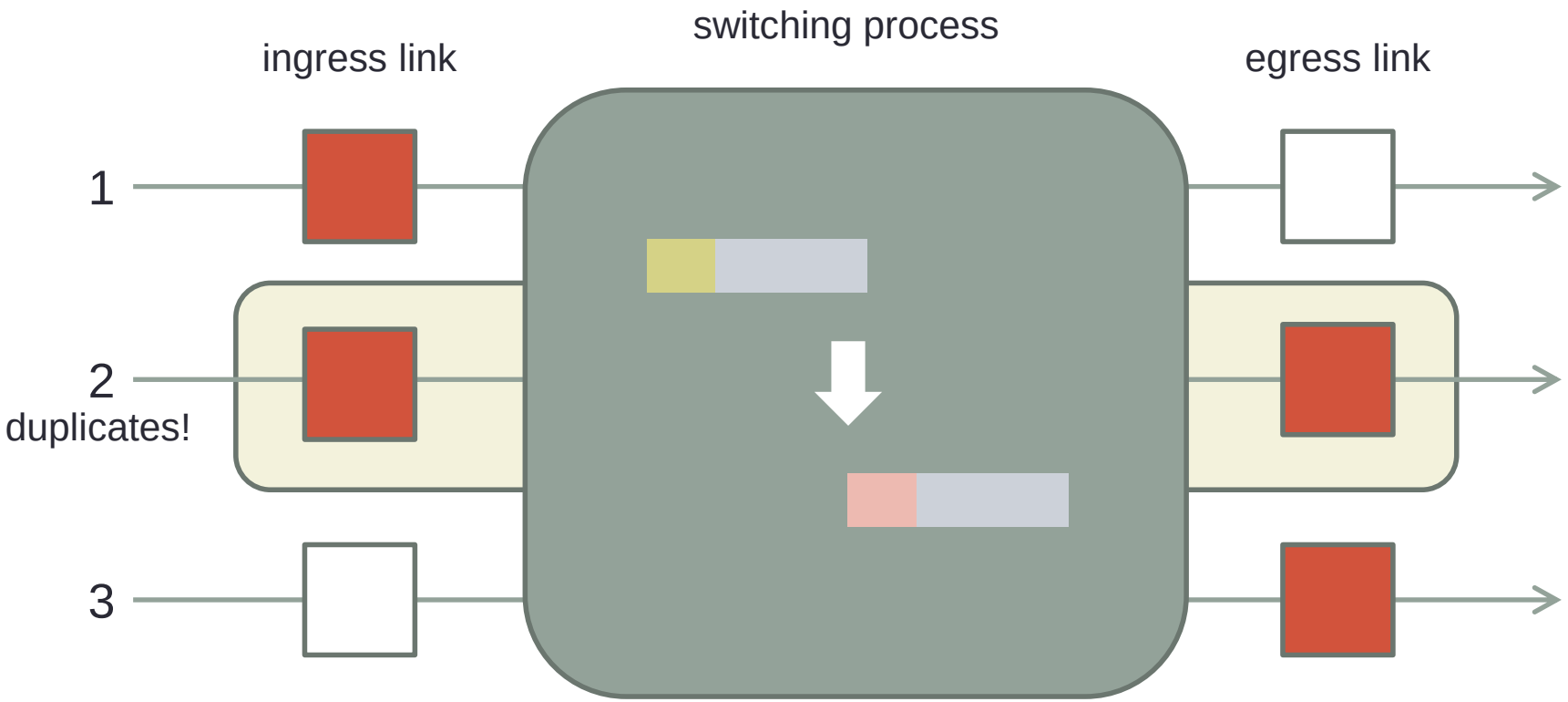


2. Theoretical analysis

monitored

not monitored

- Referred but not limited to a switched Ethernet environment
- IPv4 as layer 3 (IPv6 case is analogous)
- Packet traversing a monitored device, 3 possibilities…

