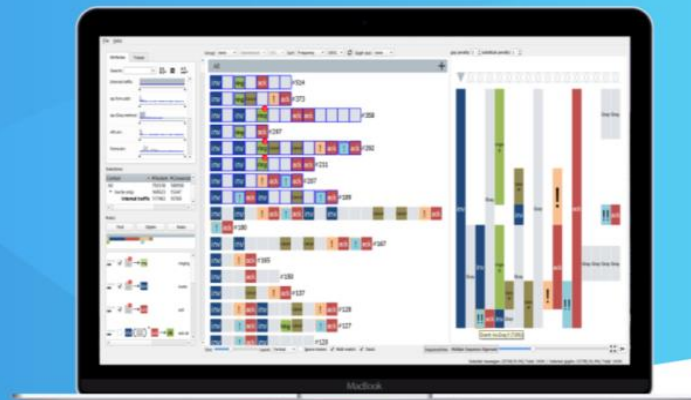




The Notepad editor for **Event Data**
Discover patterns in event data using visual analytics.

[Get started](#)


Manual

Version: 20-04-2018

PRODUCT

[Features](#)
[Plugins](#)
[Downloads](#)

CONTACT US

Email: b.c.m.cappers@tue.nl

Phone: +31 40 247 8863

Contents

Opening CSV data:	2
Analyzing PCAP traffic:	4
The system:	5
Rule Construction:	6
Basic rule construction:	7
Creation of highlight rules:	8
The constraint interface:	9
Rewrite rule creation using Regular Expressions	9
Minimap	10
Data Operations:	11
Stack	11
Sorting	11
Partitioning	11
Block scaling	12
Alignment	13
Attribute view:	15
Right-click menu options:	16
Shortcuts in the Sequence View:	18

Opening CSV data:

The first line of the CSV file, must contain the headers (comma separated).

Example file structure:

Header1,header2,header3

1,hello,12-03-2016

2,world,13-03-2016

When starting the application, click File -> Open CSV:

Input file: D:/Backup Master/Conferenties/Vis 2017/export 11gb full_0.csv Open

Group events by:

☐ None

☒ Attribute sip.Call-ID

☐ Combined

src: frame.len

dst: frame.len

sesid: frame.len

Settings (optional):

Time attribute: frame.time_epoch

Src field: frame.time_epoch

Dst field: none

	1	2
1	frame.time_epo...	Numeric
2	frame.len	Numeric
3	frame.protocols	Text
4	ip.geoip.dst_co...	Text
5	ip.geoip.src_co...	Text

Limit: 800000

Open Export Cancel

Figure 1 Open CSV Dialog

<i>Group events by:</i>	<p>This options determines how the events should be grouped in to sequences.</p> <p>In this example the phone calls will be grouped by Call-ID. This way every block sequence will correspond to a phone call and every block correspond to a message in that call.</p> <p>If you want to group the data over multiple columns (say per phone call, per day), the combined option can be chosen. The order in which the fields are selected does not matter.</p> <p>If the option none is selected, the data will be put into one large sequence.</p>
<i>Time attribute:</i>	<p>If your data contains an attribute with unix timestamps, you can specify them here. Eventpad will use this information to calculate sequence durations and plot temporal data in a linechart.</p> <p>In case there are no unix timestamps, the following datetime formats are also supported: "yyyy-MM-dd hh:mm:ss" "dd-MM-yyyy hh:mm:ss" "dd-MM-yyyy hh:mm:ss" "yyyy/MM/dd hh:mm:ss"</p>
<i>Src dst fields:</i>	Deprecated
<i>Limit:</i>	Maximum number of records Eventpad will load

When opening the CSV file, Eventpad will determine for every attribute its type. Attributes can be categorical (e.g., "Opel, Audi, Chevrolet"), numerical (e.g., "0.2, 0.5, 0.1"), or identifiers (e.g., "UI001", "UI002", "UI003") . The type of attribute will determine what you can do with the attributes (see section **Attribute View**).

Analyzing PCAP traffic:

Eventpad enables users to convert PCAP traffic to CSV provided that they have installed Wireshark. Eventpad will use the tshark application via de console to perform the conversion. Selecting the File → Convert PCAP option will result in the following window:

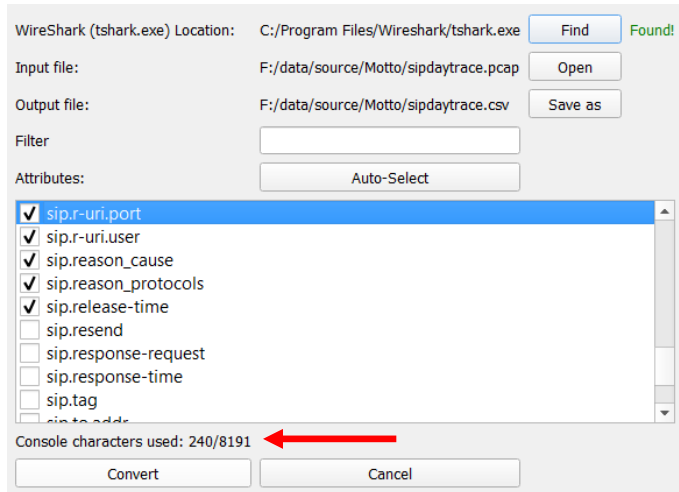


Figure 2 Convert PCAP Dialog

<i>Location:</i>	Select the location where tshark.exe is located.
<i>Filter:</i>	Deprecated: You can specify additional tshark filters in this field
<i>Attributes:</i>	Deprecated

When opening a PCAP file, Eventpad will extract all protocol fields from the first 1500 packets. These will be listed in the table. Before conversion users can select which protocol fields have to be dissected.

Since for Windows systems the console is limited to 8191 characters, there is an upperbound on the number of protocol fields that can be dissected using this method. The red arrow shows how many characters there are still available.

The system:

