Group 16
0856521 孫維崗

User Behavior Analysis

1. What model or algorithm you use? Please describe in as much detail as possible.

```
<?xml version="1.0" encoding="utf-8" standalone="yes"?>
        <Events>
          <Event xmlns="http://schemas.microsoft.com/win/2004/08/events/event">
            <System>
              <Provider Name="Microsoft-Windows-Sysmon" Guid="{5770385f-c22a-43e0-
        bf4c-06f5698ffbd9}"/>
              <EventID>22</EventID>
              <Version>5</Version>
              <Level>4</Level>
              <Task>22</Task>
              <Opcode>0</Opcode>
              <Keywords>0x8000000000000000</Keywords>
              <TimeCreated SystemTime="2020-05-18T07:29:57.611647700Z"/>
              <EventRecordID>3252</EventRecordID>
              <Correlation/>
              <Execution ProcessID="2844" ThreadID="3988"/>
              <Channel>Microsoft-Windows-Sysmon/Operational</Channel>
              <Computer>DESKTOP-P84STH6</Computer>
              <Security UserID="S-1-5-18"/>
            </System>
            <EventData>
              <Data Name="RuleName"/>
              <Data Name="UtcTime">2020-05-18 07:29:54.049</Data>
              <Data Name="ProcessGuid">{5d3d98af-2977-5ec2-0000-
        0010c6e72b00}</Data>
              <Data Name="ProcessId">3160</Data>
              <Data Name="QueryName">tr.blismedia.com</Data>
              <Data Name="QueryStatus">0</Data>
              <Data Name="QueryResults">34.96.105.8;</Data>
              <Data Name="Image">C:\Program Files
        (x86)\Google\Chrome\Application\chrome.exe</Data>
            </EventData>
          </Event>
</Events>
```

We can observe that every person owns unique Execution ProcessID, so that we can setup a rule for user behavior analysis.

```
if testcase[i].Execution_ProcessID == person[i].Execution_ProcessID:
    testcase[i] = i
```

2. Anything interesting you find or problems you encounter in the whole process.

There are lots of noise for user behavior analysis. Since the number of person is small, deep learning model is hard to outperform. I decide to use rule-based to analyze user behavior. I'm not good at coding in python. This is my first time to preprocess xml and json files. It took me a long time.