Visual Network Analysis

Alessandro **Giannetti**

Gianluca **Tasciotti**

Introduction

With this project, we are going to illustrate a possible visualization of a dataset in which are present a huge quantity of attacks. We will see some techniques to represent it and which problems we crossed to show it in the best way.

Pipeline

DATASET

TECHNOLOGIES

VISUALIZATIONS

ACTIONS

FILTERS

Dataset

Intrusion Detection Evaluation Dataset (CICIDS2017)

- **Format**: CSV
- **Dimension**: 48.00 GB
- **Protocols**: HTTP, HTTPS, FTP, SSH, and E-mail protocols
- **Days**: 3rd of June 2017 at 9:00 to 7th of June 2017 at 17:00
- **Parameters**: Flow ID, Source IP, Source Port, Destination IP, Destination Port, Protocol, Timestamp, Label, TotalLengthOfFwdPackets, Flow Duration, Total Fwd Packets, Total Bwd Packets etc.
- **Example**: **172.16.0.1-192.168.10.50-17560-80-6, 172.16.0.1, 17560, 192.168.10.50, 80, 6, 7/7/2017 3:59, 1341015, 5, 0, 30, 0, 6, 6, 6, 0, 0, 0, 0, 0, 22.37111442, 3.72851907, 335253.75, 669835.666, 1340007, 1, 1341015, 335253.75, 669835.666, 1340007, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 100, 0, 3.72851907, 0, 6, 6, 6, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 7.2, 6, 0, 100, 0, 0, 0, 0, 0, 0, 5, 30, 0, 0, 256, -1, 4, 20, 0, 0, 0, 0, 0, 0, 0, 0, DDoS**

Dataset Manipulation

- We have selected only 7 of the most significant out of 85 parameters:
  - **Source IP**
  - **Destination Port**
  - **Destination IP**
  - **TotalFwdPackets**
  - **TotalLengthOfFwdPackets**
  - **Timestamp**
  - **Label**
- We have removed all samples with **Label** equal to **BENIGN**
- The most frequently used **Destination Port**, **Source IP** and **Destination IP**
- The dataset proposed respects the **AS rule: #tuples * #dimensions between 10.000 – 50.000**
  - ~ **5,16** secs computation time **(Proposed Solution)** → **AS = 5.510 * 7 = 38.570**
  - ~ **30,5** secs computation time (**Discarded Solution**) → **AS = 63.028 * 7 = 441.196**

Dataset → Manipulated Dataset

Example of Initial Dataset Line

Flow ID: 172.16.0.1-192.168.10.50-17560-80-6,

Source IP: 172.16.0.1,

Source Port: 17560,

Destination IP: 192.168.10.50,

DestinationPort: 80,

Timestamp: 7/7/2017 15:59,

+ 79 fields…

Label: DDoS

Example of Final Dataset Line

Source: 172.16.0.1,

DestinationPort: 80,

Target: 192.168.10.50,

Timestamp: 7/7/2017 15:59,

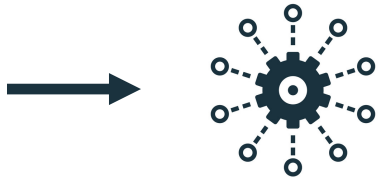
TotalFwdPackets: 8,

TotalLengthOfFwdPackets: 56,

Label: DDoS

Pipeline

DATASET



TECHNOLOGIES

VISUALIZATIONS

ACTIONS

FILTERS

Technologies

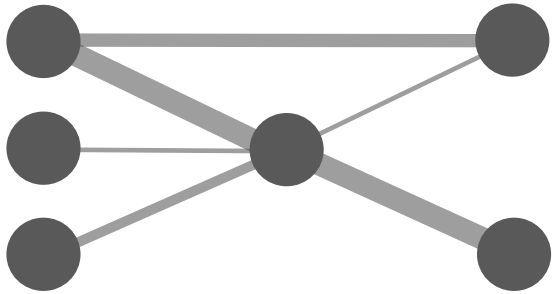
- **Python**
  - Dataset manipulation
- **D3 + JS**
  - Dataset Visualization