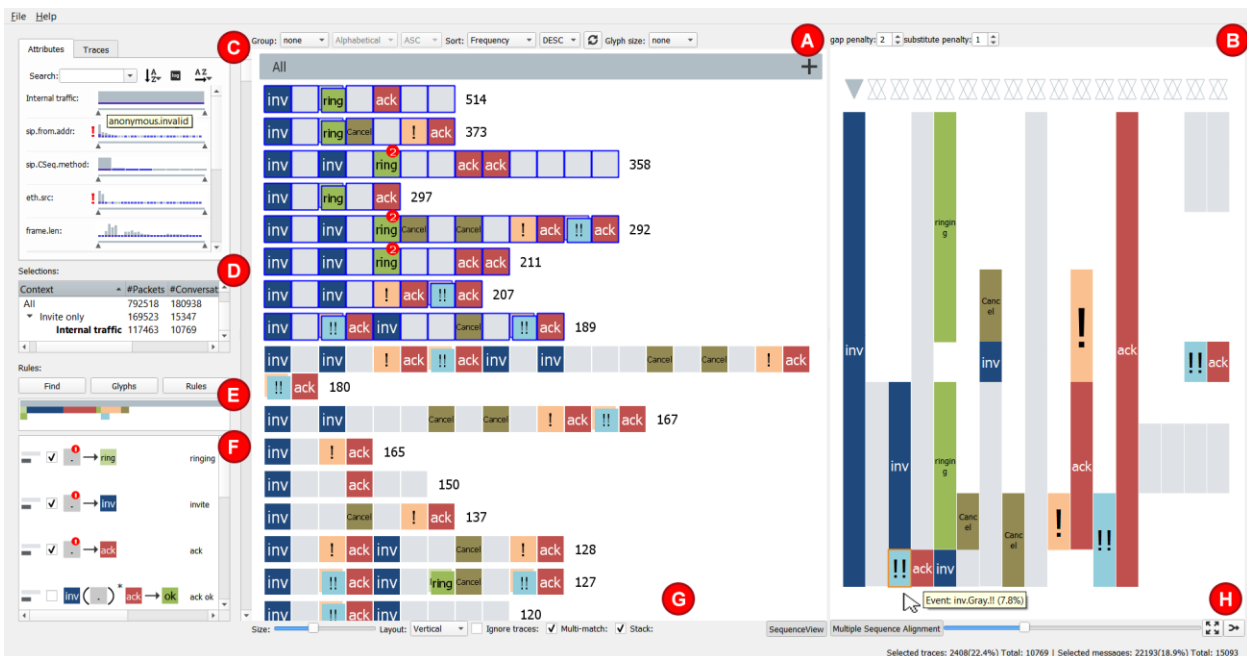


Figure 3 Screenshot Eventpad System

The Eventpad system consists of the following components (See figure 3):

A) The sequence view visualizes event sequences as series of glyphs, starting every sequence on a new row. Event sequences that do not fit on a single line are wrapped over multiple rows. Scroll bars are used to inspect the entire data set. Since Eventpad focusses on the reduction and analysis of event patterns, time-intervals between events are disregarded in the visualization.

B) The alignment view tries to discover patterns between different sequences through *alignment*. The iced plot on the right shows how often events

C) The attribute view shows a histogram of every event property in the data. The histogram shows how often event values occur with respect to that attribute and are interactive. Selecting a bar in the histogram for instance will highlight all events in the sequence view with that property. Vice versa, selecting a group of events in the sequence view will show in the Attribute view what values they have in common. The sliders under the histograms can be used to restrict selections within certain boundaries. In case of categorical attributes, the red exclamation mark shows up whenever that attribute has more 40 unique values (more about this in Section **Attribute View**)

D) The context view enables users to save selected events of interest into a new context by assigning a name to them. The creation of a new context results in a new attribute separating the selected events from the non-selected. This attribute is added to the data and can be used for further analysis and drill down, enabling analysts to tag the data with more domain-specific information throughout exploration.

E) The Rule overview widget shows how much percent of the dataset is affected by a certain rule. This can help you to understand the coverage of a certain rule. Hovering the mouse over one of the block

shows the exact coverage percentage.

F) The Rule view shows a list of all rules that are applied in the system. The ordering for rewriting is controlled via drag and drop operations. Rules can be toggled on or off along with longest and shortest match settings.

G) and H) are additional options to change the visualizations.

Every Rule in the Rule View is constructed as follows:

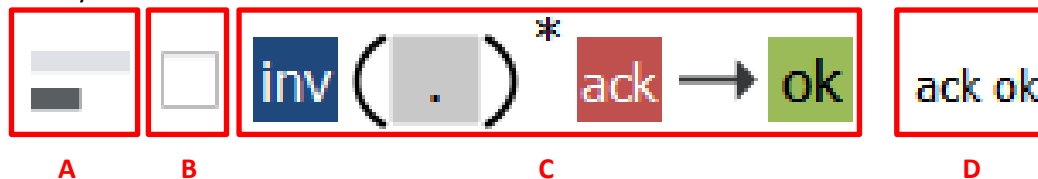


Figure 4 Visual summary of the rule options in the Rule view

A) The first icon enables the user to switch between longest and shortest matching.

Example:

Suppose we have the sequence: **abab** and we want to replace the data with the following rule **a.*b → c** (i.e., a sequence that starts with an 'a' and ends with a 'b' and I don't care what is in between should be replaced by a 'c')

Shortest match (indicated with the short bar, default) results in the sequence: cc

Longest match (indicated with the long bar) results in the sequence: c

B) The checkbox enables or disables a particular rule.

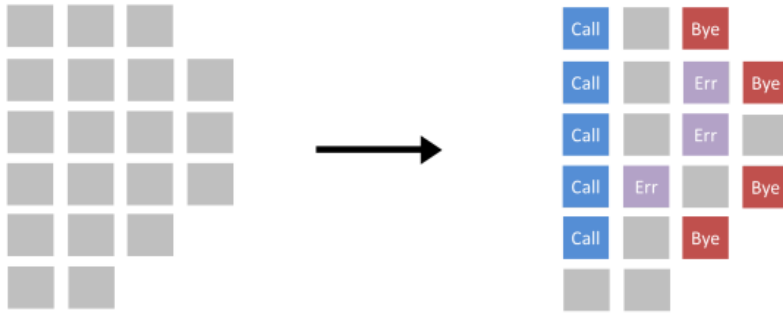
C) The third part is a thumbnail of the constructed rule.

D) The fourth part represents the name of that rule

Rule Construction:

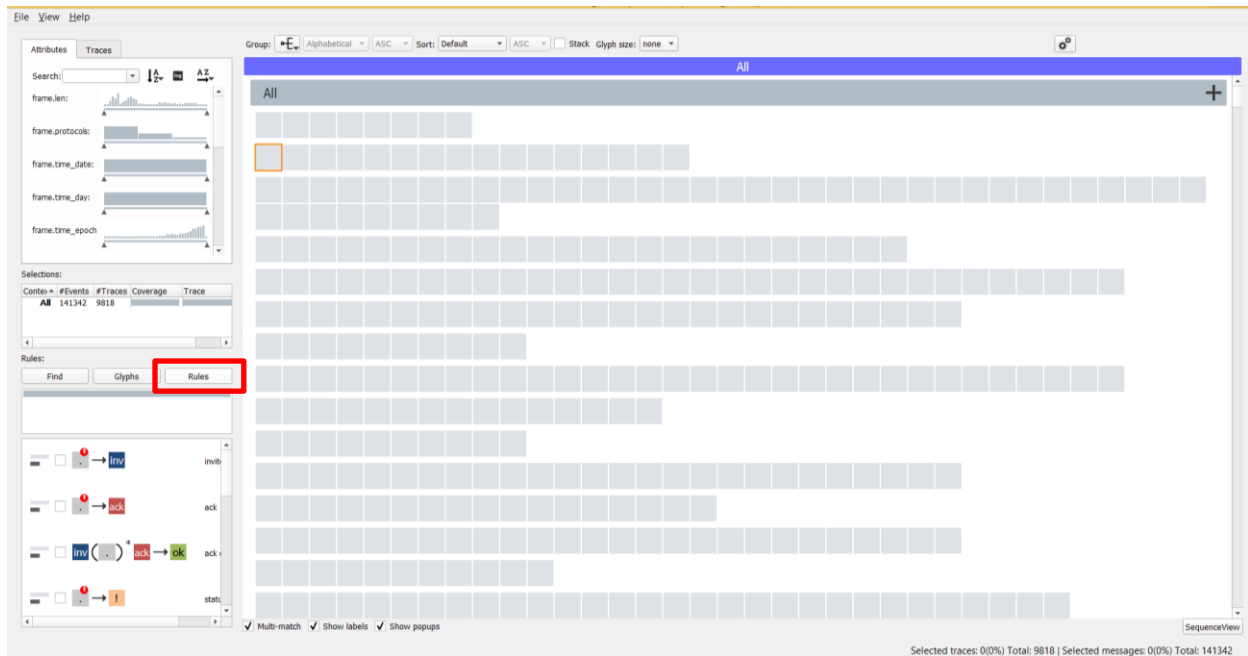
Rules enable users to highlight event properties that are of interest and to simplify data using find&replace functionality (similar to a notepad editor). Operators such as sequential composition, choice, and iteration (0 or more times) can be used to construct patterns. Figure 2 shows how these operators are visually encoded in the interface. Similar to Word's equation editor, operators can be combined to construct more complex patterns.

