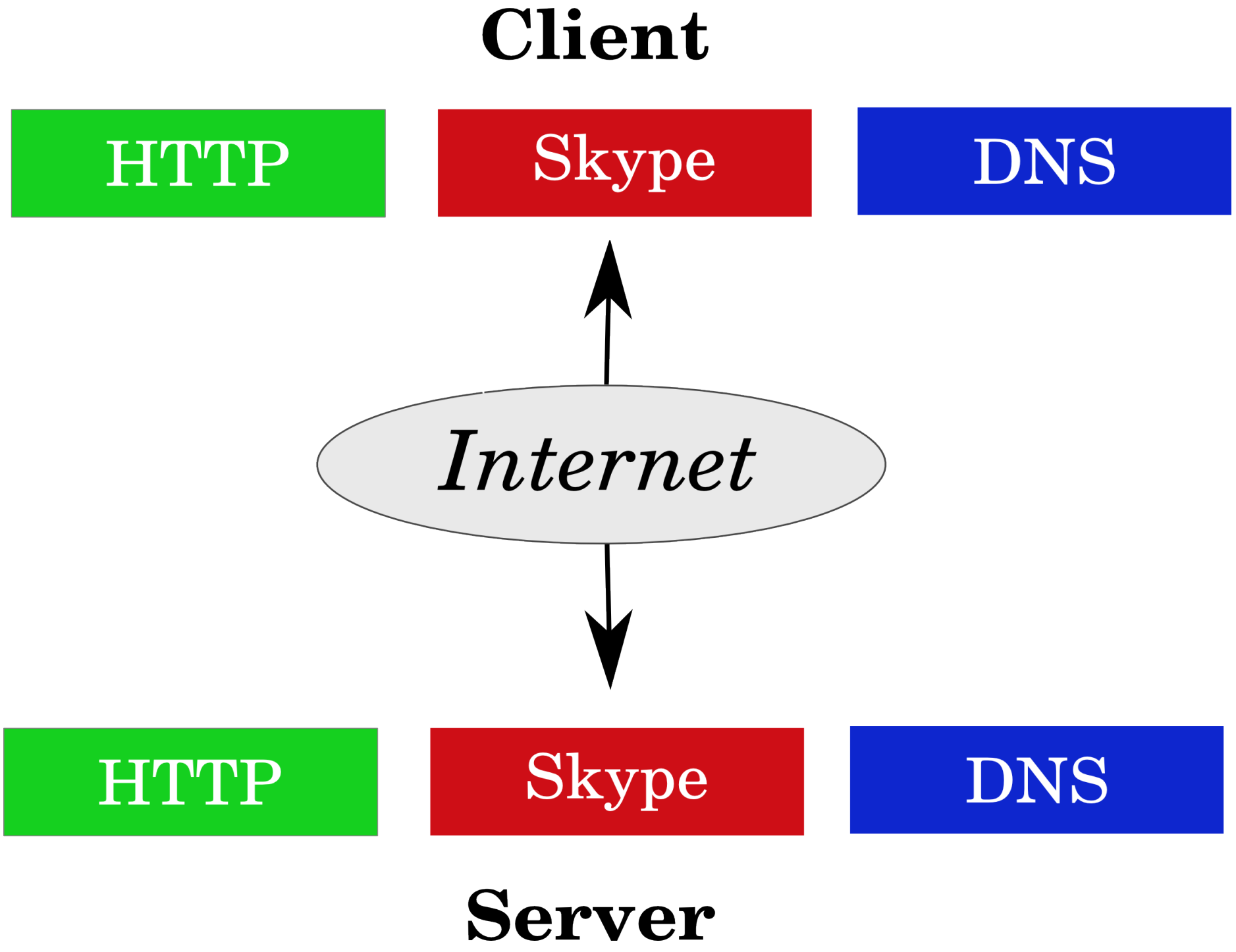
Statistical, real-time classification of IP traffic in Linux operating system

