Evaluation of Anomaly Detection Method based on Pattern Recognition

- Romain Fontugne
  The Graduate University for Advanced Studies

- Yosuke Himura

  The University of Tokyo

- Kensuke Fukuda

  National Institute of Informatics

Outline

- Motivation
- Temporal-spatial structure of anomaly
- Pattern-recognition-based method
  - Hough transform
- Parameter space
- MAWI database
- Study case
- Conclusion

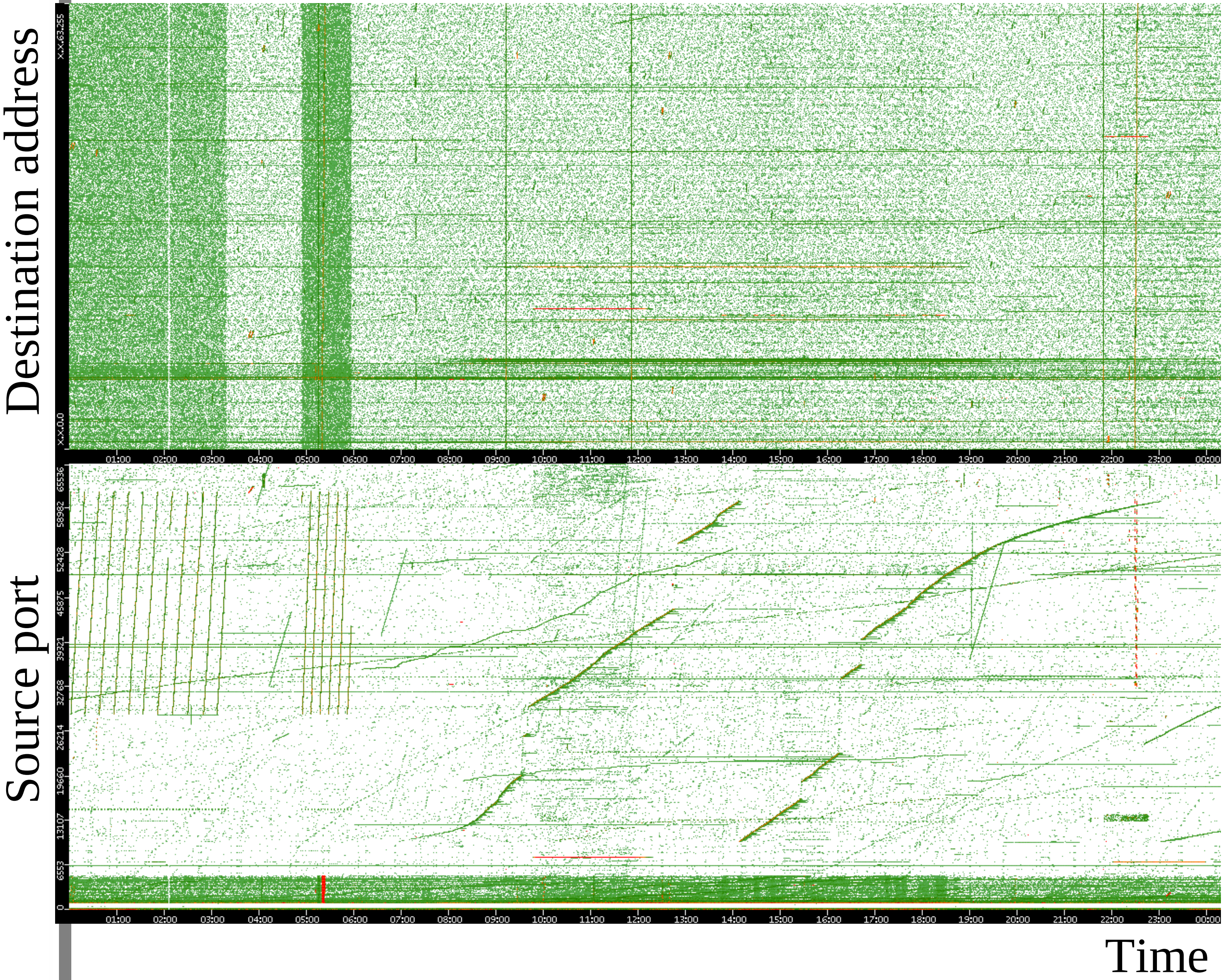
Motivation (1)

- Network traffic anomaly:
  - Misconfigurations, failure, network attacks
- Side effects:
  - Bandwidth consuming
  - Weaken network performance
  - Harmful traffic
  - Alter the traffic's characteristics

Motivation (2)

