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 [5]).*
2. *one or more characteristics of the packet itself (e.g., number of MPLS labels)*
3. *one or more fields derived from packet treatment (e.g., next hop IP address, output interface)*

A packet is defined as belonging to a Flow if it completely satisfies all the defined properties of the Flow.

(Specification of the IPFIX Protocol (RFC 7011))

Problems of the IPFIX definition:

- Unclear meaning of *packet header* and *characteristics of a packet*.
- Limited to IP traffic. Flows are often used for non-IP connections 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 record):
A flow record *is a tuple which describes a particular* flow *containing values of:*

1. *the* flow key *used to specify the* flow*,*
2. *other properties of the* flow *derived from:*
   (a) *data contained in the packets of the* flow*,*
   (b) *the packet treatment of the* flow *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

	Aggregated properties	Non-aggregated properties
Packet data	Number of bytes TCP flags Time to Live	Source IP address Destination port Transport protocol
Packet treatment	Number of packets Flow start timestamp	Input interface number Next-Hop IP address

Formalized definition

Definition (flow selection function):
Let $\widehat{P}^*$ be a set of all finite sequences of extended packets, $\widehat{P}$ be a set of extended packets. We say that a function of type

$$\varphi : \widehat{P}^* \times \widehat{P} \to \{true, false\}$$

is a flow selection function*.*
Let $\mathbb{S}$ be a set of indexes of packets observed at an observation point*:*

$$\mathbb{S} = \{1, \ldots, n\} \vee \mathbb{N},$$

where $n \in \mathbb{N}$ is the number of observed packets when the number is finite.

A flow $\mathcal{F}$ is a sequence of packets defined by a sequence of extended packets with indexes in $\mathbb{S}$ and a *flow selection function* φ. We require that a packet belongs to a flow if it is determined by all previous packets of that flow.

Use of formalized definition

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

Construction of a flow

1: Denote $\mathbb{I}$ the set of packet indexes that belong to the flow $\mathcal{F}$
2: Start with $\mathbb{I} = \emptyset$
3: **while** An index k of the first extended packet $\widehat{p}_k$ for which $\varphi((\widehat{p}_n)_{n \in \mathbb{I}}, \widehat{p}_k) = true$ exists **do**
4: Add k to $\mathbb{I}$
5: **end while**
6: The flow $\mathcal{F}$ is a sequence of packets with indexes from $\mathbb{I}$

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.

Standard flow monitoring process can be described as follows:

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 $\mathcal{F}$ in flow cache **do**
6: Apply flow selection function ϕ to $\mathcal{F}$ and M
7: **if** $\phi(\mathcal{F}, M) = true$ **then**
8: Aggregate M to $\mathcal{F}$
9: Set **found** = **true**;
10: **break**
11: **end if**
12: **end for**
13: **if not** found **then**
14: Create new flow record $\mathcal{F}$ from M
15: Insert $\mathcal{F}$ 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 / 1 VS) granted by the Ministry of the Interior of the Czech Republic under No. VI20162019029 The Sharing and analysis of security events in the Czech Republic.

https://csirt.muni.cz/
@csirtmu
https://sabu.cesnet.cz/
@CESNET_CERTS

CSIRT-MU