Guidelines to Annotating Malware Reports With BRAT V2.01

Contents

1 Introduction

There are 3 main steps to annotating APT reports with BRAT.

1. **Label Terms**
   This refers to labelling word-phrases with Term labels, such as Subject, Object, Action, Modifier.

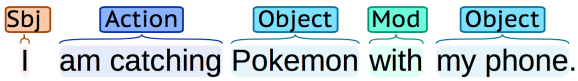


Figure 1: Example of Term labels: Subject (Sbj), Object (Obj), Action (Act) and Modifier (Mod)

2. **Label Relations**
   This refers to labelling links between pairs of Term labels.

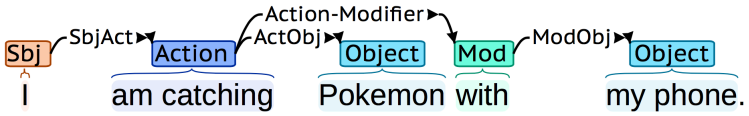


Figure 2: Example of Relation labels

3. **Label Attributes**
   This refers to labelling Terms (Actions only) with Attribute labels. Note: This can be done in conjunction with Step 1.

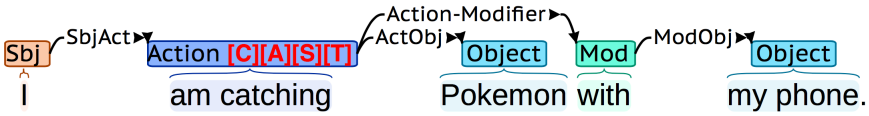


Figure 3: Example of Attribute labels ([C][A][S][T])

2 Details

2.1 STEP 1 - Label Terms

2.1.1 What to label

First of all, this is done only for sentences relevant to the following:

- Technical Capabilities of the malware
- Technical Activities of the malware

Example of sentence to be annotated:

> *Abilities include uploading and downloading files to and from the target, using the file-retrieval tool Wget to download files from the Web to the target.*

As a general guide, the sentence should imply a particular malware action or capability, with reference to the list of attribute labels.

DO NOT label sentences related to the following:

- Geopolitical or commercial effects of the malware
- Investigations into origins of the malware (aka attribution)
- Advertisements for security products

Example of sentence NOT to be annotated:

> *North Korea, with a population of 25 million, has an active duty force of 1.19 million personnel, the fourth largest in the world.*

2.1.2 What are the Term Labels

Most sentences that we deal with involve a description of some form of action. Take the following as a simple example.

I am catching Pokemon with my phone.

- **Subject**
  This is the initiator of the action aka the *do-er*.
  In the above case, *I* is the Subject.
  As a rule of thumb, in ambiguous cases, try replacing the Subject with the word *it* or *they*. The sentence should still sound correct and retain its original meaning.
  For example in the given sentence:

  > *The RCS sample sent to Ahmed adds a Run registry key.*

  The RCS sample sent to Ahmed will be tagged as the Subject. Replacing the phrase with *it* gives the following sentence:

  > *It adds a Run registry key.*

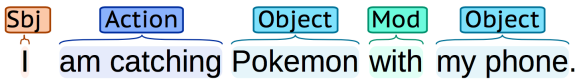
  which still sounds correct and retains the original meaning.

- **Object**
  This is the recipient of the action aka the *do-ee*.
  In the above case, *Pokemon* and *my phone* are the Objects.
  The same rule of thumb used in Subject labels applies for Object labels in ambiguous cases.

- **Action**
  This is the event.
  In the above case, *am catching* is the Action.
  Note: label an entire progressive verb phrase as Action, for example, label the entire phrase *am catching*, instead of just *catching*. Other examples include *is opening* and *are carrying*.

- **Modifier**
  This refers to words that link to other word-phrases that elaborate on the Action.
  In the above case, *with* is the Modifier. Modifiers are typically, but not always, prepositions.

For the sentence above, the following should be the resultant labels.



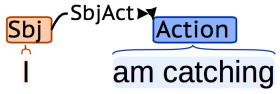
2.2 STEP 2 - Label Relations

2.2.1 What are the Relation Labels

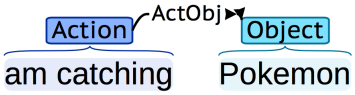
Essentially we want to connect all the labelled Terms together. Each Term should have at least one Relation, possibly more. Using the below sentence as a simple example.

I am catching Pokemon with my phone.

- **SubjAction** - Subject-Action
  This connects a Subject to the Action that it is performing.

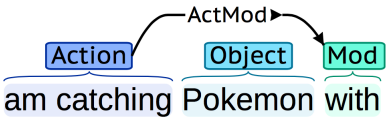


- **ActionObj** - Action-Object
  This connects the Action to the Object that it is acting on.

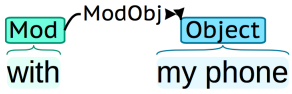


  Note: NOT *catch* and *my phone*, since there is a Modifier between these.

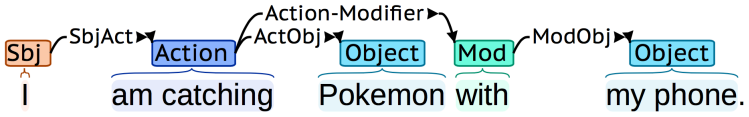
- **ActionMod** - Action-Modifier
  This connects the Action to the relevant Modifier.



- **ModObj** - Modifier-Object
  This connects the Modifier to the relevant Object.



For the above sentence, the following should be the resultant labels.



2.3 STEP 3 - Label Attributes

This is the hard part.

Attributes are labelled for Actions.

Essentially we want to assign the Action to a Technical Classification under the Malware Attribute Enumeration and Characterization (MAEC) vocabulary. See `https://maec.mitre.org/language/version4.1/MAEC_Vocabs_Spec_v1_1.pdf` for more information on MAEC.

There are four categories of Attribute Labels: ActionName [A], Capability [C], StrategicObjectives [S] and TacticalObjectives [T].

For each Action, try to tag as many relevant Attributes as possible, although in some cases not all four categories may apply.

As a quick example, see the sentence below.

Stuxnet has the ability to hide copies of its files copied to removable drives.

This describes Stuxnet's ability to prevent itself from being detected. The Action *hide* should be tagged with the following Attribute Labels:

Table 1: Example of Attribute-Labeling

ActionName	032:File-hide_file
Capability	002:MalwareCapability-**anti-detection**
StrategicObjectives	007:**AntiDetection**-hide_malware_artifacts
TacticalObjectives	023:**AntiDetection**-hide_file_system_artifacts

*Note: words in bold should be the same

This is difficult because it requires a certain amount of domain knowledge and there are a huge number of labels.

To make this easier, we have provided AttributeLabels_V1.01.pdf.

This file contains all the Attribute Labels from the four categories. For each Label, there is an associated Description and a list of possible keywords related to the label. You can then use Ctrl-F or Cmd-F to quickly search through the labels.