Paweł Foremski
Advisor: dr inż. Arkadiusz Biernacki

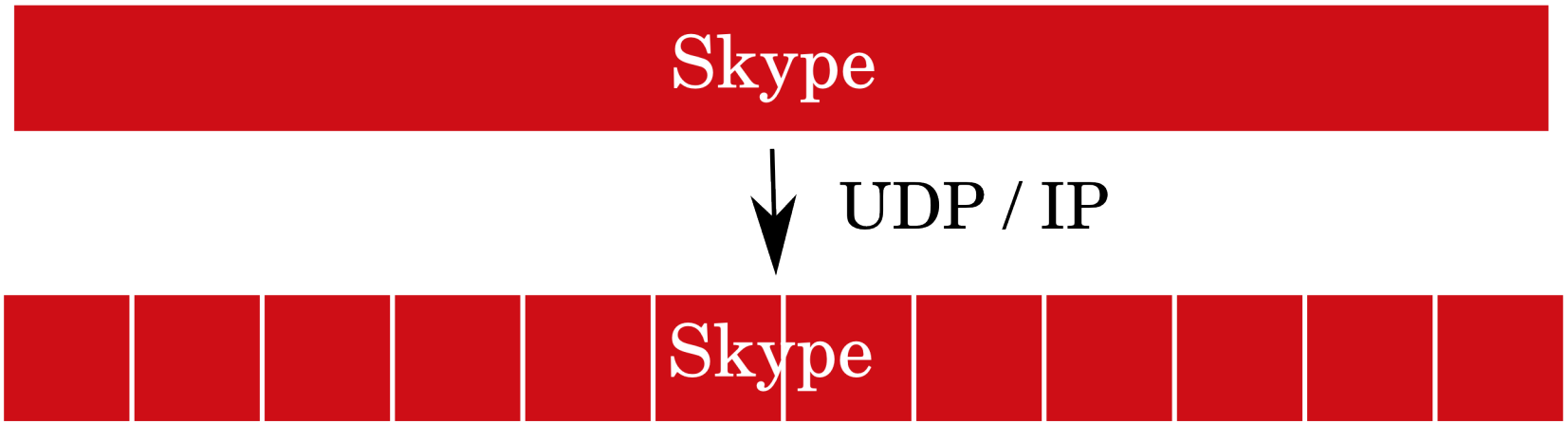
Agenda

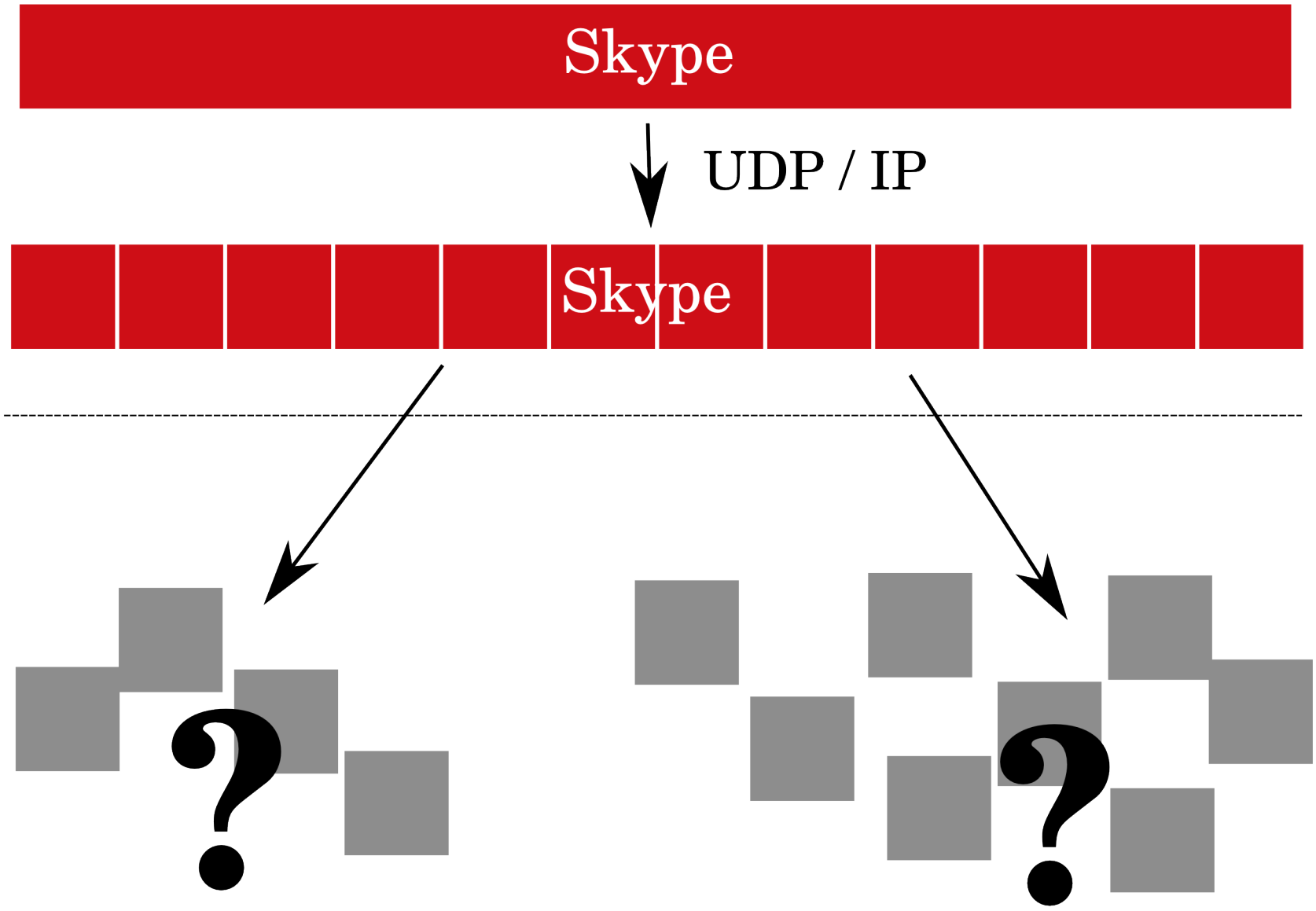
- IP traffic classification
- Statistical approach
- The KISS algorithm
- My work

