SignalTreeExplorer: A Multi-Tree Comparison System for Visually Exploring Time Sequential Communication Data

Guozheng Li[1] Yu Zhang[1] Siming Chen[1] Haocheng Zhang[2] Jie Liang[3] Xiaoju Dong[2]
Jinsong Wang[4] Xiaoru Yuan[1] *

1) Key Laboratory of Machine Perception (Ministry of Education), and School of EECS, Peking University
2) Department of Computer Science, Shanghai Jiao Tong University
3) School of Software, University of Technology, Sydney
4) Southwest Electric & Telecom Engineering Institute

Abstract

Communication data have hierarchical structure and could be built as signal trees. However, the quick generation, considerable diversity, and complex patterns/anomalies of signal trees presents a great challenge for understanding and exploring for analysts. We propose SignalTreeExplorer to facilitate visual analysis for time sequential communication data, through comparison of multiple signal trees at different granularity levels concurrently. Furthermore, usage scenarios in this paper which help analysts gain insights into patterns and anomalies of communication data demonstrate the usefulness of SignalTreeExplorer.

1 Introduction

Determined by the communication protocol, the data during communicating have a hierarchical structure. More specifically, at the bottom all data in communication process are transmitted through the physical layer, but at the top they are generated from different applications. Obtained these information by parsing the captured data package and then signal trees can be built. One of the most important attributes during communication is flow size, existed in every node of signal trees. Also, It contains many significant patterns and anomalies, so timely situation awareness of communication data has attracted a lot of analysts' attention in recent years.

Previously, a variety of techniques have been proposed to compare hierarchical data. [1,3]. Nevertheless, there still remain three major challenges for communication data analysis, namely, (1) design a reasonable analysis procedure from both overview and detail perspective to adapt its rapid generation speed. (2) take the time sequential characteristic into consideration during analysis. (3) support the discovery of patterns/anomalies from both topological structure and inner nodes attributes aspects concurrently.

To address the first challenge, we propose a novel framework which analyzes the communication data efficiently from multiple granularity levels, including overview, topological structure and inner nodes attributes distributions. It defines an analysis procedure to help the analysts detect anomalies and patterns effectively. For the second challenge, arc-link histogram and linked projection are proposed to visualize the sequential variation of signal tree dataset. For the third challenge, the multi-tree comparison view facilitates analysts to compare topological structure and nodes attributes for several trees simultaneously.

*e-mail: guozheng.li@pku.edu.cn, yuzhang94@pku.edu.cn, siming.chen@pku.edu.cn, zhanghaocheng@sjtu.edu.cn, christy.jie@gmail.com, xjdong@sjtu.edu.cn, jinsong.wang@126.com, xiaoru.yuan@pku.edu.cn

2017 IEEE Pacific Visualization Symposium (PacificVis)
18-21 April, Seoul, Korea
978-1-5090-5738-2/17/$31.00 ©2017 IEEE

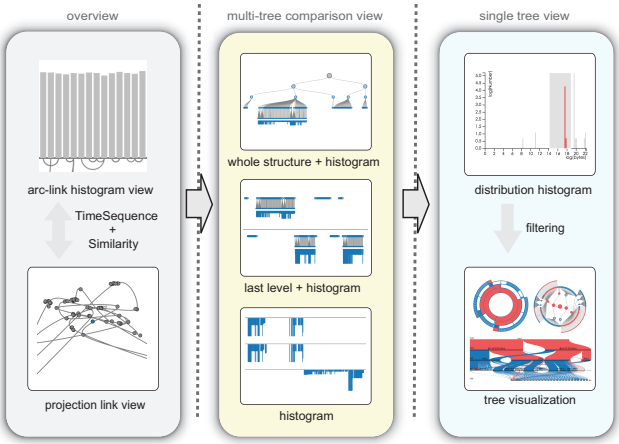

Figure 1: SignalTreeExplorer Analysis Framework: This framework is divided into 3 parts: overview, multi-tree comparison view and single tree view. Overview provides insights about the topological differences and sequential attributes. Multi-tree comparison view support comparison of topological structure and node attributes for multi-tree. Single tree view facilitates to explore nodes attributes distributions.

2 Approach

The analytical framework enables analysts to explore communication data from 3 different granularities as shown in Figure 1. Firstly, the arc link histogram and projection link view provide overview and facilitate the discovery of patterns and anomalies from the granularity of each tree as a whole. Next, analysts continue exploration from the structure and node attributes perspective in multi-tree comparison view. Finally, specific to a certain tree, SignalTreeExplorer supports analysts to explore the distributions of nodes attributes.

Both arc-link histogram view and projection link view concern topological similarity and time sequential characteristics. The similarity of signal trees is calculated by edit distance. The histogram link view encodes each tree as a bar and its horizontal location follows its temporal order. Arc links connect the signal trees with top N similar topological structure, the N can be dynamically tuned by analysts. Unlike arc-link histogram, projection link view encodes the similarity relationship by node location, which is computed using MDS projection of distance matrix. The connection among the projection nodes could be regarded as a distorted timeline. Projection link view focuses on the node similarity of the entire dataset and histogram link view pays more attention to the time sequence characteristics. These two views are interlinked together to help

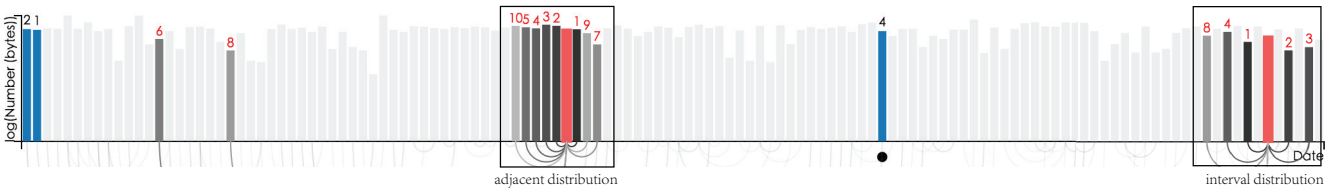

Figure 2: Arc-Link Histogram: Each bar represents a signal tree, and height of bars visualize the whole flow size during communication. The arc link below the bar connect the most similar signal trees together. Hovering on a bar, which is highlighted in red, shows more similar signal trees, with rank of similarity displayed above.

