Network Traffic Classification Based on Flow Characteristics

Pavel Piskač

piskac@mail.muni.cz

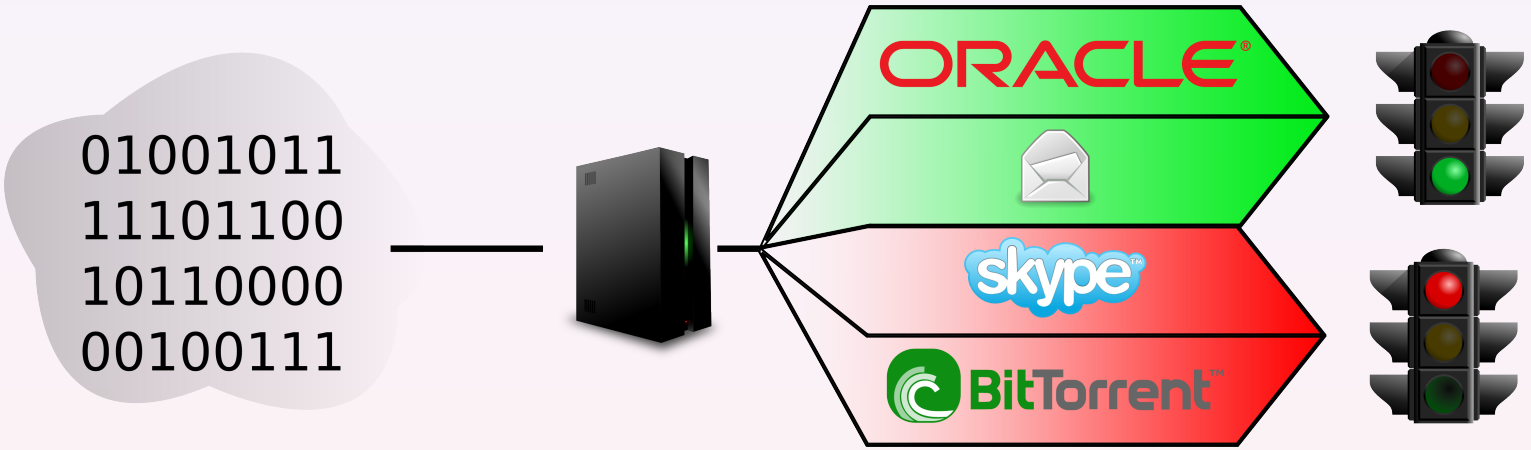
DTEDI Presentation, November 14, 2011

Part I

Introduction

Usage of Protocol Detection

- Protocol detection is used in many situations
  - Suspicious activity detection



Protocol Detection Methods

	Port numbers	DPI	Behavior	Flows
Fast	✓	✗	✓	✓
Precise	✗	✓	✓	✓
Power consuming	✗	✓	✗	✗
Encryption	✓	✗	✓	✓
Groups	✗	✗	✓	✓ / ✗
Probability	✗	✗	✓	✓

Work Goals

	Port numbers	DPI	Behavior	Flows
Fast	✓	✗	✓	✓
Precise	✗	✓	✓	✓
Power consuming	✗	✓	✗	✗
Encryption	✓	✗	✓	✓
Groups	✗	✗	✓	✓ / ✗
Probability	✗	✗	✓	✓

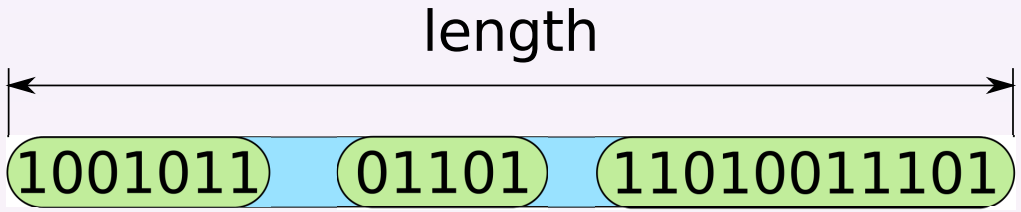
Work Goals

- Achieved in the following steps
  - One protocol detection
    - Explore one protocol
    - Detect selected protocol
    - Implement detection method
  - General protocol detection
    - Take advantage from previous research
    - Find and test clustering algorithm
    - Implement detection method