The "Wildcard" block (indicated in Gray with a ".") corresponds to any block. Double-clicking a block in the query interface results in a textual interface (Figure 3). This enables users to specify constraints over attributes and values in the data. Now only blocks that satisfy those constraints are replaced by the rules right hand side. For the right hand side of a rule, users can design their own blocks by choosing a particular shape, color and/or label. Earlier defined glyphs can be reused for the creation of new rules. In general, rules are used for three purposes, namely **highlighting**, **filtering**, and **data rewriting**.

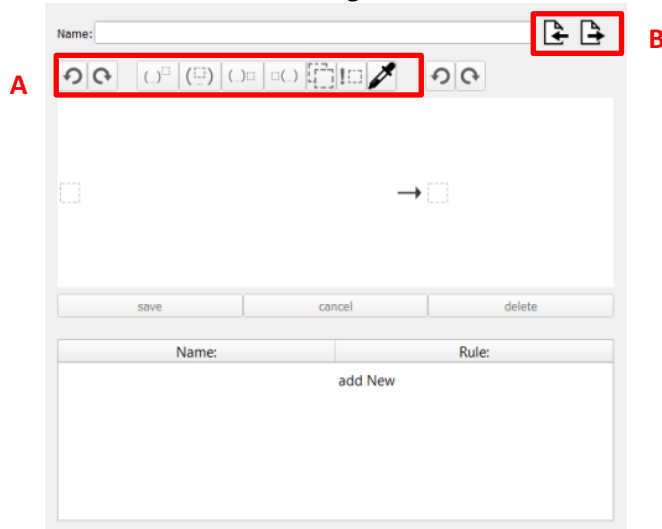


Basic rule construction:








Click on the rule button in the GUI.




This result in the following interface:



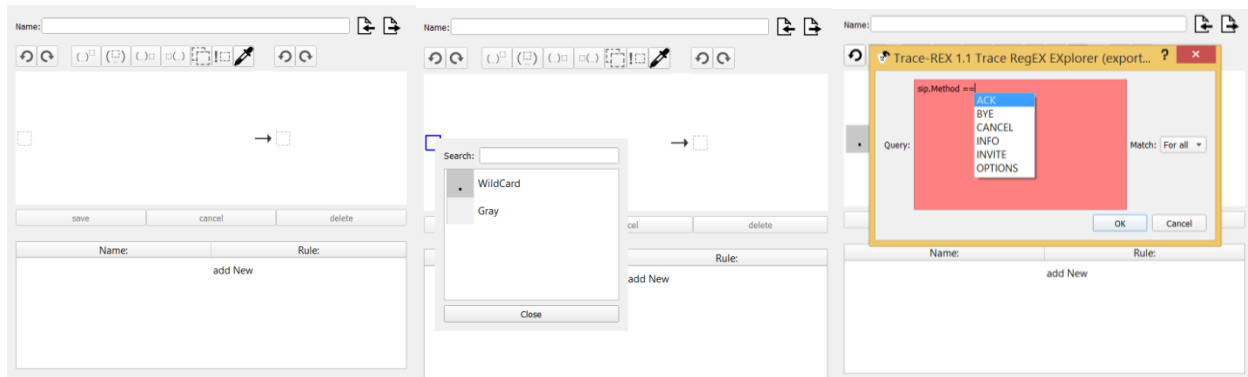
Controls in box A:

Buttons:	Description:
	Undo and redo functionality when constructing block patterns
	Adds an iterator around the currently selected blocks. This enables users to specify how often the pattern should occur
	Adds a block above the selected block expression. Vertical block represent a choice (either one or the other should happen)
	Next and previous buttons can be used to put a new block to the right or left of the selected block expression. This enables users to create sequential patterns
	The stack button enables users to search for events that correspond to two or more blocks at the same time.
	The negation operator can be used to match "anything but" the selected block pattern
	The pipet tool enables users to quickly select blocks in the Sequence view instead of manually constructing them in the interface.

Controls in box B:

Buttons:	Description:
	Eventpad can import and export rule sets respectively

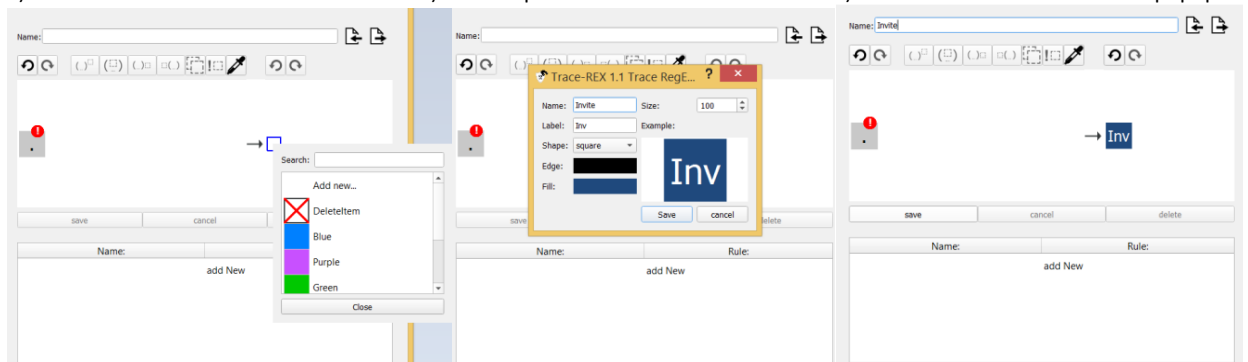
Creation of highlight rules:



a) Initial interface

b) Select square to choose block

c) Double click block results in a popup



d) Select Add new to create a new block

e) Choose shape, color etc. for the new block

f) Resulting rewrite rule.

The constraint interface:

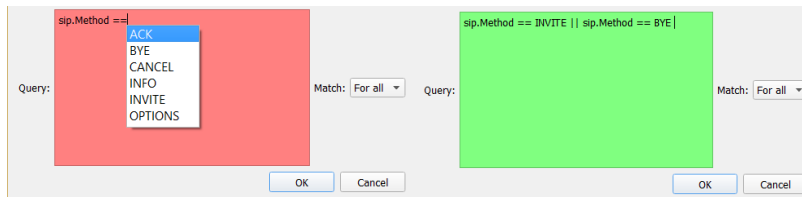


Figure 5 Invalid Query

Figure 6 Valid Query

The constrain interface enables users to reason about block with specific properties. The query can be an arbitrary complex Boolean expression.

