Improving Network Flow Definition: Formalization and Applicability

Petr Velan
Institute of Computer Science, Masaryk University
Brno, Czech Republic
velan@ics.muni.cz

Abstract
Network flow monitoring has been used for more than 20 years and has become an important part of network accounting and security. A significant effort was invested into the standardization of flow monitoring by the Internet Engineering Task Force (IETF). The flow monitoring has steadily evolved to satisfy new requirements created by the demand for increased visibility and accuracy. Therefore, it is not surprising that even the most recent flow definition created by the IETF does not consider several specifics of the flow monitoring process as it is used nowadays. We present a revised flow definition that is more generic and is designed to accommodate more specific flow monitoring requirements. Moreover, we formalize our definition to avoid ambiguity and imprecision introduced by the use of natural language. An additional benefit of formalizing the flow definition is that it implicitly describes the flow creation process as well.

Why is a revised flow definition needed?
A flow is defined as a set of IP packets passing an Observation Point in the network during a certain time interval. All packets belonging to a particular flow have a set of common properties. Each property is defined as the result of applying a function to the values of:

1. one or more packet header fields (e.g., destination IP address), transport header fields (e.g., destination port number), or application header fields (e.g., RTP header fields);
2. one or more characteristics of the packet itself (e.g., number of MPES-headers).

Problems of the IPEX definition:
- Unclear meaning of packet header and characteristics of a packet.
- Limited to IP traffic. Flows are often used for non-IP traffic as well.
- Does not allow to work with fragmented traffic.

Proposed definitions
Definition (flow): A flow is defined as a sequence of packets passing an observation point in the network during a certain time interval. All packets that belong to a particular flow have a set of common properties derived from the data contained in the packet, previous packets of the same flow, and from the packet treatment at the observation point.

Definition (flow key): A flow key is a set of common properties that is used to specify a flow.

Example of flow properties

<table>
<thead>
<tr>
<th>Aggregated properties</th>
<th>Non-aggregated properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet data</td>
<td></td>
</tr>
<tr>
<td>Number of bytes</td>
<td>Source IP address</td>
</tr>
<tr>
<td>TCP flags</td>
<td>Destination port</td>
</tr>
<tr>
<td>Time to live</td>
<td>Transport protocol</td>
</tr>
<tr>
<td>Packet treatment</td>
<td></td>
</tr>
<tr>
<td>Number of packets</td>
<td>Input interface number</td>
</tr>
<tr>
<td>Flow start timestamp</td>
<td>Next-Hop IP address</td>
</tr>
</tbody>
</table>

Use of formalized definition
An algorithm describing a creation of a single flow from a set of packets:

Construction of a flow
1. Denote I the set of packet indexes that belong to the flow, F.
2. Start with $I = \emptyset$;
3. while An index $k$ of the first extended packet $\hat{p}_k$ for which $\phi(\hat{p}_k) = \text{true}$ exists do
   4. Add $k$ to $I$;
   5. end while;
7. The flow $F$ is a sequence of packets with indexes from $I$.

The algorithm can be used to construct a sequence of flows as well:
- Create a flow.
- Remove all packets in the flow from the original set of packets.
- Create new flow from the new set of packets, repeat.

Construction of flow records
1. loop;
2. Get new packet $P$;
3. Extract packet metadata $M$;
4. Set found = false;
5. for all flow record $F$ in flow cache do
   6. Apply flow selection function $\phi$ to $F$ and $M$;
   7. if $\phi(F, M) = \text{true}$ then
      8. Aggregate $M$ to $F$;
   9. Set found = true;
10. break;
11. end if;
12. end for;
13. if not found then
14. Create new flow record $F$ from $M$;
15. Insert $F$ into flow cache;
16. end if;
17. end loop;

Acknowledgement
This research was supported by the Security Research Programme of the Czech Republic 2015-2020 (BV III/1V5) granted by the Ministry of the Interior of the Czech Republic under No. VZ0060160929 The Sharing and analysis of security events in the Czech Republic.