+

Pipeline

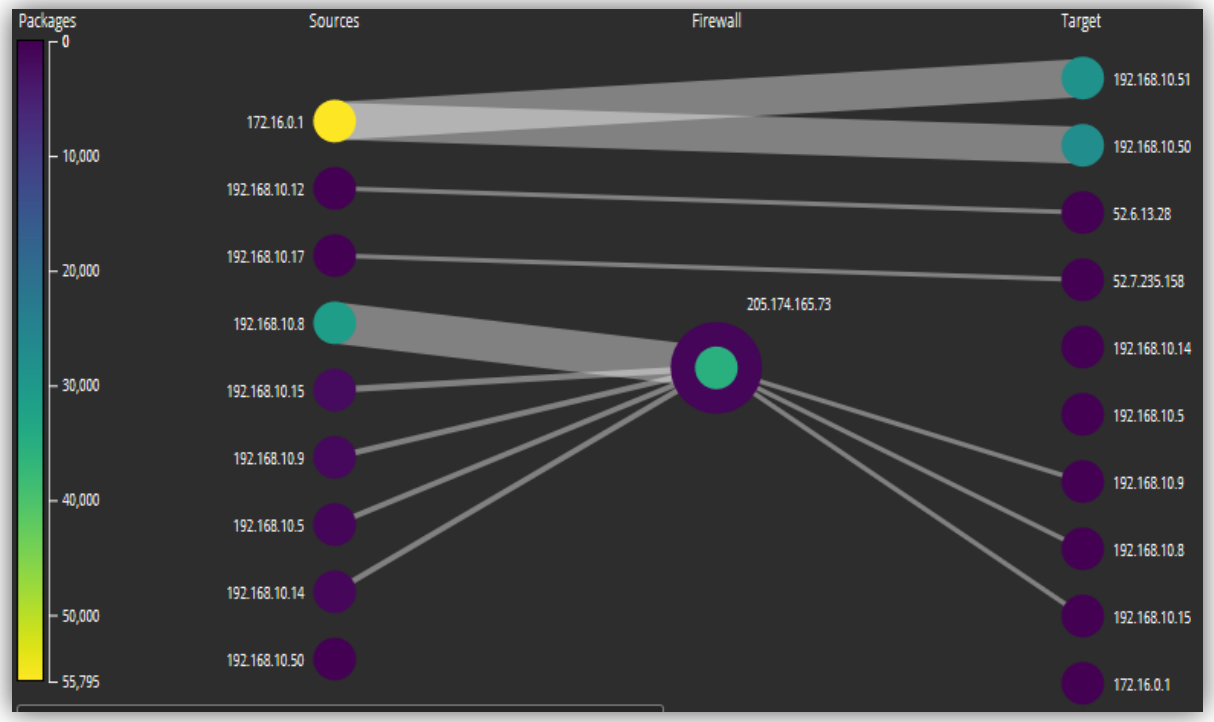
Visualizations

- **Graph**



Graph

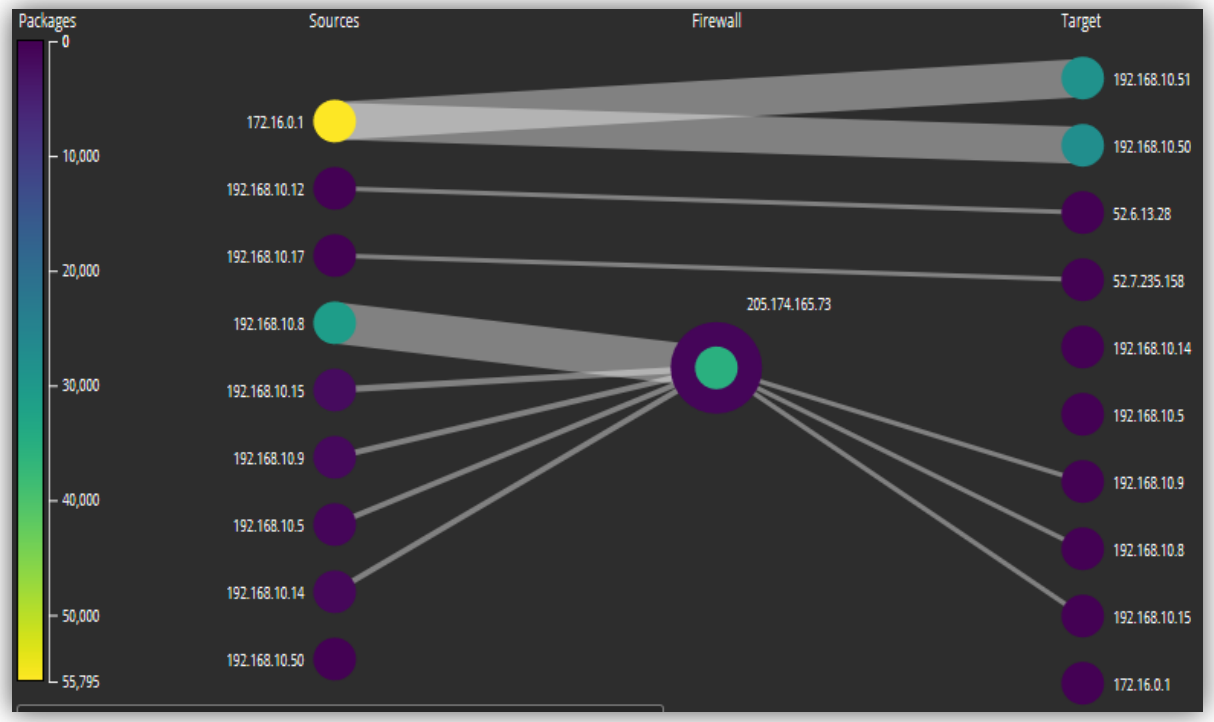
- **Nodes**: IP address
- **Edges**: Packages exchanged between two parties
- Bipartite Graph With 3 Layers:
  1. Attackers
  2. Firewall
  3. Targets



Graph

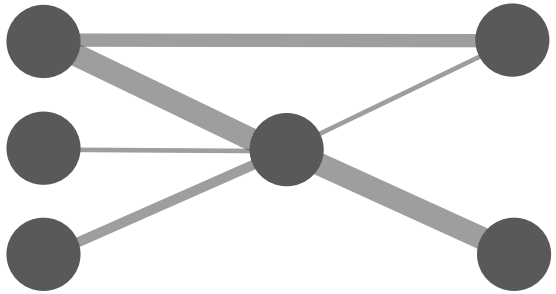
- **Colour Of Nodes** :
  Based on the number of **packets they sent and received**

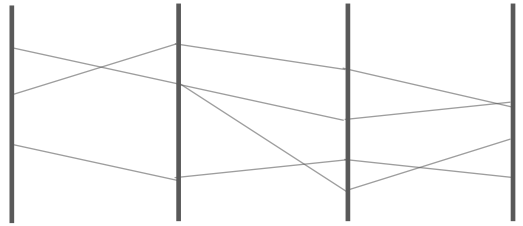
- **Edge Thickness** :
  Based on the number of **packets passing through the source and target**



Visualizations

- **Graph**
- **PCA**

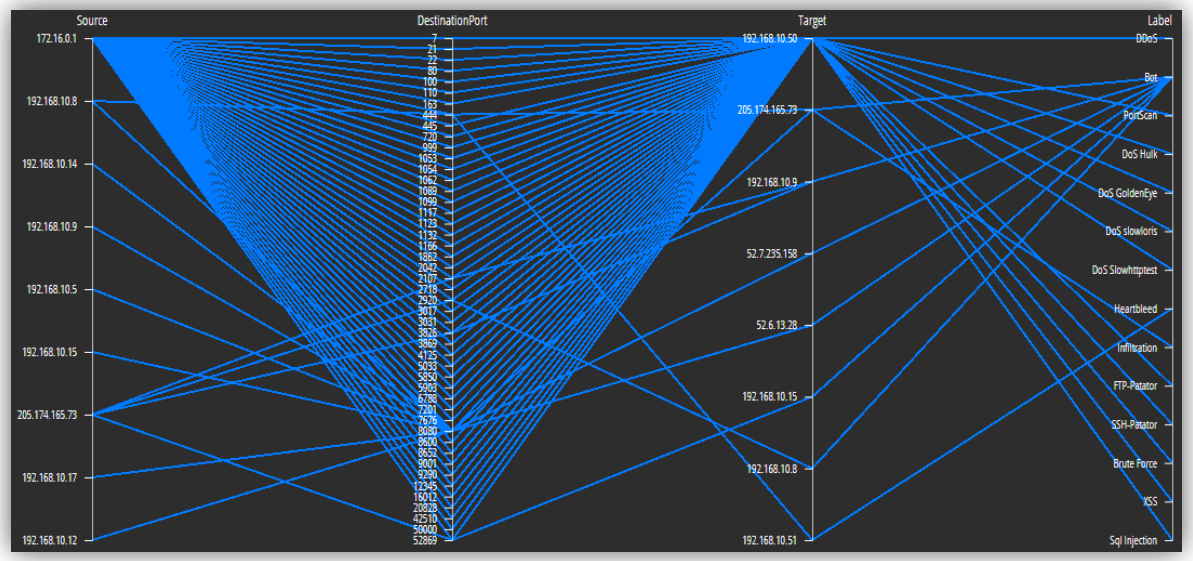




PCA

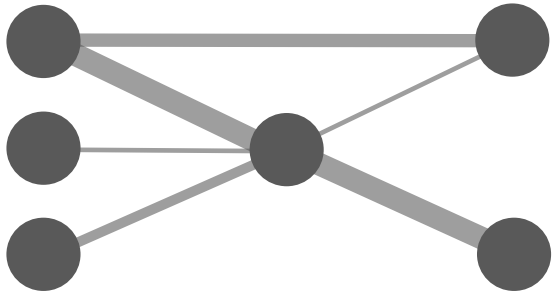
- **Axis**:
  - Source
  - Destination Port
  - Target
  - Label

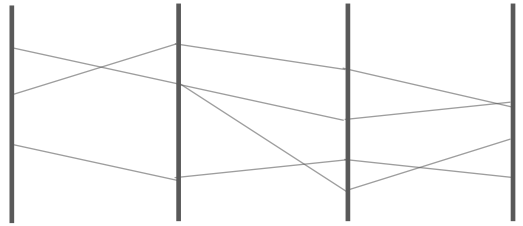
Each tuple in the dataset has been represented by a line, where this line intersects a specific value of these four parameters.

