Anomaly Detection in Industrial Networks

PSE GRUPPE

KOMPETENZZENTRUM FÜR ANGEWANDTE SICHERHEITSTECHNOLOGIE
Advisor: M.Sc. Ankush Meshram

This Document outlines the requirements (both functional and non-functional), environment, target audience, and use cases of the software described below.

Version 0.5

Contents

1 Purpose

The goal of this project is to create a software visualization tool for industrial network traffic to simplify the analysis of anomalous behaviour, both in realtime and from captured stored data.

This software is part of the ADIN framework and is referered to as the "ADIN Inspector".
One component to achieve this goal is a web interface built with modularity in mind so as to make it easily extendable.

The Web view is able to display a series of diagrams and charts to easily identify the behaviour of the network. Within the Web view the user has the ability to zoom, select, highlight, and filter out data to better understand the aforementioned behaviour in different OSI layers, as well as visualize the flow rate between network nodes.

To support this Web view a back-end messaging solution is needed. This allows the user to easily switch between multiple streams of captured data.

2 Overall Description

While computers can analyze raw data easily, for humans it's easier to recognize behavior using visual aids. Visual analytics is utilizing graphics to enhance human cognition to understand a problem better or find 'a needle in the haystack' within a huge amount of data. Making it an invaluable tool for security analysis.

Industrial network security aims to understand the communication network traffic generated in an industrial production system. Analyzing the traffic generated by underlying industrial protocols is the primary step.

Real-time visualization of analyzed network data will help the end-user to understand the system's communication behavior and changes within it more clearly. Deviations or incidents can be detected visually by operators as they are occurring or already persisting.

Supplemental to this, the ability of operators to run an analysis of past anomalous behavior can help strengthen infrastructure for future failure or attacks. As well as provide useful information for training of new security analysts.

3 Interfaces

The environment of the project is a modern browser with LAN access. The project is OS-agnostic. Therefore the underlying Operating system is irrelevant on the users end.

3.1 Software

- Client
  - web-browser of the latest generation.
- Server:
  - Java
  - Kafka messaging framework
  - MongoDB database

3.2 Hardware

- Client:

  System capable of network connectivity
- Server:
  - Network capable system
  - System capable of running all backend-software components
  - System with adequate storage

4 Functional Requirements

4.1 Must Have

FR100 When opening the Web view the user has to be greeted by a login screen.

FR110 The user has to be able to log out of the system.

FR200 There need to be at least two levels of access for different account types (aka security roles). Level of access is defined as: the specific set of data streams the user is able to view and select to analyze.

FR300 Once logged in an user has to be able to select a data stream to be visualized.

FR400 The user has to have the ability to select multiple diagrams to visualize the selected data stream.

FR500 The user can use at least these diagram types:

- A timeline plot

- A scatter plot
- A Network diagram

FR600 The user is able to dynamically change which components of the data are used for the X and Y axis of the diagram.

FR700 The user should be able to add new diagrams to the GUI and configure them (i.e. setting diagram type and axis) both at creation and at a later time.

FR710 The GUI has to support a minimum of 4 different diagrams at once.

FR720 Each diagram should be able to be maximised to take on the full size of the diagram container.

FR800 Each drawn diagram can be connected to a different data stream.

FR900 The amount of data can be limited via a slider, effectively setting a limited time window, to which all diagram must update to.

FR910 Within the slider the user is able to scroll through the timeline and the diagrams need to react in real-time.

FR1000 There needs to be an auto scroll function (play button) which automatically scrolls through the selected time window and whose speed is adjustable.

FR1100 ADIN Inspector has to have a function to pick any data point and show all its information.

FR1110 ADIN Inspector has to support, for node-link diagrams, both picking of nodes and links

FR1200 ADIN Inspector has to offer a function to select one or more data points.

FR1210 ADIN Inspector has to offer a function to create a new diagram showing the selected data points.

4.2 Should Have

- The GUI should be able to support an undeterminate number of diagrams and scrollbar.??

4.3 Nice To Have

- Interactive filter options (Sliders, drop down menus etc.)
- Extended selection of diagram types (e.g. temporal raster plot and scatterplot matrix).
- Save the session state and display it in the user profile.

5 Data Requirements

5.1 User Data

- User ID *(Unique)*
- User Name *(Unique)*
- User Password
- User role

5.2 Event stream data for Web UI

- Timestamp
- 2 types of events for both notifications (eg. warning or errors) and data points
  - Notification
    - Severity level
    - Notification content
  - Data Point
    - Packet type, eg. protocol or network layer
    - Packet summary (incl. packet length)
    - Source
    - Destination

5.3 Raw packet metadata

Raw packet metadata include Time, Source, Destination, Protocol and Packet Length.

6 Non-Functional Requirements

NF100 The rendering latency should be no longer than 2 seconds.

NF200 The web UI should be viewable on modern web browsers.

NF300 The web UI should not crash when network connection is unstable.

NF400 The framework should be able to handle data streams from at least 100 physical nodes in the network.

NF500 The system should have a mechanism in place that is able to deny access from unauthorized personnel.

NF600 The system should be able to be recovered with ease from a crash.