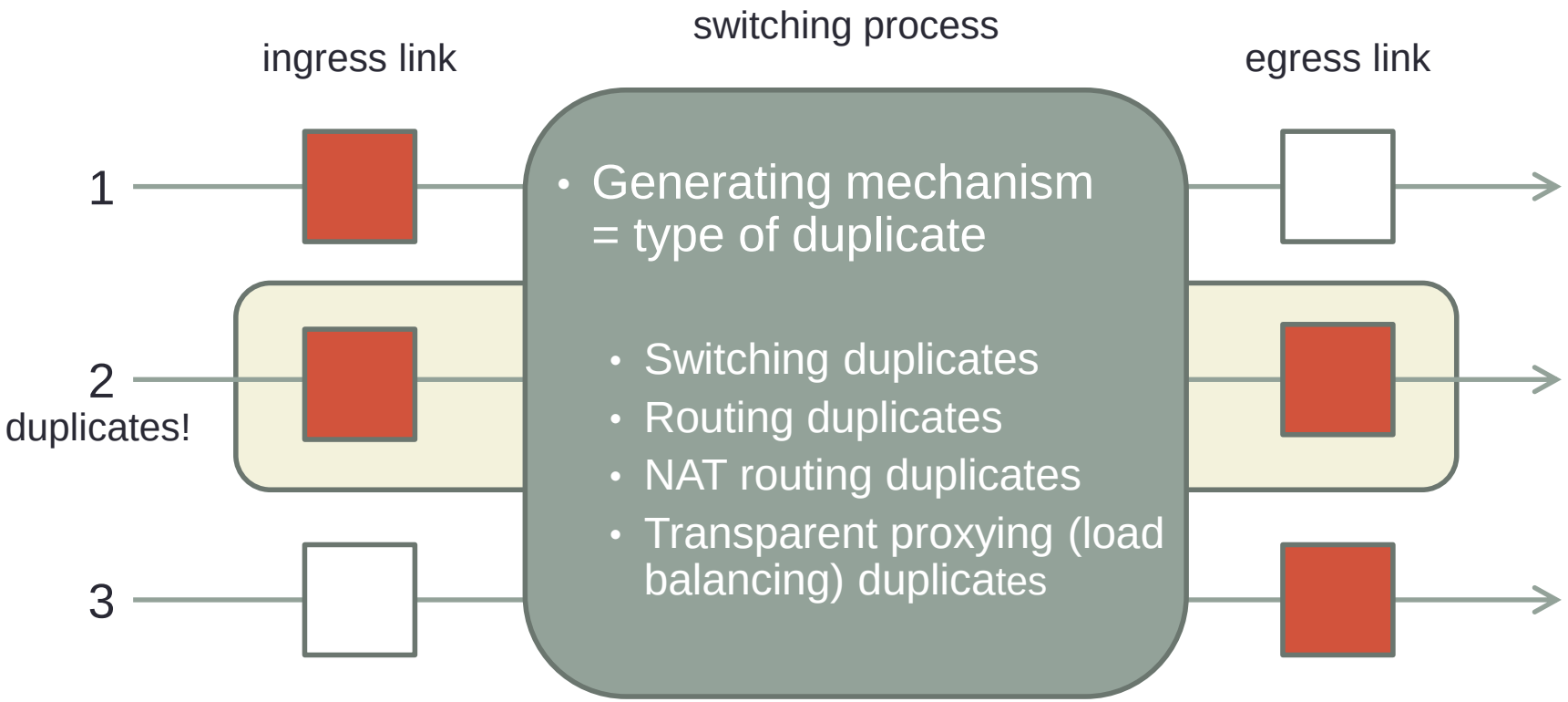


2. Theoretical analysis

monitored

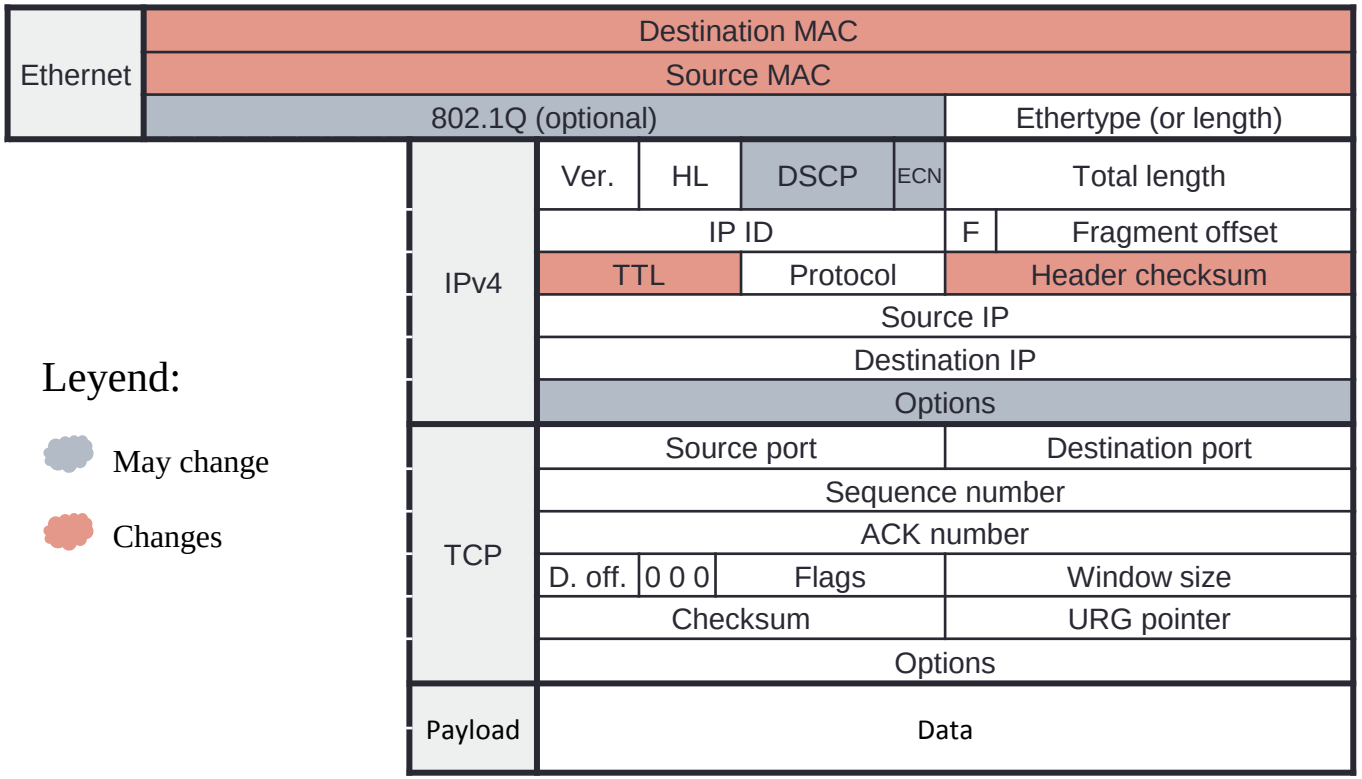
not monitored

- Referred but not limited to a switched Ethernet environment
- IPv4 as layer 3 (IPv6 case is analogous)
- Packet traversing a monitored device, 3 possibilities…