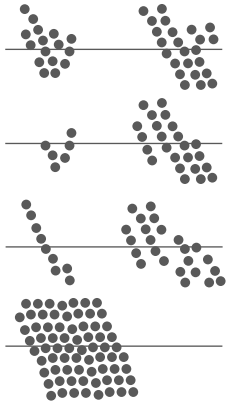


Visualizations

- Graph
- PCA
- Scatterplot of attacks (compared to time)





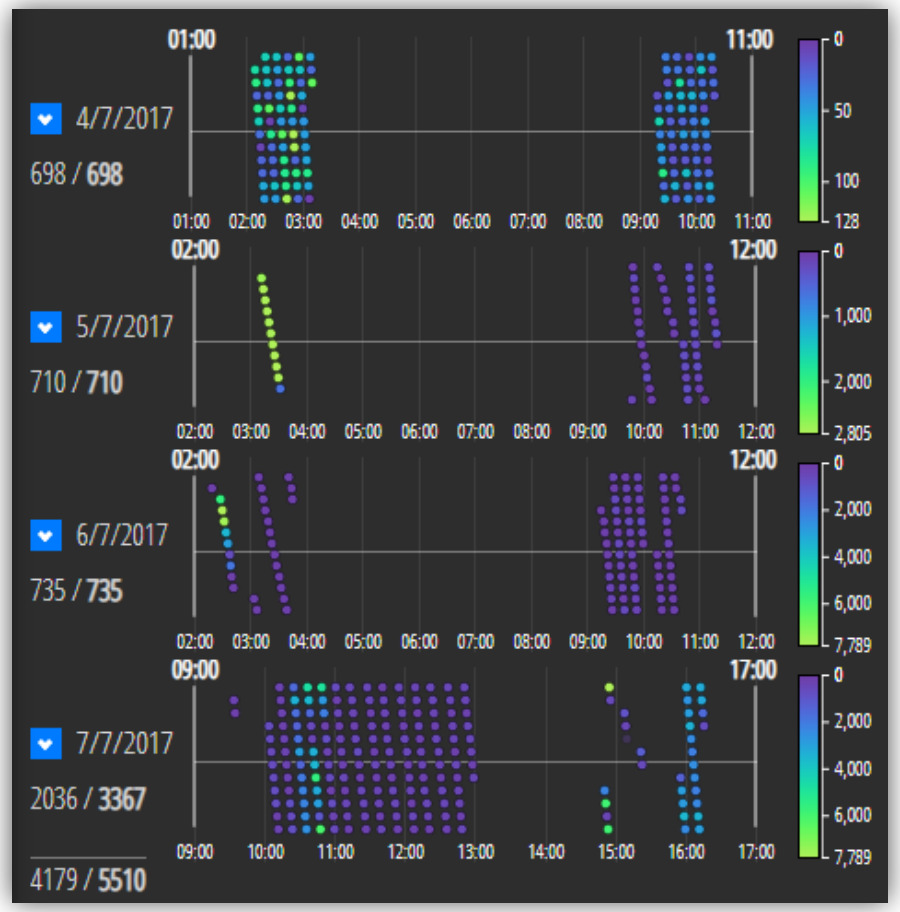


Scatterplot TIMESTAMP/ATTACK

- **Axis**:
  - **X:** Timestamp
  - **Y:** (algorithm to avoid overlapping)

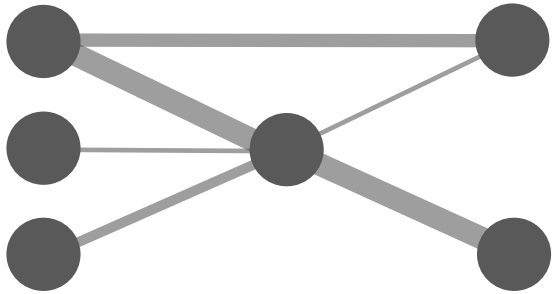
In this scatterplot, we have an overview of the attacks during the day. The Y-axis is calculated automatically so that the dots do not overlap.

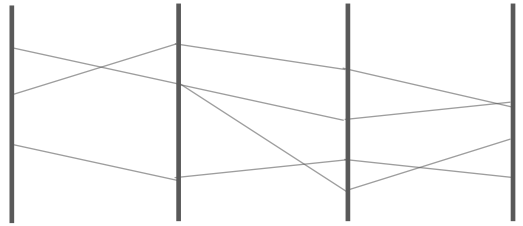
Colour Of Dots :
Based on the number of attack in that timestamp.