NF700 The data visualization should be easily understandable or learnable for non-professionals.

NF800 The web UI should be accessible to all user groups, including people with conditions like color blindness or myopia.

NF900 The system should not crash when malformed or incomplete data are received.

6.1 Quality Requirements

Product Quality	really good	good	normal	not relevant
Functionality				
Appropriateness x				
Accuracy	x			
Interoperability			x	
Security	x			
Reliability				
Error tolerance			x	
Recoverability			x	
Usability				
Understandability x				
Learnability				x
Usability	x			
Efficiency				
Time behaviour		x		
Consumption behaviour		x		

7 Essential Test cases

T100 Successful login

Precondition Open browser window.

Action The user enters the URL for the ADIN web server in the address bar and presses Enter.

Reaction The browser loads the ADIN website and shows the login screen.

Precondition ADIN website had been opened and show the login screen.

Action The user enters a valid username and the corresponding password.

Reaction Successfully logged in. The browser loads the ADIN main screen with one empty diagram.

T101 Failed login (wrong username or wrong password)

Precondition Open browser window.

Action The user enters the URL for the ADIN web server in the address bar and presses Enter.

Reaction The browser loads the ADIN website and shows the login screen.

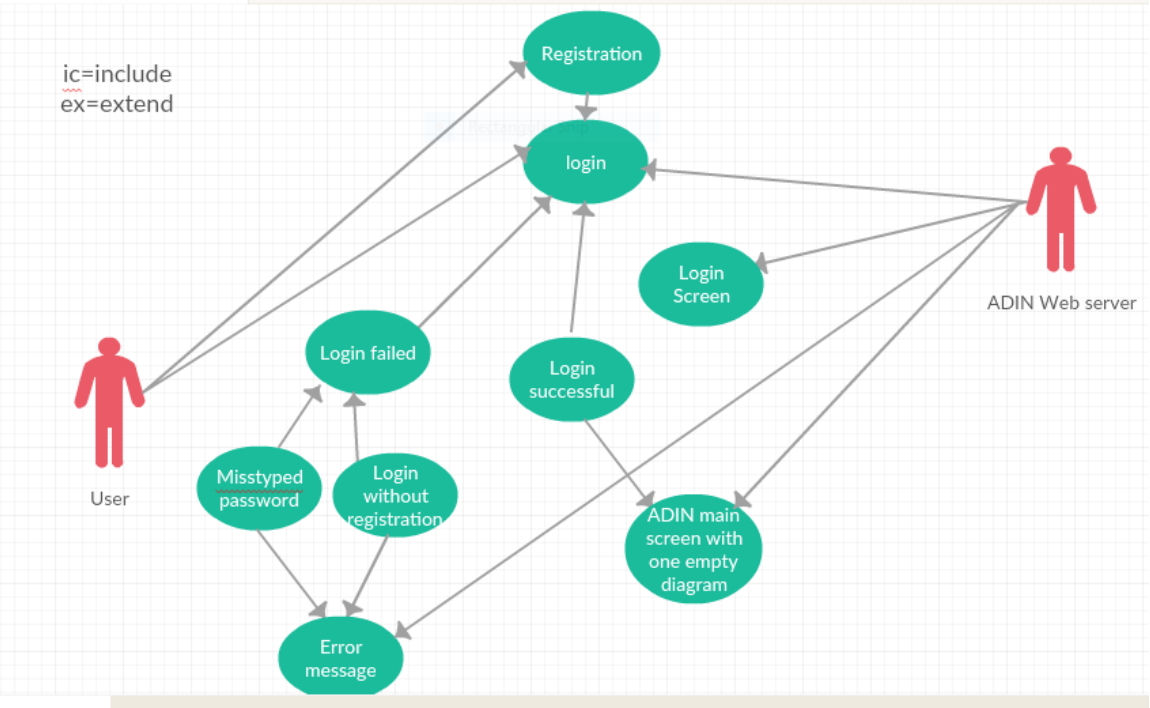

Figure 1: Log-in Cases

Precondition ADIN website has been opened and shows the login screen.

Action The user enters either a valid username and an incorrect password or a non existing username and an arbitrary password.

Reaction The login screen shows an error message that the username or password is incorrect and the entry field for the password is cleared.

T200 Open second diagram.

Precondition The browser has been logged in to ADIN and shows the main screen with one empty diagram.

Action The user presses the "New Diagram" button on the top right side of the screen.

Reaction The browser opens a modal with diagram settings

Action The user presses the create button at the bottom left

Reaction The browser opens a second diagram, splitting the diagram panel in to two.

T210 Open multiple diagrams

Precondition The browser has been logged in to ADIN and [T200] has been passed.

Action The user repeats [T200] two more times.

Reaction The diagrams grid now displays four diagrams in a 2 x 2 formation.