IP traffic classification

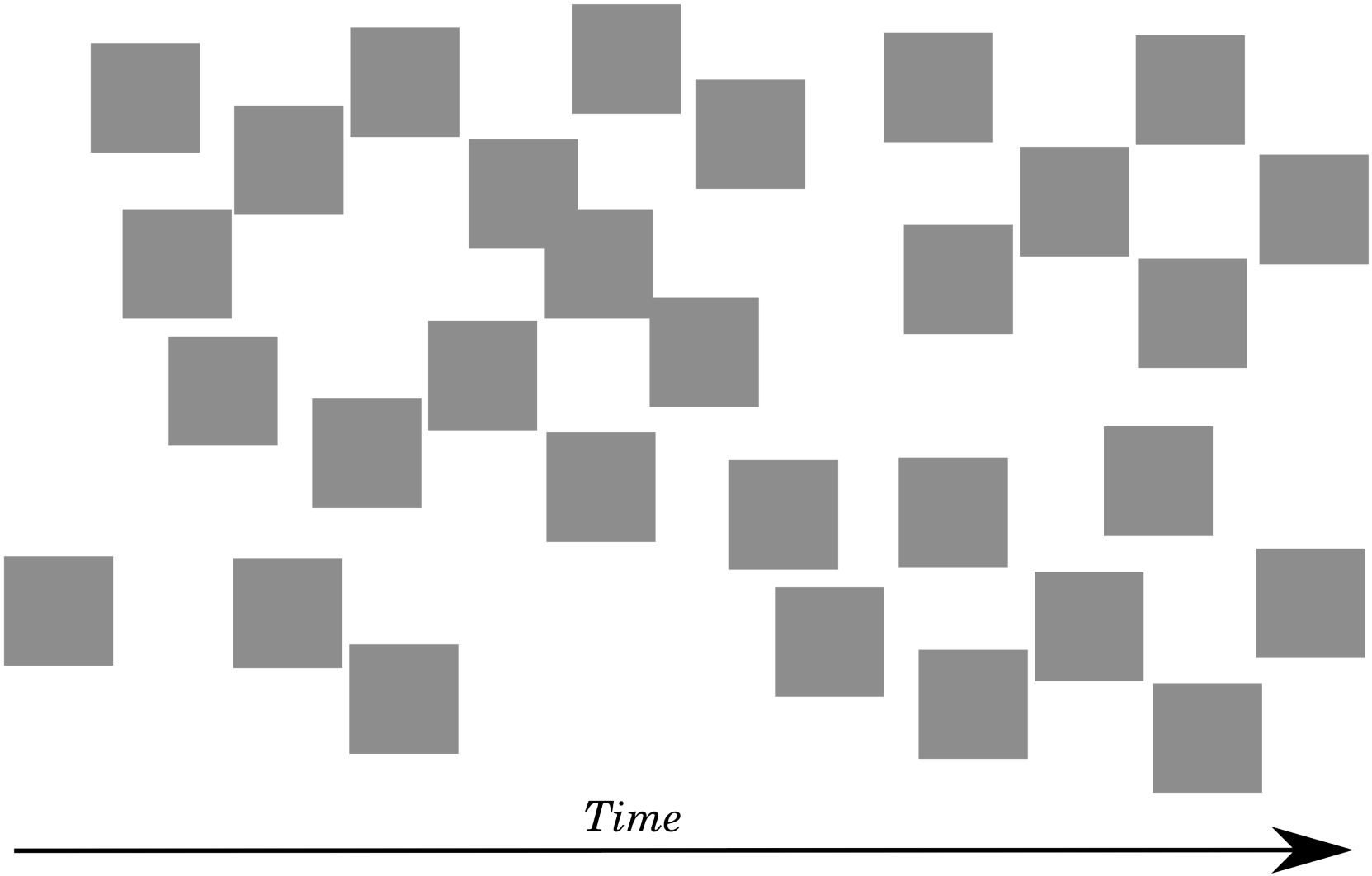

Client
HTTP
Skype
DNS
Internet
HTTP
Skype
DNS
Server

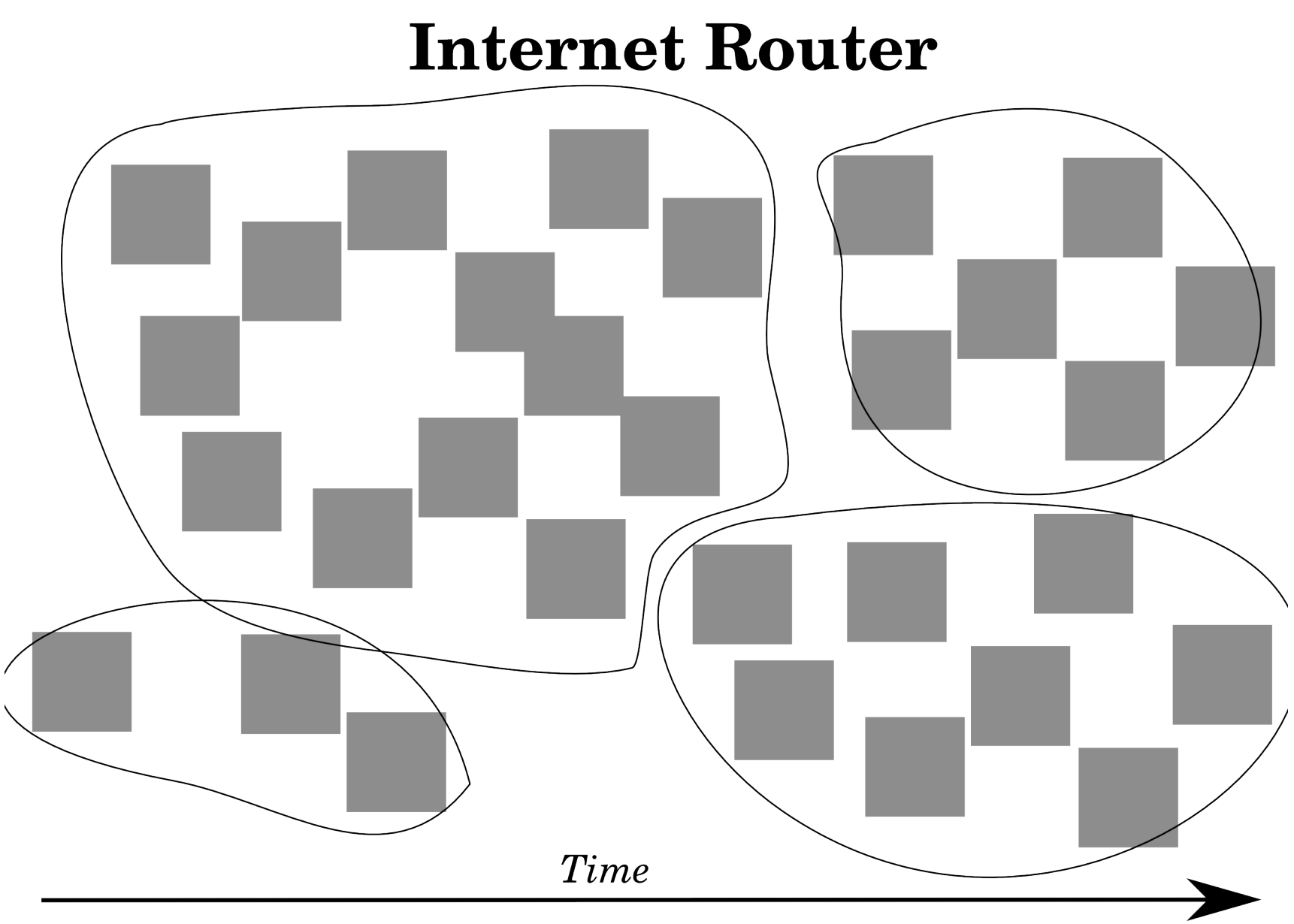
Skype

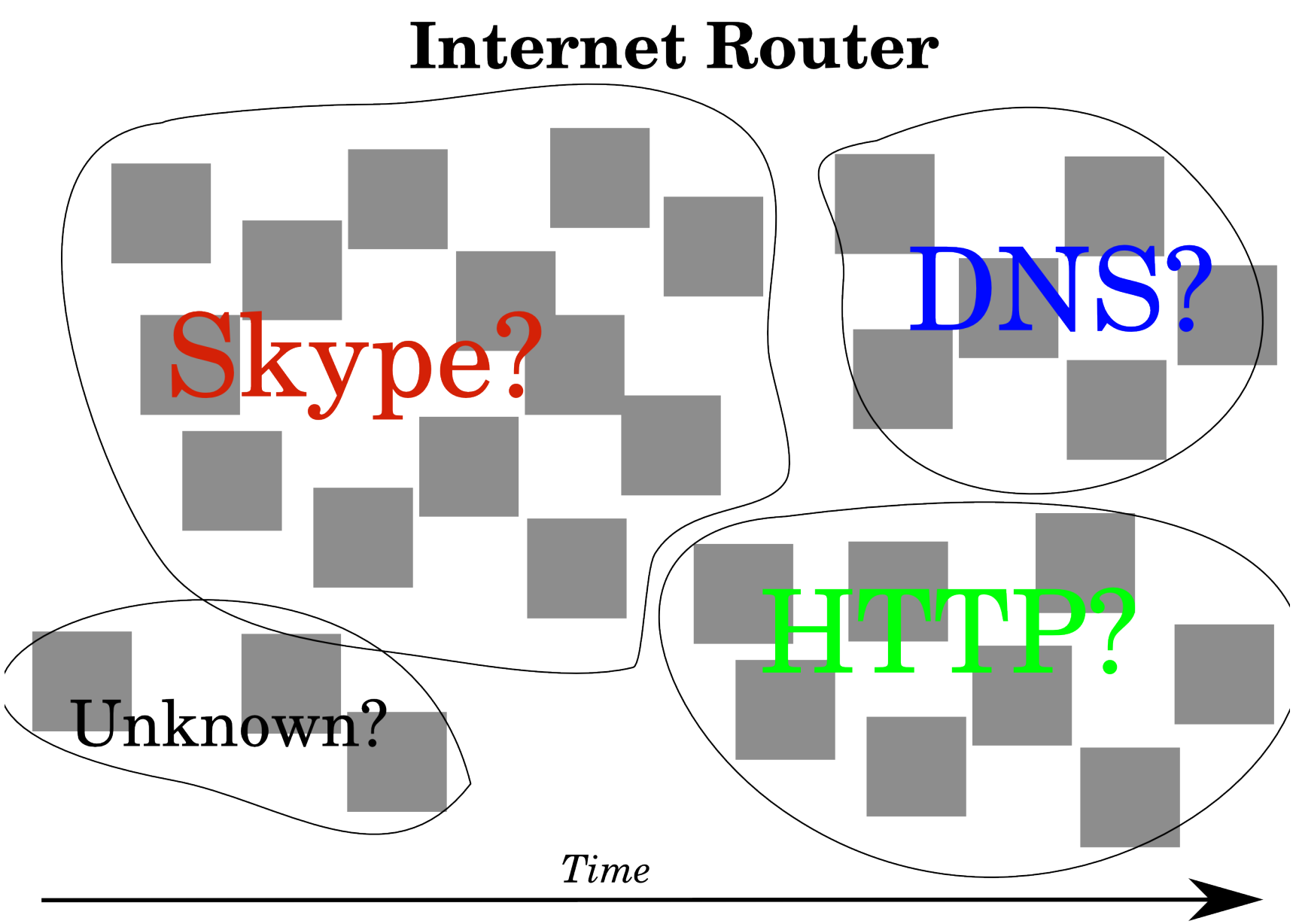

Skype
UDP / IP
Skype


Skype
UDP / IP
Skype
?
?

