Network Security
389.159 - SS 2018
Lab Exercise 3 & Lab Exercise 4

TEAM 02
Corentin Bergès (11741629) (066 506)
Christoph Echtinger-Sieghart (00304130) (066 938)

June 6, 2018

List of Corrections

Error: Include row descriptions and expand . . . 1
Error: Answer question . . . 1
Error: Wording! . . . 2
Error: tcp . . . 3
Error: icmp . . . 3
Error: udp . . . 3
Error: Wording . . . 4

1 Lab Exercise 3

1.1 rep-10

Matlab Code (Listing 2)
Figure 1 shows the stem plots for packets, bytes, IP sources and IP destinations per hour.

Optional Figure 2 shows all signals from Figure 1 combined, normalized and smoothed with a moving average filter.

1.2 rep-11

Matlab Code (Listing 3)
The signal that shows the lower correlation to the other signals is **IP sources**. The minimum linear correlation coefficient is **0.588568** between the signals **IP sources** and **IP destinations**. See Table 1 for the raw data.

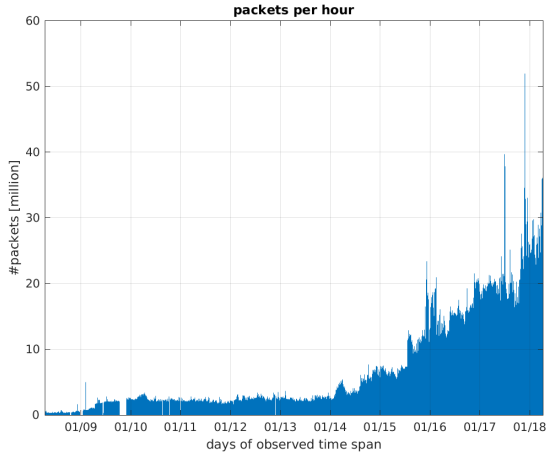
FiXme Error Include row descriptions expand

Bytes	Packets	IPs	IPd
0.9655	0.9655	0.7203	0.9340
0.7203	0.6105	0.6105	0.9732
0.9340	0.9732	0.5886	0.5886

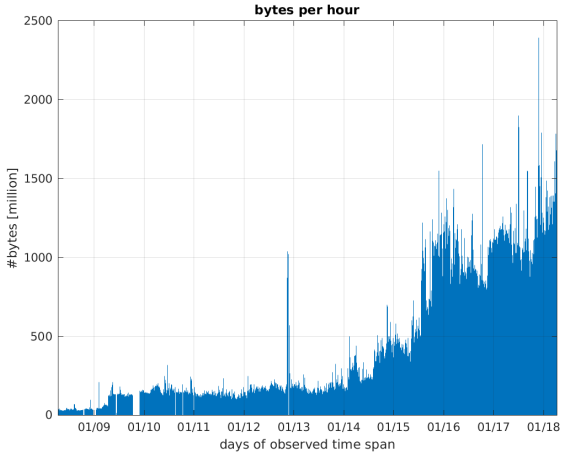
Table 1: Correlation coefficients between signals

The reason why the drop in unique IP sources does not cause a proportional drop ...