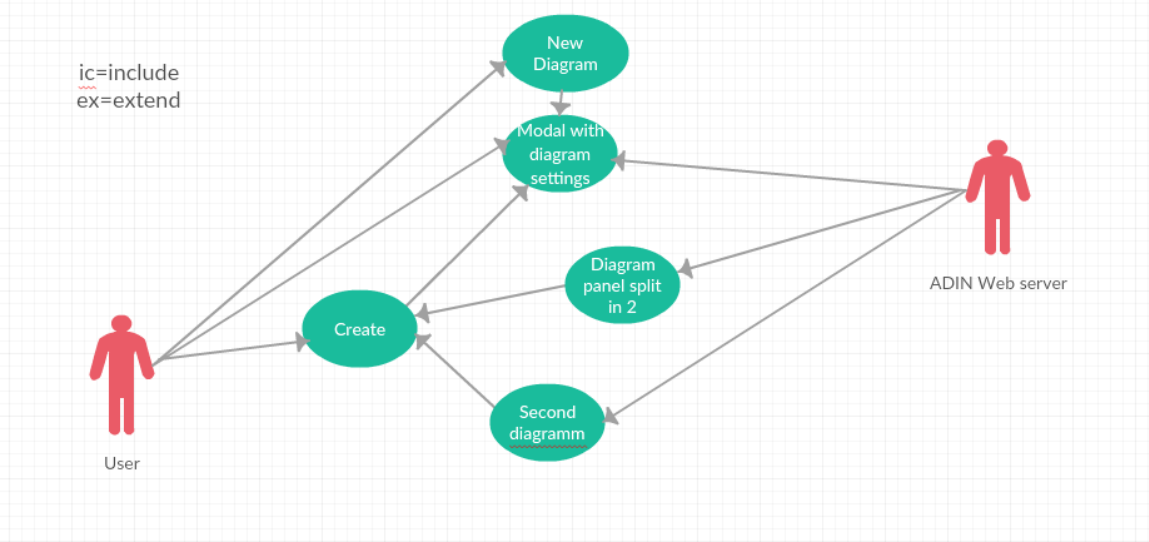

Figure 2: Second Diagram

Figure 3: Enable/Disable Filter

T220 Configure a diagram. (Concrete case: timeline diagram of packet size)

Precondition The browser has been logged in to ADIN and shows the main screen with one u

Action The user selects "Timeline Diagram" from the Diagram selection box.

Reaction The diagram changes to a timeline diagram. The x-axis is labeled with "time [s]".

Action The user selects "Packet size" in the "Y-Axis" selection box.

Reaction The diagram's y-axis is labeled with "size [bytes]".

T300 Filtering

Precondition The browser has been logged in to ADIN and shows at least one diagram

Action The user enables a filter from the global filters section.

Reaction The diagram now only shows filtered data

Action The user disables the same filter.

Reaction The diagram shows original data again

T310 Filter chaining

Precondition Logged in to ADIN and at least one diagram and one global filter is active

Action The user enables another filter from the global filters section.

Reaction The diagram now only shows relevant data

T400 Full screen a diagram

Precondition Logged in to ADIN and shows at least two diagrams

Action The user presses the full screen button on the top right corner of the diagram

Reaction The diagram's window is maximized to the diagram container of the web page.

T450 Full screen and exit Full screen

Precondition Logged in to ADIN and shows at least two diagrams

Action The user presses the full screen button on the top right corner of the diagram

Reaction The diagrams window is maximized to the diagram grid of the web page.

Action The user presses the exit full screen button on the top right corner of the diagram

Reaction The diagram grid of the web page is restored.

T500 Play button basic functionality

Precondition Logged in to ADIN and shows at exactly one diagram.

Action The user presses the play button at the bottom left of the web page.

Reaction The play button turns into a stop button.

Reaction The diagram updates according to the time shown at the play head.

Action The user presses the stop button.

Reaction The diagram remains static with the currently displayed data.

T510 Time window(s)

Precondition Logged in to ADIN with one diagram. The play head is in its default position all the way to the left

Action User drags the left diagram time window key-frame to the right.

Reaction The timestamp above the keyframe updates according to it's position

Action User drags the right diagram time window key-frame to the right.

Reaction The timestamp above the keyframe updates according to it's position

T511 Multiple Time windows

Precondition Logged in to ADIN At least 2 diagram windows open

Action User clicks on one of the diagrams

Reaction Corresponding keyframes highlight in the timeline and display their timestamp on top of themselves

Reaction All NOT corresponding keyframes gray out and hide their timestamps

Action User now clicks on a different diagram

Reaction Previously highlighted keyframes gray out and keyframes corresponding to selected diagram highlight and display their timestamp

Reaction All NOT corresponding keyframes gray out and hide their timestamps

T512 Time Window dragging restriction

Precondition Logged in to ADIN At least 2 diagram windows open Test case [T711] passed One diagram has been selected

Action User tries to drag a not highlighted keyframe from the timeline

Reaction Unable to drag and drop not highlighted keyframes

T600 Investigate a data point

Precondition A diagram showing data points.

Action The user hovers with the cursor over a data point.

Reaction ADIN Inspector opens a tooltip which shows all values or key-value pairs of this data point.

T610 Investigate nodes and links in a node-link diagram

Precondition A node-link diagram showing data points.

Action The user hovers with the cursor over a node.

Reaction ADIN Inspector opens a tooltip which shows information about this node.

