# Heterogeneous Logs Graph Visualization and Clustering for Attack Traces Discovery

Laetitia Leichtnam[1], Éric Totel[1], Nicolas Prigent[2], and Ludovic Mé[3]

[1]CentraleSupelec (Cidre team, IRISA)
[2]LSTI
[3]Inria, Univ. Rennes, IRISA

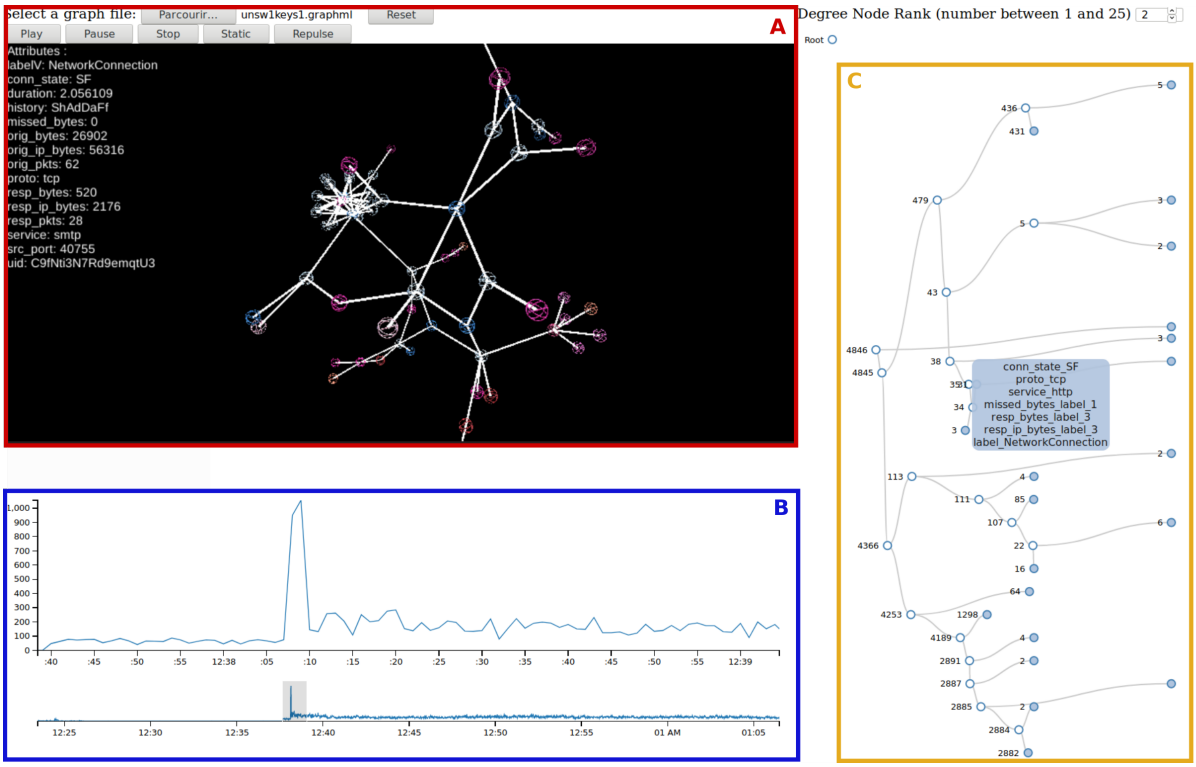


Figure 1: Global Interface - Control Command and Graph Visualization (A), Timeline Plot(B), Dendrogram (C).

**Index Terms:** H.5.2 [Information Interfaces and Presentation]: User Interfaces—Graphical user interfaces (GUI); C.2.0 [Computer-Communication Networks]: General Security and Protection

## 1 Introduction

Large networks are difficult to monitor due to the quantity and heterogeneity of security-related information that is collected. Current automated tools are generally accepted as imperfect and security analysts benefit from visualization tools that provide an overall view of the information system (through relations between events gathered from various sources and sensors) and emphasize abnormal activities. These tools help them increase their situational awareness.

A natural way to display relations between events is to use graph representations. For example, Percival [1] proposes to the analyst to correlate events by representing attack graphs. In a previous work, we used a graph to connect system objects that appear as attributes of events [2]. We proposed a clustering method to cope with the large volume of data. Given the lack of representative and labelled data set, we relied on unsupervised machine learning algorithm to build our clusters. One of the known limitations of unsupervised learning is the difficulty to understand its results. In this paper, we propose a dendrograms representation to help the security analyst define the criteria useful to aggregate events. The analyst handles the dendrograms manually as he or she knows contextual information that would be most often difficult or even impossible to model for an automatic treatment. The resulting dendrograms drives the graph clustering method.

This paper is structured as follows: Section 2 describes the design and implementation and Section 3 presents conclusion and future work.

## 2 Design and Implementation

We designed and implemented a web application to explore security-related data. It follows the principle of Baldonado et al. [3] which states that multiple views can help users understand relationships among different data sets with multiple attributes.

