Network Flow based botnet detection using supervised learning

Praveen Keshavamurthy

CTU-13 Dataset

- Total Flows ~20M from 13 scenarios
  - Background Flows - 19175568
  - Botnet labelled Flows - 444699
  - Normal labelled Flows - 356433
- Directional Flow Stats
  - Bidirectional flows - 374130
  - Client to Server flows - 426996
  - Server to Client flows - 6
- Final Labelled dataset per destination IP
  - Botnet Flows: 48479
  - Normal Flows: 4101
  - Multiple Label Flows: 145 - To be excluded

Feature Extraction

- Identifying following features for each Destination IP
  - Total Source IPs involved
  - Number of different protocols used
  - Number of Bidirectional/Unidirectional flows
  - Average & Standard Deviation per Source IP w.r.t
    - No. of flows
    - No. of Packets
    - No. of Bytes
    - No. of Source Bytes
  - No. of IPs in the same destination subnet (/24)
- Temporal features
  - No. of flows which periodically communicate with the destination IP

Machine learning models

- Simple logistic regression model
- 2 layer neural net model
- Random Forest