Action The user hovers with the cursor over a node.

Reaction ADIN Inspector opens a tooltip which shows all values or key-value pairs of the data point that this line represents.

T620 Select data points

Precondition A diagram showing data points.

Action The user clicks on a data point.

Reaction ADIN Inspector marks this data point, e.g. by highlighting.

Action The user clicks on a different data point.

Reaction ADIN Inspector marks this second data point, e.g. by highlighting, and unmarks the first data point.

T630 Create a new diagram based on a selection, where the new diagram has a different type than the first diagram

Precondition A diagram showing data points.

Action The user selects one or more data points.
The user right-clicks the selection.
In the pop-up menu the user selects "create new diagram from selection"

Reaction ADIN Inspector opens a new diagram.

Action The user selects a diagram type.

Reaction ADIN Inspector displays the selected data points in the new diagram.

8 Software Modeling

8.1 GUI

The basic data structure needed for graphs are a given set of nodes and a given set of edges, as they are often drawn as node-link diagrams. In the postal data set, nodes could represent the origins and destinations of postal flows. Edges represent the flows between the respective origins and destinations.In intelligence analysis, investigators use semantic graphs to organise concepts and relationships as graph nodes and links in hopes of discovering key trends, patterns, and insights." A key issue in graph visualisation is the size of the graph, i.e. the size of the data to visualise. With a growing amount of data to display, graphs can become too complex and overburdening for the analyst's cognitive capacity. It thus becomes difficult for the user to conduct significant analysis. Because of the issues described above, research often focuses on ways to solve the problems of visual clutter, e.g. by aggregation or clustering techniques, which is also one of the main topics of cartographic generalisation.
In this subsection we'll discuss the GUI, it's components and the workflow of a typical

user.
As the programm is opened the user is greeted with a login screen shown in Figure 4. After the user has entered it's credentials and clicked on the login button he is shown Figure 5.