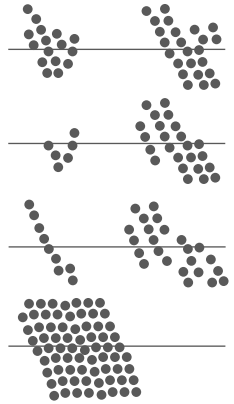


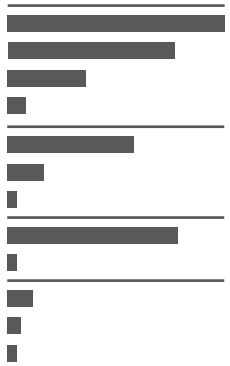
Visualizations

- Graph
- PCA
- Scatterplot of attacks (compared to time)
- Bar chart





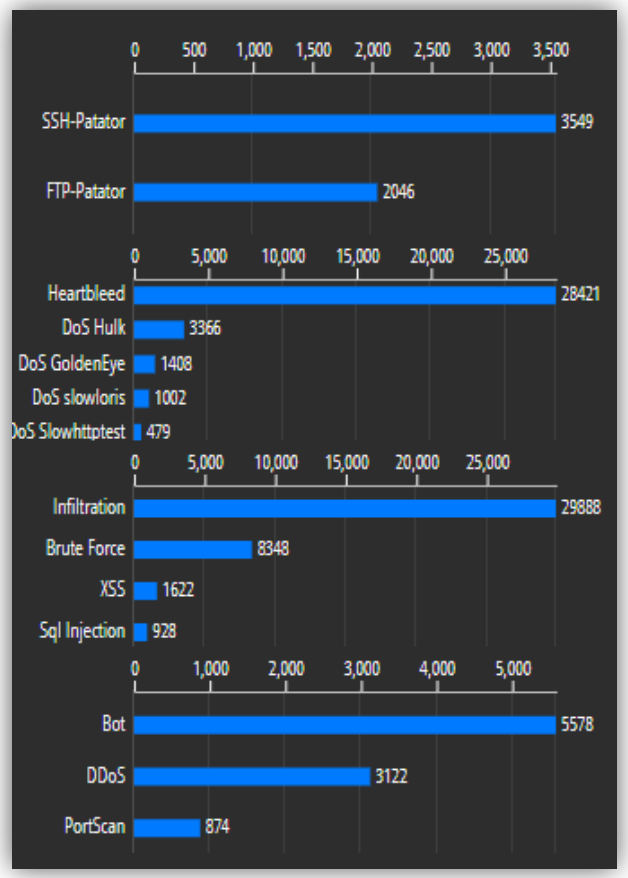




Bar Chart

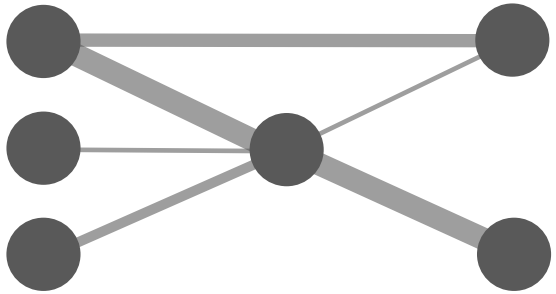
- **Axis**:
  - **X:** Number of attacks
  - **Y**: Type of attack

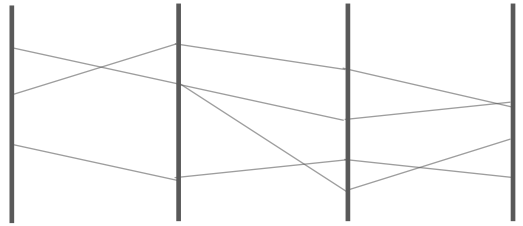
This graph gives us the opportunity to understand for each day the types of attack and the number of malicious packages sent.

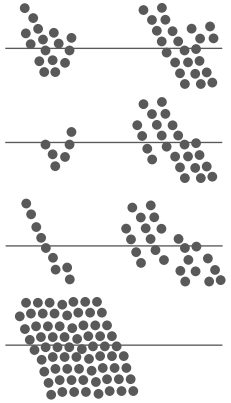


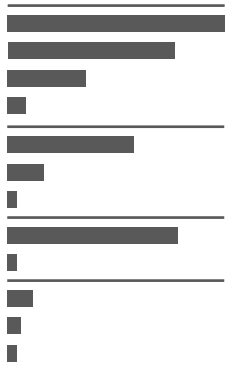
Visualizations

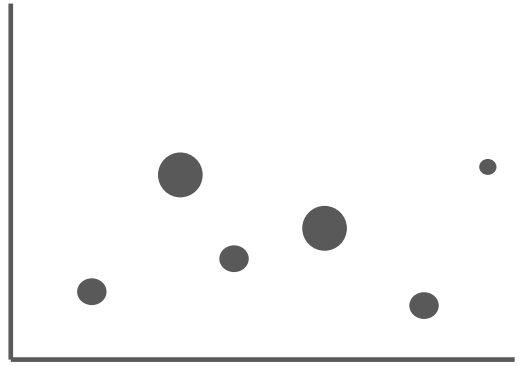
- Graph
- PCA
- Scatterplot of attacks (compared to time)
- Bar chart
- Scatterplot of Ports







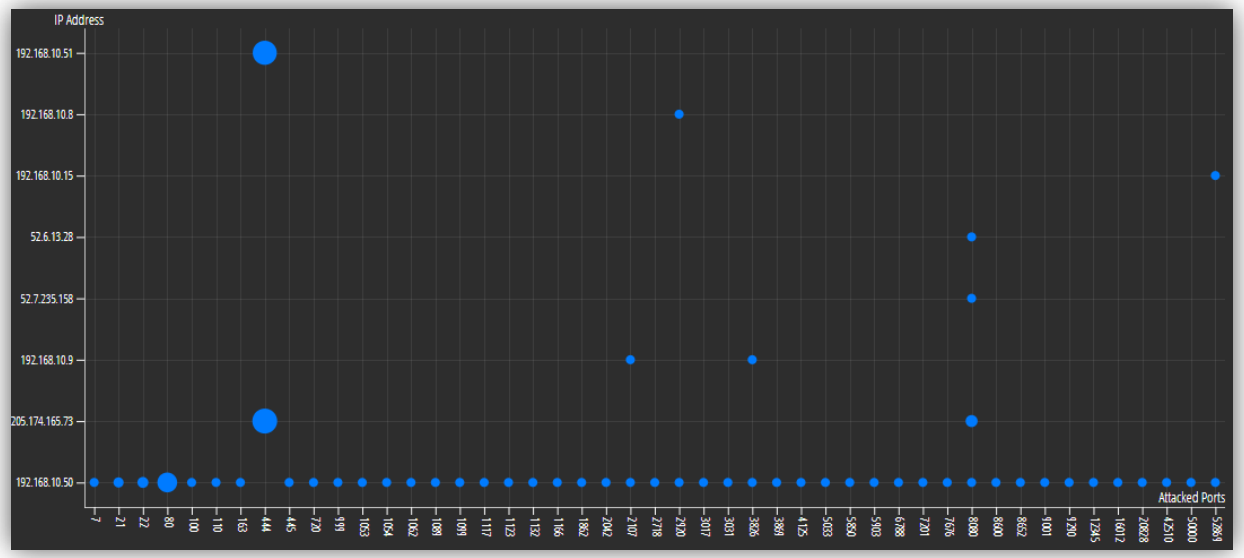




Scatterplot DESTINATION/PORTS

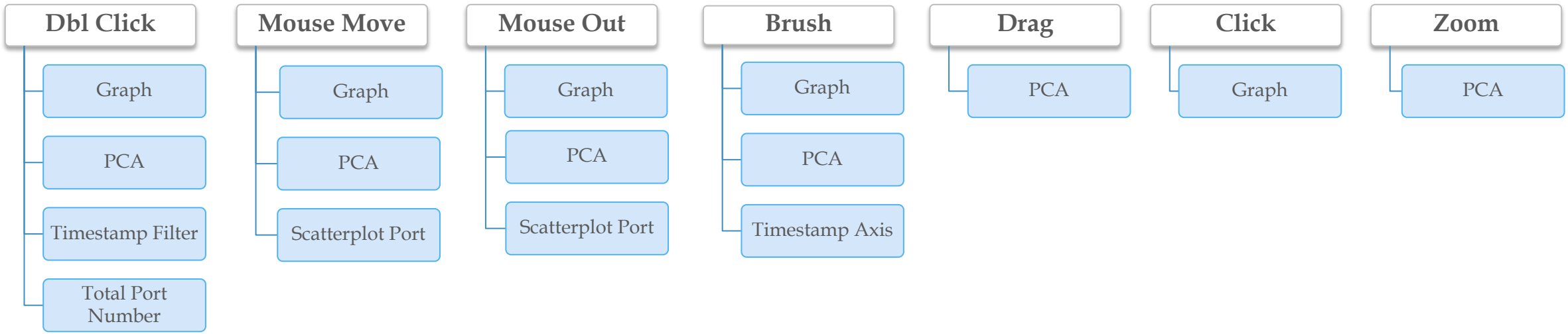
- **Axis**:
  - **X**: Destination Port
  - **Y**: Targets IP address

