The Community Labeling and Sharing of Security and Networking Test datasets (CLASSNET)

Jelena Mirkovic, John Heidemann, Wes Hardaker, Michalis Kallitsis

To support network and security research, the Community Labeling and Sharing of Security and Networking Test datasets (CLASSNET) project provides new, labeled, rich and diverse datasets to the research community. The project develops a framework for collaborative, community-driven enrichment and labeling of data, enabling use of our datasets for machine learning in networking and security. Further, the CLASSNET project makes data available to researchers through multiple methods, ensuring privacy of data while enabling flexible data computation. The project also generates diverse continuous (constantly, automatically updated) and curated (selected by human) datasets for research use. The CLASSNET project innovates in dimensions of data labeling, data distribution and data sources. In data labeling, the CLASSNET's collaborative portal, called COMUNDA, will provide a low-friction framework for sharing annotations among researchers. The framework will incentivize labeling with feedback mechanisms and user credits, and support bulk, automatic, algorithmic labeling. In data distribution, CLASSNET supports multiple ways of data access, ranging from downloading anonymized data to processing data in cloud, on provider machines or via the code-to-data approach. Finally, CLASSNET data sources provide new, diverse, continuous, and curated datasets that are useful for network and security research, including traffic packets and flows, network telescope data, DNS data and Internet topology data.

The project website is at https://ant.isi.edu/classnet/ and the portal to access the datasets is at https://comunda.isi.edu. We invite the research community to use the COMUNDA portal to engage with the datasets.