The expression is always of the form:

<Expression> → <attribute name> <operator> <value> [<&& (And) || (Or) >> <Expression>]

Autocompletion is provided when typing the query.

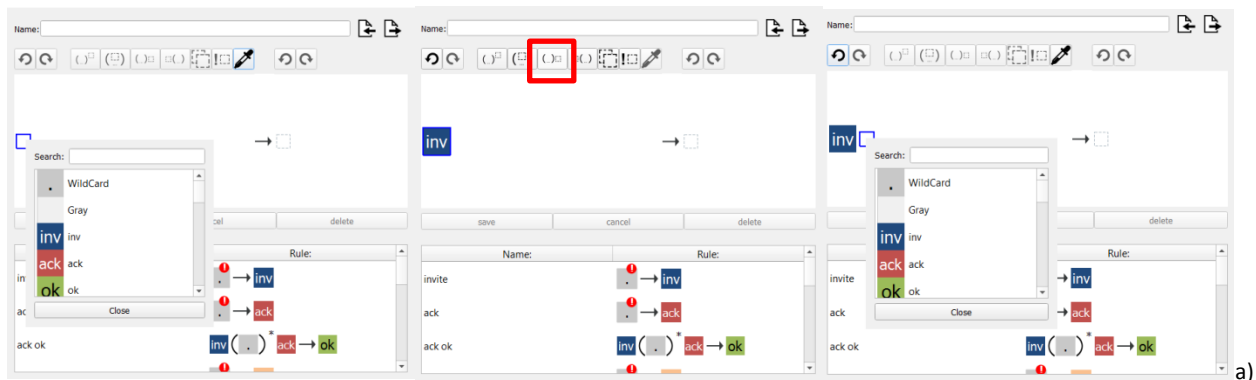
The following operations are supported:

<=	Less than or equal
>=	Greater than or equal
<	Less than
>	Greater than
==	Equal
!=	Not equal
?=	Contains (e.g., sip.method ?= "vite" → true if the value of the event in the column sip.method contains vite)

In case values contain spaces, the value must be surrounded by quotes ("").

Make sure that there are no spaces in the attribute names!!!

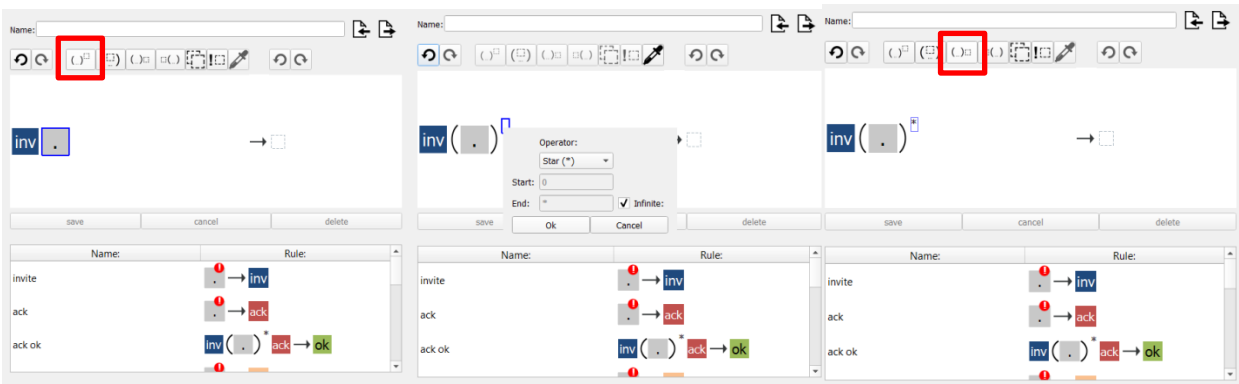
Rewrite rule creation using Regular Expressions



Select square to choose block

b) Click red rectangle for next block

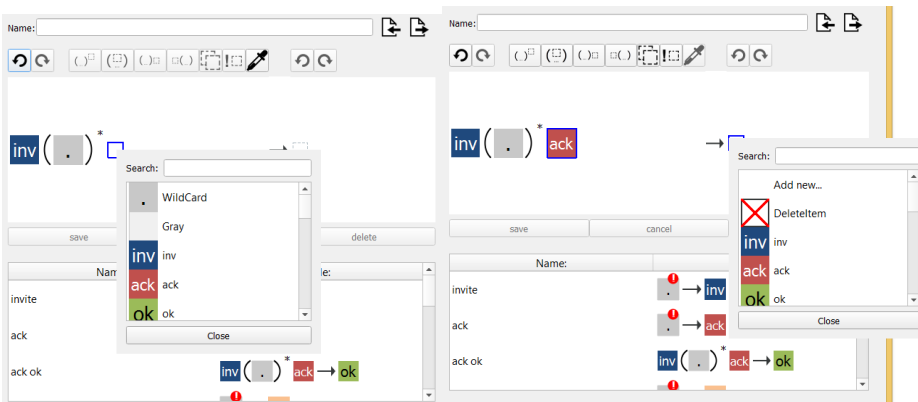
c) Select next block



d) Select iteration for block repetition

e) Block occurs 0 or more times

f) Click square for next block

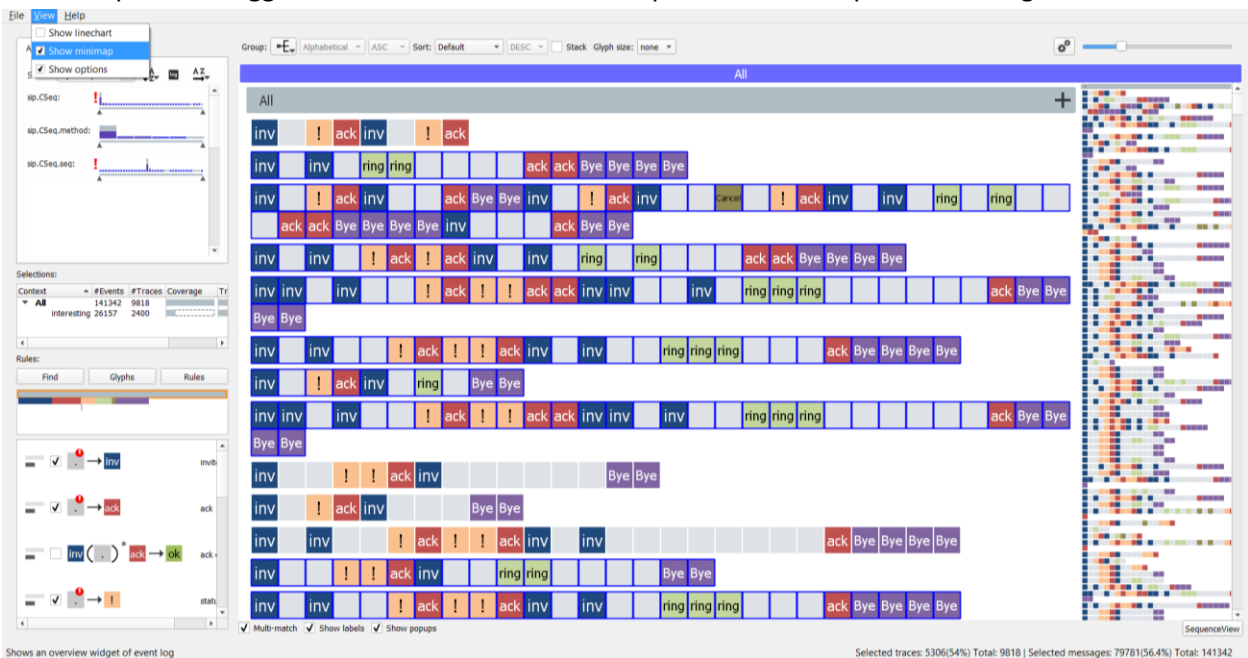


h) Select block of interest

i) Select block to rewrite the block pattern to

Minimap

A minimap can be toggled in the View → Show Minimap. This will show patterns in larger collections