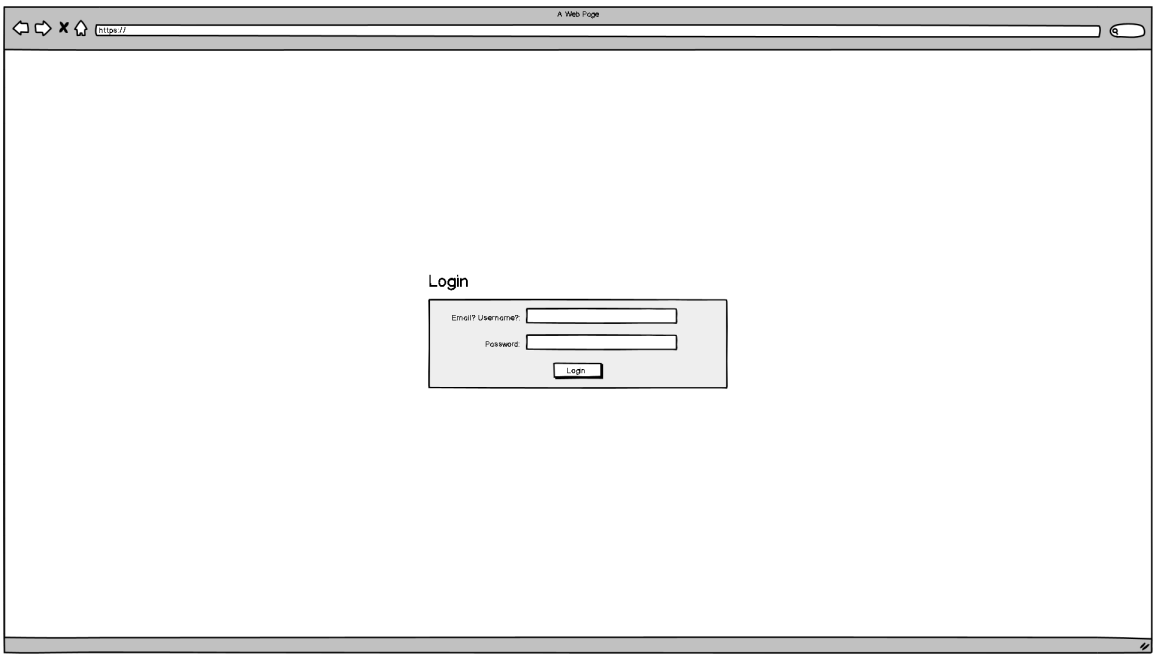

Figure 4: The Login Windows shown as the page is first loaded

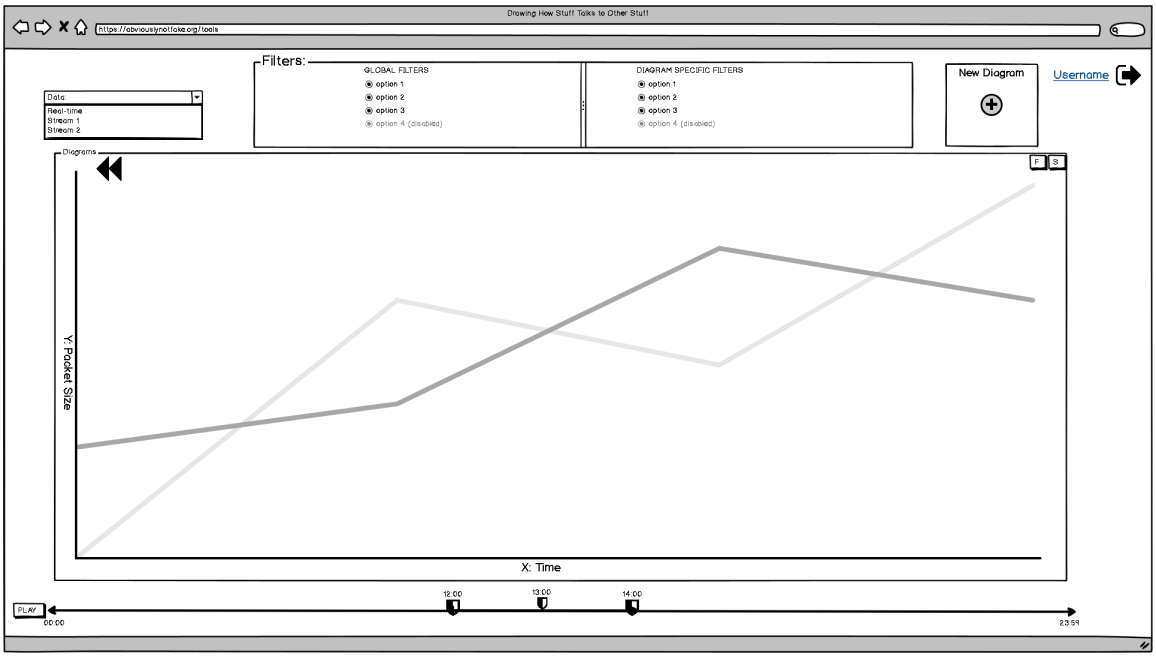

Figure 5: The Main window the user sees once he logs in

To discuss in depth this interface it has been divided into subsections labeled in Figure 6, these are:

1. The Data stream drop down menu allows the user to select which stream he is currently listening to. This dropdown menu shows the selected diagram's selected data stream and by default is set to real-time.
2. Here the user can set filters affecting all diagrams, namely limiting which layers and protocols are currently being shown.
3. Here are the filters that can further restrict the data being shown in the selected diagram.
4. By clicking this button the user can create a new diagram shown in number 5. The creation workflow is shown in Figure 7 to Figure 6.
5. Here is the currently logged in user, by clicking the button to the right of his name the user can log out.
6. This is the diagram container, inside are all diagrams the user has created with all the set constraints. At most 4 diagrams are shown in the container, and more can be made visible via the slider on the right of the container.
7. This is a diagram.
8. By hovering with the mouse over a data point a tooltip is shown with all data associated with this data point.
9. These buttons control whether a diagram is fullscreen and the diagram settings. By clicking the button labeled F the diagram grows to take on all available space in the container, like shown in Figure 5.
   Clicking on the button labeled "S" replaces the diagram with the Settings for this diagram like shown in Figure 7
10. The play button auto scroll along the selected time frame
11. This timeline shows on the left and right bottom labels the beginning and end of all data streams, the labels on top show the currently selected time window and are movable to increase,decrease the time window, the middle shield icon shows the current time while being played, and is also movable to scroll by data manually. (TBD)

