Skype Traffic Detection and Characterization

Master´s Thesis Presentation: 4.9.2007
Author: Andrea Buonerba
Supervisor: Professor Jorma Virtamo

Agenda

- Background and Objective
- Overview of Skype Application
- Time Domain Analysis
- Statistical Fingerprinting for Skype Classification
- Conclusions and Outlook

Background

- VoIP telephony is gaining tremendous popularity
- Skype is one example of this evolution
- Skype traffic is obfuscated and protocol is propietary
- Skype traffic is raising security concerns

⇩

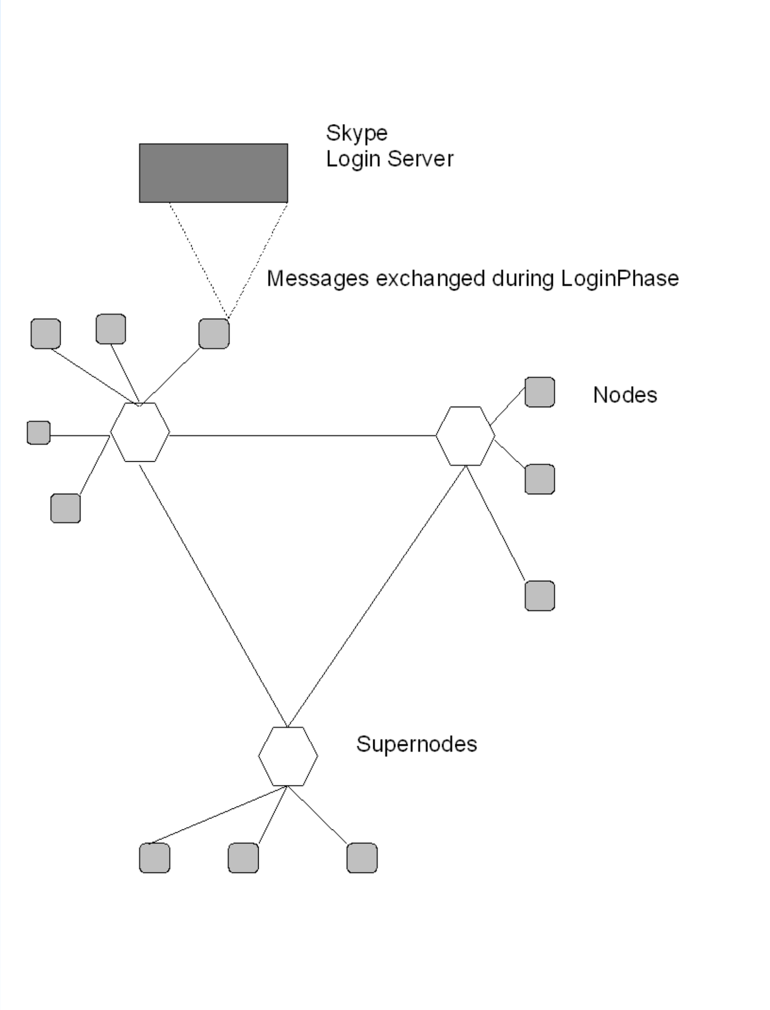
Skype Traffic Identification is fundamental

Objectives

The primary goals of this thesis are:

