## CTU-13 dataset

**Dataset files**

* <https://stratosphereips.org/category/dataset.html>

We checked 2013-10-01\_capture-win8.netflow.csv file. Files with .netflow are in text format.

**Dataset entry description**

Each entry is a traffic flow, description:

Attribute 1 (StartTime) : Time of capture

Attribute 2 (Dur) : Duration of

Attribute 3 (Label) : Labelling field

Attribute 4 (Proto) : Protocol

Attribute 5 (SrcAddr) : Source address (MAC or IP)

Attribute 6 (Sport) : Source port

Attribute 7 (Dir) : Direction of flow

Attribute 8 (DstAddr) : Destination address (MAC or IP)

Attribute 9 (Dport) : Destination port

Attribute 10 (State) : State of connection

Attribute 11 (sTos) : Not Known

Attribute 12(dTos) : Not Known

Attribute 13 (TotPkts) : Total packets

Attribute 14 (TotBytes): Total bytes

**Example entry**

1970/01/01 01:00:00.000000,0.000000,flow=Background,llc,00:00:00:00:00:00,0, ->,00:00:00:00:00:00,0,INT,,,1,60

**Labelling field**

Attribute 3 (Label)

Possible values (Background,

Background-ARP,

From-Botnet-V1-UDP-Attempt,

From-Botnet-V1-UDP-Established,

From-Botnet-V1-DNS,

From-Botnet-V1-TCP-Attempt,

From-Botnet-V1-TCP-Established,

From-Botnet-V1-SPAM,

From-Botnet-V1-TCP-HTTP-Google-Net-Established-1,

From-Botnet-V1-TCP-HTTP-SSL-Google-Net-Establish)

**Questions to answer using ML**

1. Botnet present on the network?
2. Which traffic belongs to botnet?
3. Which computer is affected and which IP/URL belong to the C&C server?
4. Botnet has already established the connection or not?

**Reference**

* <http://mcfp.weebly.com/the-ctu-13-dataset-a-labeled-dataset-with-botnet-normal-and-background-traffic.html>
* <https://stratosphereips.org/category/dataset.html>