**IPFIX与NetFlow**

**摘要**

网络流统计技术对于网络管理和维护，网络故障排查等都是十分重要的技术，而目前比较公认的技术是由思科公司所提出的NetFlow方案，目前该方案已经发展到v9版本，相较于v5版本，v9提供了更加灵活、可扩展的网络流量统计技术。之后，IETF在此基础上开发了IPFIX。本文以IPFIX为主，并结合NetFlow v9的相关内容，详细介绍了其报文结构、管理方法、传输协议等内容。

**关键词：网络流统计、IPFIX、NetFlow v9**

# 综述

Cisco 系统公司的 NetFlow 服务提供一种基于数据网络中 IP 网络流信息的网络管理方式。该方式由网络中的设备节点（路由器交换机）收集网络流的相关数据并输出给收集器，这些收集的数据具有高度的灵活性和扩展性，能够提供精细的统计信息。

NetFlow v9相较于之前的版本v5增加了网络流信息统计的灵活性。相较于v5的固定消息格式，v9采用了模板的方式来记录所要提取的流信息，这使得用户可以根据自己的需要定义网络流信息统计模板。

IETF基于NetFlow v9开发了IPFIX。IPFIX报文格式与NetFlow v9报文格式几乎完全相同，其不同之处在于对于自定义字段给出了更加标准方便的定义和使用方案，另外对于模板管理和报文的传输方式、安全性进行了更加完备的介绍。

本文之后的介绍中将以RFC7011作为标准进行介绍，对于NetFlow v9（RFC3954）中不同的部分将会以注释的方式进行介绍。本文之后按照以下顺序进行介绍：在定义部分介绍了在IPFIX中所使用的相关定义。之后在报文格式部分详细介绍IPFIX的消息格式，同时介绍IPFIX消息中与NetFlow v9中不同的部分。然后介绍消息管理的内容，再介绍Information Element，之后介绍IPFIX报文的传输方式。最后对本文进行总结。

# 定义

该部分介绍了IPFIX的主要定义，我将这些定义分成了四类分别进行介绍：IPFIX硬件相关的定义，网络流相关定义、IPFIX处理架构相关定义、IPFIX报文相关定义。对于IPFIX报文相关定义我们将在下一章介绍IPFIX报文时再详细介绍。

## 2.1 IPFIX硬件相关定义

该部分定义与IPFIX硬件部分有一定的相关性，故取该名字。

### 观察点

观察点是网络中进行数据流观测的一个位置。一个观察点属于一个观察域，一个观察点可以是一些观察点的超集。比如：一个完整的线卡是一个观察点，线卡上的接口也是一个观察点，则线卡是接口的超集。为了简便，我们可以简单的将观察点看作是网络设备上的一个接口。

### 观察域

观察域是一系列观察点的集合，比如一个线卡。每一个观察域有一个单独的ID。在每一个输出过程中，观察域ID是唯一的，同时RFC7011也建议每一个IPFIX设备的观察域也是唯一的。

### 2.1.3 输出器

含有一个或多个输出过程的设备。

### 2.1.4 IPFIX设备。

同时具有生成过程和输出过程的设备。

### 2.1.5 收集器

包含多个收集过程的设备。

## 网络流相关定义

### 网络流

网络流是具有相同特性的数据包的集合。网络中流的定义是非常精细化的，可以采用包括 IP 地址，包和字节数，时间戳记，服务类型（ToS），应用端口，输入和输出接口等指标来记录或区分一个特定的流。

一般来说，一个流存在失效老化的情况，数据流在下列情况下将进入老化：

* 当该类型数据流传送完毕时，输出器将输出数据流记录。例如：在一个 TCP【参阅 RFC793】会话连接过程中出现了 FIN（No more data from sender）或者 RTS（Reset the connection）【参阅 RFC761】的请求时。
* 在一个时间段内某个数据流处于不活跃的状态，该数据流将被老化。这个时间段应该是可以由输出器进行配置的，当最小值设置成 0 时数据流立刻会进入老化状态。
* 对于一个大的（或者长时间活跃的）数据流收集器将按照规则输出数据流记录，并强制将该数据流老化，这个时间段应该在输出器上进行配置。
* 在收到输出器内部条件限制的情况下，数据流也会被强制进行老化。例如：计数器归零或内存不足。