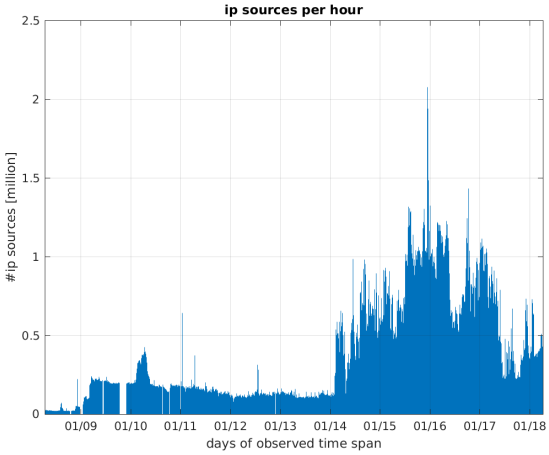
FiXme Error Answer ques



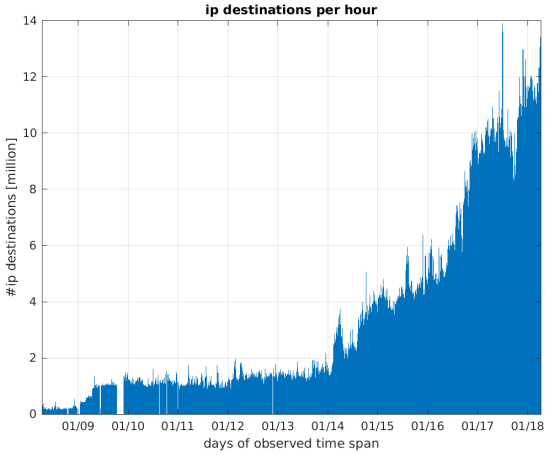
(a) Packets per hour



(b) Bytes per hour



(c) IP sources per hour



(d) IP destinations per hour

Figure 1

1.3 rep-12

Matlab Code (Listing 4)

FiXme Error Wording!

There are around ten times more IP sources than IP destinations. It makes sense that the number of IP sources is significantly bigger than the number of IP destinations because the darkspace is only a part of the whole internet.

1.4 rep-13

Matlab Code (Listing 5)

The main peak in IP sources starts at 14-Dec-2015 and lasts until 16-Dec-2015. See Table 2 for the detailed data.

Date	# IP sources
14-Dec-2015	2075358.074306
15-Dec-2015	1704892.012500
16-Dec-2015	1942072.404167

