In this post, we will look at network traffic analytics, which is an important subject for cybersecurity and network administration. In this study, we want to shed some light on the usage of machine learning-based network analytics on traffic flows for both network management and forensic investigation. Our goal is to create and develop technologies that will assist the human analyst in assessing "unknown background" communications. Because of the increased traffic volume, as well as the ever-changing nature of normal and malicious activity, data analytics methodologies for network analytics in both streaming and forensic modes must be investigated.

Offline and streaming approaches are employed: Stream-GP (streaming) and Random Forest are two machine learning approaches (off-line). To do this, a fixed training partition is employed. In off-line classification, a predefined training partition is utilized to create a machine learning model. After then, the model is applied to previously unknown test data. In streaming categorization, the stream is continuous. The current stream window is always accessible to the classifier.

The following datasets and assessments were used: CTU-13 botnet datasets were combined to create the botnet datasets used. These thirteen datasets were pooled to form a single dataset containing all types of botnet assaults. The data analysis includes IP addresses and port numbers.

Discussions on the findings and future work: The majority of known background traffic is classified as valid by Stream-GP, while a considerable minority is classified as malicious behavior. This research is centered on Stream-GP-labeled hazardous activities. The background traffic in the CTU-13 botnet dataset is investigated using Stream-GP, a streaming ML classifier, and Random Forest, an off-line classifier. The outputs of the two models are analyzed and compared to known real-world traffic.