### 2.2.2 Flow Key

Flow Key指的是用来区分不同流的参数。比如一般的流是以5元组为划分依据，则5元组便是Flow Key。

### 2.2.3流记录

流记录指的是使用IPFIX或是NetFlow所记录的流的信息，比如说流传输的包长，源和目的IP等。

## IPFIX处理架构相关定义

### 记录过程

记录过程用来产生流记录，对于进入观察点的每一个数据包根据所要观察的数据包的特征进行统计形成相应的记录信息。

### 输出过程

输出过程将记录过程的流记录通过IPFIX报文发送给收集过程。

### 收集过程

将输出过程中的IPFIX报文收集下来用于分析处理。

# IPFIX消息格式

IPFIX报文格式如图3.1所示，整个IPFIX报文被称为一个IPFIX消息（IPFIX Message），它由两部分组成，分别是消息头（Message Header）和一系列消息集合（set）组成。下面我们将分别对这两部分进行介绍。

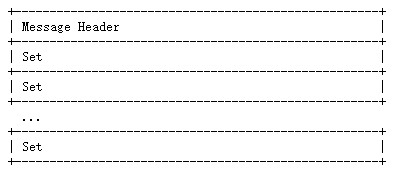


图3.1 IPFIX消息格式

## IPFIX消息头

IPFIX的消息头格式是固定的一共有5个字段，如图3.2所示，每一个字段意义如表3.1所示。

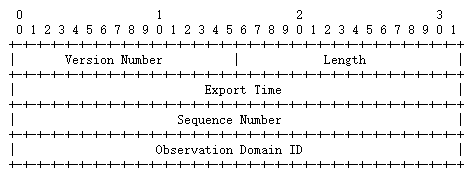


图3.2 IPFIX消息头

|  |  |
| --- | --- |
| 字段 | 介绍 |
| Version Number | IPFIX版本，其值为10。 |
| Length | IPFIX消息的长度，包括消息头，以字节为单位，最大长度为65535。 |
| Export Time | 消息输出时间，以秒为单位表示自1970年1月1日以来的时间。 |
| Sequence Number | 由输出过程从当前观测域发送到当前流中的所有IPFIX数据记录的增量序列。每个SCTP流分别计算序列号，而TCP连接或UDP会话中的所有消息都被认为是同一流的一部分。收集过程可以使用此值来确定是否遗漏了任何IPFIX数据记录。模板和选项模板记录不增加序列号。 |
| Observation Domain ID | 观察域的32位标识符，对输出过程在本地是唯一的。输出过程使用观察域ID来唯一地向收集过程标识对流进行计量的观察域。建议这个标识符对于每个IPFIX设备也是唯一的。收集过程应该使用传输会话和观察域ID字段来分离来自同一输出过程的不同输出流。当没有特定的观察域ID与整个IPFIX消息相关时，观察域ID应该为0，例如，在输出输出过程统计信息时，或者在输出聚合数据记录时在收集器的层次结构中。 |

表3.1 IPFIX消息头字段

## 消息集合（set）

IPFIX的消息集合由消息集合头（Set Header）、消息记录（record）和padding组成，如图3.3所示。。其中padding由全0组成，是为了保证消息的长度为4个字节的整数倍。一个消息集合中包含多个消息记录。

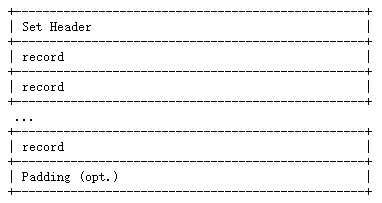


图3.3 IPFIX消息集合

### 3.2.1 消息集合头（Set Header）

消息集合头由Set ID和Length组成，如图3.4所示。其每个字段的意义如表3.2所示。

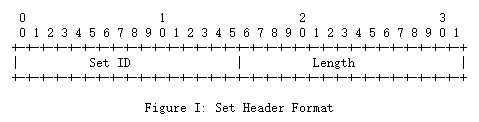


图3.4 消息集合头

|  |  |
| --- | --- |
| 字段 | 介绍 |
| Set ID | 值为2是模板消息集合，值为3为可选模板消息集合，小于255的留做将来扩展使用，大于255的值为数据消息集合使用。 |
| Length | 整个消息集合的长度，单位是字节。 |

表3.2 消息集合头字段

### 3.2.2 消息记录（record）

目前IPFIX共有三种消息记录：分别是模板消息记录、可选模板消息记录和消息记录。下面将分别介绍这三种记录形式。

#### 3.2.2.1 模板消息记录

一个模板消息集合的例子如图3.5所示。一个模板消息集合可由多个模板消息记录组成，一个模板消息记录由模板记录头和若干域声明组成，模板消息记录主要包括5个字段，各字段介绍如表3.3所示。