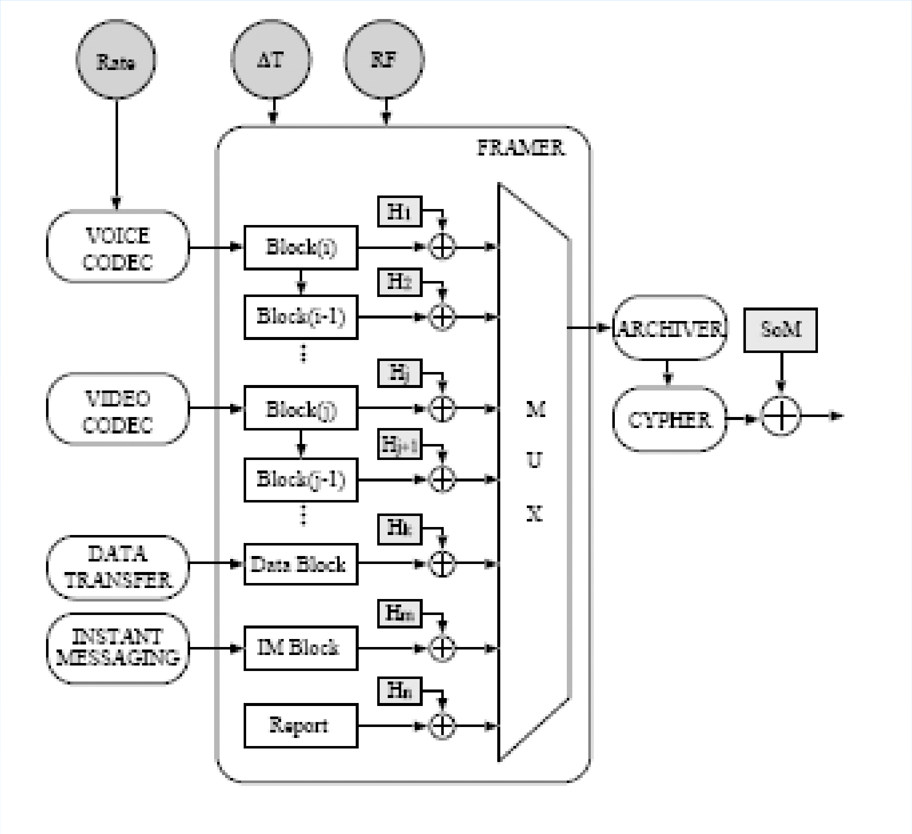
- To analyze Skype protocol
- To analyze Skype traffic in Time and Frequency Domain
- To propose and test a new classification method