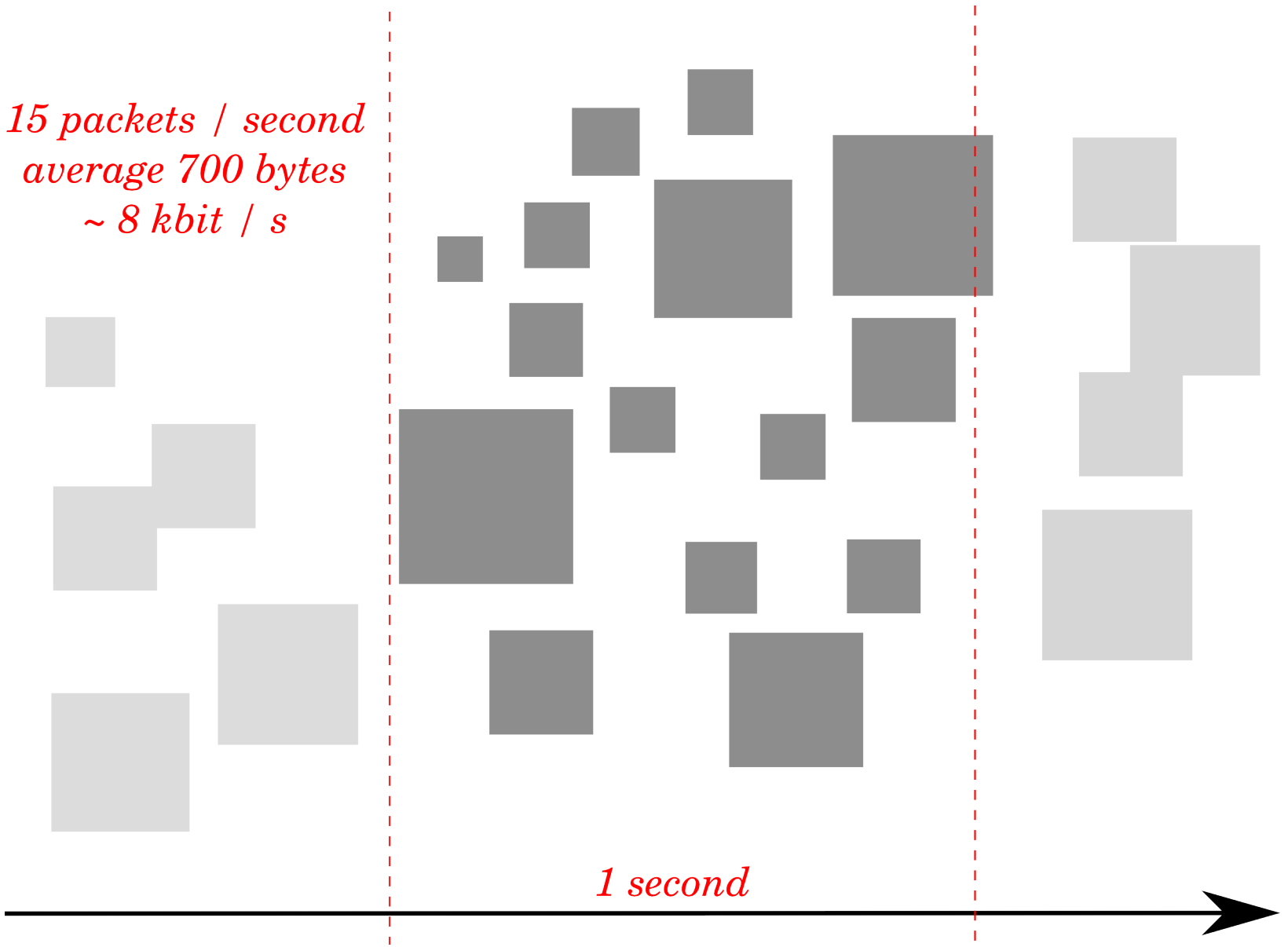
Internet Router

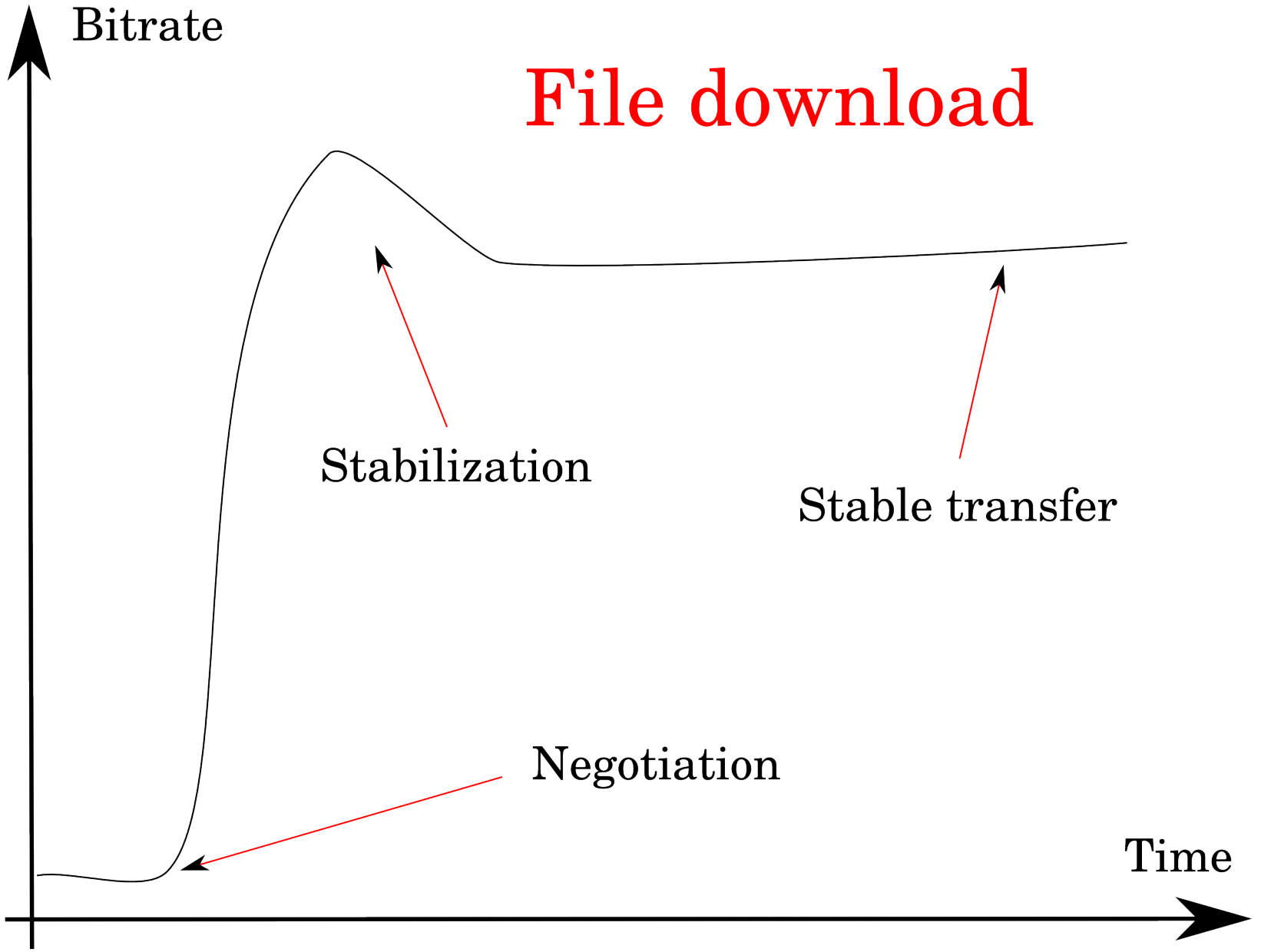


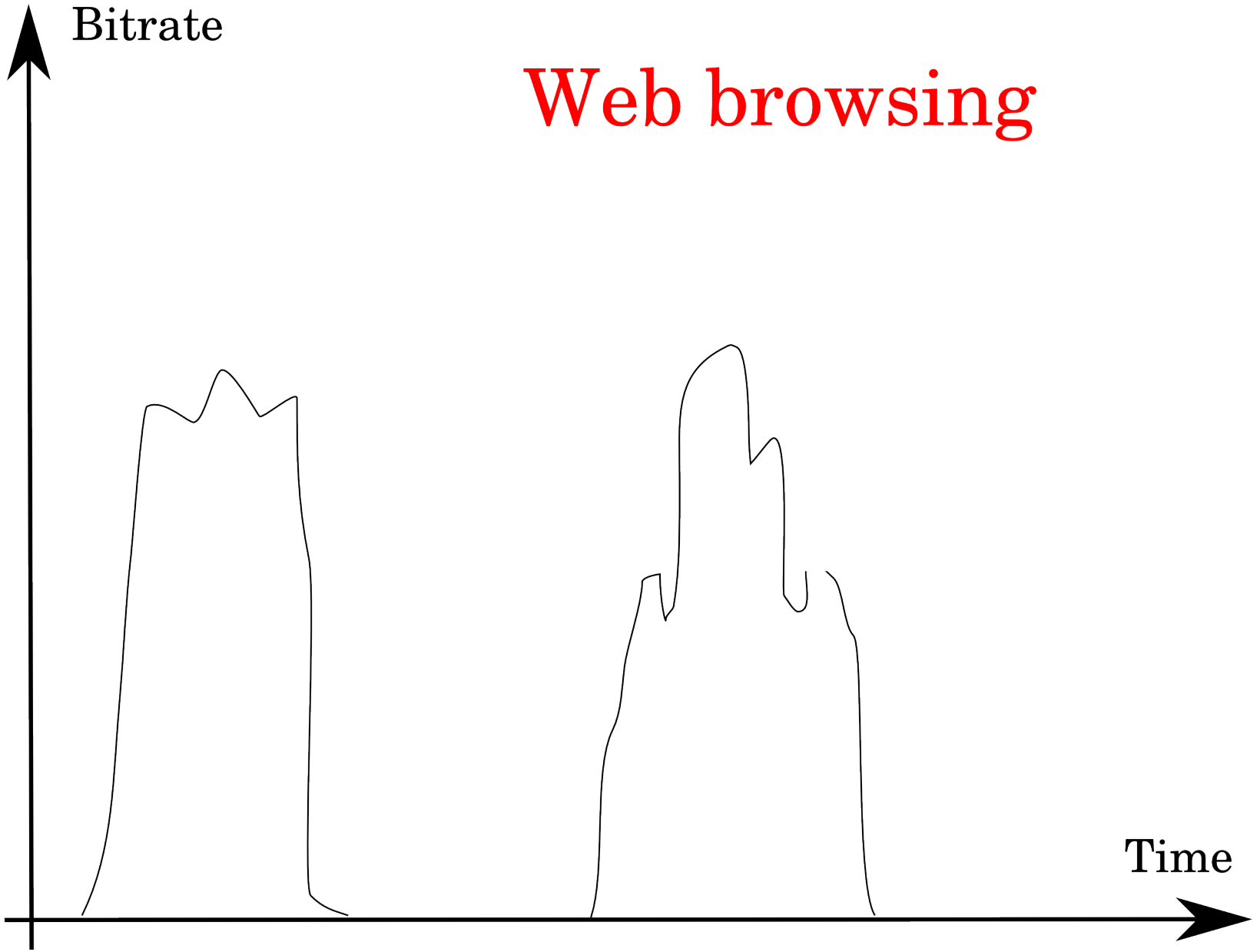

Internet Router
Time

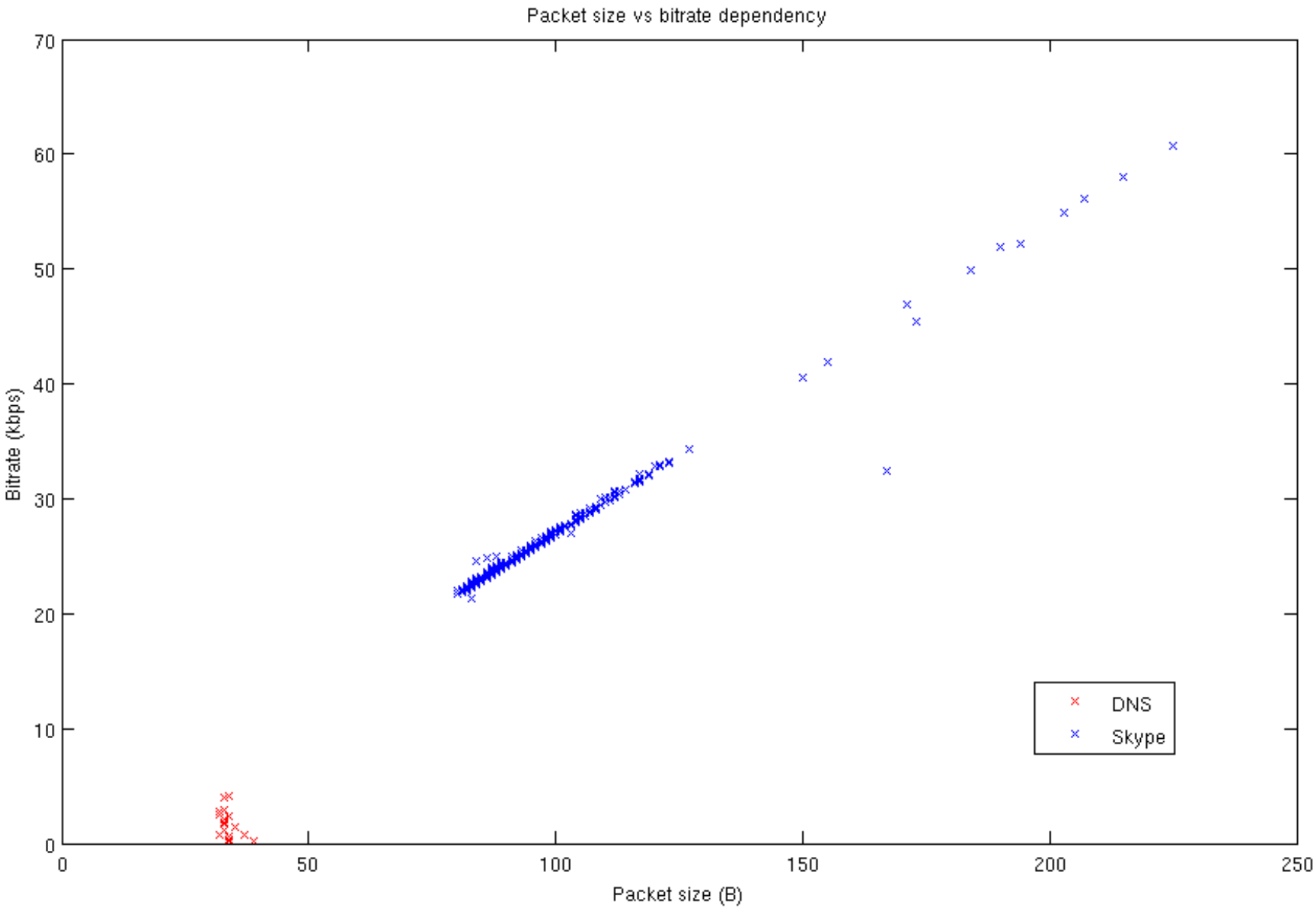

Internet Router
Skype?
DNS?
HTTP?
Unknown?
Time

Statistical approach


15 packets / second
average 700 bytes
~ 8 kbit / s
1 second


Bitrate