The dot changes size based on how many attacks has been taken by the couple IP address-Port



Pipeline

Actions


Dbl Click
Graph
PCA
Timestamp Filter
Total Port Number
Mouse Move
Graph
PCA
Scatterplot Port
Mouse Out
Graph
PCA
Scatterplot Port
Brush
Graph
PCA
Timestamp Axis
Drag
PCA
Click
Graph
Zoom
PCA

Actions

Double Click

With the double click, we give the opportunity to the user to reset the filters carried out. In addition, double-clicking on the graph's nodes allows you to deselect it.

Zoom

On the PCA we have given the possibility to zoom, this is useful when you want to make a brush on the axis and the values are very close.

Drag

The PCA axis can be changed with a drag: the user can see the relationships between the axis where it is not possible, as they are not close to the axis of interest.

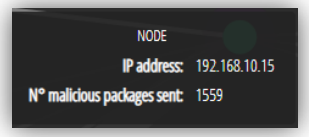
Mouse Out

When the mouse is no longer over objects that support the "mouse move", all actions triggered by it will be reset

Actions

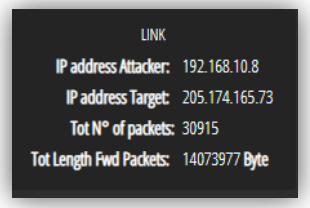
Mouse Move - LOCAL CHANGES

- **Graph:**
  - Node – Show tooltip with extra info:
    - IP address
    - Number of malicious packages sent or delivered



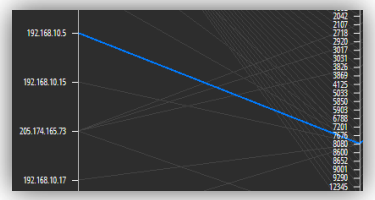
Node tooltip

  - Edge – Show tooltip with extra info:
    - IP address Attacker
    - IP address Target
    - Tot number of packets
    - Tot length of Fwd Packets [Byte]



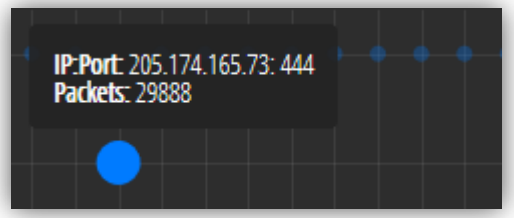
Link tooltip

- **PCA:**
  - Edge – Focuses on the edge, reducing the opacity of all the others.



Focused edge

- **Ports Scatterplot**
  - Dots – Focuses on the dot, reducing the opacity of all the others.
    - Show tooltip with extra info:
      - IP address
      - Port
      - Packets