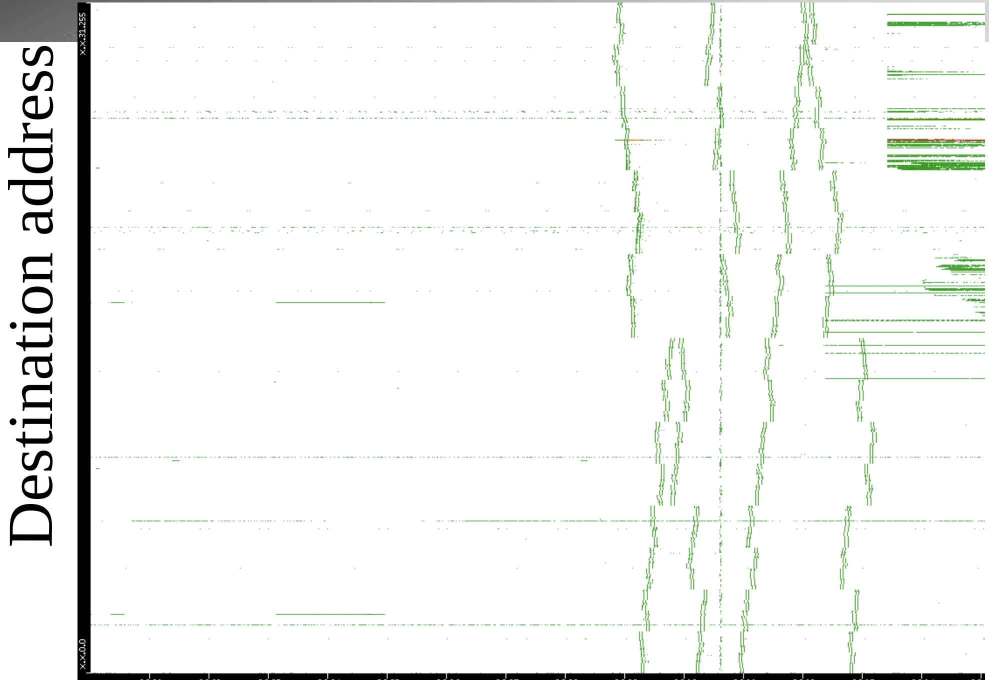
- Difficulties:
  - Huge amount of data
  - Variety of anomalous traffic
  - Identification of tiny flows
- Anomaly detection method:
  - Usually treated as a statistical problem
    - Evaluate the main characteristics of traffic
    - Discriminate traffic with singularities

Temporal-spatial structure of anomaly (darknet)

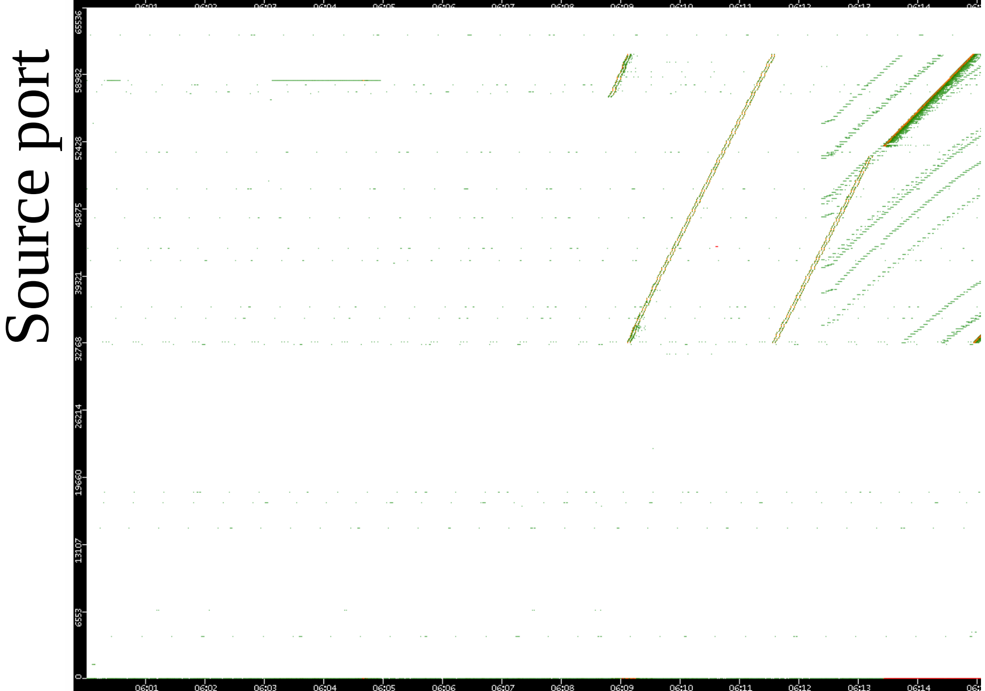


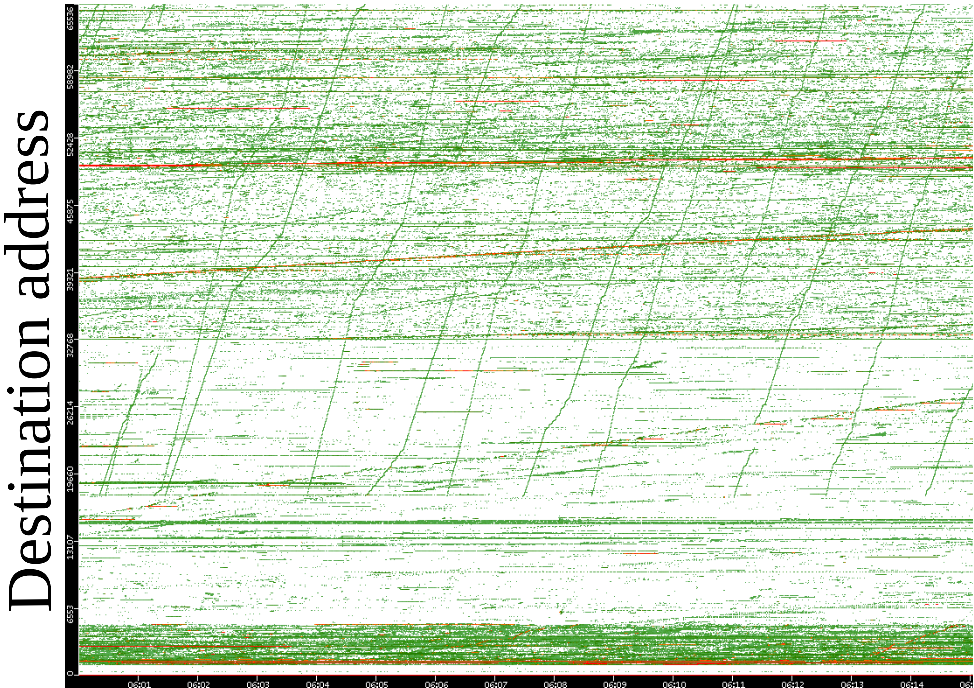
- Unwanted traffic
- Linear structures
- Unusual distribution of traffic feature