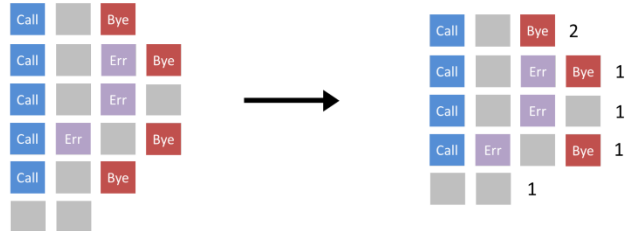
Data Operations:

After loading in the data, Eventpad enables users to perform the following operation:

Stack

Group: Alphabetical ASC Sort: Default ASC ☒ Stack Glyph size: none

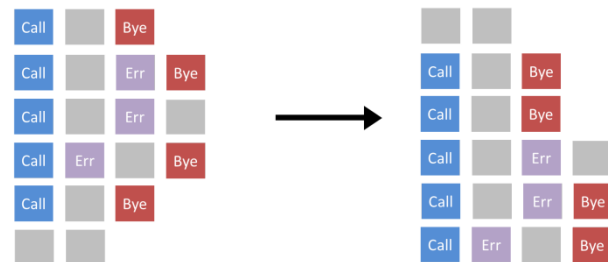
Result:



Sorting

Group: Alphabetical ASC Sort: Default ASC ☐ Stack Glyph size: none

Result:

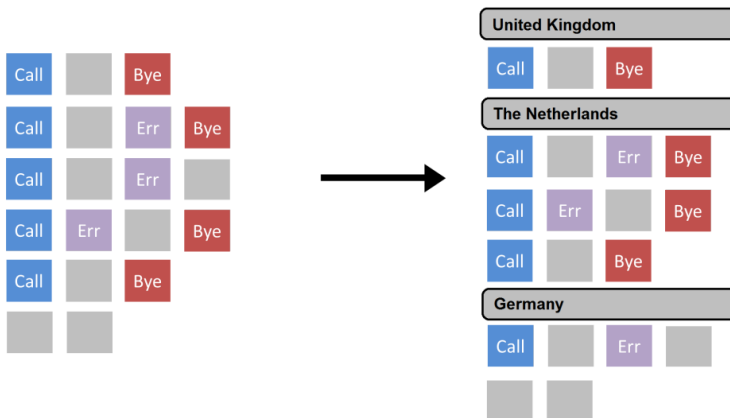


Event sequences can be s

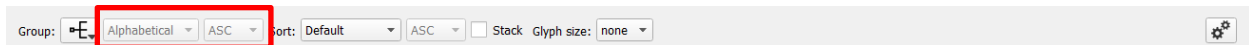
Partitioning

Group: ☒ Alphabetical ASC Sort: Default ASC ☐ Stack Glyph size: none

Result:

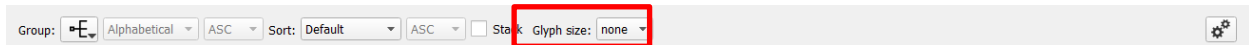


Groups can be sorted using controls:



Block scaling

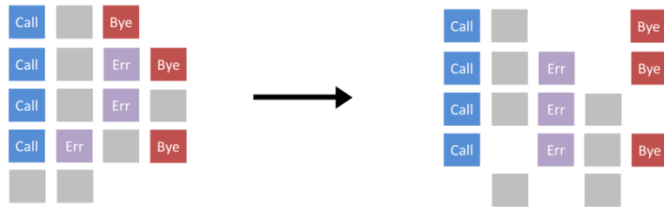
The size of the blocks can be set proportional to some numeric attribute of chose.



Result:



Alignment



To enable the Alignment view, click on the button indicated in the red square (Figure 5):