2. Theoretical analysis

- Example: **routing duplicates** over IPv4 and TCP



3. Duplicate detection methodology

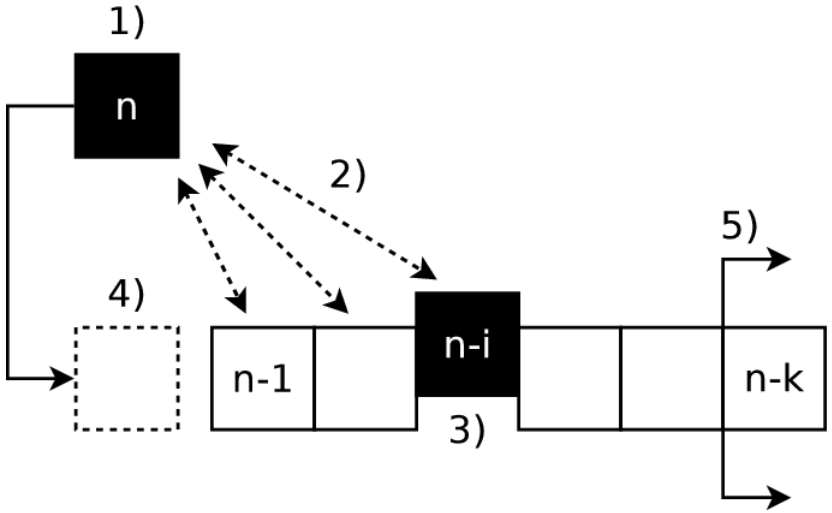
3. Duplicate detection methodology

- To compare only the payloads is not an option
  - There will be many packets without data
  - The type of duplicate is a valuable information

	Switching	Routing	NAT	Proxying
Utilization factor per VLAN				
Transport level statistics				
To study both sides of a NAT or proxy				

3. Duplicate detection methodology

- Intended to work offline on previously saved captures
- Sliding window
- Packet comparison
  1. Highest layer payload
  2. Fields that do not change
     - All of them must be compared
  3. Fields that change
     - TTL and checksums are not compared
     - Source and destination MACs must be compared to ensure that they change
  4. Fields that may change
     - Trunking encapsulation, DSCP value and options are not compared
     - The pairs src/dst IPs (NAT, proxy), src/dst ports (NAT) and TCP sequence/ACK (proxy) must be compared to ensure that only one changes



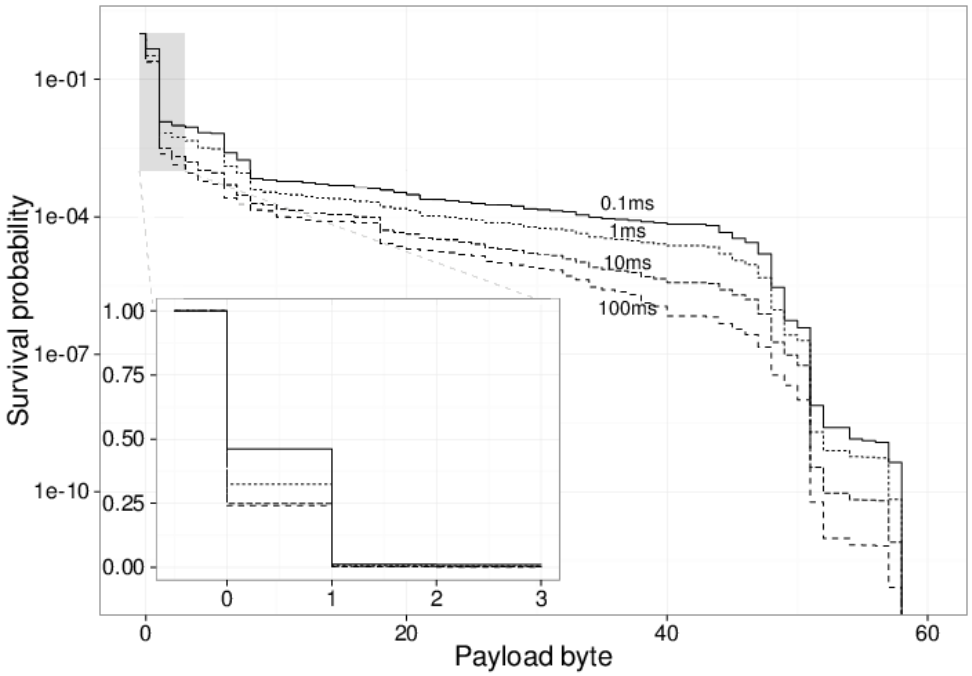
- Implementation available at Github: https://github.com/Enchufa2/nantools