File download
Stabilization
Stable transfer
Negotiation
Time


Bitrate
Web browsing
Time


Packet size vs bitrate dependency
Bitrate (kbps)
Packet size (B)
0
10
20
30
40
50
60
70
0
50
100
150
200
250
DNS
Skype

Existing statistical approaches

- Flow-level analysis
  - Language detection in Skype
- Behavioral (e.g. P2P)
- Packet-level inspection

The KISS algorithm

KISS algorithm

- KISS: **Chi-S**quare **S**ignatures
- Idea: each UDP packet must have an additional application-level header
  - Counters
  - Constants
  - Random numbers

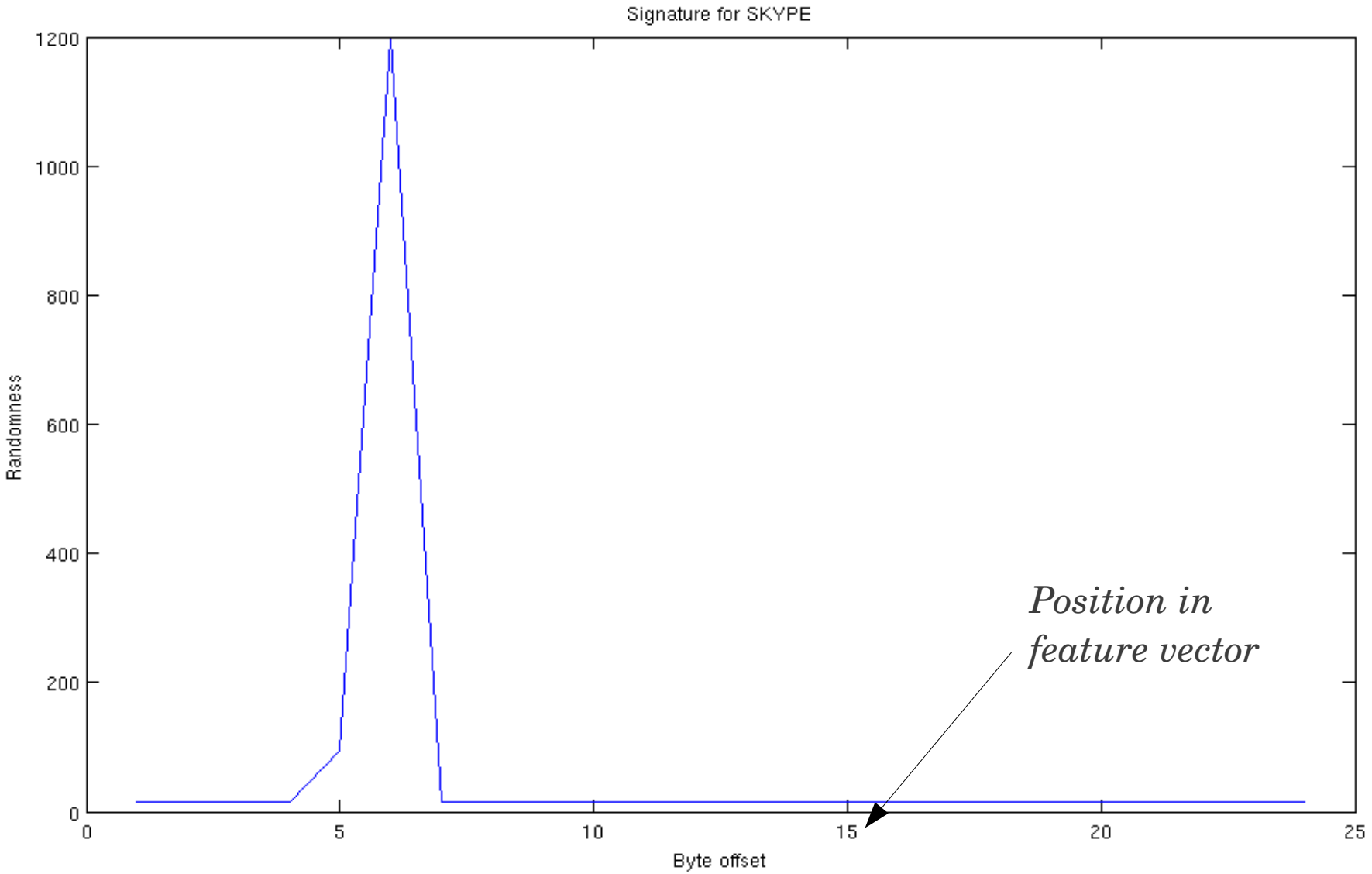
KISS algorithm

- First 12 bytes as protocol header
- Measure randomness on each byte position $p = 1...12$