Temporal-spatial structure of anomaly (MAWI)



- Samplepoint-F:
  - 2009/02/21





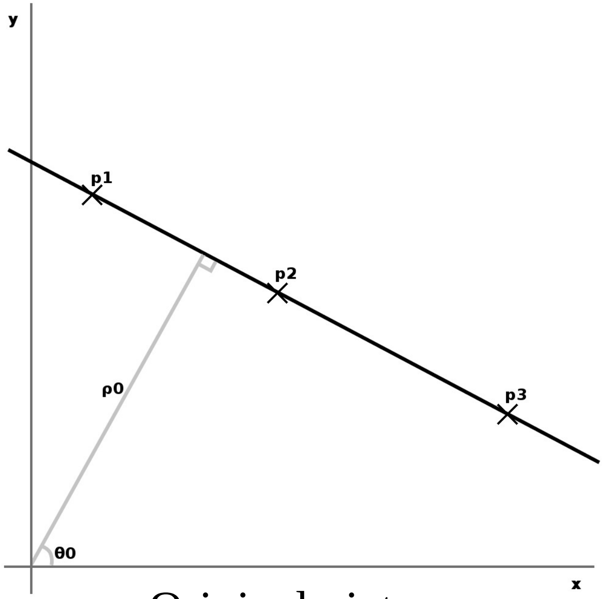
Pattern-recognition-based method

- Identification of linear structures in pictures:
  - Generate pictures from traffic
  - Hough transform
  - Retrieve packet information
  - Report anomalies

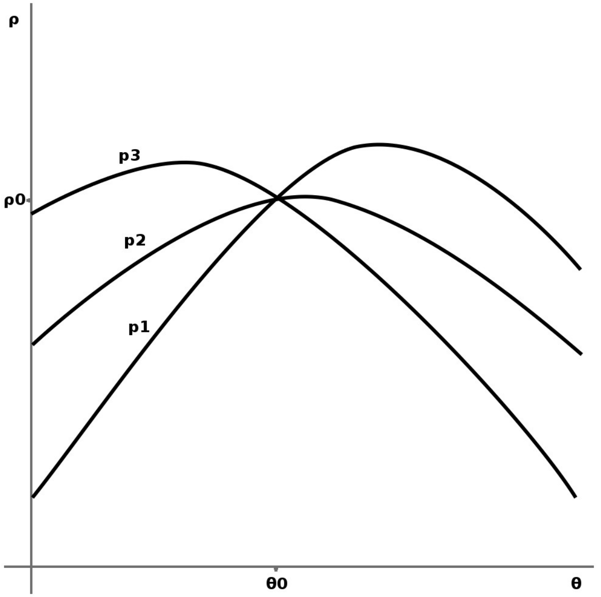
Hough transform

- Voting procedure
  - Points elects lines
  - Polar coordinates
    $\rho = x \cdot \cos \theta + y \cdot \sin \theta$
  - Hough space