4. Efficiency aspects

Single comparison | Number of comparisons

4. Efficiency aspects

- Efficiency of a single comparison
  - Resolve a non-duplicate pair using the smallest possible number of fields
  - The payload constitutes the most significant difference



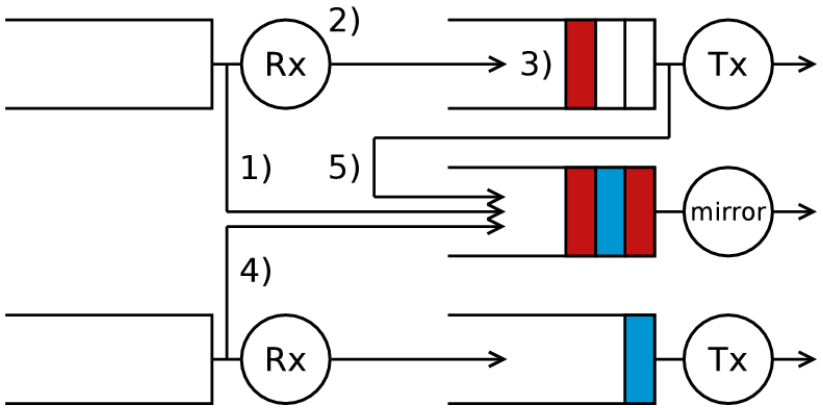
- Experiment with a deduplicated trace of real Internet traffic
- Comparisons over a sliding window, 4 window sizes
- Number of bytes compared until the mismatch was found
- More than the 99 % falls within the first byte

4. Efficiency aspects

- Reducing the total number of comparisons
  - Duplicates are expected to be close
  - Using the smallest possible window is desirable (without losing duplicates)
  - Enclosing the distance between duplicates…
  - **Window size in terms of time or number of packets?**

4. Efficiency aspects

- Model
  1. Ingress copy
  2. Switching time
  3. Queueing time (Tx)
  4. Other packets
  5. Egress copy



1 2 3 4

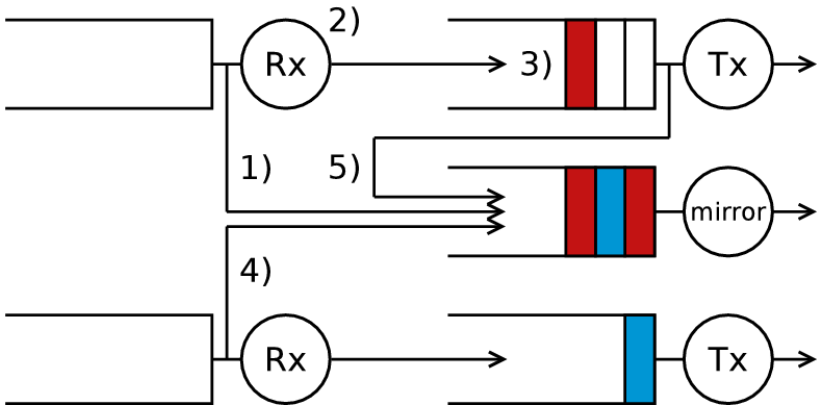
- Time between copies: $\Delta t_n = -w'_n + x_n + w_n + w''_n = s_n + (w''_n - w'_n)$

$$\overline{\Delta t} = \bar{x} + \bar{w} = \bar{s}$$

- Packets between copies: $\overline{\Delta n} = \sum \mu_i \bar{s}$

4. Efficiency aspects

- Model
  1. Ingress copy
  2. Switching time
  3. Queueing time (Tx)
  4. Other packets
  5. Egress copy