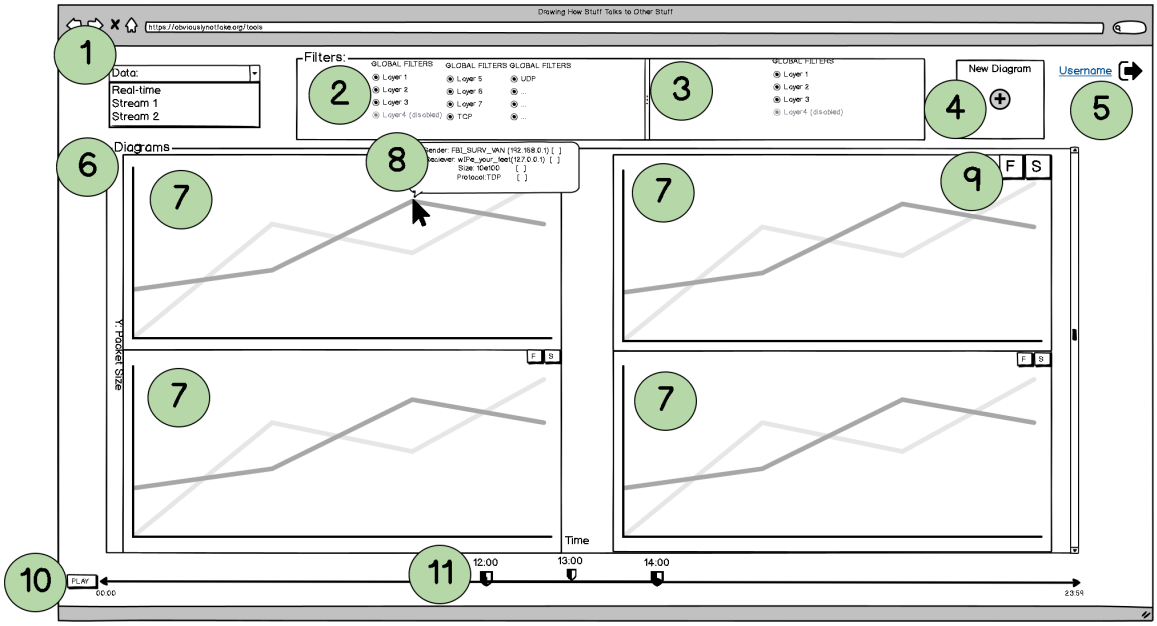

Figure 6: The GUI divided into relevant sections

Next we'll take a look at the workflow a user will go through when opening a new diagram.
As referenced above Figure 5 is what the user first sees. When he wants to create a new diagram, by clicking the button labeled new diagram, the Main window is split (shown in Figure 7) and the new diagram settings are shwon where the new diagram will be.
The user sets here the settings for the new diagram, these are:

- Which data stream it draws its data from
- Which data is pulled to be represented as the x and y axis
- Any filters the user wants to apply at the start

Afterwards, the screen looks like Figure 8, the diagram container split, by going through this process two more times the diagram container fills up, containing four diagrams total (shown in Figure 9), at this point the diagrams have their minimum size.
Any diagrams created afterwards are spawned underneath the existing ones and the user can scroll up and down to view all of them (shown in Figure 10) (TBD)

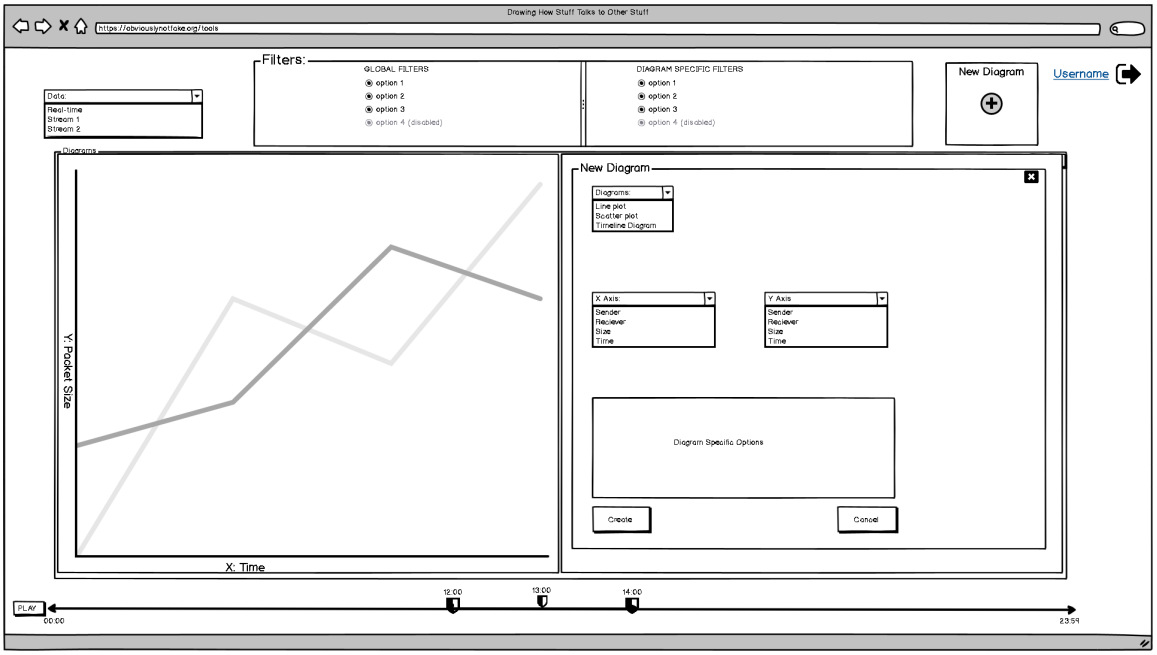

Figure 7: Main window split in two after user has clicked on the "New diagram" button. New diagram options are set up on the right.

TODO:MISSING SECTION ABOUT TOOLTIP FILTERING, NEED TO TALK ABOUT FINAL CHANGES TO THIS! IMAGES ARE NOT FINAL , STILL UP FOR DEBATE