- Identify line means extract max in the Hough space
  - Relative threshold



Original picture



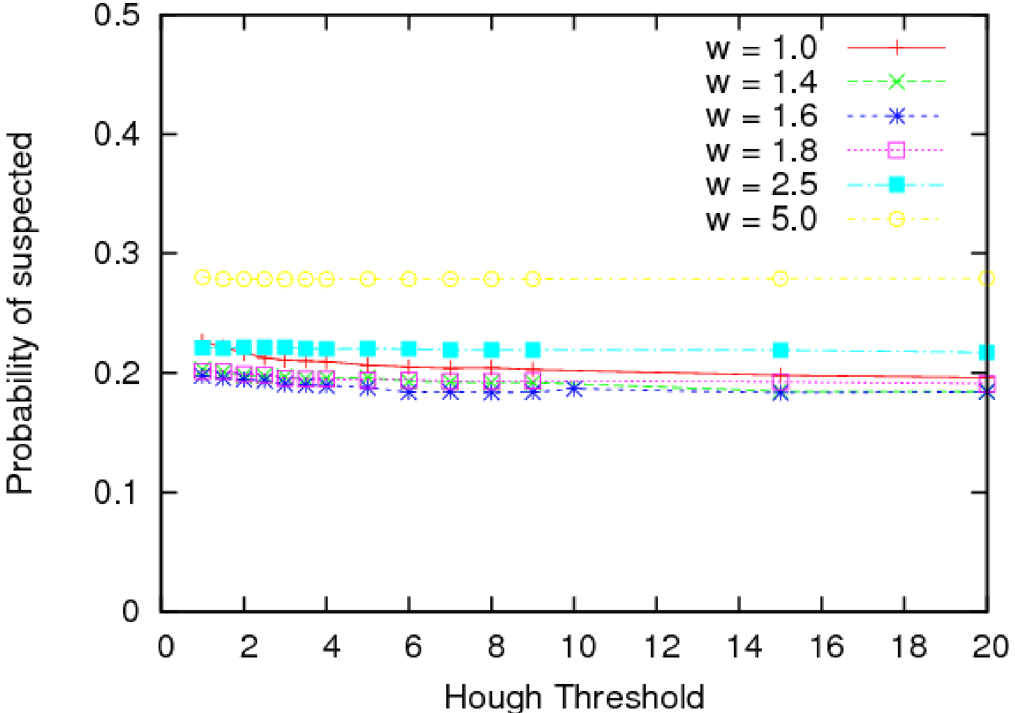
Hough space

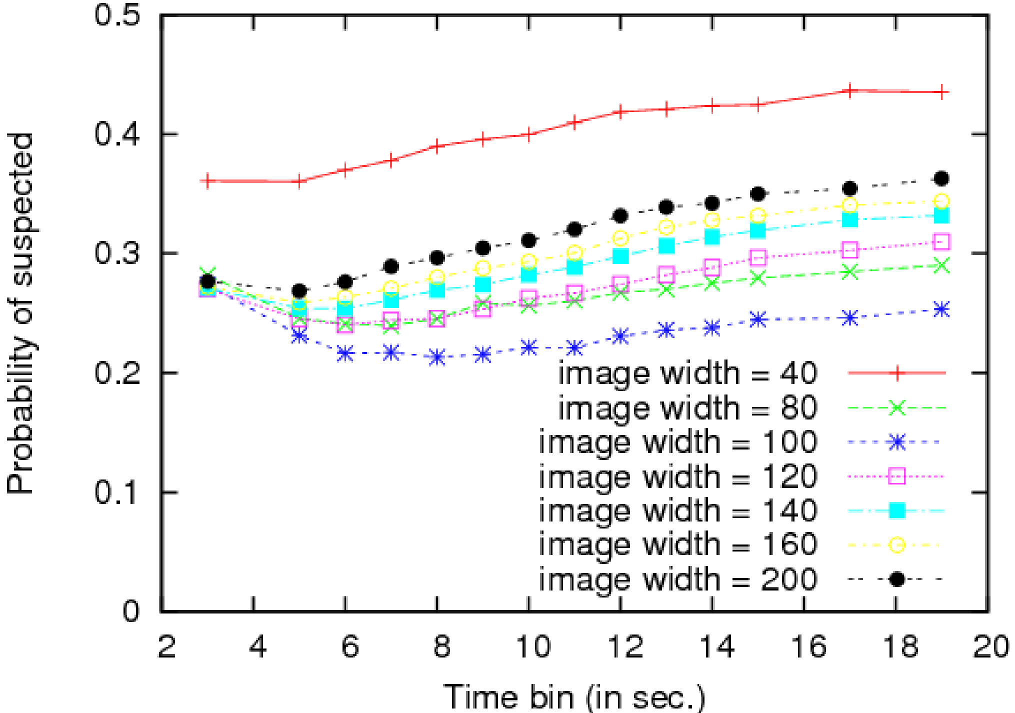
Parameter space

- Hough parameter:
  - Weight for the voting procedure
  - Threshold to determine candidate line
- Picture resolution:
  - Time bin
  - Size of pictures

Evaluation of parameter space

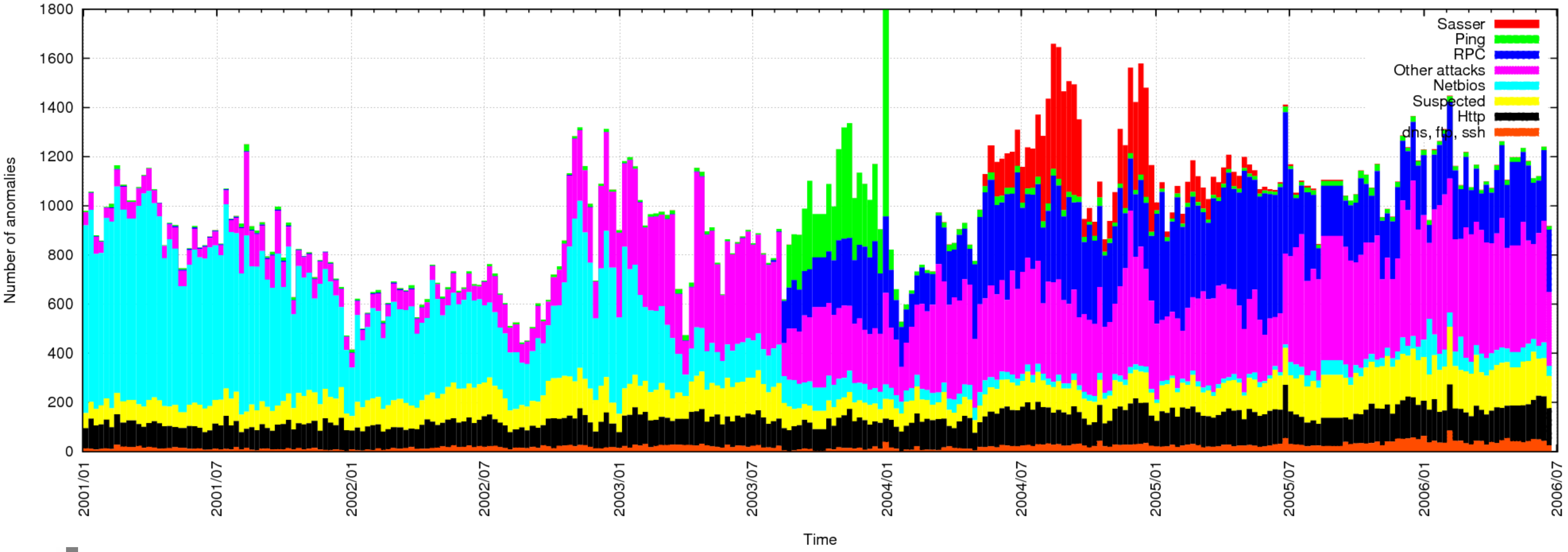
- Heuristics:
  - suspected = false positive + unknown
- Prob. of suspected = suspected / total anomalies
  - Lower is better