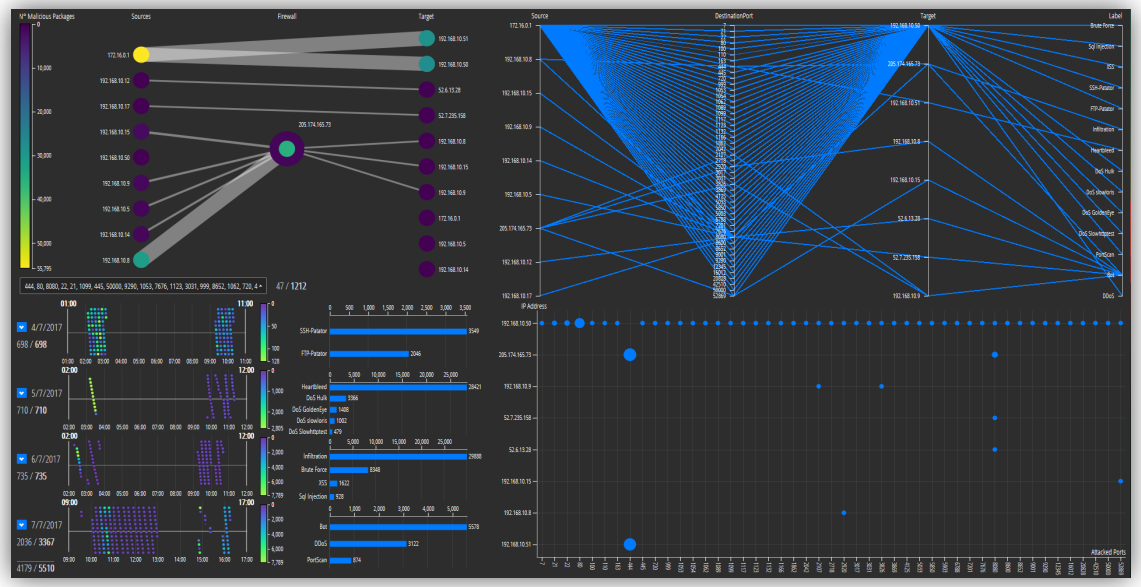


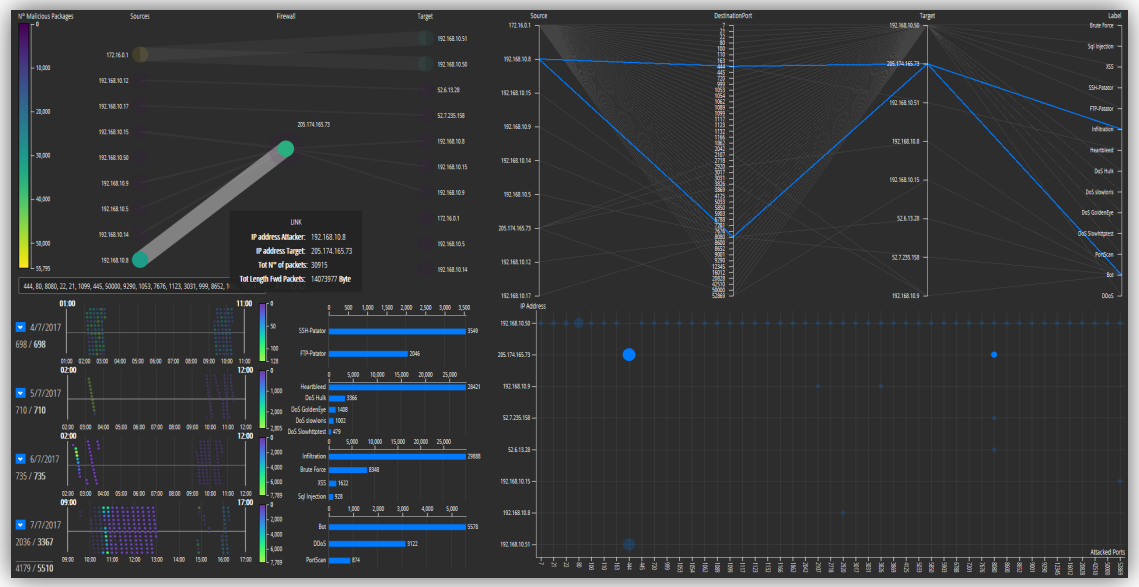
Focused dot + tooltip dot

Actions

Mouse Move - GLOBAL CHANGES

- **Graph:**
  - **Node** – PCA, Scatterplot Port, Scatterplot Over Time
  - **Edge** – PCA, Scatterplot Port, Scatterplot Over Time
- **PCA:**
  - **Edge** – Graph, Scatterplot Port, Scatterplot Over Time
- **Scatterplot Port:**
  - **Dots** – Graph, PCA, Scatterplot Over Time





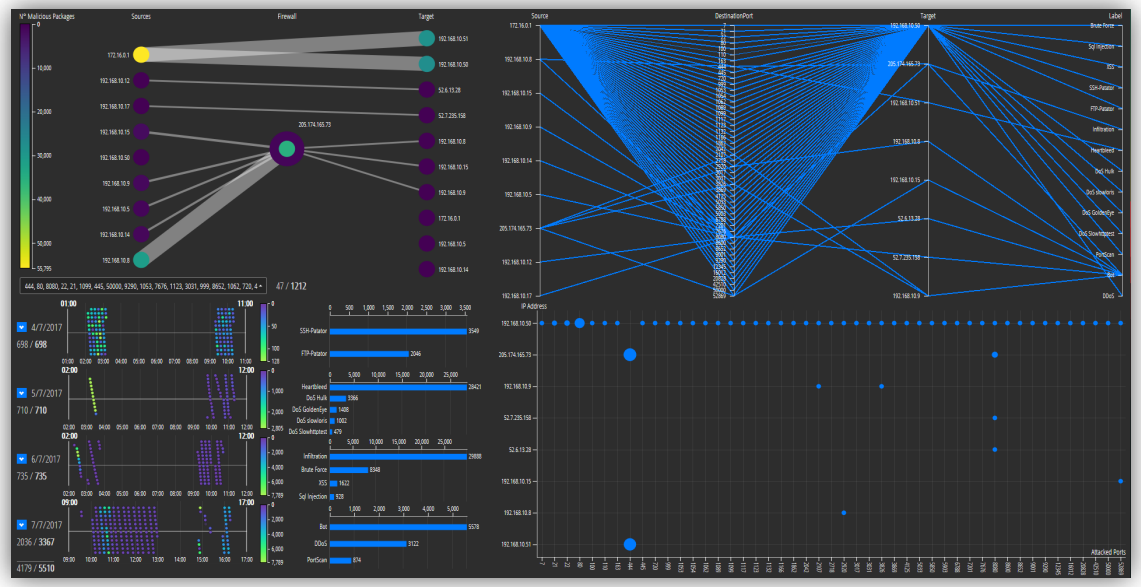
Actions

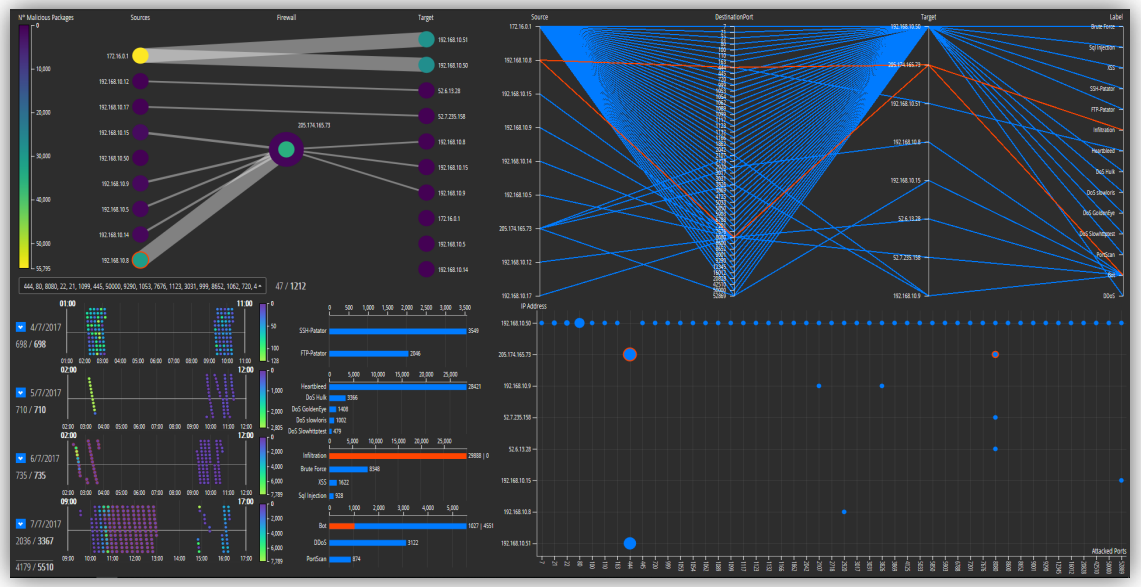
Click – GLOBAL AND LOCAL CHANGES

- **Graph:**
  - **Node –** PCA, Scatterplot Port, Scatterplot Over Time, Bar Chart