Our tool is made of two views. The *3D-graph view* comes from our previous work [2]. Each object (e.g., a given IP address, a given port number) that appears in the log entries is represented by a node in the graph. If two objects appear in the same log entry, the two

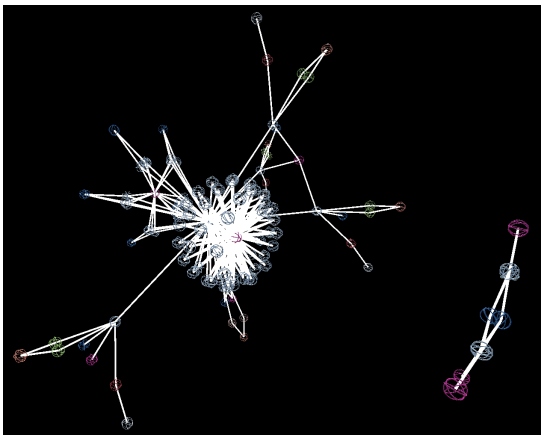

Figure 2: The complete 3D-graph before aggregation.

corresponding nodes are linked by an edge. An example of graph representation is depicted in Figure 2.

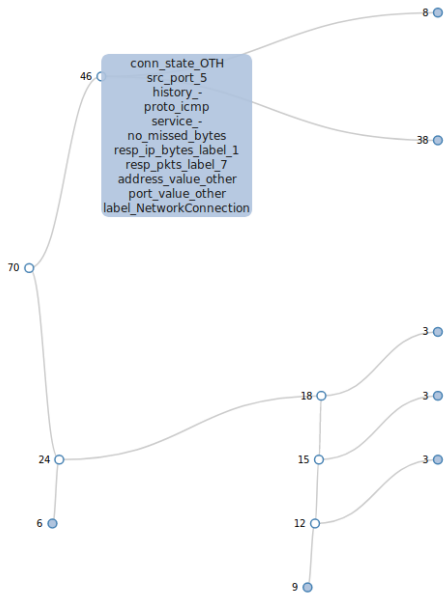


Figure 3: The dendrogram build on the objects with the highest degree: the top branch groups all network connections linked to a network scan whereas the bottom branch groups unrelated network connection.

For large data sets, the huge numbers of nodes and edges prevented efficient understanding by analysts. In response to this limitation, we propose the *dendrogram view*. A dendrogram is a hierarchical tree representation that graphically shows the distance between its nodes in a efficient way. By construction, the closer two leaves of a dendrogram, the more similar the two objects they represent. Non-terminal nodes of the dendrogram represent a cluster and each branch is a criterion of separation of two clusters. Such a dendrogram can be built for each node of the 3D graph from the characteristics of all the neighbors of this node.

Our tool allows the analyst that focuses on a given node (e.g., an IP address of interest, a node with many neighbors) of the 3D graph to build a dendrogram that corresponds to this node. A first strategy to identify an interesting node is to consider the ones with high degree (see Figure 3). The dendrogram obtained for such a

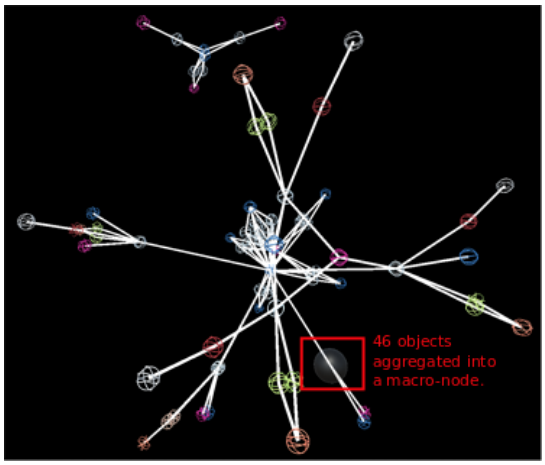


Figure 4: One macro-node represented in the graph view.

node allow the analyst to determine, among the neighbors of the node, those that can be usefully grouped together. Clicking on the dendrogram's node creates a macro-node on the 3D-graph view (see Figure 4). This is therefore a mean to cluster the 3D-graph through human analysis that would be hard to perform automatically because it requires contextual knowledge that only analysts have.

## 3 Conclusion and Future Work

We proposed a tool to group what likely corresponds to normal behavior. As a result, remaining nodes in the 3D-graph that have not been grouped using with the dendrogram view require an additional investigation by the analyst.

The prototype of this tool highlights potential for further improvements:

- While an intuitive first approach is to select nodes with high degree as they have the greatest potential for aggregation, other approaches should be explored that may improve the relevance of the dendrograms and accelerate the aggregation of objects.
- A test of our tool by security analyst will help us validating its added-value compared to other tools, and will help us improve its global design.

## References

[1] M. Angelini, N. Prigent, and S. Giuseppe. PERCIVAL: Proactive and rEactive attack and Response assessment for Cyber Incidents using Visual AnaLytics. *Vizsec - IEEE Symposium on Visualization for Cyber Security*, pp. 1–9, 2015.

[2] L. Leichtnam, E. Totel, N. Prigent, and L. Mé. STARLORD: Linked Security Data Exploration in a 3D Graph. *Vizsec - IEEE Symposium on Visualization for Cyber Security*, pp. 1–4, 2017. doi: 10.1109/VIZSEC.2017.8062203

[3] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. *Proceedings of the working conference on Advanced visual interfaces*, pp. 110–119, 2000.