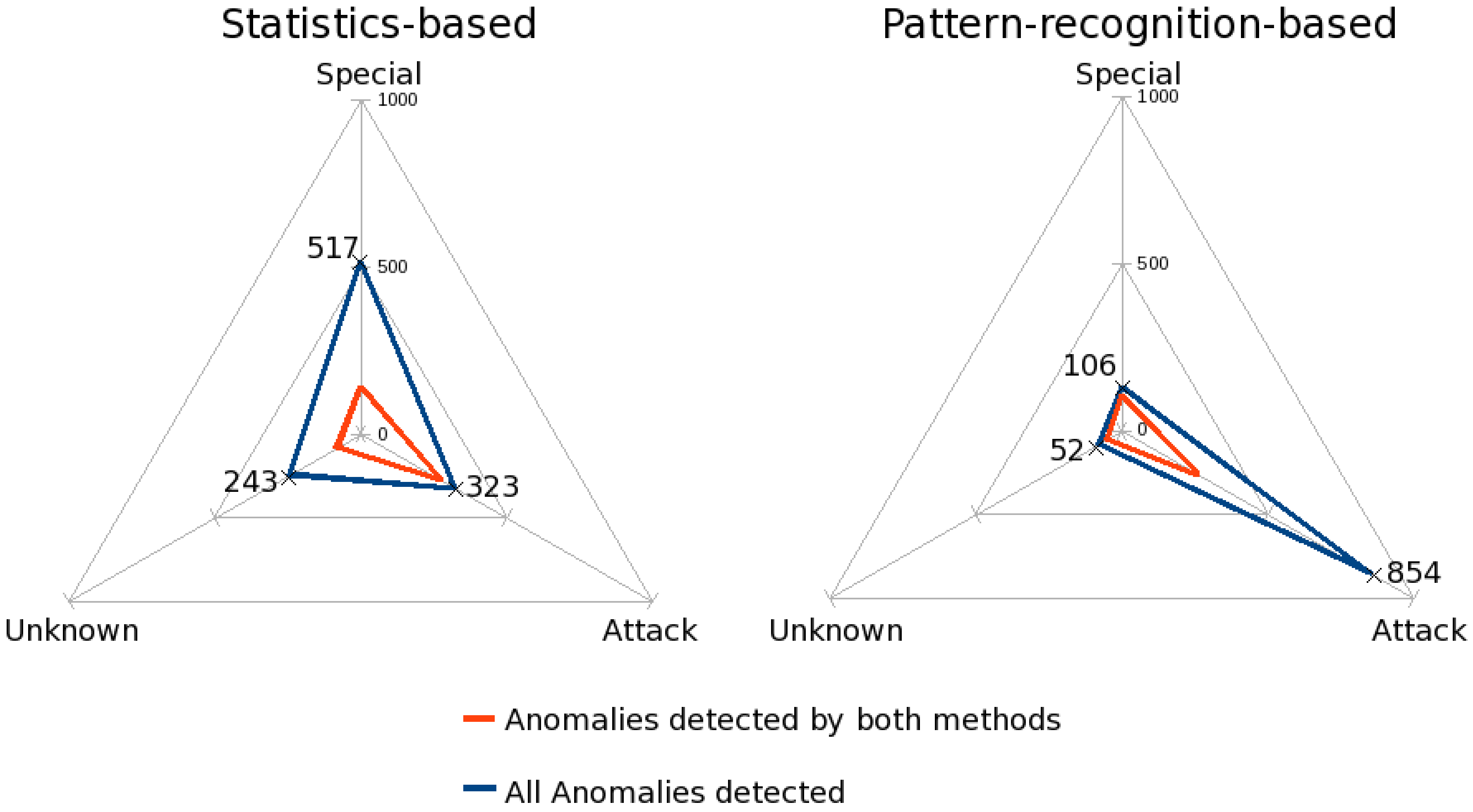
MAWI database

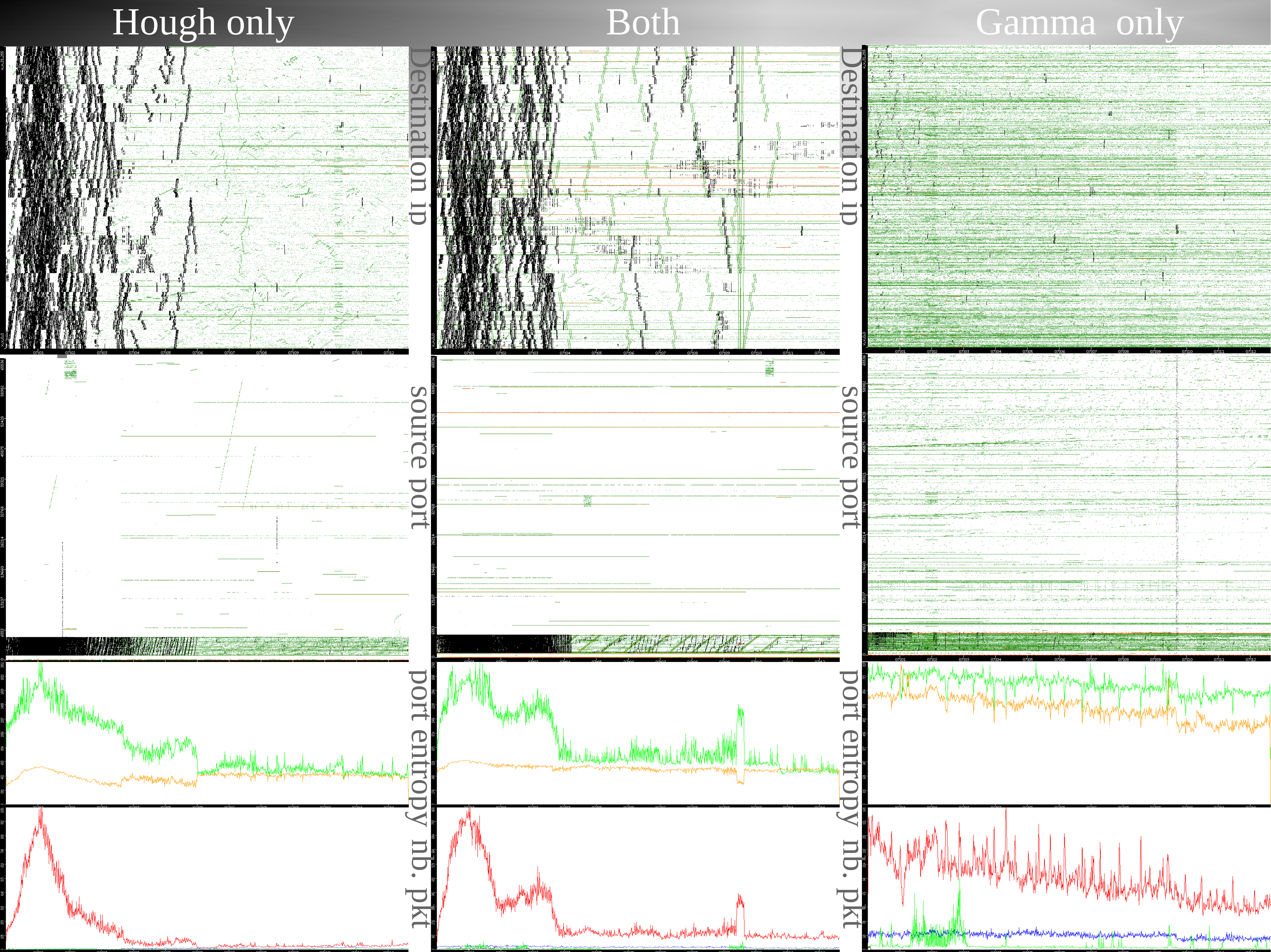
- Samplepoint-B:
  - From 2001/01 to 2006/06



Study case: sasser infection

- Gamma modeling vs. Pattern recognition (2004/08/01)
- Gamma modeling-based method tuned to detect the same number of anomalies (Includes many false positives)




Hough only
Both
Gamma only
Destination ip
Destination ip
source port
source port
port entropy
port entropy
nb. pkt
nb. pkt

Discussion