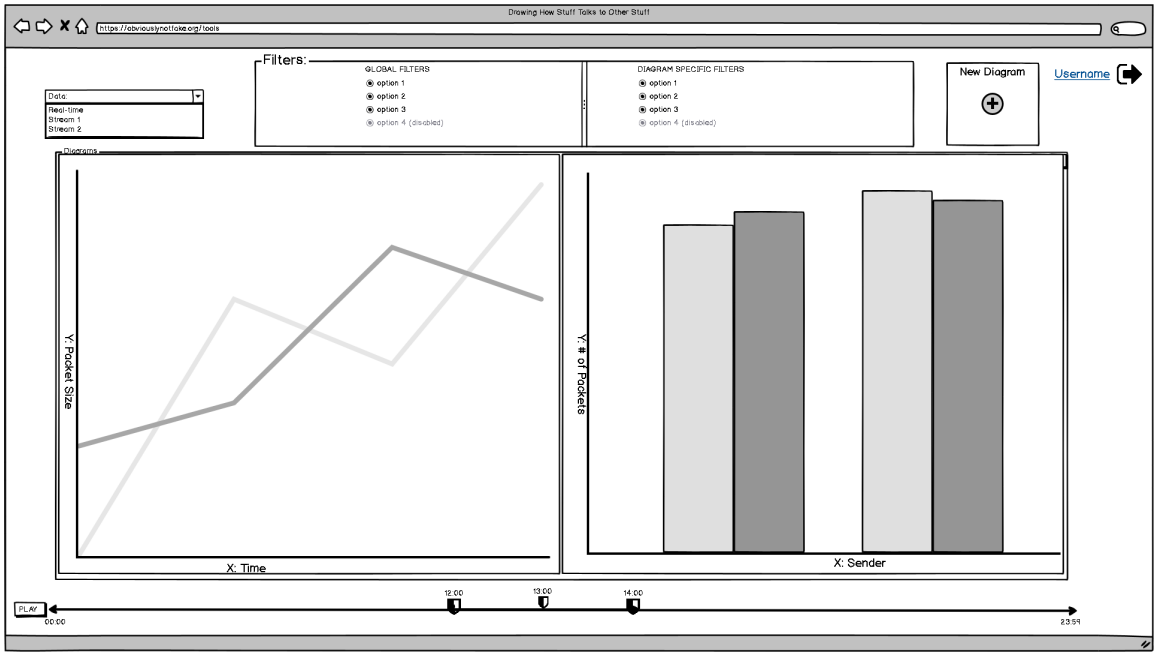

Figure 8: Main Window with two diagrams side by side.