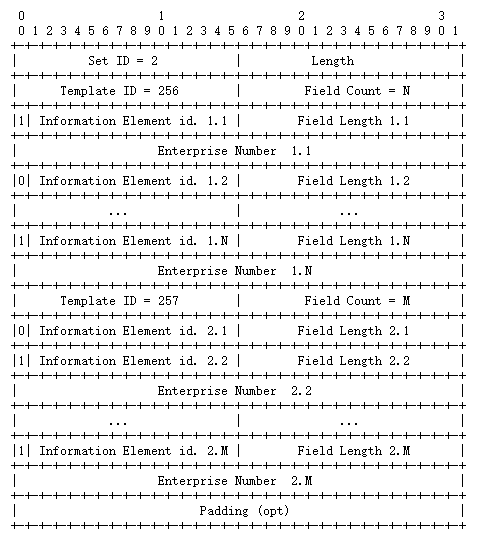


图3.5 模板消息集合实例

|  |  |
| --- | --- |
| 字段 | 介绍 |
| Template ID | 模板ID，该值>255，与某个数据消息集合中的Set ID相同。 |
| Field Count | 记录模板中每个域声明的数量。 |
| E | 为域声明的一部分，一位。0表示其在IANA中注册的标准声明，此时该域声明不包含Enterprise Number；其值为1表示自定义的声明，必须在Enterprise Number中声明自己的组织号。 |
| Information Element identifier | 表示该域声明表示什么信息。 |
| Field Length | 域信息的长度，以字节为单位。 |
| Enterprise Number | 组织号。 |

表3.3 模板消息记录字段

#### 3.2.2.2 可选模板消息记录

一个可选模板消息集合的例子如图3.6所示。一个可选模板消息集合可由多个可选模板消息记录组成。可选模板消息记录与模板消息记录相似，只是多了scope部分，可选模板记录域声明部分是由scope域声明和一般域声明部分组成的。其中Field Count记录了两部分域声明的总数，Scope Count记录了scope域声明部分的总数。

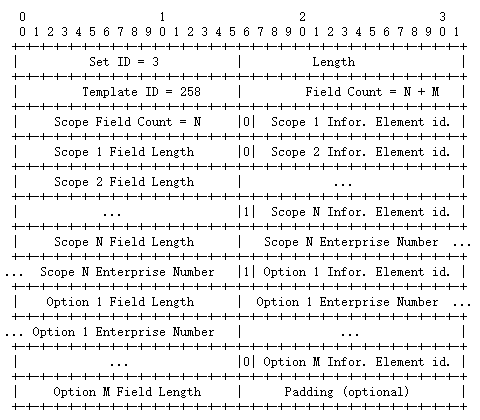


图3.6可选模板消息集合实例

#### 3.2.2.3数据消息记录

每一个数据消息记录对应着一条流的一条记录信息，其长度和类型由模板消息记录和可选模板消息记录中的域声明部分给出。

## IPFIX消息格式

IPFIX共有三种消息集合，针对三种消息集合，共有三种消息格式。我们将在该部分进行详细介绍。

### 混合格式

该消息格式如图3.7所示。该格式在一个消息中既包含了模板也包含了模板对应的数据。如果模板和数据不是特别长（整个消息总长度不超过65535），且传输方式采用的是有连接的传输方式（如TCP）适合使用该方式传输。

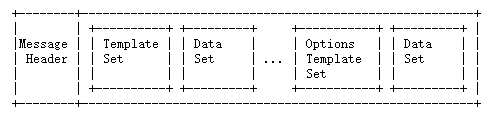


图3.7 混合模式

### 只有模板消息的格式

该消息格式如图3.8所示。消息中只包含模板消息，该消息主要与只包含数据的消息结合使用来传输相应的统计信息，这种方式适合无连接的形式（如UDP），或是传输的统计信息太多的情况。此外这种消息格式还可以用来进行模板管理。