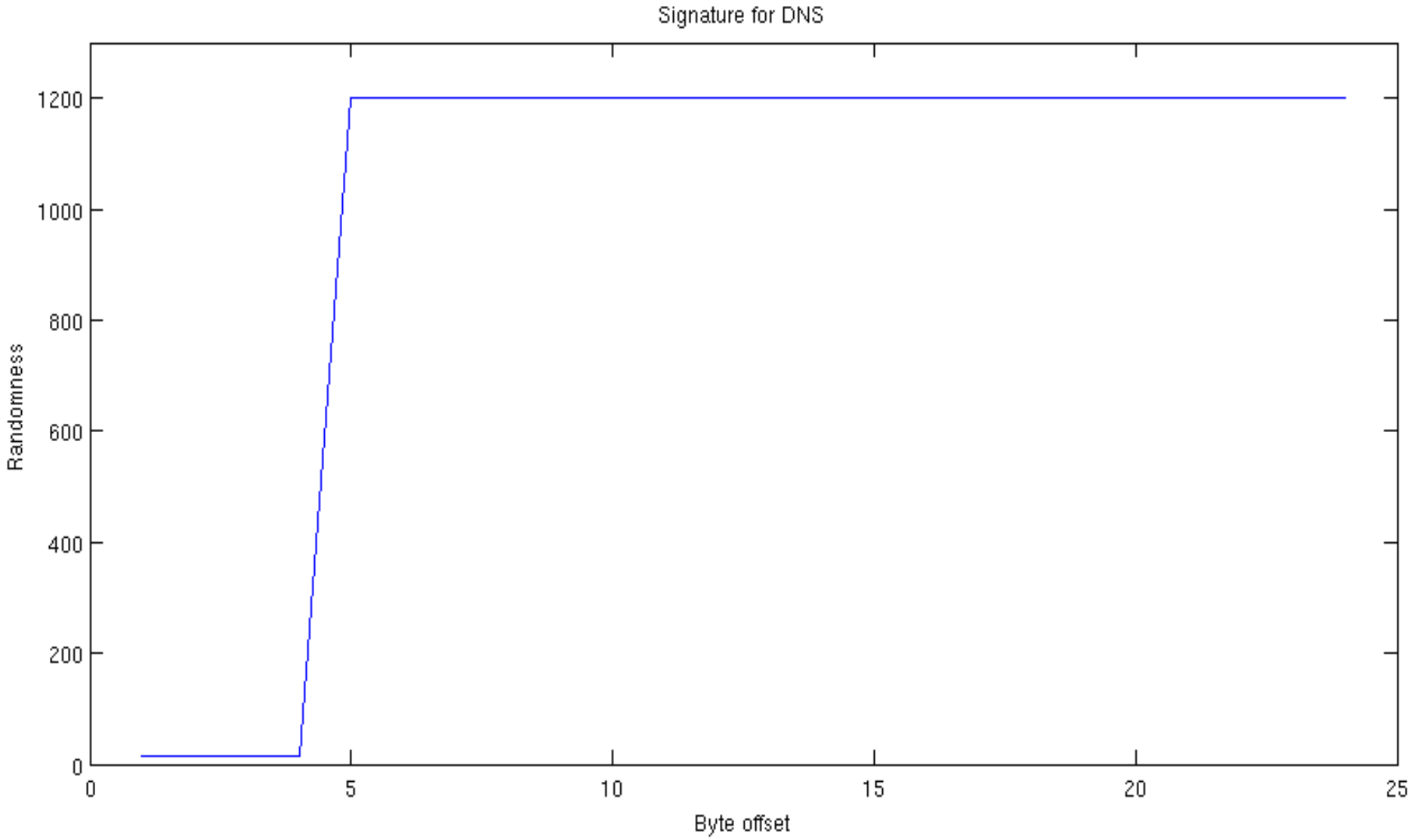
$$X_p = \sum_{i=0}^{2^b - 1} \frac{(O_i - E)^2}{E}$$

- Result: feature vector – a protocol “fingerprint”