- Two different backgrounds
  - 50% of their results in common
- Detection of anomalies involving a tiny number of packets
- Identify easily network/port scans (dispersed distribution)
- Intensive uses of source port
- Gamma modelling = deeper analysis of the traffic's characteristics (highlight singular traffic)

Conclusion and future work

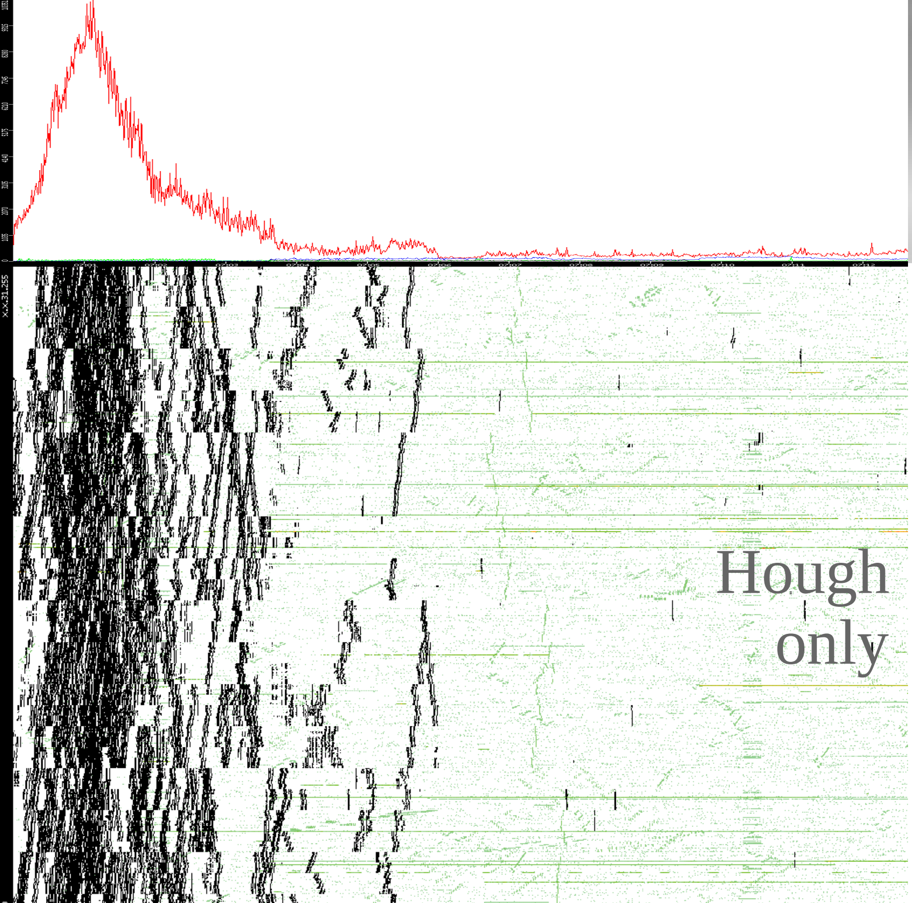
- No perfect method
- Combination of several methods
- Need of methods with different backgrounds