Overview of Skype Application/1



- Peer-to-peer network
- Skype Login Server
- Supernodes
- Stun Protocol

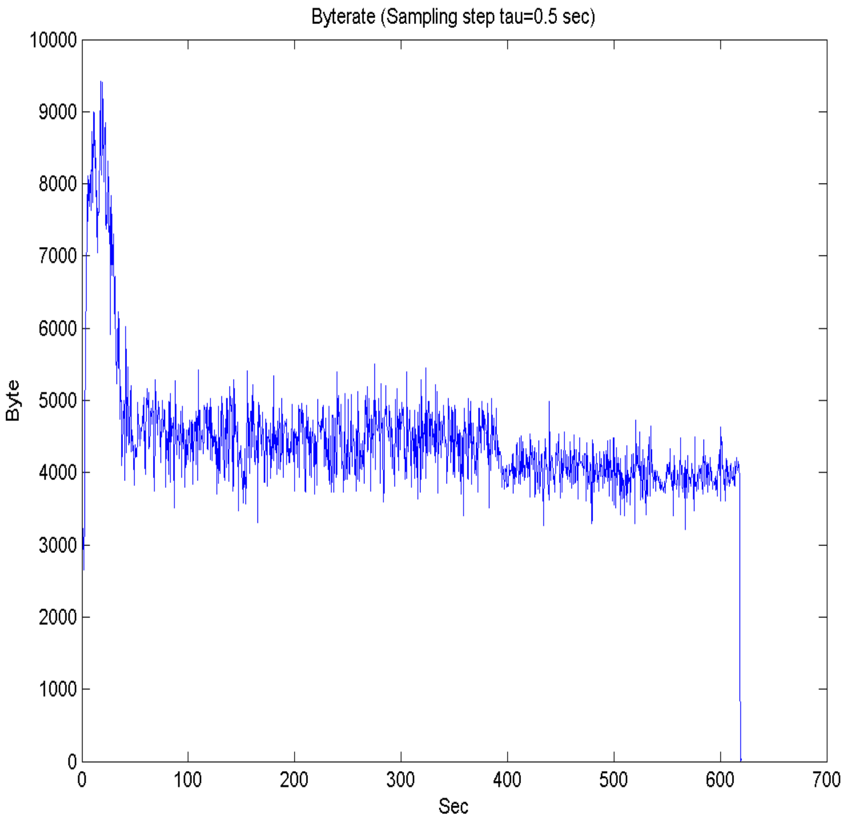
Overview of Skype Application/2



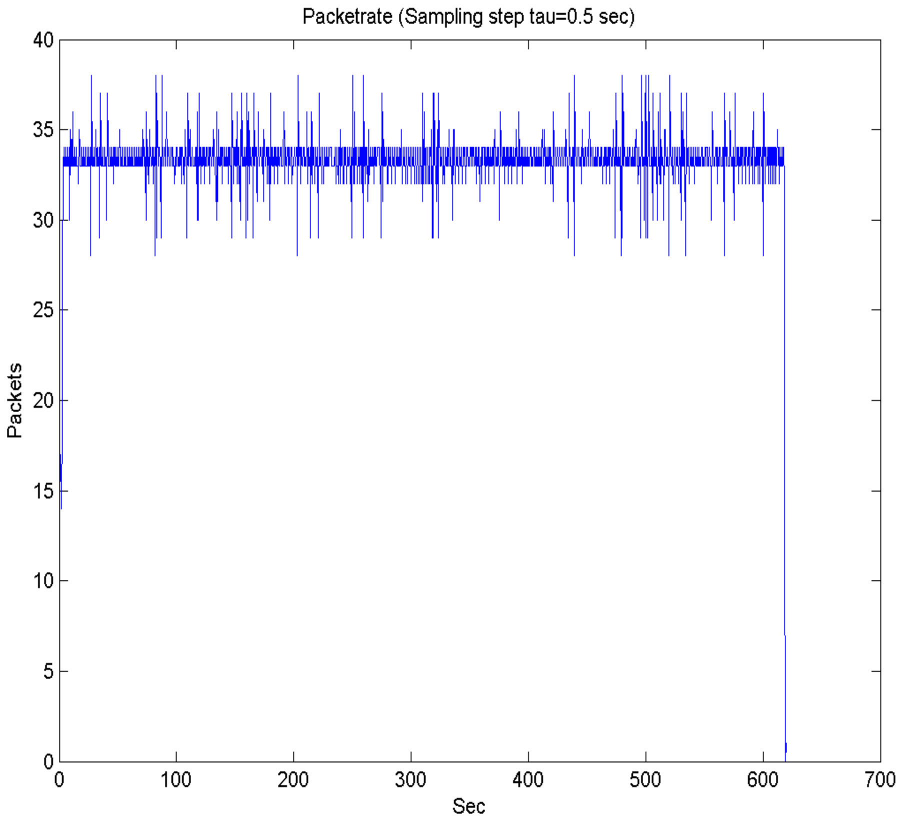
- Different codecs
- Archiver
- Multiplexer
- Cypher