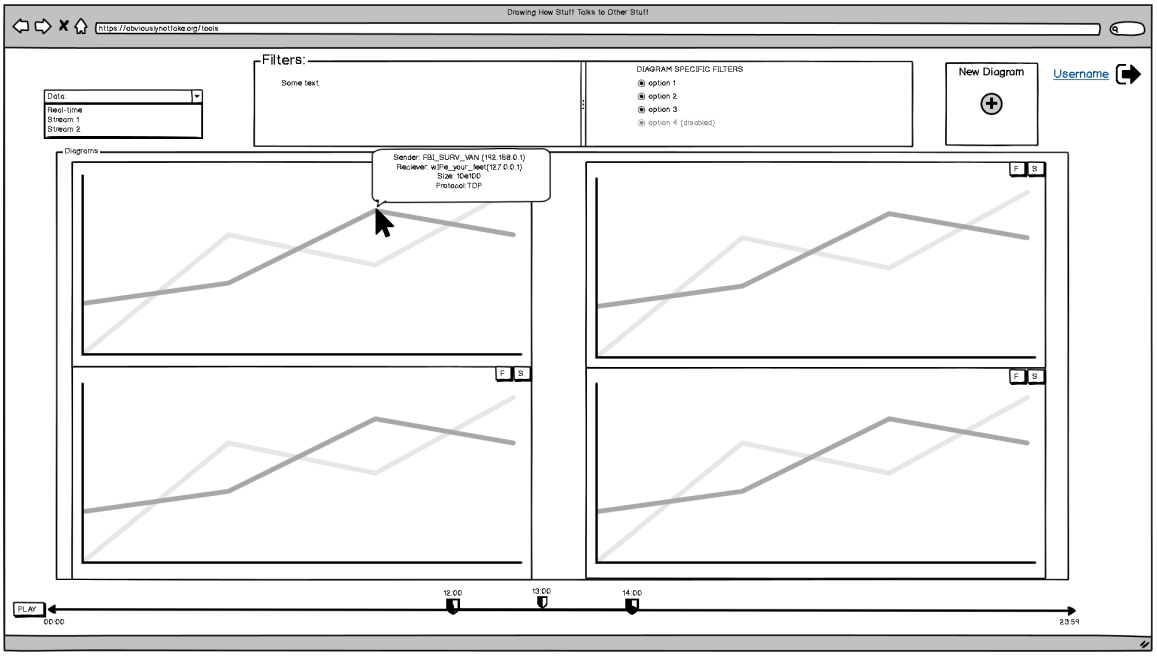

Figure 9: Main window with four diagrams open

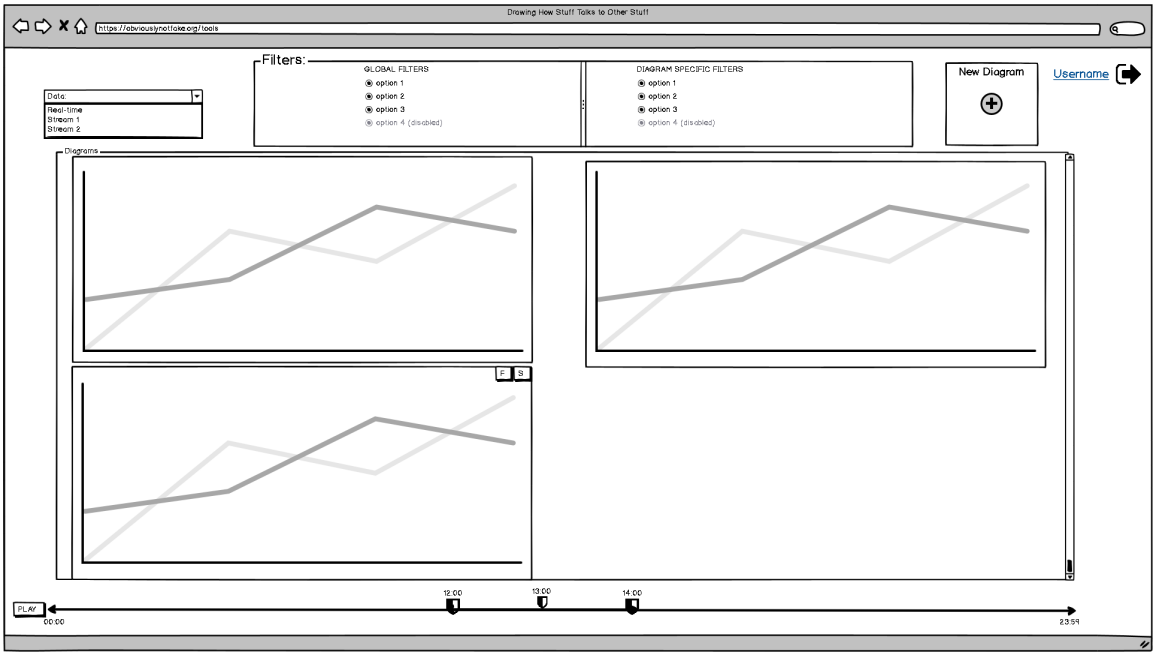

Figure 10: Main window with more than four diagrams open, more diagrams can be shown using the scroll bar

8.2 Scenario

S100: An operator wants to check manually/visually whether network nodes appeared or disappeared over the last day

- the operator opens the web page
- the operator selects the database as data source
- the operator selects a time-line-based diagram type
- the operator selects node addresses as the data to be displayed
- the operator moves to or selects the last 24 hours as the range of data to display
- the operator closes the web page

S200: A security analyst wants to look at the current flow rates between network nodes to see whether they change / there are trends

- the analyst opens the web page
- the analyst selects a source of live data
- the analyst selects an appropriate visualization type
- the analyst selects node addresses as the independent variable
- the analyst selects flow rates as the data to be displayed

S300: A security analyst wants to examine a specific point of data
Precondition: the analyst has already selected the relevant dataset and visualization type

- the analyst selects a data point
- the GUI displays a small pop-up window with all the data of this data point
- the analyst right-clicks one of the attibutes in the pop-up window and selects "Display all matching types"
- the GUI marks all data points that have the same value in this attribute

S400: The user wants to look at alarms/notifications (TBD)

- the user opens the web page
- the user selects the database as data source
- the user selects the data stream from the relevant dissector
- the GUI diplays the notifications along a timeline, according oder of occurance
- the user right-clicks on the x-axis and selects "use record number"
- the GUI diplays the notifications along a timeline adjacently

S500: The user wants to look at normal data together with alarms/notifications
Precondition: Scenario /S100/ apart from closing the web page

- the user selects menu "data", entry "sources"

- the GUI displays a list of all known data sources with a checkbox in front of each
- the user selects the checkboxes for the data sources they want to examine
- the GUI displays data from all these data sources within the currently active visualization

8.3 Use cases

TBD

8.3.1 Interactivity

Visual analytics methods combine interactive visualisations with automated analysis techniques. This allows the user to decide e.g. which part of the data he or she wants to explore in more detail.

A basic principle for visual data exploration was introduced by Shneiderman (1997) by what he called the “The Visual Information Seeking Mantra:

Overview first, zoom and filter, then details-ondemand”. This lets the data analyst define to a certain level what he or she wants to see and visualise.

Similar to this, Bertin (1983) specified three “levels of reading,”: The elementary level (allowing the analyst to look at the information about a single data record), the intermediate level (showing summarised information about a group of data records), and the global level (providing an overview of all data elements).

8.4 Object Modelling

8.5 Dynamic Modelling

9 Glossary

data point A tuple of values or key-value pairs.

data source A data source is a service that provides a network interface that can be accessed by the GUI and provides one or more data streams.

data stream A sequence of data points. A data stream may grow dynamically, in which case it will provide a kind of "read next" method.

diagram A graphical representation of data points as per a certain diagram type. Each data point is represented by a symbol whose position, size, color, shape etc. represent values of the data point. The diagram has labeled axes and, for other properties of the symbol, legends that allow the user to determine the values represented by each symbol.

diagram container The area of the GUI within which the diagram(s) are displayed.

diagram type A diagram type is a specific style or method of drawing a diagram and visualizing its data points.

role A security role is a list of access permissions. It determines which data streams (and, potentially, which diagram types and which data operations) are accessible to a user with a given role.