- Future work
  - Auto-tuning of parameters
  - Sampled data
  - More graphical representations
  - Study good combinations

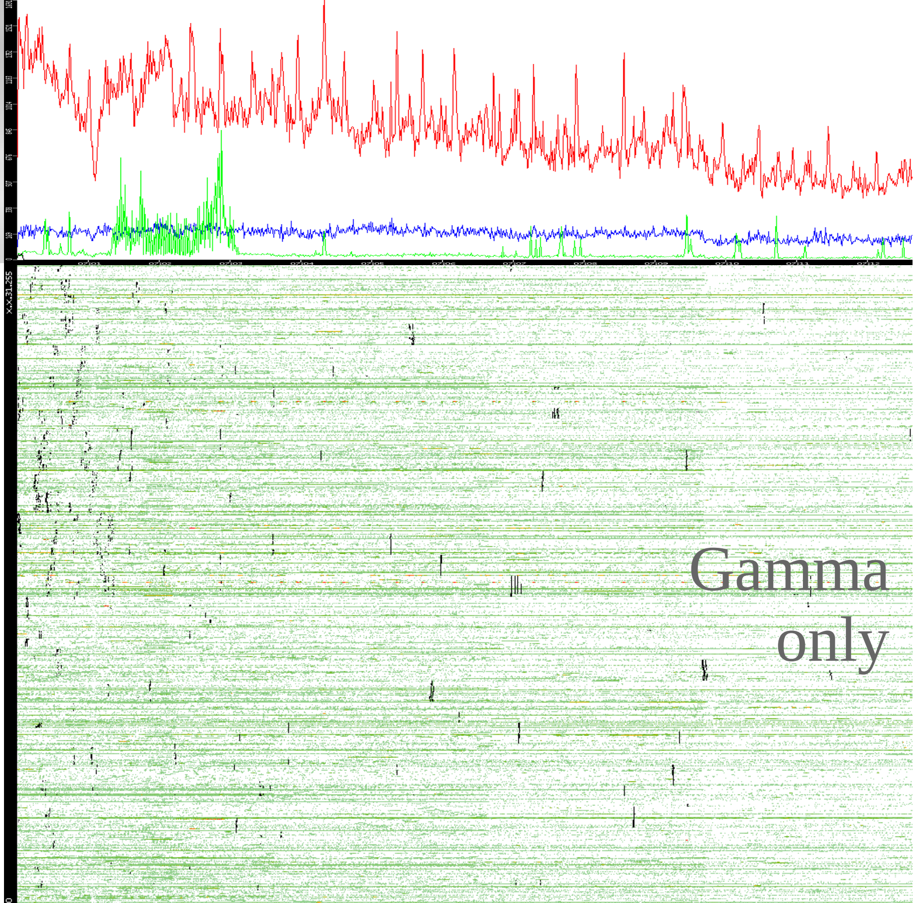
Thank you

Any questions?