1 2 3 4

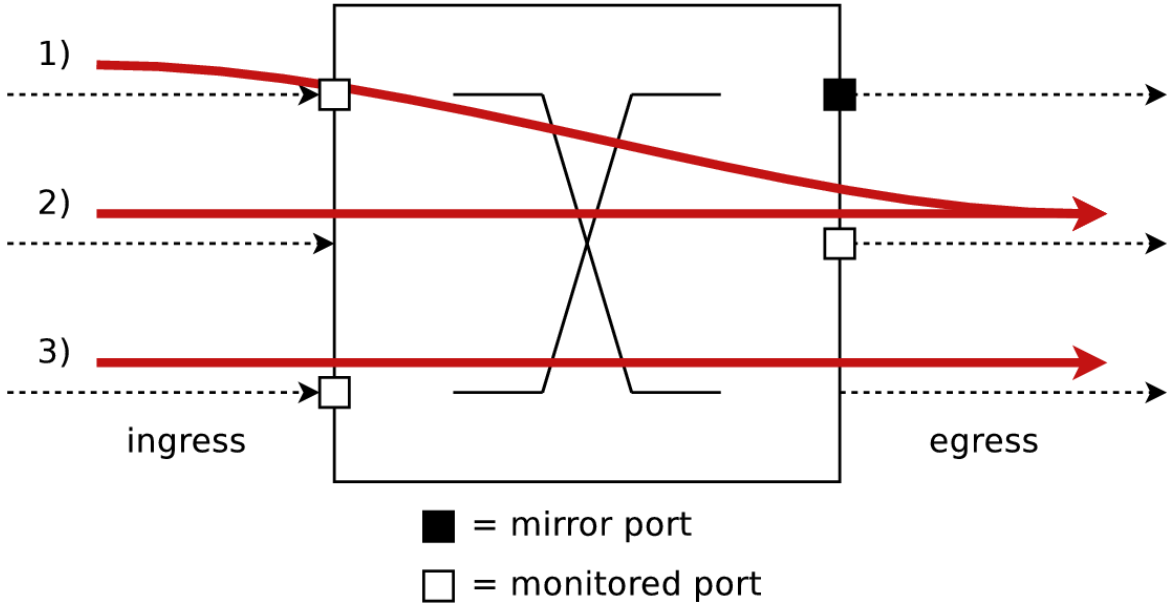
- Time between copies: $\Delta t_n = -w'_n + x_n + w_n + w''_n = s_n + (w''_n - w'_n)$

$\overline{\Delta t} = \bar{x} + \bar{w} = \bar{s}$ → system time, M/D/1

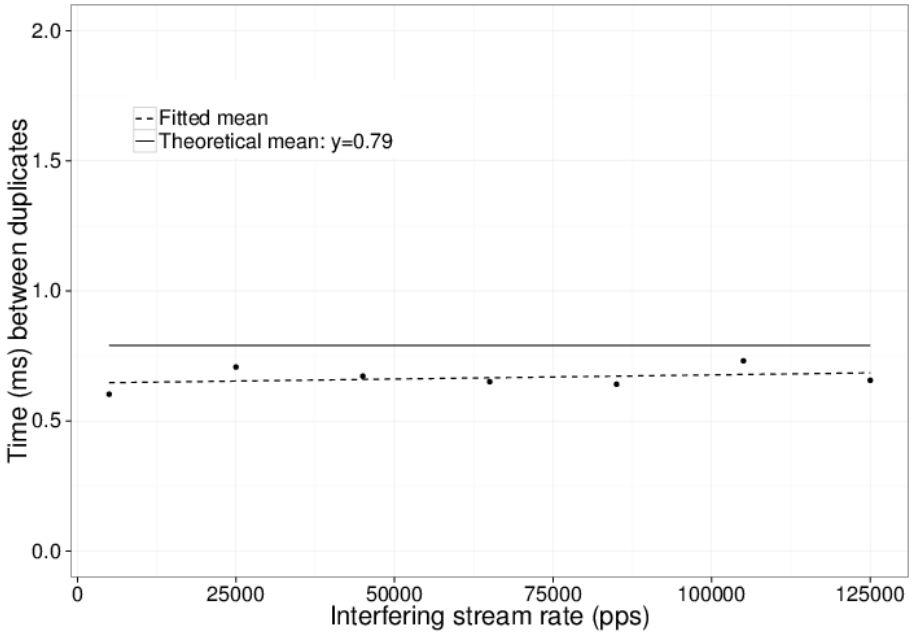
- Packets between copies: $\overline{\Delta n} = \sum \mu_i \bar{s}$