Table 2: Detailed data for peak in IP sources

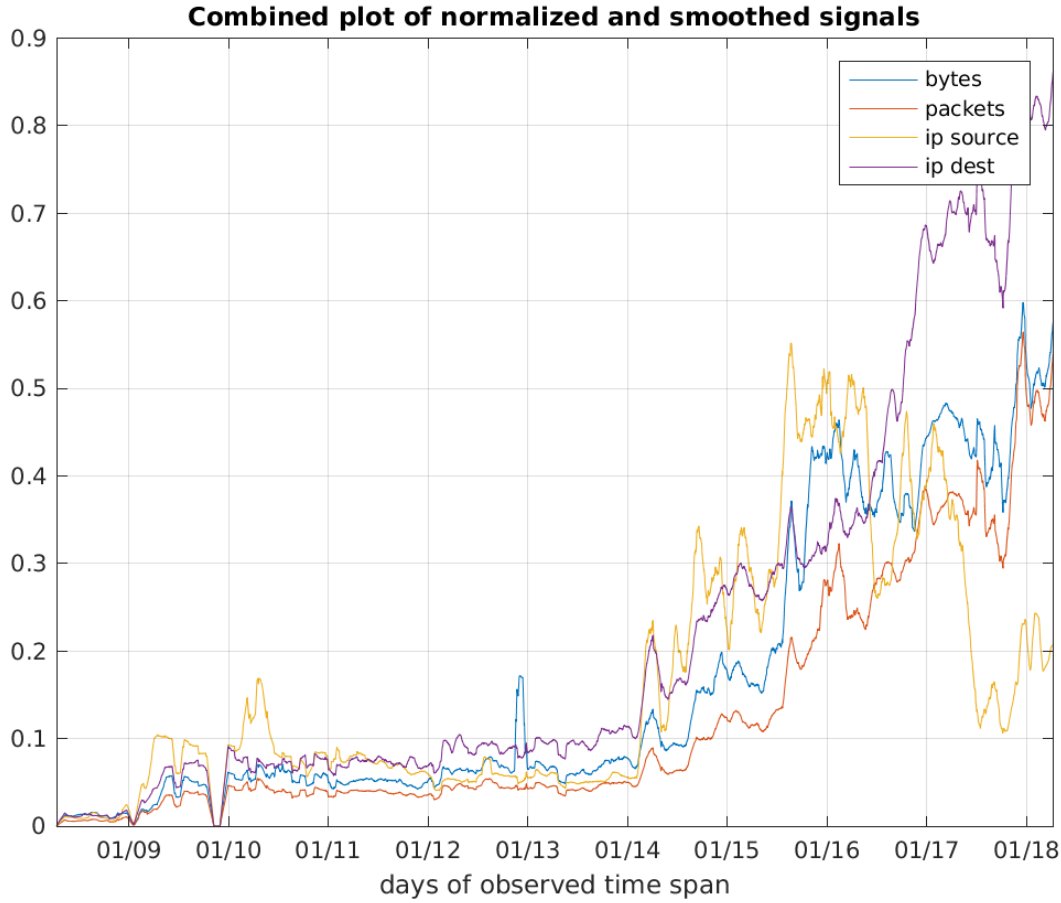


Figure 2: Combined, normalized and smoothed signals

Optional Matlab Code (Listing 6) The main peak in Bytes starts at 14-Nov-2012 and lasts until 22-Nov-2012. Note that on 19-Nov-2012 no data was available. See Table 3 for the detailed data.

1.5 rep-14

Matlab Code (Listing 7)
Table 4 gives statistics for the data from `global_last10years.csv`. Table 5 gives statistics for the data from `Feb2017_gen.csv`.

1.6 rep-15

The values do not coincide. Feb2017 seems to be a month that is not really representative for the whole span of 10 years.

optional Matlab Code (Listing 8)

1.7 rep-16

We used https://www.iana.org/assignments/protocol-numbers/protocol-numbers.xhtml to look up the protocol numbers.

Protocol 6 (TCP) The Transmission Control Protocol FiXme Error

Protocol 1 (ICMP) The Internet Control Message Protocol FiXme Error icmp

Protocol 17 (UDP) The User Datagram Protocol FiXme Error udp

Date	# Bytes
14-Nov-2012	870858582.136110
15-Nov-2012	1009586335.331900
16-Nov-2012	1038654926.456100
17-Nov-2012	1021464983.022200
18-Nov-2012	954193481.914190
20-Nov-2012	1005163238.508500
21-Nov-2012	1020526661.658000
22-Nov-2012	989613880.615110

Table 3: Detailed data for peak in Bytes

	Sum	Mean	Median	StdDev
# Packets	146373.391	41.845	17.699	40.916
# Bytes	2381.003	0.681	0.263	0.735
# IP src	123.613	0.035	0.020	0.031
# IP dst	1150.796	0.329	0.142	0.330

Table 4: Statistics for daily data [in millions]

1.8 rep-17

1.9 rep-18

We obtained negative values because of collapsing ...

FiXme Error Wording

1.10 rep-19

1.11 rep-20

1.12 rep-21

1.13 rep-22

1.14 rep-23

Listing 1: Command used to obtain IP address

```
team02@pc01:~$ ip address show dev em1
```

Port 113

```
IP 192.168.83.20.1073 > 192.168.83.33.113: Flags [S], seq 0, win 8192, length 0
IP 192.168.83.33.113 > 192.168.83.20.1073: Flags [R.], seq 0, ack 1, win 0, length 0
```

	Sum	Mean	Median	StdDev
# Packets	76871.319	114.392	113.464	7.033
# Bytes	1272.998	1.894	1.890	0.097
# IP src	59.651	0.089	0.091	0.018
# IP dst	619.875	0.922	0.931	0.070

Table 5: Statistics for hourly data [in millions]