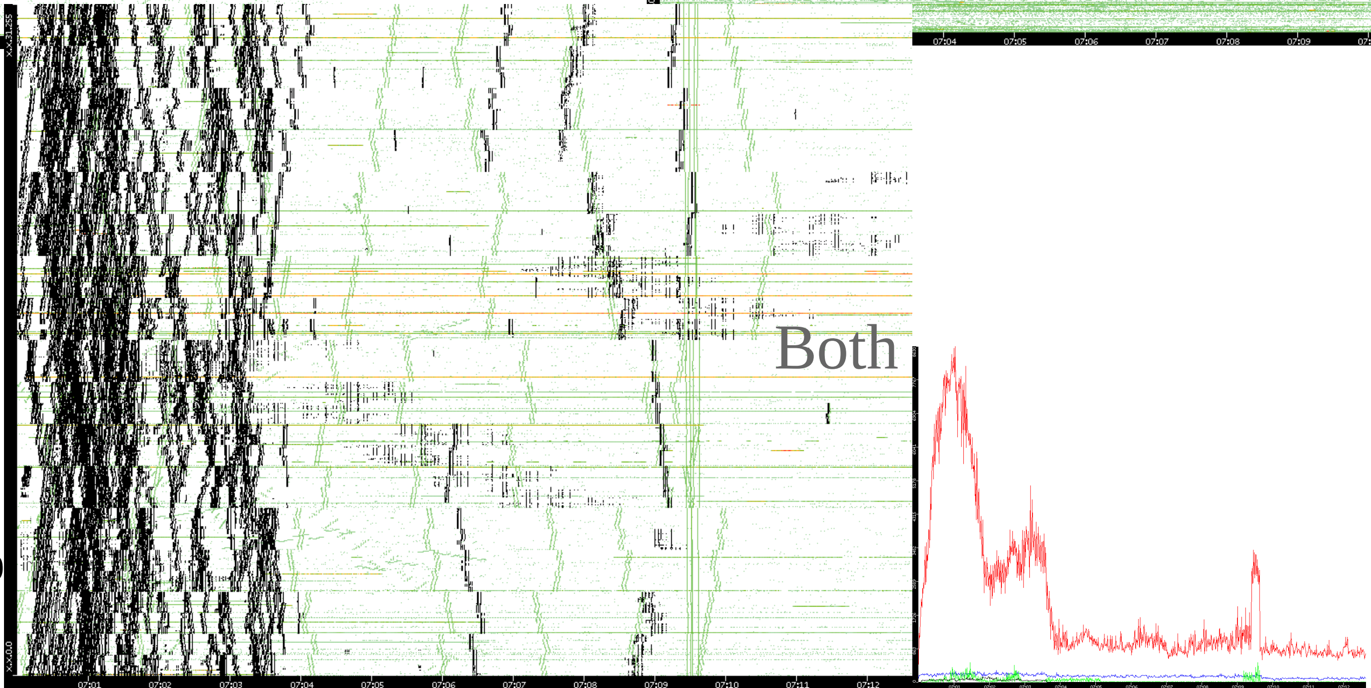
romain@nii.ac.jp



Hough only

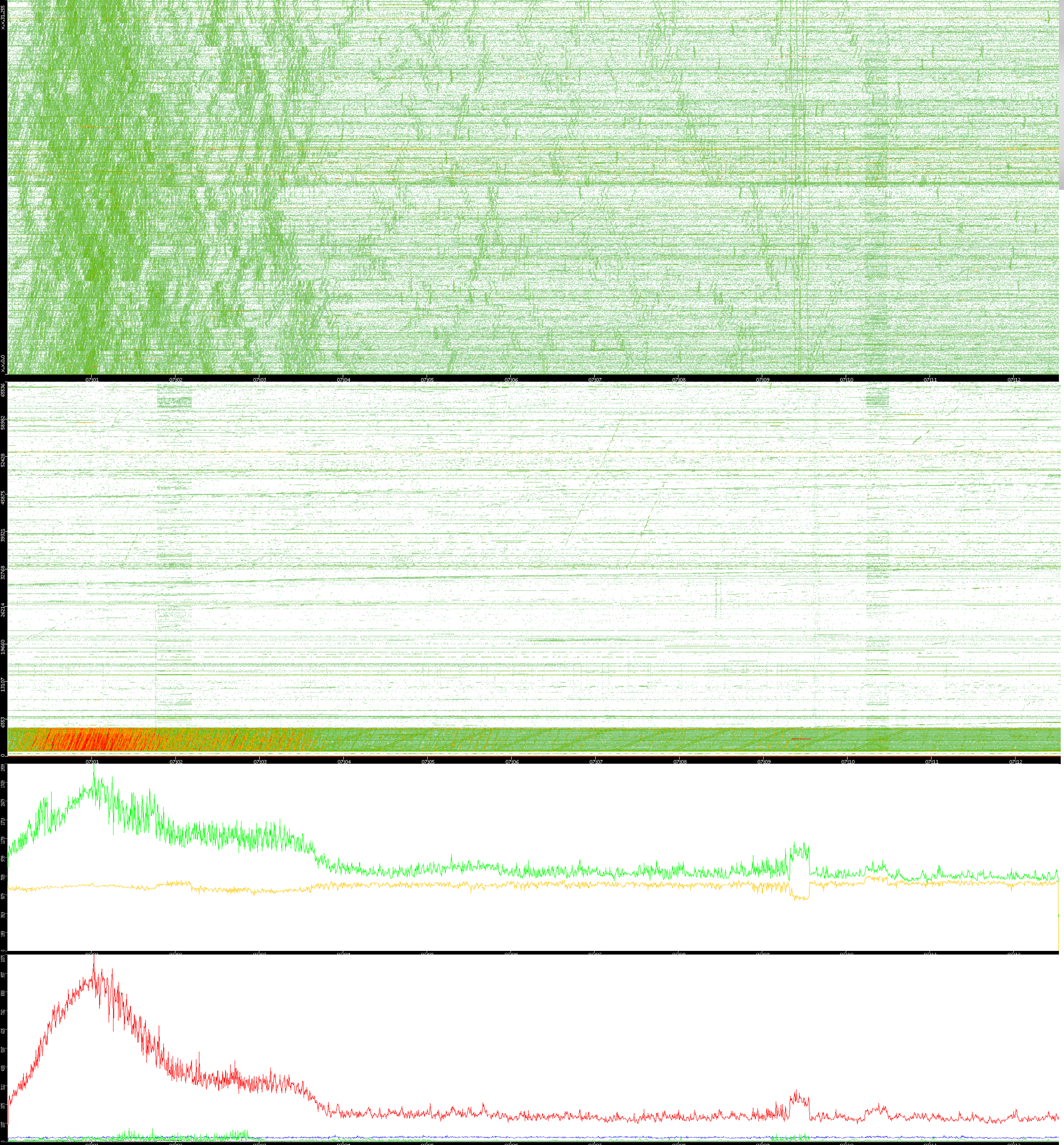


Gamma only



Both

Original data



Destination ip

source port

port entropy

volume