Selected

Not Selected

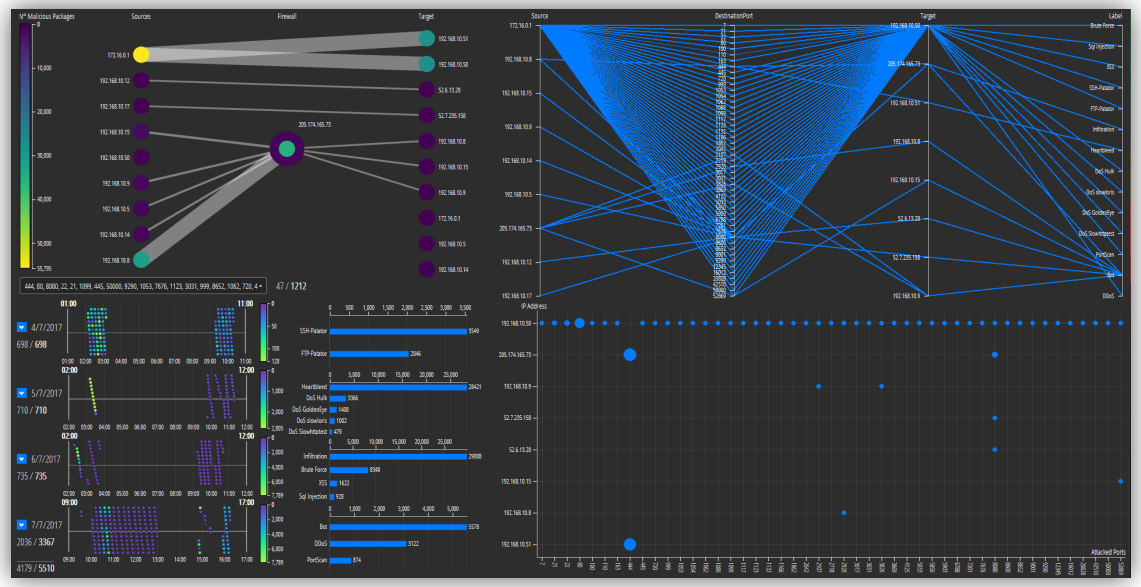


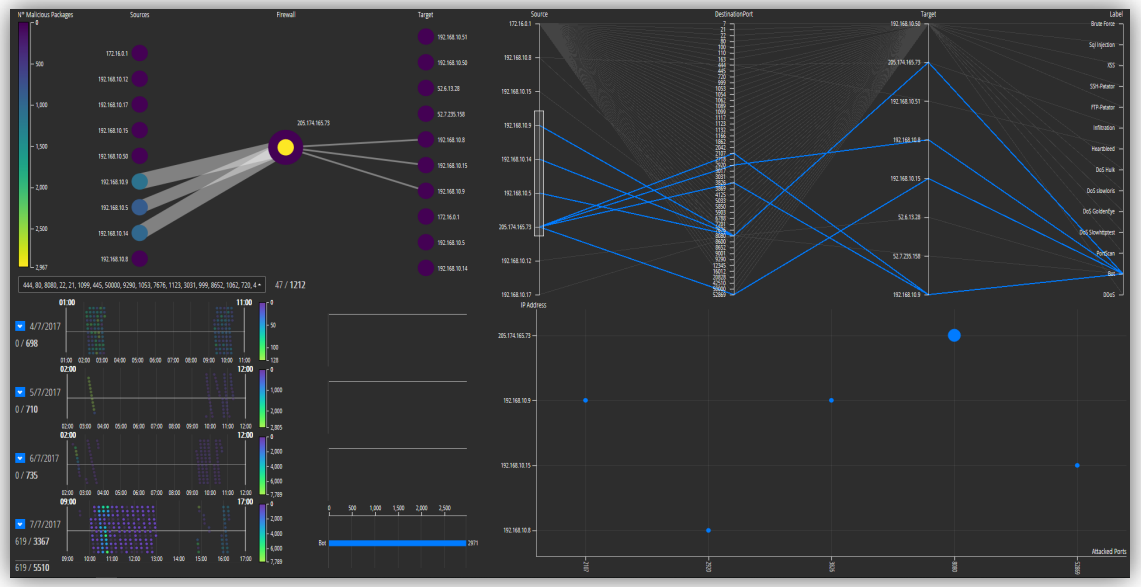


Actions

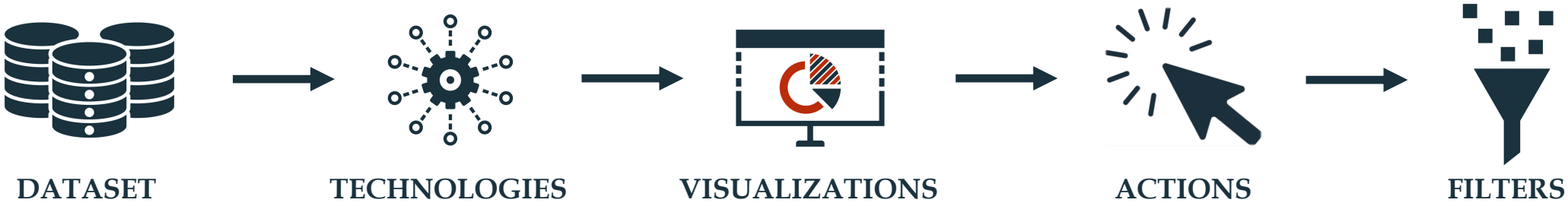
Brush – GLOBAL CHANGES

- **Graph:**
  - **Legend** CPA, Scatterplot Port, Scatterplot Over Time
- **Timestamp:**
  - **Axis** – PCA, Graph, Scatterplot Port, Bar Chart
- **PCA:**
  - **Axis** – PCA, Scatterplot Port, Scatterplot Over Time, Bar Chart