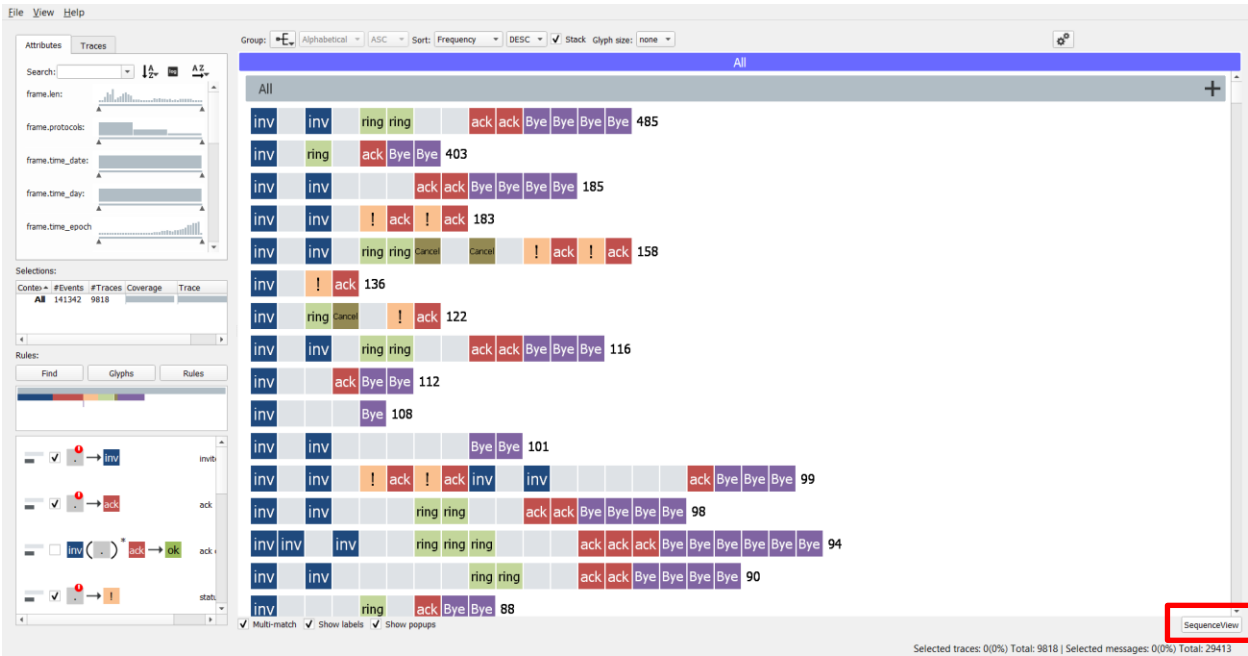
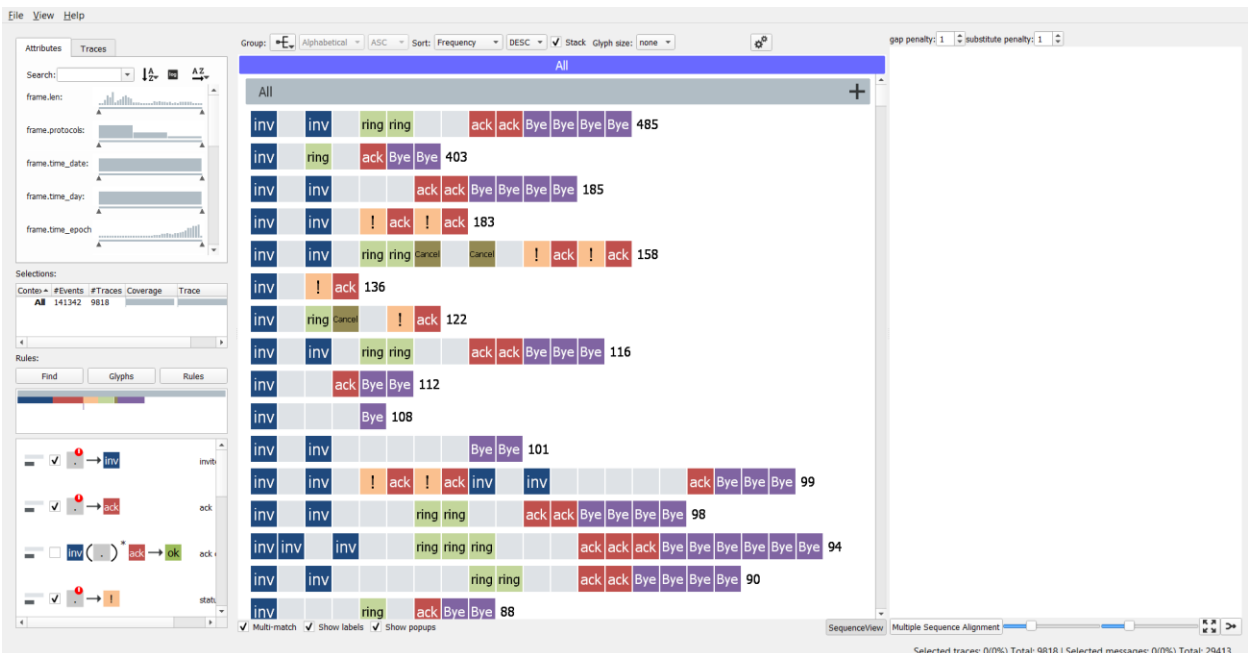


Figure 6 Sequence view

Next make a selection in the Sequence view that you want to align:



This results in an icicle plot showing for every position in the sequence how often every event occurs. When pressing the “multiple sequence alignment” button, the computer will start finding patterns in the data (Figure 7).

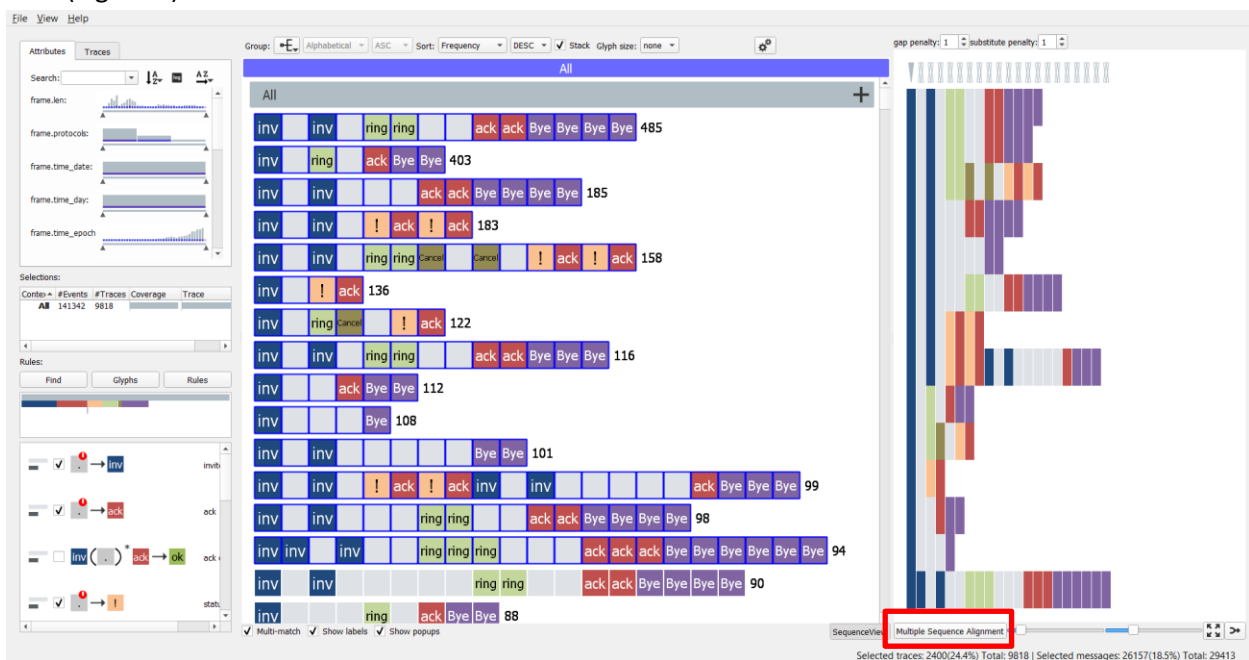


Figure 7 Example application of data alignment

Resulting Alignment is shown on the right. Users can adjust parameter settings such as gap cost to make the alignment more compact etc (red square, Figure 8).