Time Domain Analysis/1

Byterate

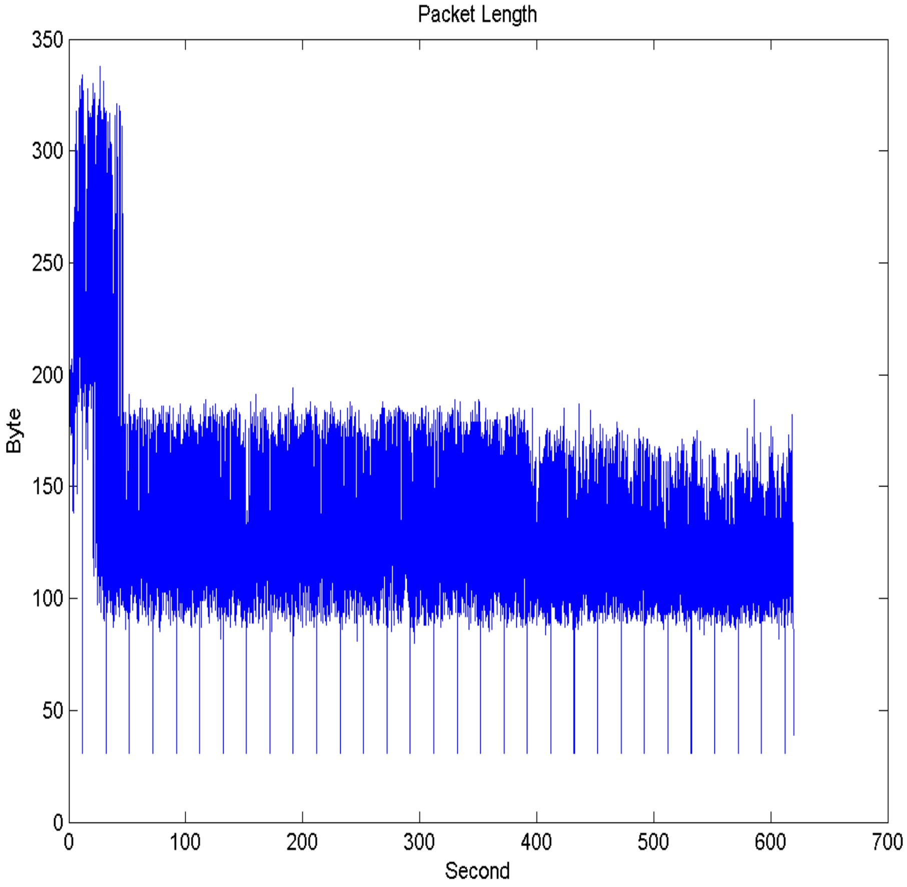


Packet-rate

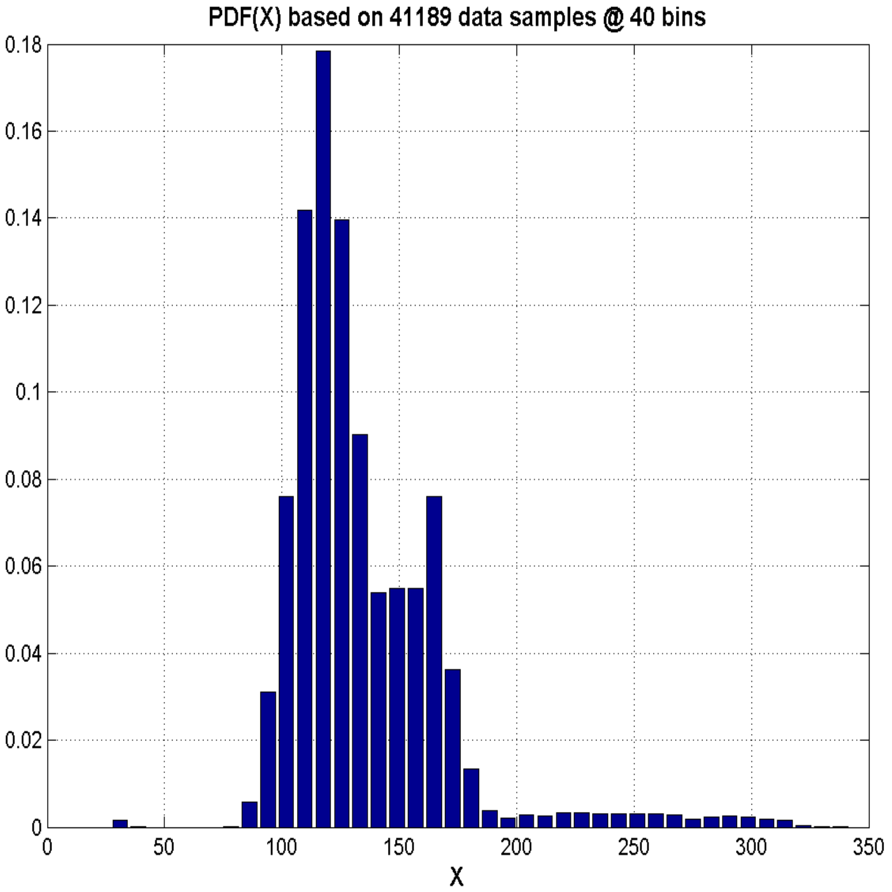


Time Domain Analysis/2

Packets Length



PDF of Packets Length



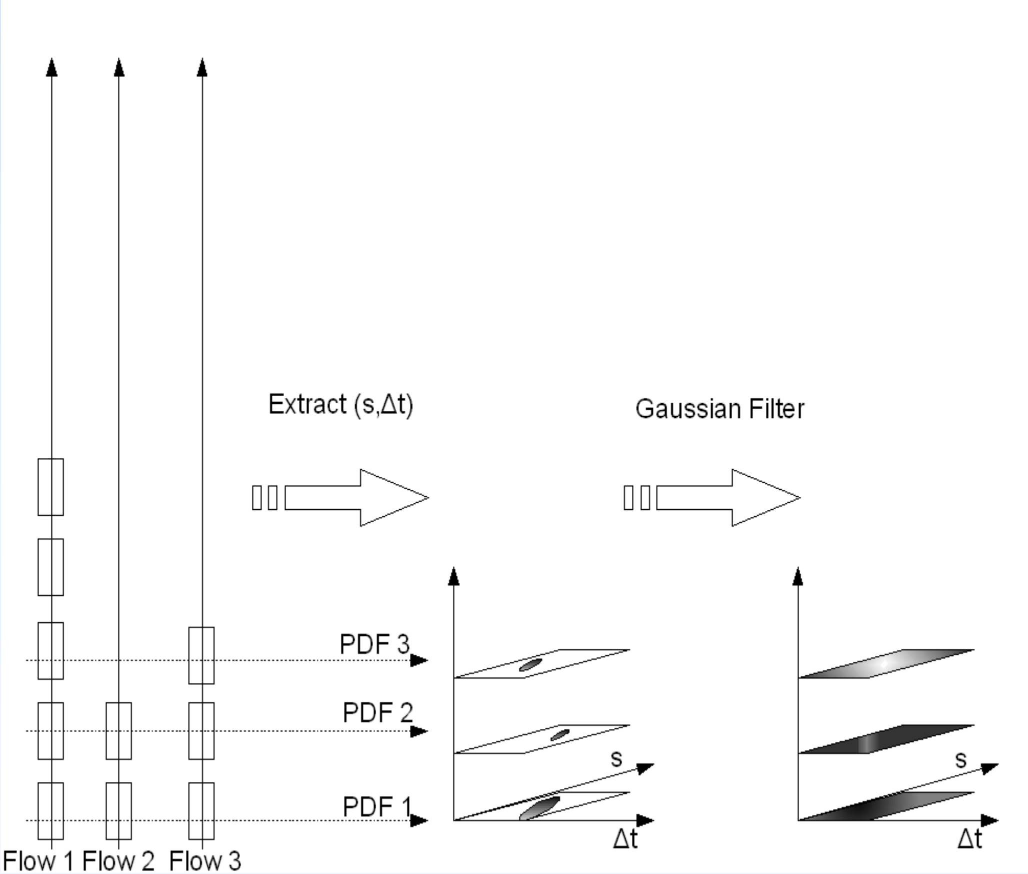
Statistical Fingerprinting for Skype Classification/1

Different classification methods:

- Header-based
- Payload-based
- Statistical classification → Protocol Fingerprinting
