Literature review:

In the paper “Toward Generating a New Intrusion Detection Dataset and Intrusion

Traffic Characterization” the authors have analysed the previously available datasets and their evaluated their effectiveness annd usefullness. They have pointed out that these datasets suffer from lack of traffic diversity and volumes, while some of them do not cover the variety of attacks, while others anonymized packet information and payload which cannot reflect the current trends, or they lack feature set and metadata.They have produced a reliable dataset containing benign and seven common attack network flows, which meets real world criteria and is publicly avaliable. Consequently, the paper evaluates performance of a comprehensive set of network traffic features and machine learning algorithms to indicate the best set of features for detecting the certain attack categories.

Also the paper “Towards Generating Real-life Datasets for Network Intrusion Detection” has proposed a systematic approach to generate unbiased full-feature real-life network intrusion datasets to compensate for the crucial shortcomings of existing datasets as they lack comprehensiveness

and completeness or are outdated.

In “Towards a Reliable Intrusion Detection Benchmark Dataset ” the authors have analysed the available benchmark datasets and have concluded that not enough research has focused on the evaluation and assessment of the datasets and hence we lack reliable dataset in this domain. In

this paper, they have presented a comprehensive evaluation of the existing datasets using their proposed criteria, a design and evaluation framework for IDS and IPS datasets, and a dataset generation model to create a reliable IDS or IPS benchmark dataset.

In the papare “ Packet and Flow Based Network Intrusion Dataset” they have pointed out that a simulated dataset cannot represent the real network intrusion scenario. So It is important to generate real and timely datasets to ensure accurate and consistent evaluation of methods. They have proposed a new real dataset to ameliorate this crucial shortcoming.

Also In the paper “Toward Instrumenting Network Warfare Competitions to Generate Labeled Datasets” they have mentioned the shortage of properly labeled datasets. here they have demonstrated that network warfare competitions can be instrumented to generate modern labeled dataset which are scientifically valuable.