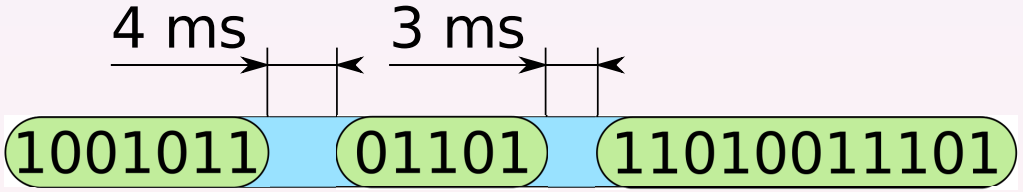
Flow Statistics

- An ordinary flow with packets and inter-packet gaps

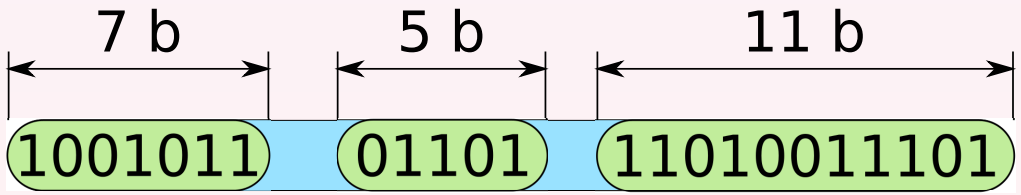
- Statistics consist of
  - Flow length



  - Information about inter-packet gap sizes



  - Infomation about packet sizes



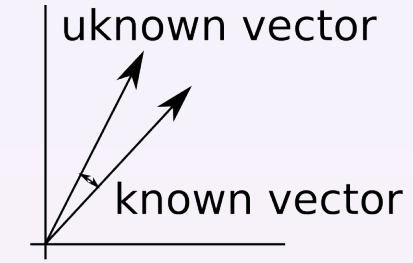
- Basic expectation: **statistics are application dependent**

Part II

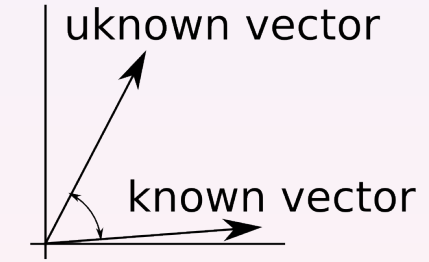
One Protocol Detection

Detection Methods

- Based on vector comparison
- Only time characteristics
- Used methods
  - Average distance between vectors
  - Root-mean-square distance
  - Euclidean distance
  - Angle between vectors
- Decision according to threshold value