2 Lab Exercise 4

2.1 rep-24

2.2 rep-25

2.3 rep-26

2.4 rep-27

2.5 rep-28

2.6 rep-29

2.7 rep-30

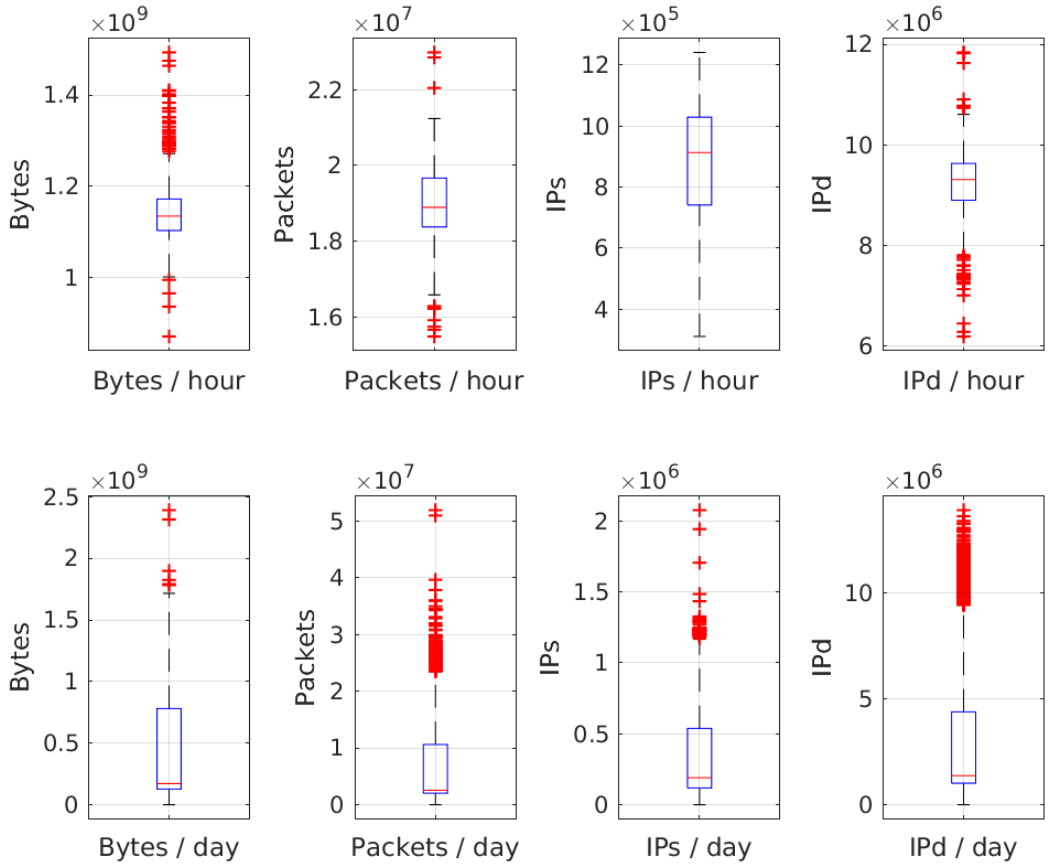


Figure 3: Boxplots for hourly and daily averaged data

A Matlab Code

Listing 2: Matlab code to solve rep-10

```
function team02_rep10
% rep-10
    [timestamps, bytes, packets, ip_s, ip_d] = read_custom_csv('~/workfiles/global_last10years.csv')

    function save_stem_plot(data, my_title, y_label, filename)
    % Do a stem plot of data in millions and write it to filename.png
        set (gca, 'fontname', 'Helvetica', 'fontsize', 20)
        figure
        stem(timestamps, data/10^6, 'marker', 'none')
        datetick('x', 'mm/yy');
        xlabel('days of observed time span');
        ylabel(y_label);
        title(my_title);
        grid on
        set(gca, 'layer', 'top');
        xlim([min(timestamps) max(timestamps)]);
        saveas(gcf, filename, 'png')
    end

    save_stem_plot(bytes, 'bytes per hour', '#bytes [million]', 'plots/rep_10_2');
    save_stem_plot(packets, 'packets per hour', '#packets [million]', 'plots/rep_10_1');
    save_stem_plot(ip_s, 'ip sources per hour', '#ip sources [million]', 'plots/rep_10_3');
    save_stem_plot(ip_d, 'ip destinations per hour', '#ip destinations [million]', 'plots/rep_10_4')

    % optional part

    function result = smooth_filter(data)
    % Moving averages filter for data
        window_size = 30;
        b = (1 / window_size) * ones(1, window_size);
        a = 1;
        % 1-D digital filter
        result = filter(b, a, data);
    end

    smooth_bytes = smooth_filter(bytes / unique(max(bytes)));
    smooth_packets = smooth_filter(packets / unique(max(packets)));
    smooth_ip_s = smooth_filter(ip_s / unique(max(ip_s)));
    smooth_ip_d = smooth_filter(ip_d / unique(max(ip_d)));

    figure
    plot(...
        timestamps, smooth_bytes, '-', ...
        timestamps, smooth_packets, '-', ...
        timestamps, smooth_ip_s, '-', ...
        timestamps, smooth_ip_d, '-' ...
    );
    legend('bytes', 'packets', 'ip source', 'ip dest');
    datetick('x', 'mm/yy');
    xlabel('days of observed time span');
    title('Combined plot of normalized and smoothed signals');