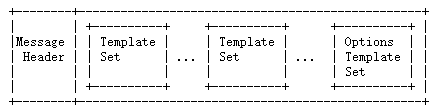


图3.8 只有模板消息的格式

### 3.3.3 只有数据消息的格式

该消息格式如图3.9所示。消息中只包含数据消息，该消息与只包含模板的消息结合传输相应统计信息。

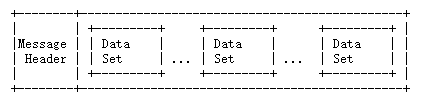


图3.9 只有数据消息的格式

## 与NetFlow v9的不同

### 3.4.1 消息头的不同

* NetFlow v9版本号为9。
* NetFlow v9的Length字段成为Count，记录的是消息集合（set）的数量，而不是消息长度。
* NetFlow v9中多了一个sysUpTime字段用来记录系统时间。

### 3.4.2 消息集合的不同

消息集合的不同之处主要体现在模板消息上。主要包括以下几点：

* IPFIX中Set ID为0时表示模板消息集合，为1时表示可选模板消息集合。
* IPFIX对自定义的模板进行了更加精细化的定义，使用E和Enterprise Number字段来声明自定义域，而在NetFlow v9中这两个字段是没有的，且对自定义模板并没有细化的规定，使得定义和使用自定义模板不方便。
* IPFIX相比NetFlow v9定义了更多的公共的域类型。

# 模板管理

IPFIX定义了详细且灵活的模板管理方式，由于IPFIX支持无连接协议UDP和有连接协议（TCP和SCTP），所以他有两种模板管理的方案。下面将详细介绍这两种方案的实现方式。

## 基于TCP和SCTP的模板管理

对于这种面向连接的协议，每一个session即是一个消息传输过程。在一个session中IPFIX采用模板撤销报文来撤销相应的模板，在该报文之后的数据消息将不会再以此模板的格式来进行传输和解释，并且之后可以重用该模板的ID。该方式的撤销报文共有两种，一种是撤销某一模板，如图4.1所示，另一种是撤销所有的模板，如图4.2所示。

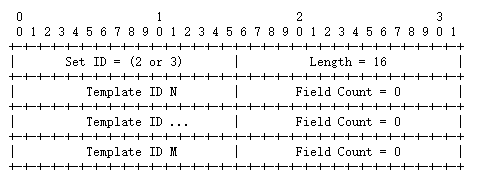


图4.1 模板撤销格式——撤销单个模板

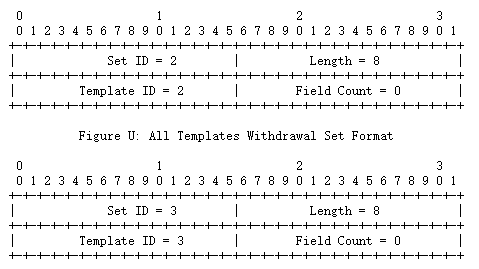


图4.1 模板撤销格式——撤销所有模板

对于这种模板管理方式来说，收集过程对于所收到的模板报文处理方式为：

* 如果收到的是一个模板撤销报文，但是其撤销的模板不存在，则收集过程将忽略这个撤销报文。
* 如果收到的是一个重复的模板报文，收集过程认为该报文是一个重传报文，不进行任何处理。
* 如果收到的是的模板报文中模板ID与当前可用的模板ID相同，但是其内容不同，则出现了一个错误，收集过程应当记录该错误。

## 基于UDP的模板管理

由于UDP没有连接的概念，传输的报文是乱序的，不能采用模板撤销报文的方式来管理模板。基于UDP的IPFIX模板管理采用如下方式：

输出过程在输出模板时同一报文需要输出多次，以确保收集过程能够正确接收模板报文。输出过程需要配置一个模板报文的重传时间，在固定时间中重传模板报文；收集过程需要配置一个模板报文生存时间。对于一个模板来说，模板报文需要定时重传确定其保活，而对于超过模板报文生存时间的模板，收集过程可认为该模板已过期，此时该模板的ID是可重用的。模板报文生存时间需要大于3倍模板报文重传时间。

对于这种模板管理方式来说，收集过程对于所收到的模板报文处理方式为：

* 如果收到一个重复的模板报文，则该报文为重传报文，不进行处理。
* 如果收到的是的模板报文中模板ID与当前可用的模板ID相同，但是其内容不同，则收集过程将新模板替换旧模板。

# Information Element

该部分是IPFIX中与NetFlow的最大区别，表示的是在IPFIX中一条流的一种记录的类型，在NetFlow中被称为field。二者在报文中的区别请参见3.4.2。

由其报文形式可以看出，Information Element可以分为两种类型，其中E为0的字段由IANA定义，若要添加需要向IANA申请，该部分的范围为1-32767，其中1-127与NetFlow v9兼容。如果E为1则可由相应组织自己定义，但必须在报文中给出组织号。