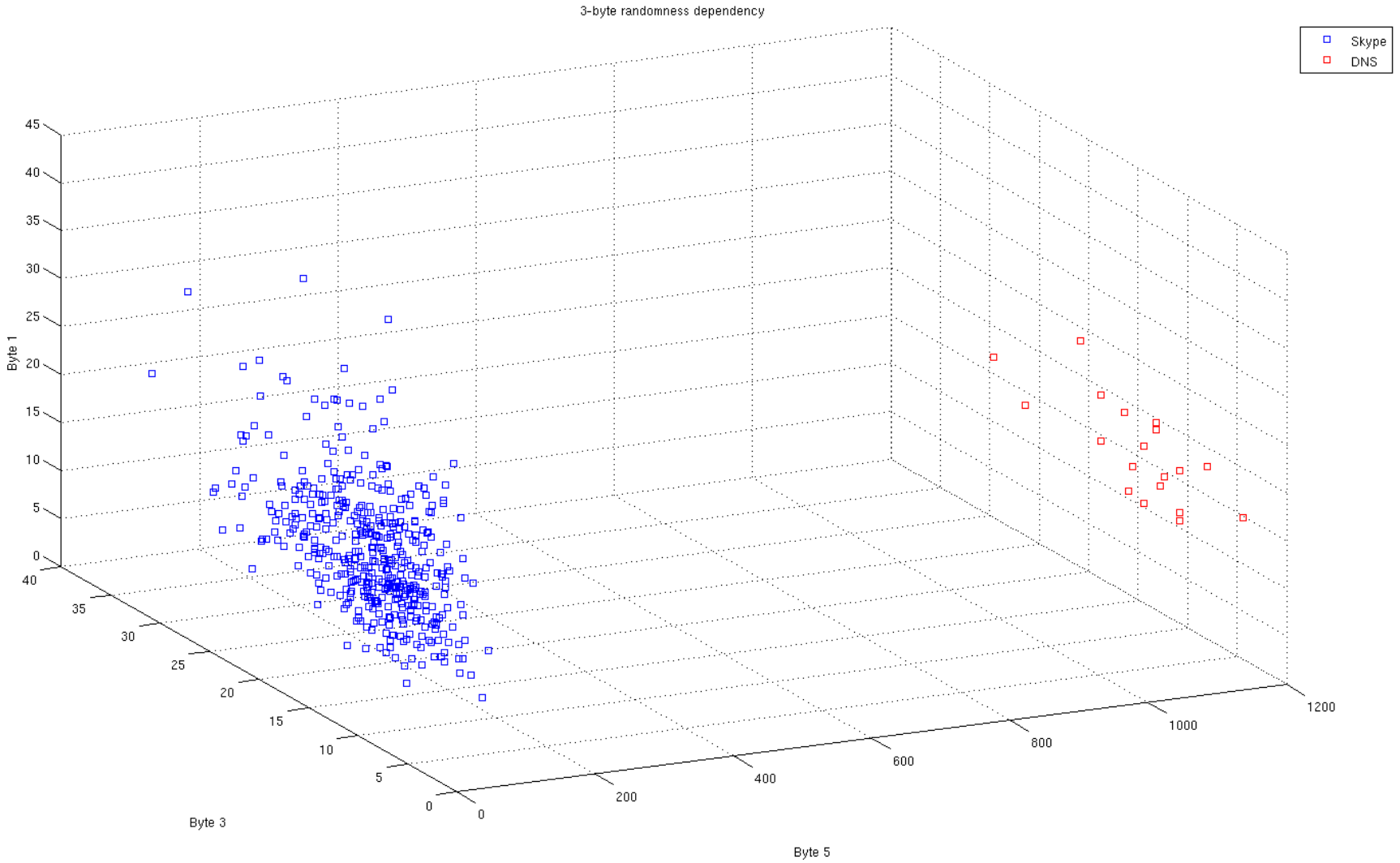
Example: Skype



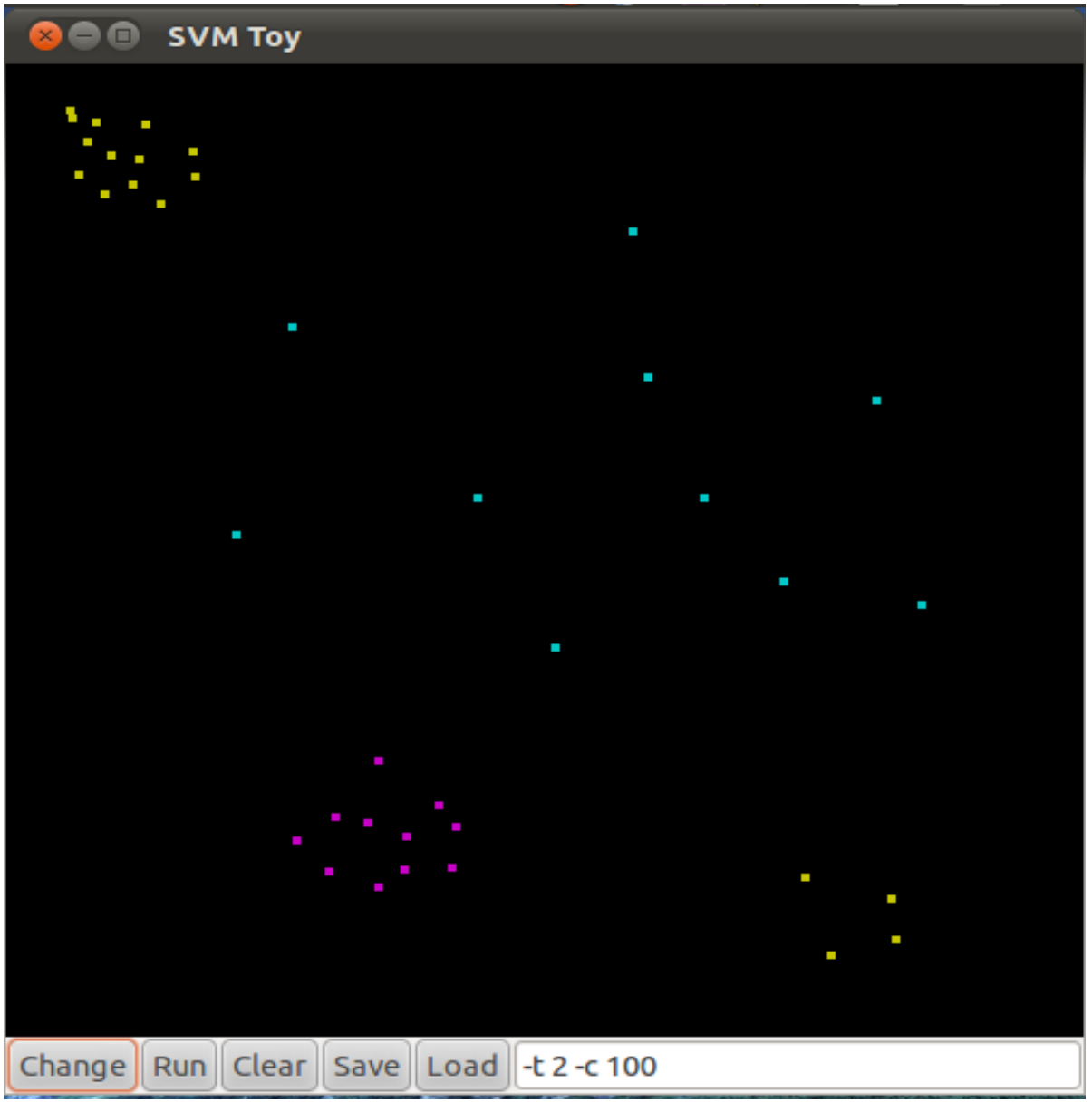
Example: DNS


Signature for DNS
Randomness
Byte offset
0
200
400
600
800
1000
1200
0
5
10
15
20
25

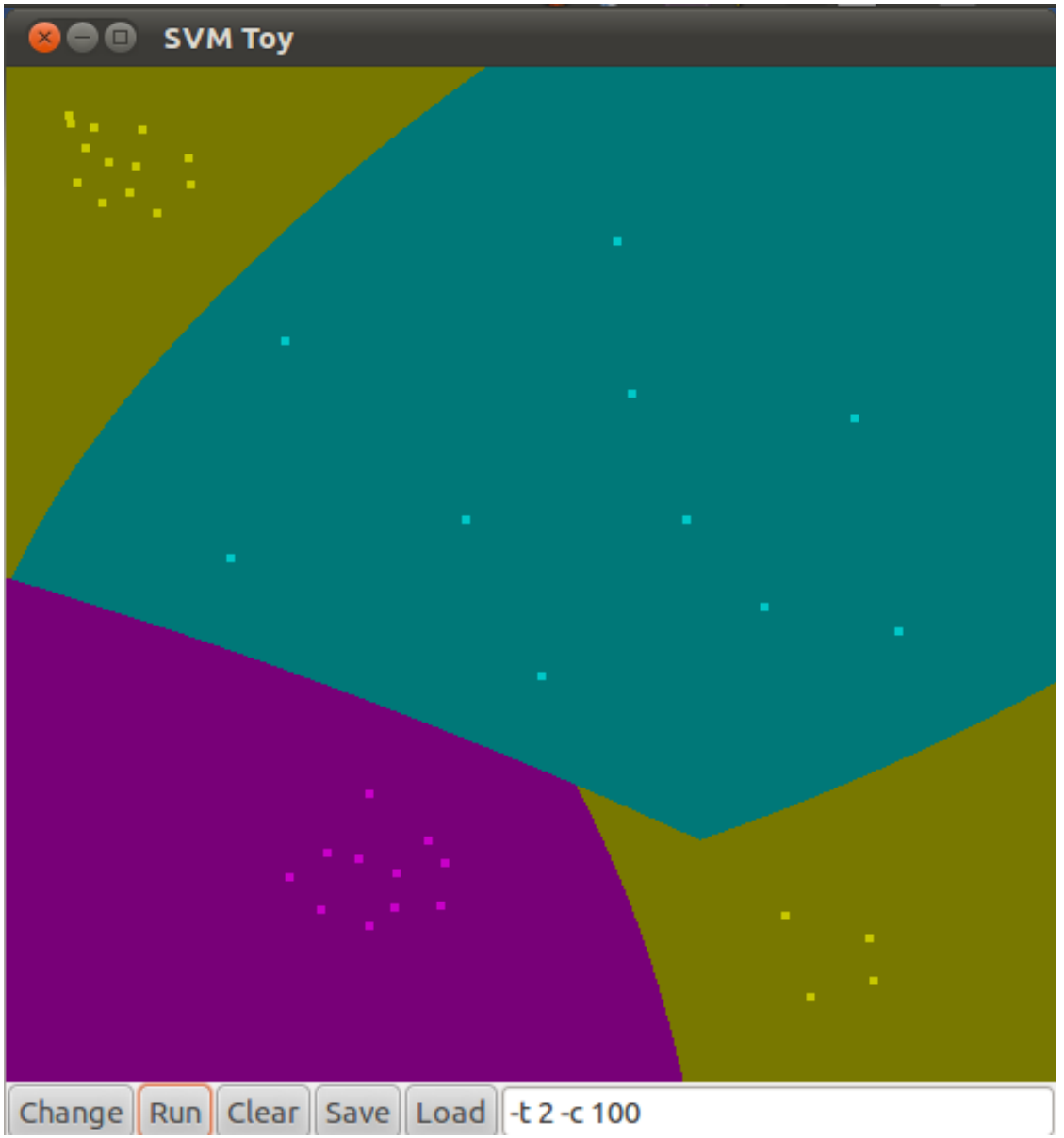
Cross-byte dependency



Support Vector Machines



Support Vector Machines



My work

MATLAB prototype

- ~1k lines of source code
- Packet sniffer
- Grouping in flows
- Computation of KISS signatures
  - Flow-level additions
- Preliminary SVM classification
  - libsvm

Current objectives

- Representative traffic samples
- Evaluation of flow-level extensions in feature vector
- Tuning of SVM classification

Target

- Implementation in C
- Real-time
- Work as live packet sniffer
- Work as Linux firewall
  - netfilter NFQUEUE target

Questions?

Statistical, real-time classification of IP traffic in Linux operating system

Paweł Foremski
Advisor: dr inż. Arkadiusz Biernacki