    grid on
    set(gca, 'layer', 'top');
    xlim([min(timestamps) max(timestamps)]);
    saveas(gcf, 'plots/rep_10_optional', 'png')
end
```

Listing 3: Matlab code to solve rep-11

```
function team02_rep11
% rep-11
    [~, bytes, packets, ip_s, ip_d] = read_custom_csv('~/workfiles/global_last10years.csv');

    function result = correlation(a, b)
        result = unique(min(corrcoef(a, b)));
    end

    names = { ...
        'Bytes <-> Packets', 'Bytes <-> IPs', 'Bytes <-> IPd', ...
        'Packets <-> IPs', 'Packets <-> IPd', 'IPs <-> IPd' ...
    };
    correlations = [ ...
        correlation(bytes, packets), correlation(bytes, ip_s), ...
        correlation(bytes, ip_d), correlation(packets, ip_s), ...
        correlation(packets, ip_d), correlation(ip_s, ip_d) ...
    ];
    [minimum_coeff, idx] = min(correlations);
    fprintf('Minimum linear correlation coeff: %f (%s)\n', minimum_coeff, names{idx});

    names_signal = {'Bytes', 'Packets', 'IPs', 'IPd'};

    means = [ ...
        % Bytes         | Packets         | IPs           | IPd
        correlations(1), correlations(1), correlations(2), correlations(3); ...
        correlations(2), correlations(4), correlations(4), correlations(5); ...
        correlations(3), correlations(5), correlations(6), correlations(6)
    ];
    disp(names_signal);
    disp(means);
end
```

Listing 4: Matlab code to solve rep-12

```
function team02_rep12
    [~, ~, ~, ip_s, ip_d] = read_custom_csv('~/workfiles/global_last10years.csv');
    ip_s(ip_s==0) = NaN;
    ip_d(ip_d==0) = NaN;

    fprintf('Ratio IPs to IPd: %f\n', nanmean(ip_s) / nanmean(ip_d));
end
```

Listing 5: Matlab code to solve rep-13

```
function team02_rep13
    [timestamps, ~, ~, ip_s, ~] = read_custom_csv('~/workfiles/global_last10years.csv');
    % from visual inspection
    cutoff = 1.5*10^6;
    peak_locations = ip_s>cutoff;

    peak_timestamps = timestamps(peak_locations);
    peaks = ip_s(peak_locations);

    dates = arrayfun(@datestr, peak_timestamps, 'UniformOutput', false);
    result = dates';
    result(2,:) = num2cell(peaks);
    fprintf('%s: %f IPs\n', result{:});
end
```