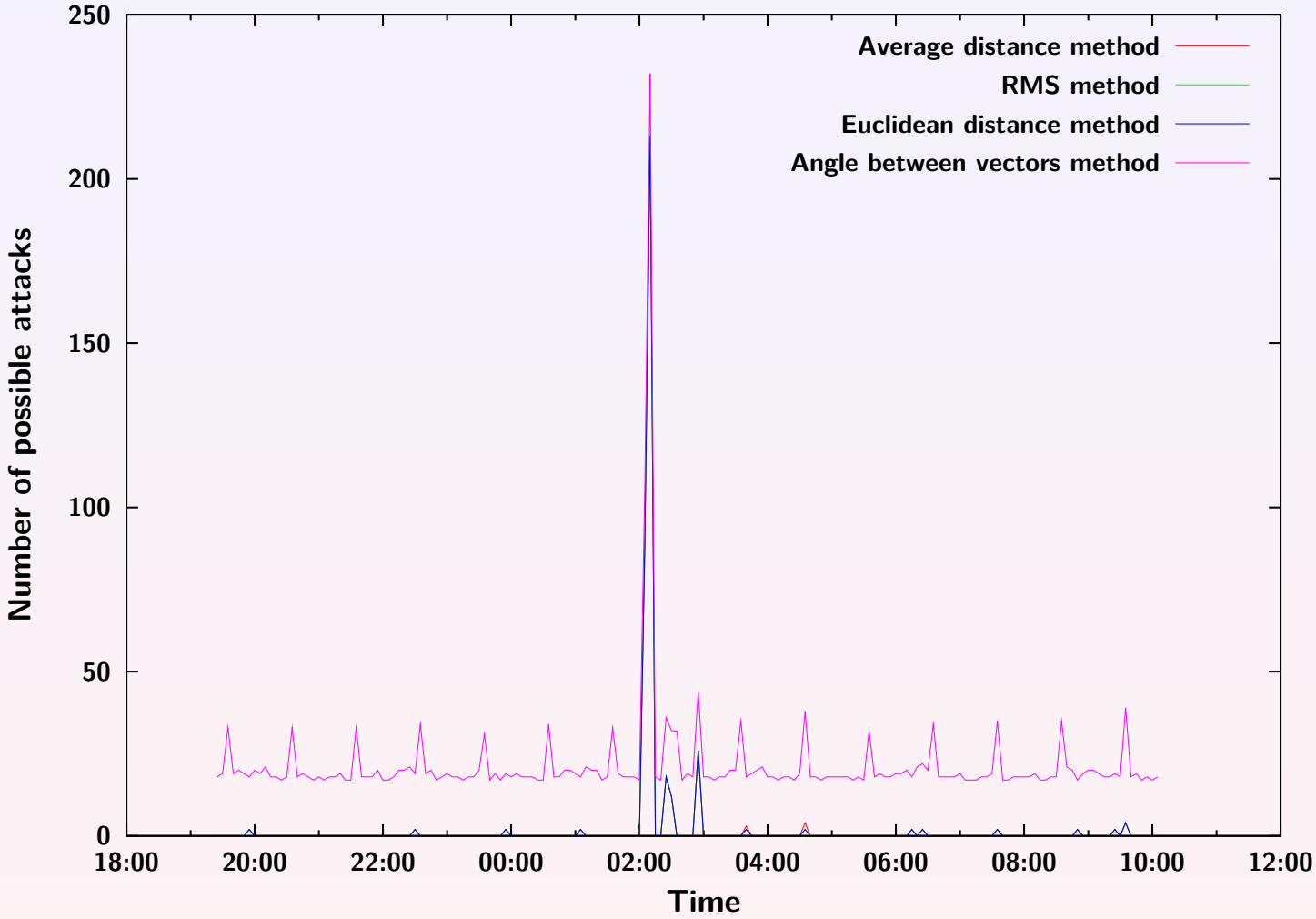
SIMILAR



DISSIMILAR

SSH Protocol Detection

- First step of our work
- Training and learning phase
- Based only on time characteristics
- Results:
  - \+ Dictionary attack detection
  - \+ Accuracy about 90 %
  - \+ Usefulness of time characteristics
  - \- Unable to detect all situations


Average distance method
RMS method
Euclidean distance method
Angle between vectors method
Number of possible attacks
250
200
150
100
50
0
18:00
20:00
22:00
00:00
02:00
04:00
06:00
08:00
10:00
12:00
Time

Part III

General Protocol Detection

Clustering Algorithms

- Automatized division into groups
- Based on vector comparison
- QT clustering algorithm
  - \+ First evaluation
  - \+ Nonrandom
  - \- Slow
- K-Means clustering algorithm
  - \+ Widespread
  - \+ Faster than QT
  - \- Random
  - \- Cannot detect number of clusters

Main Issues

- Minimal set of vector components
- Minimizing influences in time characteristics caused by network
- Finding **the core** of flows

Part IV

Future Work and Conclusion

Future Work

- Precise protocol detection
- Programmable hardware probes
- All data in IPFIX format

Conclusion

- Dictionary attacks on SSH protocol detection
- Minimizing influences in time characteristics cased by network
- The main issue **finding the core of flows**

Thank You For Your Attention!

Pavel Piskač
piskac@mail.muni.cz

Network Traffic Classification Based on Flow Characteristics