- Hybrid

Some statistical properties of basic elements of each network flow should be sufficient to determine which application has generated the traffic

Statistical Fingerprinting for Skype Classification/2



Three properties are taken into consideration:

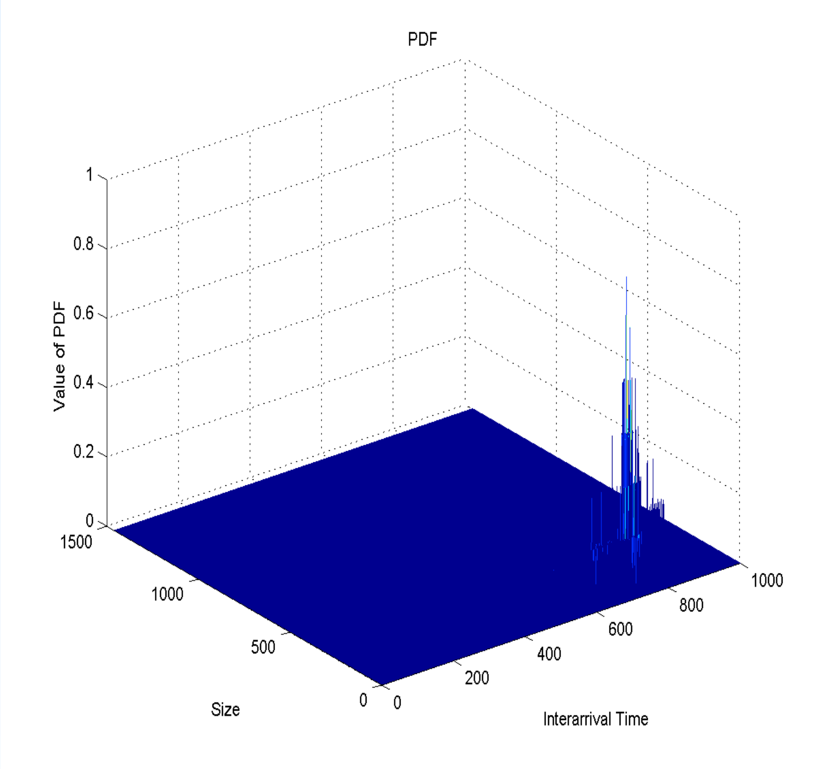
- Packet size
- Interarrival Time
- Arrival order of Packets.