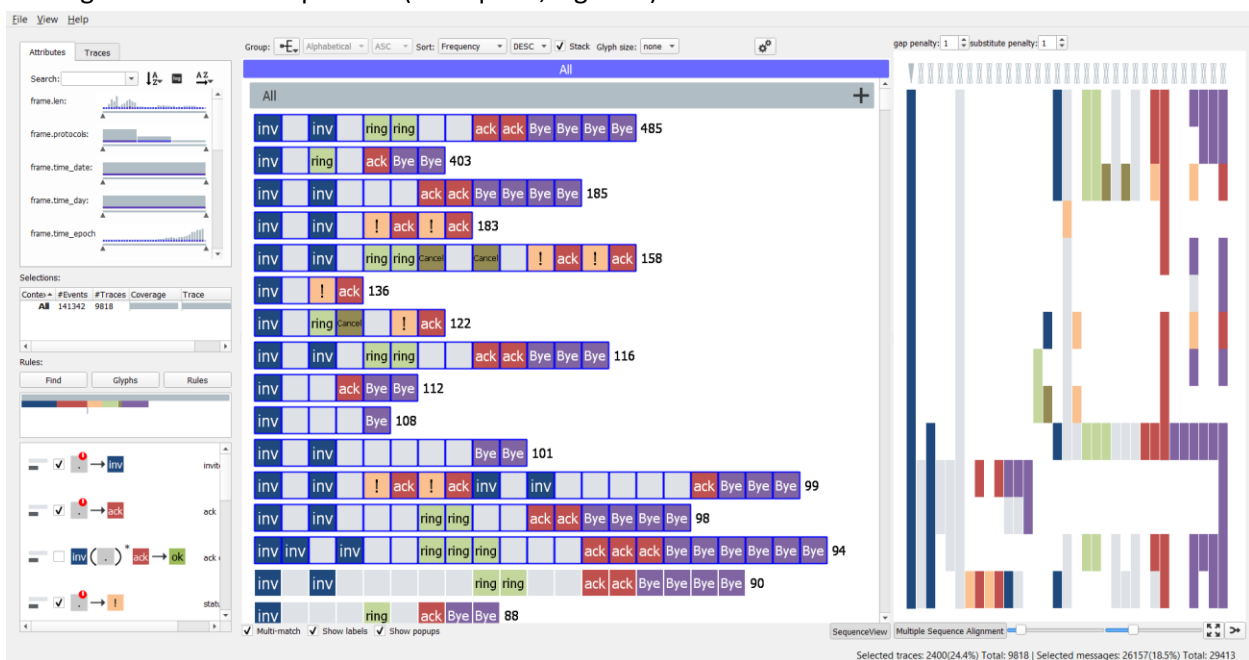


Figure 8 Parameter settings are indicated in red.

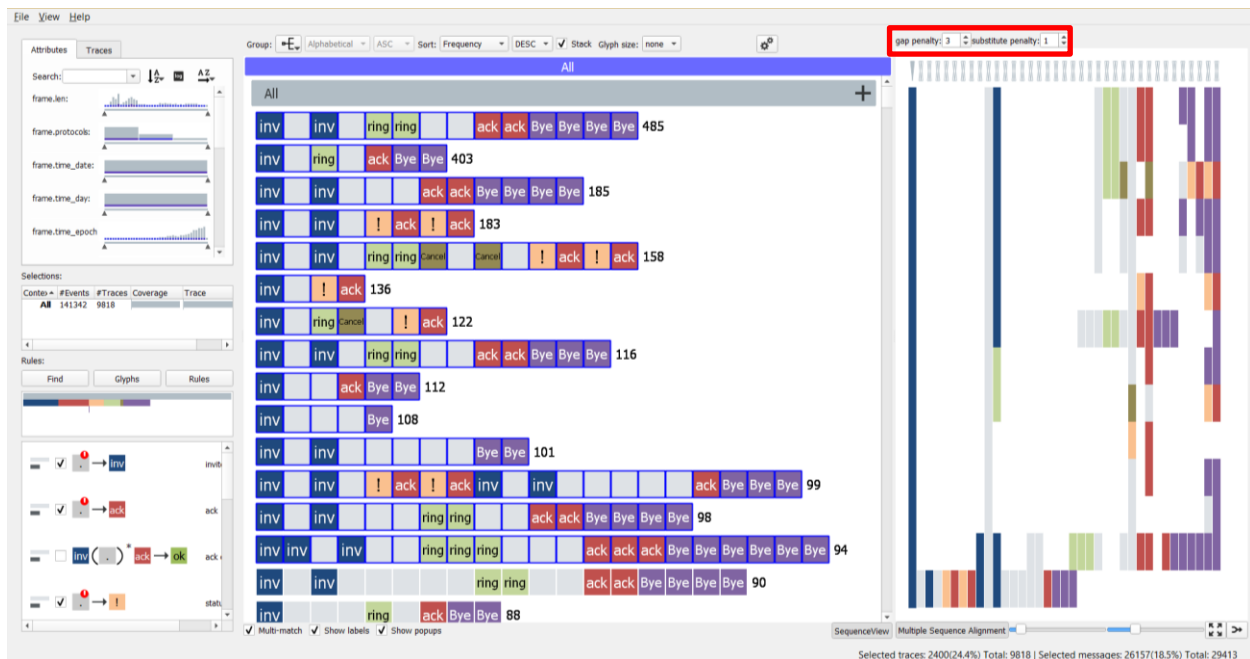
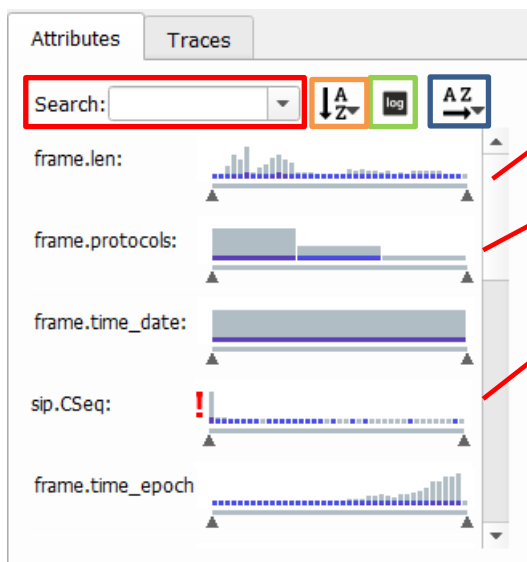


Figure 9 Less compact data alignment by adjusting the gap settings

Attribute view:

The attribute view show a histogram for every column in the CSV table. Depending on the type of attribute (categorical,numeric,identifier) there are three different types of histograms:



1 frame.len is a numeric attribute. This histogram shows the value distribution of that attribute.

2. frame.protocol is a categorical attribute → every bar in the histogram corresponds to a value

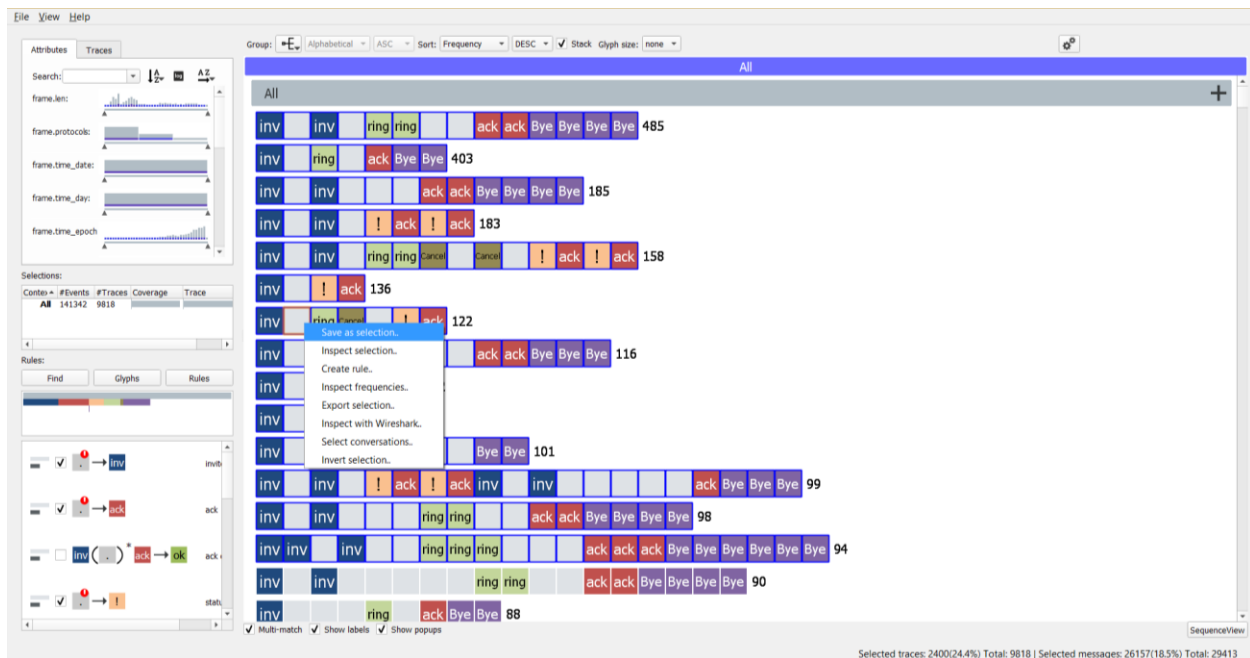
3. sip.CSeq is an **identifier** → again every value is shown in a separate bar, but since there are more values than pixels on the screen we cannot show everything. The red exclamation mark indicates that only the first 40 most frequent values are shown. Alternatively, users can also sort the bars differently (see box)