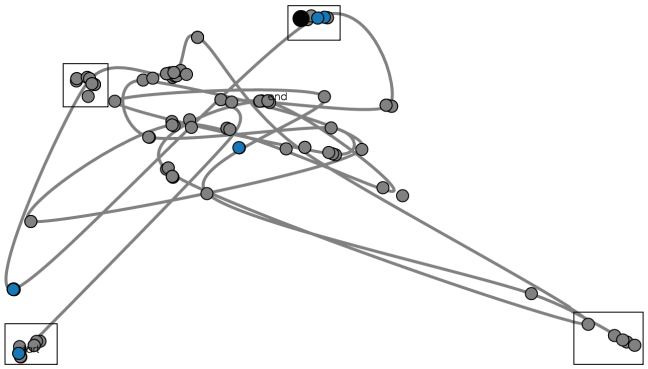

Figure 3: Projection Link View: Each node represents a signal tree, and Euclidean distance between two nodes visualize their similarity. The connection line connects the projection nodes according to their time sequence.

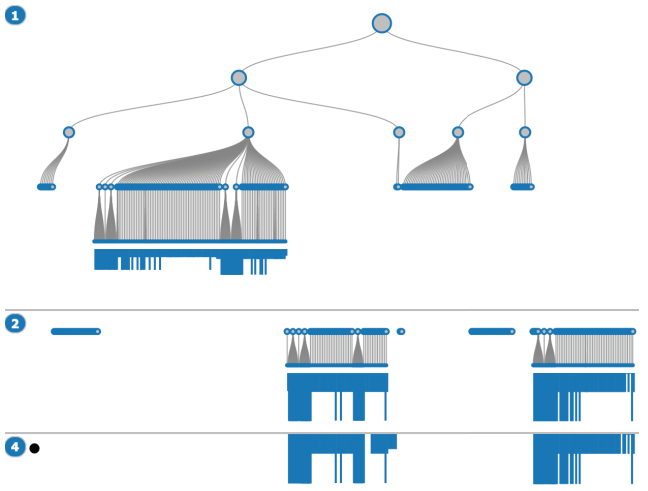

Figure 4: Multi-tree Comparison View: Multi-tree comparison view supports 3 different states of signal trees as shown above. **Tree1** - the entire signal tree, **Tree2** - last level of signal tree with flow histogram, **Tree4** - histogram

analysts gain more insights.

After gaining an overview of the entire dataset, analysts could select the signal trees of interest into multi-tree comparison view, which supports comparison of topological structure and node attributes for multi-tree simultaneously. The approach of comparison is to combine the juxtaposition [1] and superposition [2] together, more specifically, the system get the superposition of selected trees to determine the locations of each node, and then make the juxtaposition vertically for each signal tree but only display the nodes belonging to this signal tree. Considering more efficient space utilization, comparison view supports different granularity levels of signal trees, including the whole tree, leaf nodes and flow histogram.

3 Prelimary Findings

SignalTreeExplorer supports analysts to make exploration for signal tree dataset to get some significant findings easily and effectively.

- **Finding I:** From arc-link histogram view, two kinds of distributions of similar signal trees are displayed in Figure 2, adjacent distribution and interval distribution. Adjacent distribution means that users continue one communication behavior for a lone time. Interval distribution appears when users switch between different applications regularly.
- **Finding II:** From the projection link view, there are several clusters as shown in Figure 3. Each cluster represents a class of signal trees with similar topological structure. We could gain insights about the transformations between different states.
- **Finding III:** We could find that Tree2 and Tree4 are similar but different with Tree1 in Figure 4. Also, attributes of similar nodes are almost same but different nodes vary greatly. It means that same applications always have same flow size and anomalies may occur when flow size changes greatly.

4 Conclusion and Future Works

In this paper, we present SignalTreeExplorer, a visual analytical system to help communication data analysts to analyze the patterns/anomalies of users' communication behaviors. This system introduces an analytical framework which defines an efficient analysis process from multiple granularities. Arc-link histogram view and projection link view provide analysts with overview for whole dataset. Multi-tree comparison view make comparisons about its topological structure and node attributes distribution for several signal trees. The distribution of inner nodes attributes also is visualized. For the next step, we try to help analysts to gain the overview and detailed structure simultaneously with a single visualization.

Acknowledgments

This work is partially funded by PKU-Qihoo Joint Data Visual Analytics Research Center.

References

[1] S. Bremm, T. von Landesberger, M. He, T. Schreck, P. Weil, and K. Hamacherk. Interactive visual comparison of multiple trees. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 31–40, Oct 2011.

[2] G. W. Furnas and J. Zacks. Multitrees: enriching and reusing hierarchical structure. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 330–336. ACM, 1994.

[3] D. Holten and J. J. Van Wijk. Visual comparison of hierarchically organized data. *Computer Graphics Forum*, 27(3):759–766, 2008.