对Information Element进行描述的RFC为RFC7012，之前的版本为RFC5102。在RFC5102中对Information Element进行了类别上的划分，尽管该划分方式已经弃用，但是对Information Element的字段类型给出了较好的分类，故在此给出，如表5.1所示。

|  |  |
| --- | --- |
| 类别 | 意义 |
| Identifiers | 主要是一些用来标记IPFIX架构、IPFIX设备、IPFIX的协议的字段。 |
| Metering and Exporting Process Configuration | 记录和输出过程配置信息。 |
| Metering and Exporting Process Statistics | 记录和输出过程统计信息，主要是输出的IPFIX消息的包数等信息。 |
| IP Header Fields | IP头信息。 |
| Transport Header Fields | 传输层头信息。 |
| Sub-IP Header Fields | 子网信息，包括vlan、wlan的信息等。 |
| Derived Packet Properties | 可以看作是数据包payload中提取的信息。 |
| Min/Max Flow Properties | 包的最大长度，最小长度等信息。 |
| Flow Timestamps | 时间戳信息，比如包的开始时间，结束时间等。 |
| Per-Flow Counters | 流统计信息，流中包长、包数等。 |
| Miscellaneous Flow Properties | 流其他信息。 |
| Padding | padding信息。 |

图5.1 RFC5102对Information Element的划分

# 传输协议

IPFIX支持的传输协议有TCP、SCTP、UDP，它们之间的区别如表5.1所示。

|  |  |  |  |
| --- | --- | --- | --- |
|  | SCTP | TCP | UDP |
| 拥塞避免 | 较好的拥塞避免方式。 | 有拥塞控制，但是需要处理记录过程和输出过程在拥塞过大时不一致的情况 | 没有拥塞避免方案，UDP方式应该是在输出器与收集器直连的情况 |
| 可靠性 | 可靠。 | 可靠。 | 不可靠，可以使用IPFIX的Sequence Number字段来确认丢包。 |
| MTU | 根据PMTU进行IPFIX消息分片。 | TCP为可靠传输，IPFIX消息长度可为其最大值。 | UDP不可靠，需要对一个大的IPFIX消息进行分包后发送。 |
| session建立和关闭 | 由于SCTP支持多流传输，当关闭一条流时，IPFIX消息仍然可以在其他流上传输。 | 一个TCP连接为一个session，用来传输一次完整的IPFIX记录结果。 | session是基于时间间隔的，当超过某一间隔收集器没有收到IPFIX消息，则认为session结束。 |
| 连接故障 | 如果收集过程不承认输出过程试图建立连接，则SCTP将使用指数回退自动重试连接建立。。 | 如果收集过程不承认输出过程试图建立连接，则TCP将使用指数回退自动重试连接建立。 | 无法知道连接故障，但可以将消息发送到多个收集器来进行备份。 |

表5.1 SCTP 、TCP、UDP传输IPFIX消息时的不同

# 总结

本文针对目前流行的网络流信息统计方法——IPFIX，详细介绍了它的主要定义，协议格式和与NetFlow v9的不同，模板管理方法，Information Element以及传输协议。希望对该协议有一个比较深入的了解。此外，有关IPFIX安全性考虑、一些基本的可选模板格式、收集过程的介绍请参考RFC70114，有关Information Element的更加详细介绍请参考RFC51026与RFC70125，有关IANA已经注册的公共Information Element定义请参考“IP Flow Information Export (IPFIX) Entities”1，有关Enterprise Number的详细信息请参考“Private Enterprise Numbers”2，NetFlow v9请参考RFC39543。

# 参考文献

1. IANA,”IP Flow Information Export(IPFIX) Entities”,http://www.iana.org/assignments/ipfix
2. IANA, "Private Enterprise Numbers"，http://www.iana.org/assignments/enterprise-numbers/.
3. Claise, B., Ed., "Cisco Systems NetFlow Services Export Version 9", RFC 3954, October 2004.
4. Claise, B., Ed., Trammell, B., Ed., and P. Aitken, "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information", STD 77, RFC 7011, September 2013.
5. Claise, B., Ed., and B. Trammell, Ed., "Information Model for IP Flow Information Export (IPFIX)", RFC 7012, September 2013.
6. Quittek, J., Bryant, S., Claise, B., Aitken, P., and J.Meyer, "Information Model for IP Flow Information Export",RFC 5102, January 2008.