Values can also shown in logarithmic scale using

In case there are too many CSV columns, users can search attributes based on keywords (See)

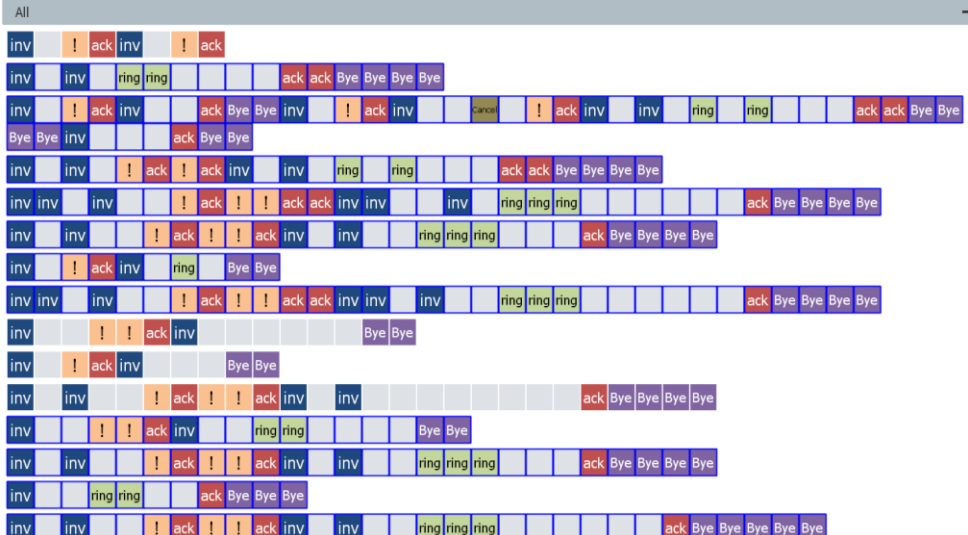
Finally, attributes can be sorted according to various metrics (see).

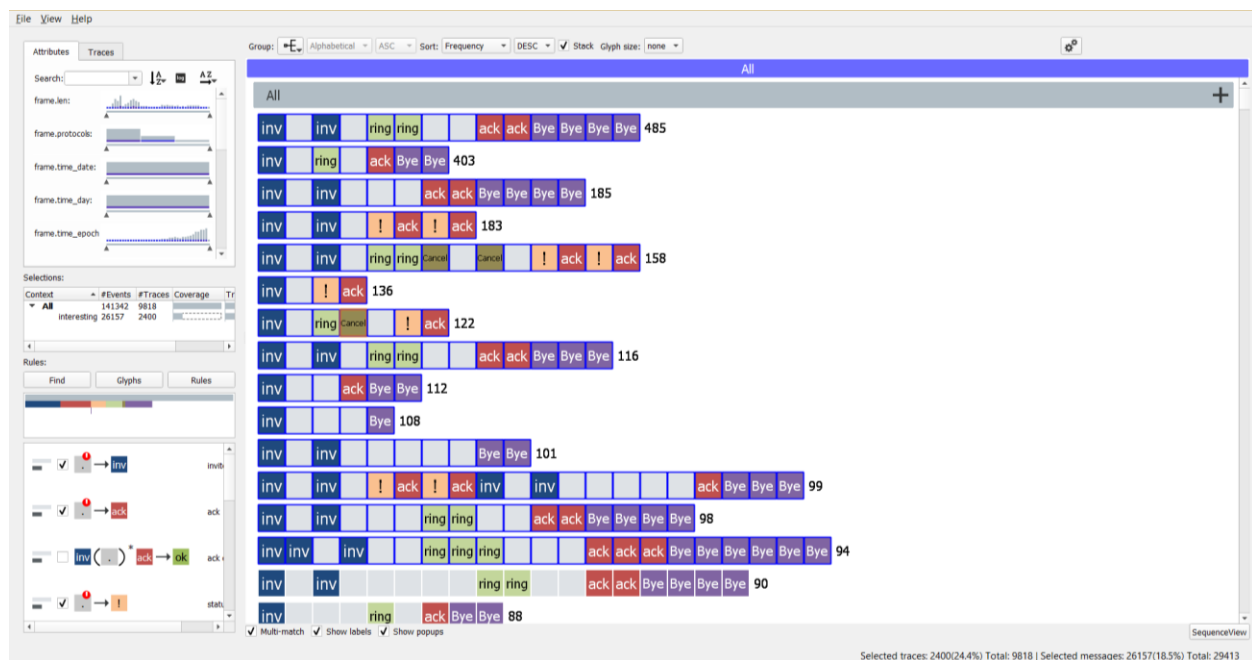
Right-click menu options:



Save Selections...	<p>Creates a new column in the CSV data that separates the selected events from the non-selected.</p>
Inspect Selections...	<p>Inspect selected events in a tabular view</p>
Create Rule...	<p>This opens the Rule view where selected events are automatically put into a filter rule</p>
Inspect Frequencies	<p>Opens an enlarged histogram to show distributions of selected events in greater detail</p>

	
Inspect with Wireshark...	<p>In case you are analyzing PCAP traffic, it is possible to view an event selection in greater detail in the Wireshark interface.</p> <ol style="list-style-type: none"> 1) Select events that you find interesting 2) Right-click Inspect with Wireshark... 3) Select the PCAP where you extracted the CSV data from 4) Done (Wireshark will start and you can view your selection) <p>This option only works if you have Wireshark installed on your computer!</p>
Select conversations...	<p>This option enables user to select all sequences that were partially selected.</p> <p>Partial selection (only ring events are selected)</p>  <p>Selection after applying this option:</p>

	
Invert selection...	<p>This options inverts the selection. In other words everything that was selected becomes deselected and every event that was previously not selected becomes selected.</p>



Shortcuts in the Sequence View:

- Double click on an event will select the entire sequence.
- Double click on white space to select the entire data set.