A PDF vector is built as in figure

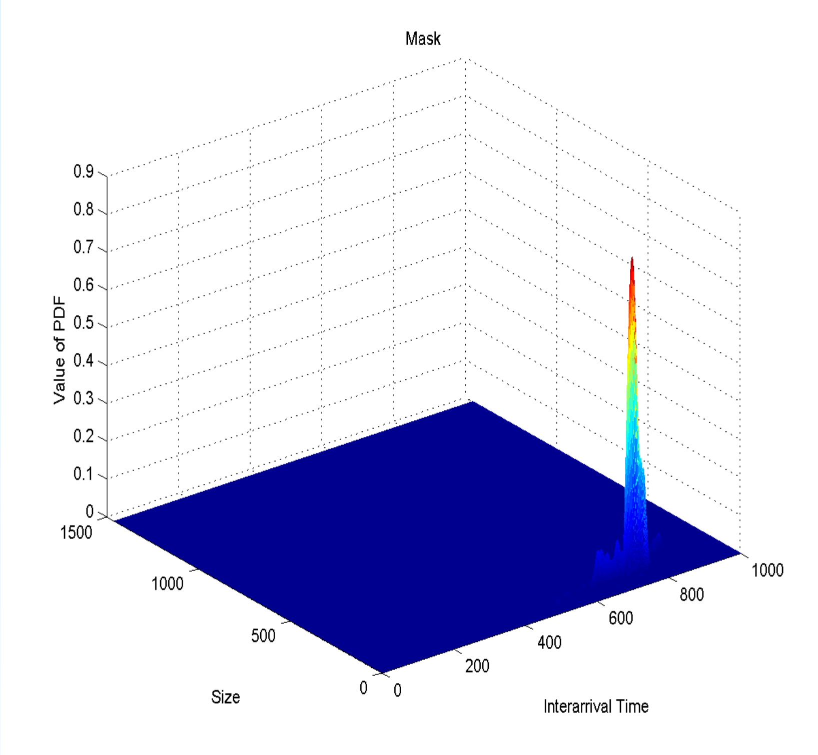
A Gaussian Mask is applied to each PDF element of the vector

Statistical Fingerprinting for Skype Classification/3

PDF before and after the application of the Smoothing filter



P_i



$M_i(P_i)$

Statistical Fingerprinting for Skype Classification/4

Protocol Decision:

Computation of each single Anomaly Score for each packet (and so for each Mask Vector element):