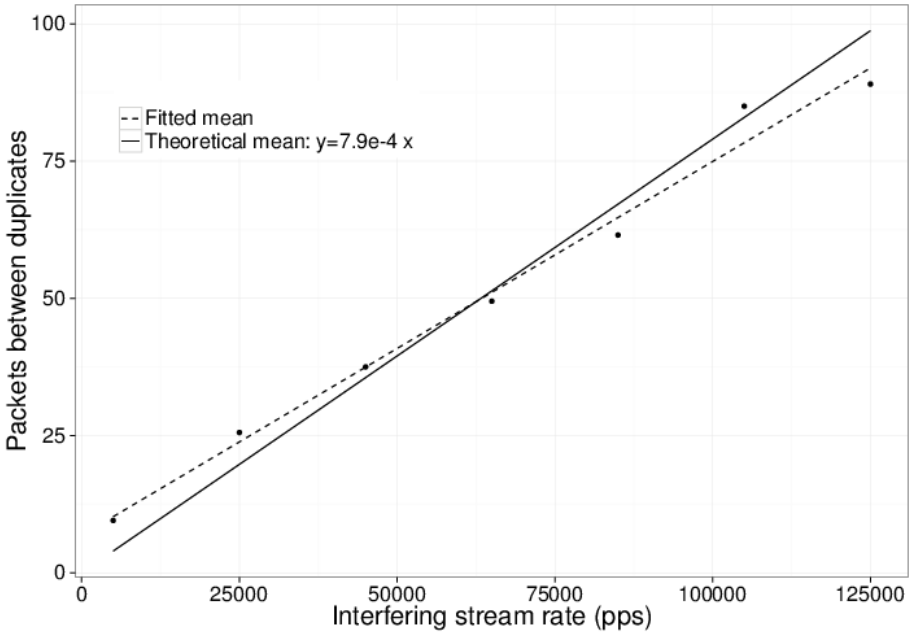
4. Efficiency aspects

- Scenario
  1. Main stream generates duplicates
  2. Auxiliary stream forces queueing
  3. Interfering stream inserts packets between duplicates at different rates



4. Efficiency aspects





4. Efficiency aspects

- A time-based sliding window is the best option
- 3 contributions to the time difference:
  - Main contribution: queueing time at the transmission port (w_n)
  - Switching time (x_n) is negligible as compared to the queueing time
  - Queueing time at the mirror port ($w''_n - w'_n$) is zero on average
- Upper bound in terms of time as a dimensioning rule

$$WindowSize = \frac{3 \cdot \max(N_q) \cdot \max(M)}{\min(C)}$$

  - $\max(N_q)$ maximum length of the largest queue
  - $\max(M)$ maximum packet length
  - $\min(C)$ slowest link capacity

5. Conclusions

5. Conclusions

- This paper addresses an important and unattended problem
- The theoretical background has been exposed
  - Generating mechanisms / types of duplicates
- A duplicate detection methodology is proposed
- Efficiency aspects have been discussed analytically and experimentally
- Further research with other equipment is needed in order to refine these results

Duplicate detection methodology for IP network traffic analysis

Iñaki Ucar[1], Daniel Morato[2], Eduardo Magaña[2], Mikel Izal[2]

inaki.ucar@uc3m.es, {daniel.morato|eduardo.magana|mikel.izal}@unavarra.es

[1]Department of Telematic Engineering
University Carlos III of Madrid

[2]Department of Automatics and Computer Science
Public University of Navarre

upna Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa