Evaluation Criteria

We will use F1 score as our evaluation metric, following our paper in ACL 2017.

We will provide the input data for each subtask, and the participants will need to make prediction based on the given input data. Name your output files for SubTask 1, 2, 3, and 4 as Task1.out, Task2.out, Task3,out, and Task4.out respectively.

SubTask 1

We will provide a list of sentences, the participant need to predict whether the sentences are relevant for inferring the malware's actions and capabilities.

For each sentences, the participant will need to output 1 if the sentence is relevant or 0 if it is irrelevant.

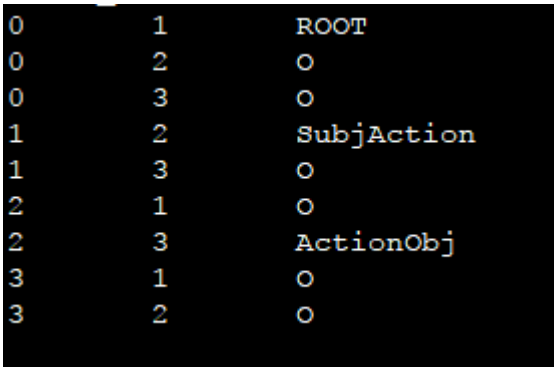
SubTask 2

Using the same sentences as provided in Task 1, the participant need to predict the token labels in the sentences. The output needs to be in BIO format. There are 3 types of token labels: "Action", "Entity", and "Modifier".

SubTask 3

We will provide a different set of sentences and their token labels. They will come in separate files according to their source document. The participant need to predict the relations between the token labels. This task will be treated as a binary classification task. For each entity pair, the participant need to output O if there is no relation between the entities, or <relation_type> if there is a relation between them. The relation types are: "SubjAction", "ActionObj", "ActionMod", and "ModObj". Token 0 will be reserved for the root. Any other tokens without a parent will be connected to this root entity with the 'ROOT' relation.

Here's an example output for SubTask 3:



0	1	ROOT
0	2	O
0	3	O
1	2	SubjAction
1	3	O
2	1	O
2	3	ActionObj
3	1	O
3	2	O

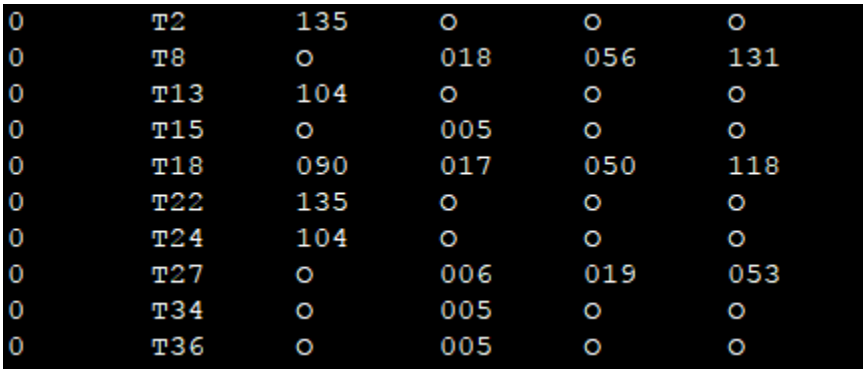
SubTask 4

We will provide a different set of sentences, their token labels, and their relation labels. The sentences will be separated into several files based on the document it came from. The participant need to predict the attributes for each 'Action' token in 4 categories: 'ActionName', 'Capability', 'StrategicObjectives', and 'TacticalObjectives'. Output 'O' if the 'Action' token doesn't correspond to any attribute in that category.

Output format (separated by tab):

<doc_id> <action_label> <action_name> <capability> <strategic_objectives> <tactical_objectives>

Here's an example of how the output looks like:



0	T2	135	O	O	O
0	T8	O	018	056	131
0	T13	104	O	O	O
0	T15	O	005	O	O
0	T18	090	017	050	118
0	T22	135	O	O	O
0	T24	104	O	O	O
0	T27	O	006	019	053
0	T34	O	005	O	O
0	T36	O	005	O	O

For Subtask 3 and 4, we will provide a file called 'doclist.txt' which contains the file names. The doc_id will be the row number of that document in the 'doclist.txt' starting from index 0.

For example, given a doclist.txt with this content:

```
Operation_Snowman
```

The_Monju_Incident

Operation_Snowman will have the doc_id 0 and The_Monju_Incident will have the doc_id 1.

For all subtasks, the output must follow the order of appearance of the sentence in the document and the order of the document in the doclist.txt (if given). Additionally for subtask 4, the action label within each document must also be output in ascending order according to their number (T2, T8, T13, ...etc.).