$$A_i(P_i, M_i) = \frac{1}{max(\varepsilon, M_i(P_i))}$$

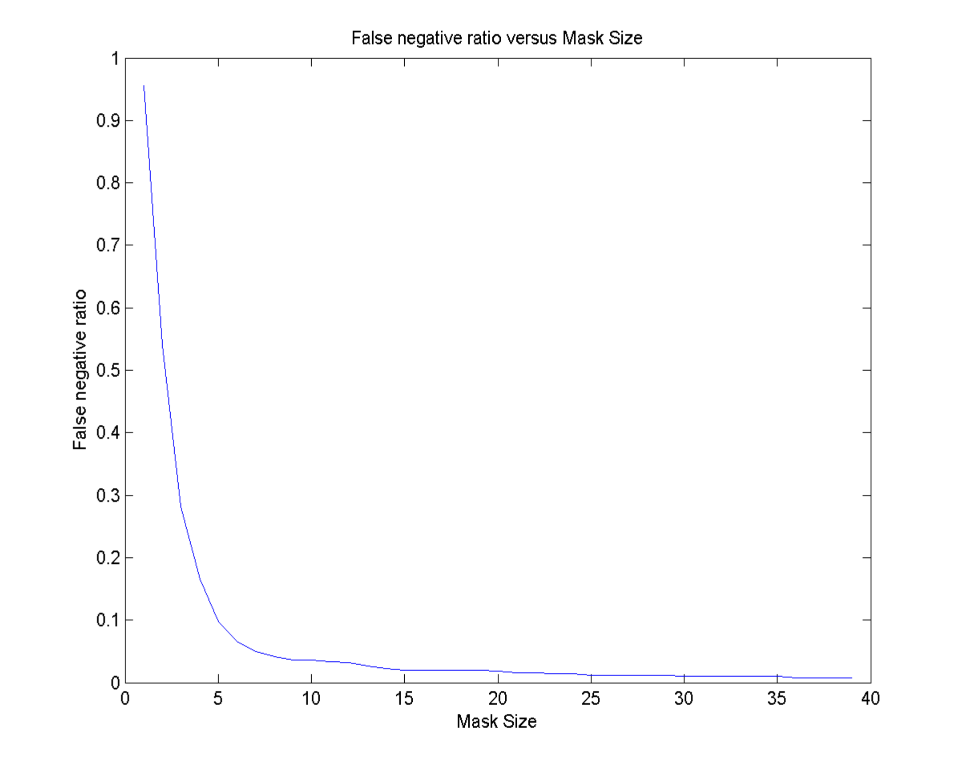
Computation of the Anomaly Score S of the unknown flow F against vector M:

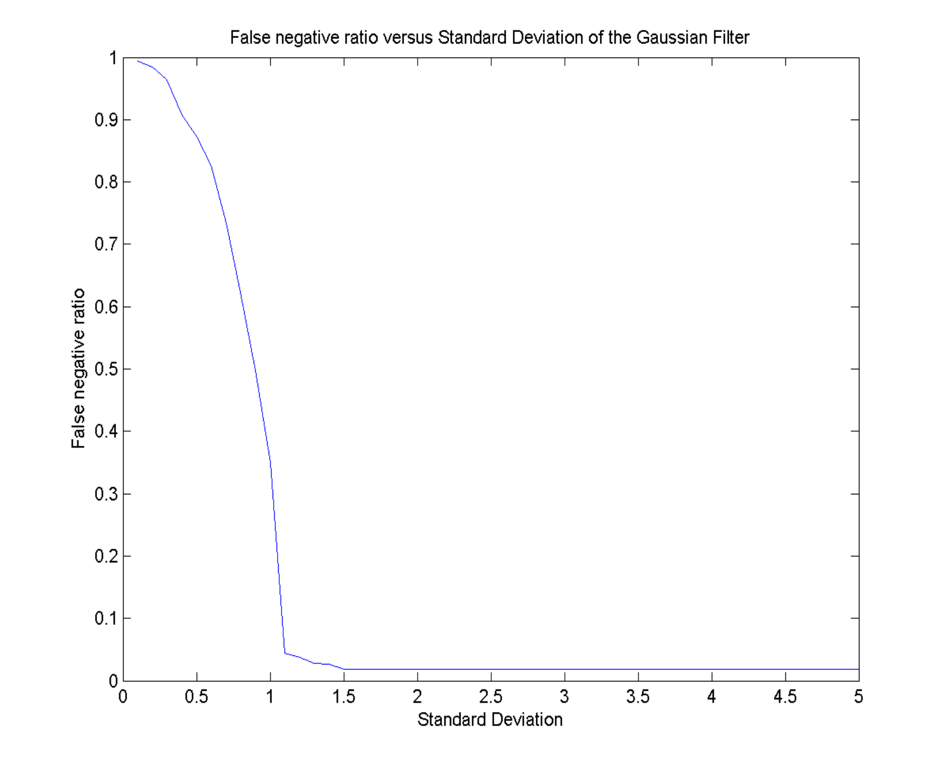
$$S_n(F, \vec{M}) = \frac{[\sum_{i=1}^{n} A_i(P_i, M_i)/n] - A_{min}}{A_{max} - A_{min}}$$

Setting of a threshold.

