Group 16
0856521 孫維崗

User Behavior Analysis

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

Sysmon.xml

```
<?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 Process ID, so that we can setup a rule for user behavior analysis.

Security.xml

```
<?xml version="1.0" encoding="utf-8"?>
<Events>
        <Event xmlns="http://schemas.microsoft.com/win/2004/08/events/event">
          <System>
            <Provider Name="Microsoft-Windows-Security-Auditing" Guid="{54849625-
        5478-4994-a5ba-3e3b0328c30d}"/>
            <EventID>4672</EventID>
            <Version>0</Version>
            <Level>0</Level>
            <Task>12548</Task>
            <Opcode>0</Opcode>
            <Keywords>0x8020000000000000</Keywords>
            <TimeCreated SystemTime="2020-05-18T07:39:28.582858700Z"/>
            <EventRecordID>5343</EventRecordID>
            <Correlation ActivityID="{9e1903ff-2cdb-0000-0b05-199edb2cd601}"/>
            <Execution ProcessID="620" ThreadID="660"/>
            <Channel>Security</Channel>
            <Computer>DESKTOP-P84STH6</Computer>
            <Security/>
          </System>
          <EventData>
            <Data Name="SubjectUserSid">S-1-5-18</Data>
            <Data Name="SubjectUserName">SYSTEM</Data>
            <Data Name="SubjectDomainName">NT AUTHORITY</Data>
            <Data Name="SubjectLogonId">0x3e7</Data>
            <Data Name="PrivilegeList">SeAssignPrimaryTokenPrivilege SeTcbPrivilege
        SeSecurityPrivilege SeTakeOwnershipPrivilege SeLoadDriverPrivilege
        SeBackupPrivilege SeRestorePrivilege SeDebugPrivilege SeAuditPrivilege
        SeSystemEnvironmentPrivilege SeImpersonatePrivilege
        SeDelegateSessionUserImpersonatePrivilege</Data>
          </EventData>
        </Event>
</Events>
```

We can observe that every person owns unique Correlation Activity ID, so that we can setup a rule for user behavior analysis.

Wireshark.json

```
  "_index": "packets-2020-05-18",
  "_type": "doc",
  "_score": null,
  "_source": {
    "layers": {
      "frame": {
        "frame.interface_id": "0",
        "frame.interface_id_tree": {
```

```
        "frame.interface_name": "\\Device\\NPF_{1BF6DE2B-07A9-4FF4-8C4F-
79DA555DBF9C}",
        "frame.interface_description": "Ethernet"
       },
       "frame.encap_type": "1",
       "frame.time": "May 18, 2020 15:30:56.810245000 Taipei Standard Time",
       "frame.offset_shift": "0.000000000",
       "frame.time_epoch": "1589787056.810245000",
       "frame.time_delta": "26.556401000",
       "frame.time_delta_displayed": "26.556401000",
       "frame.time_relative": "4203.096439000",
       "frame.number": "32284",
       "frame.len": "91",
       "frame.cap_len": "91",
       "frame.marked": "0",
       "frame.ignored": "0",
       "frame.protocols": "eth:ethertype:ip:udp:dns",
       "frame.coloring_rule.name": "UDP",
       "frame.coloring_rule.string": "udp"
      },
      "eth": {
       "eth.dst": "52:54:00:12:35:02",
       "eth.dst_tree": {
        "eth.dst_resolved": "RealtekU_12:35:02",
        "eth.dst.oui": "5395456",
        "eth.dst.oui_resolved": "Realtek (UpTech? also reported)",
        "eth.addr": "52:54:00:12:35:02",
        "eth.addr_resolved": "RealtekU_12:35:02",
        "eth.addr.oui": "5395456",
        "eth.addr.oui_resolved": "Realtek (UpTech? also reported)",
        "eth.dst.lg": "1",
        "eth.lg": "1",
        "eth.dst.ig": "0",
        "eth.ig": "0"
       },
       "eth.src": "08:00:27:be:35:e6",
       "eth.src_tree": {
        "eth.src_resolved": "PcsCompu_be:35:e6",
        "eth.src.oui": "524327",
        "eth.src.oui_resolved": "PCS Computer Systems GmbH",
        "eth.addr": "08:00:27:be:35:e6",
        "eth.addr_resolved": "PcsCompu_be:35:e6",
        "eth.addr.oui": "524327",
        "eth.addr.oui_resolved": "PCS Computer Systems GmbH",
        "eth.src.lg": "0",
        "eth.lg": "0",
```

```
      "eth.src.ig": "0",
      "eth.ig": "0"
     },
     "eth.type": "0x00000800"
    },
    "ip": {
     "ip.version": "4",
     "ip.hdr_len": "20",
     "ip.dsfield": "0x00000000",
     "ip.dsfield_tree": {
      "ip.dsfield.dscp": "0",
      "ip.dsfield.ecn": "0"
     },
     "ip.len": "77",
     "ip.id": "0x0000bbde",
     "ip.flags": "0x00000000",
     "ip.flags_tree": {
      "ip.flags.rb": "0",
      "ip.flags.df": "0",
      "ip.flags.mf": "0"
     },
     "ip.frag_offset": "0",
     "ip.ttl": "128",
     "ip.proto": "17",
     "ip.checksum": "0x00000000",
     "ip.checksum.status": "2",
     "ip.src": "10.0.2.15",
     "ip.addr": "10.0.2.15",
     "ip.src_host": "10.0.2.15",
     "ip.host": "10.0.2.15",
     "ip.dst": "192.168.100.4",
     "ip.addr": "192.168.100.4",
     "ip.dst_host": "192.168.100.4",
     "ip.host": "192.168.100.4"
    },
    "udp": {
     "udp.srcport": "50351",
     "udp.dstport": "53",
     "udp.port": "50351",
     "udp.port": "53",
     "udp.length": "57",
     "udp.checksum": "0x00003106",
     "udp.checksum.status": "2",
     "udp.stream": "964",
     "Timestamps": {
      "udp.time_relative": "0.000000000",
```

```
        "udp.time_delta": "0.000000000"
       }
      },
      "dns": {
       "dns.id": "0x00008e08",
       "dns.flags": "0x00000100",
       "dns.flags_tree": {
        "dns.flags.response": "0",
        "dns.flags.opcode": "0",
        "dns.flags.truncated": "0",
        "dns.flags.recdesired": "1",
        "dns.flags.z": "0",
        "dns.flags.checkdisable": "0"
       },
       "dns.count.queries": "1",
       "dns.count.answers": "0",
       "dns.count.auth_rr": "0",
       "dns.count.add_rr": "0",
       "Queries": {
        "settings-win.data.microsoft.com: type A, class IN": {
         "dns.qry.name": "settings-win.data.microsoft.com",
         "dns.qry.name.len": "31",
         "dns.count.labels": "4",
         "dns.qry.type": "1",
         "dns.qry.class": "0x00000001"
        }
       },
       "dns.response_in": "32288"
      }
     }
    }
```


There are some pattern to know user behavior, such as *Website they often visit*, but most of them are useless information. Since they work on virtual machine, it's hard to use those feature.

I use DNS query to classify user. First, if the domain is not IP address, it will take first two integer. For example, 192.168.1.1 -> 192.168. Second, the domain will be extracted top two level. For example, abc.google.com -> google.com. Then, I will use TF-IDF to count the domain weight for every person. If the sum(weight) is max, it means that person.

2. <u>Anything interesting you find or problems you encounter in the whole process.</u>

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.