Pipeline

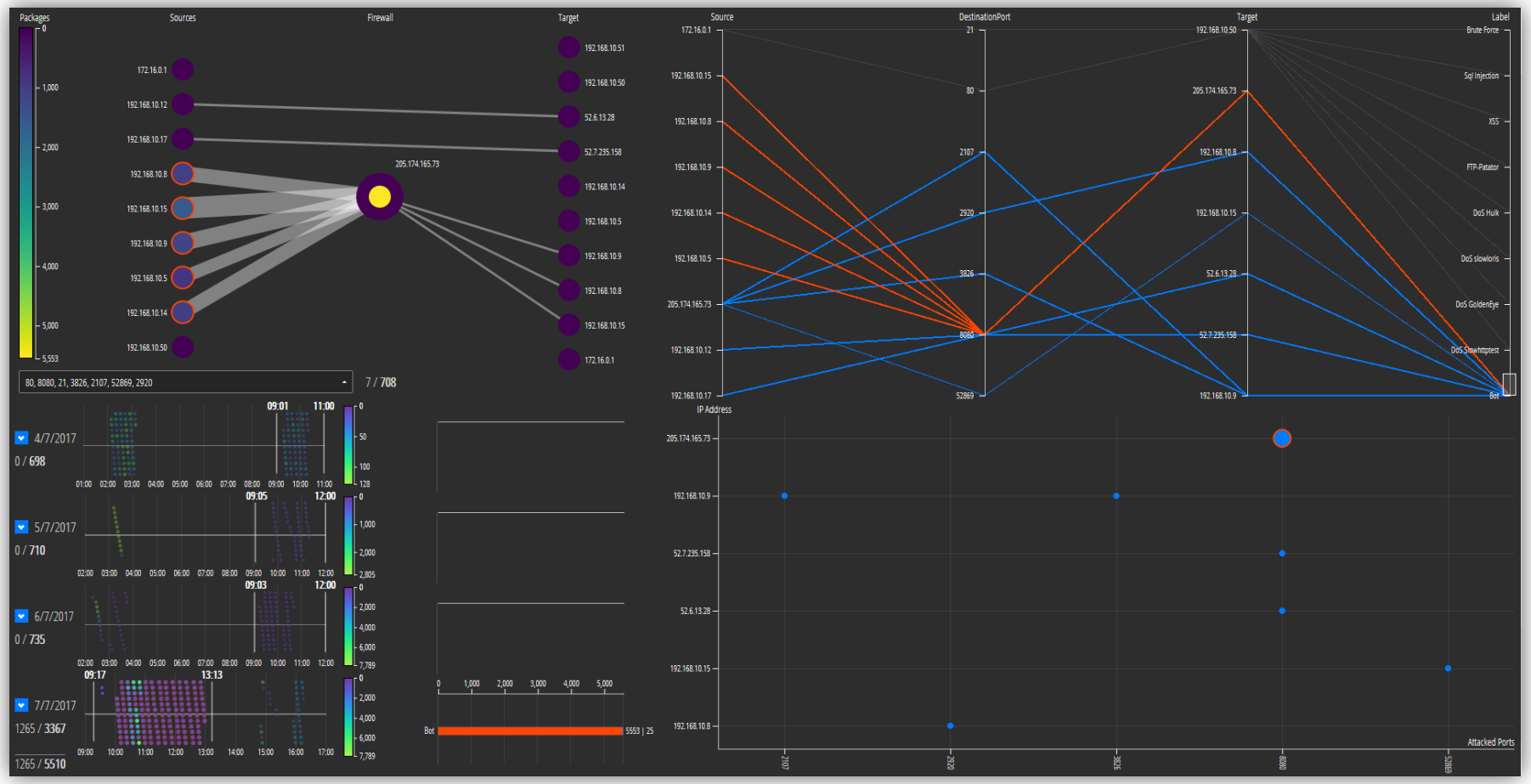


Filters

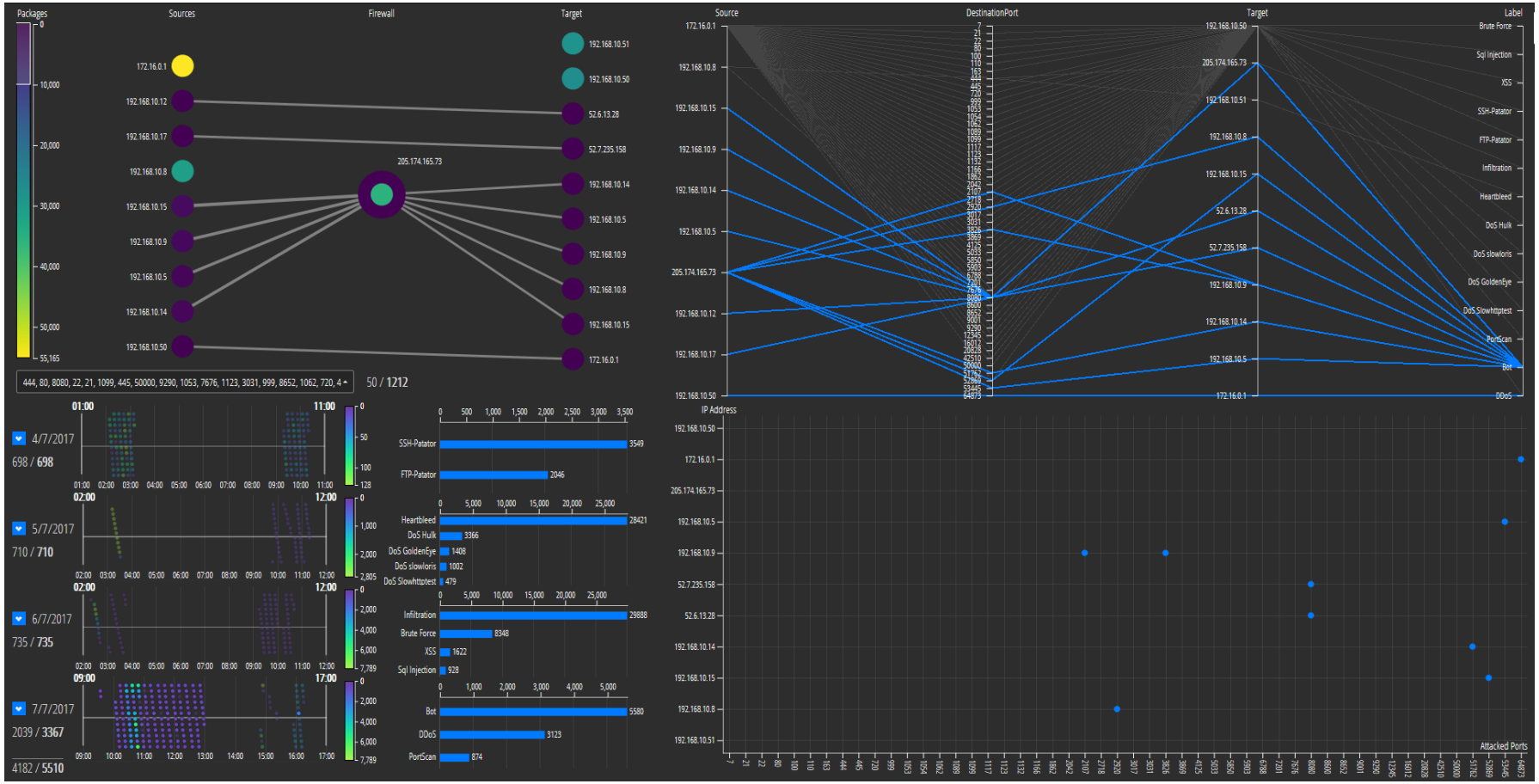
- **Timestamp:**
  - Axis – Allows you to filter the dataset by time, giving you the ability to filter by **day** and by **hours**.
- **PCA:**
  - Axis – Allows you to filter the dataset by **Source IP**, **Destination Port**, **Target IP** and **Type of Attack**
- **Graph**
  - Legend – Allows you to filter the view by number of **packets sen**t and **delivered**
- **Port filter**
  - With a multiple selection field, you can filter the **ports** and see the **percentage of use** based on the total

DEMO

1° Filtering: **Approx. 8:00 to 13:00 hours** (filtering on timestamp)
2° Filtering: **Selection of nodes that attack the firewall** (filtering on the graph)
3° Filtering: **Selection of the Bot connection to reduce the noise** (filtering on PCA)



1° Filtering: **Selection of ports 51762 53445 64873** (filtering on port selection form)
2° Filtering: **Number of packages on legend from 0 to 10000** (filtering on graph)



Observations

- The most common port is **444** (67%).
- **205.174.165.73** is a **firewall**.
- **172.16.0.1** is the attacker that sends the highest number of packets.
- the 5th is the day in which were presented the highest different types of attacks
- The attack usually appear during the **night/early morning**
- **Portscan** is the most frequent attack that appears on each port

References

- **GitHub**
  https://github.com/AlessandroGiannetti/VisualAnalytics
- **Slide**
  https://www.slideshare.net/AlessandroGiannetti3/visual-network-analysis-148780273/AlessandroGiannetti3/visual-network-analysis-148780273

- **Alessandro Giannetti**
  https://www.linkedin.com/in/alessandro-giannetti-2b1864b4/
- **Gianluca Tasciotti**
  https://www.linkedin.com/in/gianluca-tasciotti-77124b174/