Anomaly score > threshold → Training set and evaluation set do not belong to the same protocol

Statistical Fingerprinting for Skype Classification/5

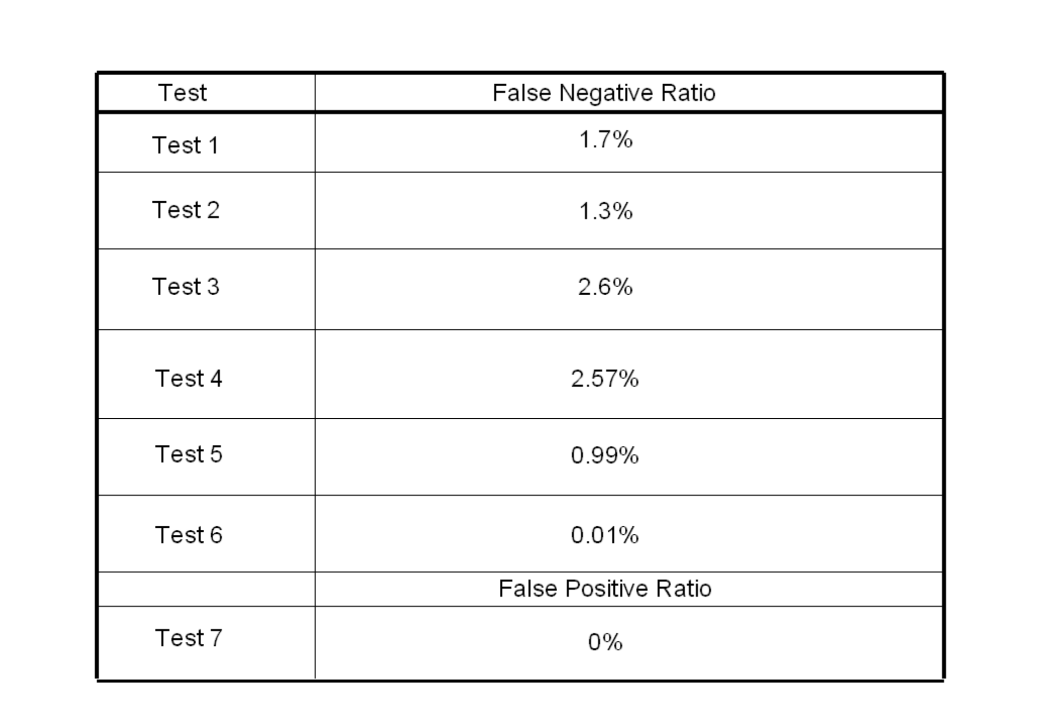




Important parameters are the mask size and the standard deviation of the Gaussian Function

As it can be seen from above graphs best results are obtained if size is bigger than 15X15 and standard deviation than 1.5

Statistical Fingerprinting for Skype Classification/6



Test	False Negative Ratio
Test 1	1.7%
Test 2	1.3%
Test 3	2.6%
Test 4	2.57%
Test 5	0.99%
Test 6	0.01%
	False Positive Ratio
Test 7	0%

Test 1 has been performed inside the campus network of Politecnico di Milano, Test 2 has been perfomed with calls between one host inside the campus of Politecnico di Milano and one host outside it. Test 3 and 4 have been performed using Mask from test 3 for valuating data from test 4 and the way around. In test 5 and 6 the mask has been created mixing data from test 1 and 2. Test 7 has been performed in order to evaluate false positive ratio (traces gathered not from Skype traffic).

Outlook

There are still many open issues:

Value of the mask size

Value of the mask standard deviation

Possibility of using different smoothing filters

Number of packets after which taking a decision

Validity of the Fingerprint

Transportability of the Fingerprint

THANK YOU

Any question?