Listing 6: Matlab code to solve rep-13 optional

```
function team02_rep13_optional
    [timestamps, bytes, ~, ~, ~] = read_custom_csv('~/workfiles/global_last10years.csv');
    % From visual inspection
    cutoff = 8*10^8;
    timestamps = timestamps(timestamps<=datenum('2014-01-01'));
    bytes = bytes(timestamps>0);

    peak_locations = bytes>cutoff;
    peak_timestamps = timestamps(peak_locations);
    peaks = bytes(peak_locations);

    dates = arrayfun(@datestr, peak_timestamps, 'UniformOutput', false);
    result = dates';
    result(2,:) = num2cell(peaks);
    fprintf('%s: %f Bytes\n', result{:});
    % NOTE: There is a gap because on 19-nov-2012 there was no data
end
```

Listing 7: Matlab code to solve rep-14

```
function team02_rep14

    function result = stats(data)
        data(data==0) = NaN;
        result = round([nansum(data), nanmean(data), nanmedian(data), nanstd(data)] ./ 10e6, 3);
    end

    disp('---- Daily avg ---');
    [~, bytes, packets, ip_s, ip_d] = read_custom_csv('~/workfiles/global_last10years.csv');
    for col = horzcat(bytes, packets, ip_s, ip_d)
        fprintf('%.3f %.3f %.3f %.3f\n', stats(col));
    end

    disp('----- Hourly avg ---');

    % WARNING: order is different
    [~, packets, bytes, ip_s, ip_d] = read_custom_csv('~/workfiles/Feb2017_gen.csv');
    for col = horzcat(bytes, packets, ip_s, ip_d)
        fprintf('%.3f %.3f %.3f %.3f\n', stats(col));
    end
end
```

Listing 8: Matlab code to solve rep-15 optional

```
function team02_rep15_optional
    [~, bytes_daily, packets_daily, ip_s_daily, ip_d_daily] = read_custom_csv('~/workfiles/global_la
    % WARNING order is different
    [~, packets_hourly, bytes_hourly, ip_s_hourly, ip_d_hourly] = read_custom_csv('~/workfiles/Feb20

    set (gca, 'fontname', 'Helvetica', 'fontsize', 20)

    ax1 = subplot(2,4,1);
    boxplot(ax1, bytes_hourly, 'Labels', {''})
    ylabel(ax1, 'Bytes');
    xlabel(ax1, 'Bytes / hour');
    grid on
    set(gca, 'layer', 'top');

    ax2 = subplot(2,4,2);
    boxplot(ax2, packets_hourly, 'Labels', {''})
    ylabel(ax2, 'Packets');
    xlabel(ax2, 'Packets / hour');
    grid on
    set(gca, 'layer', 'top');

    ax3 = subplot(2,4,3);
    boxplot(ax3, ip_s_hourly, 'Labels', {''})
    ylabel(ax3, 'IPs');
    xlabel(ax3, 'IPs / hour');
    grid on
    set(gca, 'layer', 'top');

    ax4 = subplot(2,4,4);
    boxplot(ax4, ip_d_hourly, 'Labels', {''})
    ylabel(ax4, 'IPd');
    xlabel(ax4, 'IPd / hour');
    grid on
    set(gca, 'layer', 'top');

    ax5 = subplot(2,4,5);
    boxplot(ax5, bytes_daily, 'Labels', {''})
    ylabel(ax5, 'Bytes');
    xlabel(ax5, 'Bytes / day');
    grid on
    set(gca, 'layer', 'top');

    ax6 = subplot(2,4,6);
    boxplot(ax6, packets_daily, 'Labels', {''})
    ylabel(ax6, 'Packets');
    xlabel(ax6, 'Packets / day');
    grid on
    set(gca, 'layer', 'top');

    ax7 = subplot(2,4,7);
    boxplot(ax7, ip_s_daily, 'Labels', {''})
    ylabel(ax7, 'IPs');
    xlabel(ax7, 'IPs / day');
    grid on
    set(gca, 'layer', 'top');

    ax8 = subplot(2,4,8);
    boxplot(ax8, ip_d_daily, 'Labels', {''})
    ylabel(ax8, 'IPd');
    xlabel(ax8, 'IPd / day');
    grid on
    set(gca, 'layer', 'top');

    saveas(gcf, 'plots/rep_15_optional.png', 'png')
end
```