

# IOAM Update

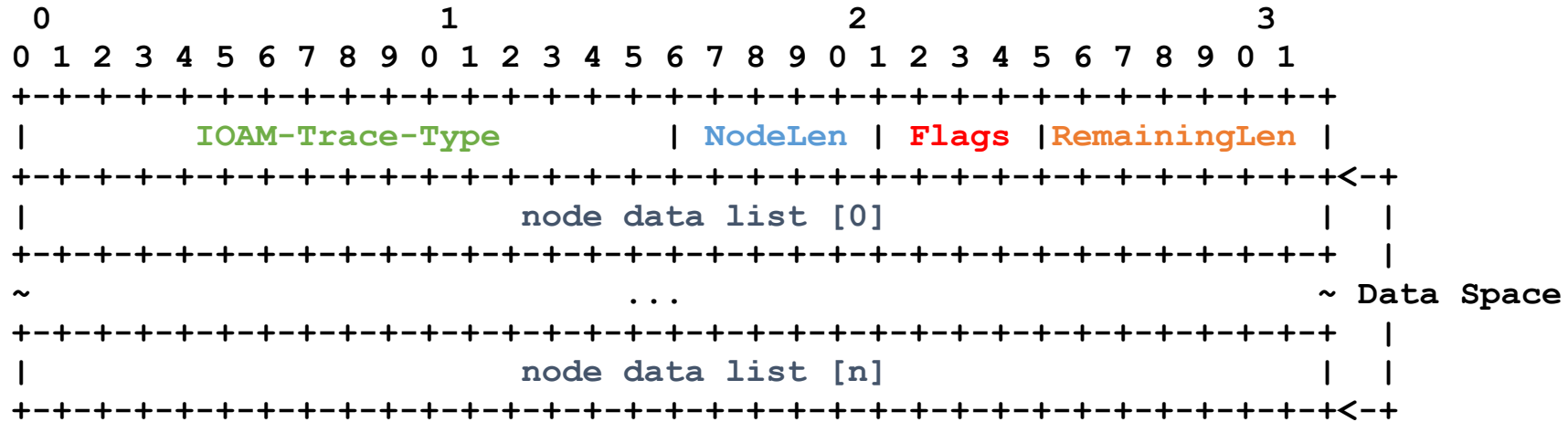
**P4 Apps Working Group**

**July 12, 2018**

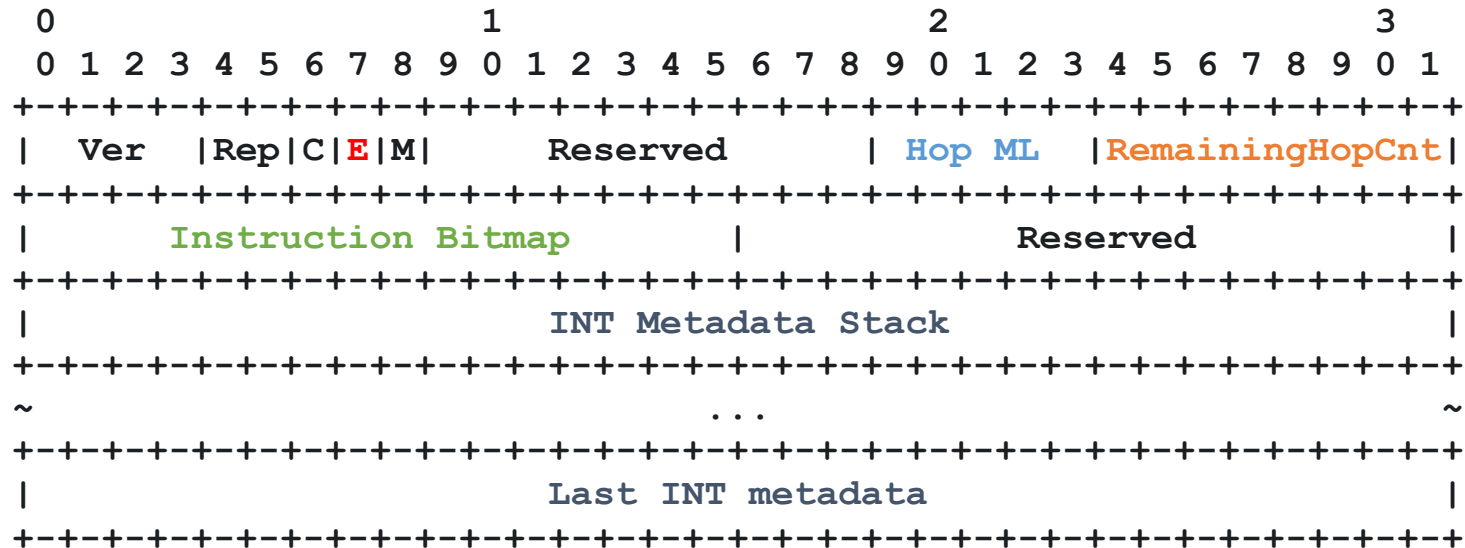
# Data Fields for In-situ OAM

draft-ietf-ippm-ioam-data-03

# In-situ OAM (IOAM) Trace Option



# INT v1.0 Hop-by-Hop Metadata Header Format



# IOAM Trace Option vs INT Hop-by-Hop Metadata Header

- IOAM-Trace-Type

- Bit 0 (MSB): Hop\_Lim + node\_id

- Bit 1: ingress\_if\_id (16 bits) + egress\_if\_id (16 bits)

- Bit 2: timestamp seconds

- Bit 3: timestamp subseconds

- Bit 4: transit delay

- Bit 5: app\_data (32 bits)

- Bit 6: queue depth

- Bit 7: variable length Opaque State Snapshot

- Bit 8: Hop\_Lim + node\_id (wide: 64 bits)

- Bit 9: ingress\_if\_id (32 bits) + egress\_if\_id (32 bits)

- Bit 10: app\_data (64 bits)

- Bit 11: Checksum Complement

- The remaining bits are undefined.

- NodeLen excludes length of Opaque State Snapshot

- RemainingLen in multiples of 4-octets

- Flags includes Overflow and Loopback bits

- Pre-allocated and Incremental Trace Options

- INT instruction bitmap:

- Bit 0 (MSB): Switch ID

- Bit 1: Level 1 Ingress Port ID (16 bits) + Egress Port ID (16 bits)

- Bit 2: Hop latency

- Bit 3: Queue ID (8 bits) + Queue occupancy (24 bits)

- Bit 4: Ingress timestamp

- Bit 5: Egress timestamp

- Bit 6: Level 2 Ingress Port ID + Egress Port ID (4 bytes each)

- Bit 7: Egress port Tx utilization

- Bit 15: Checksum Complement

- The remaining bits are reserved.

- Hop ML includes length of all metadata inserted at each hop

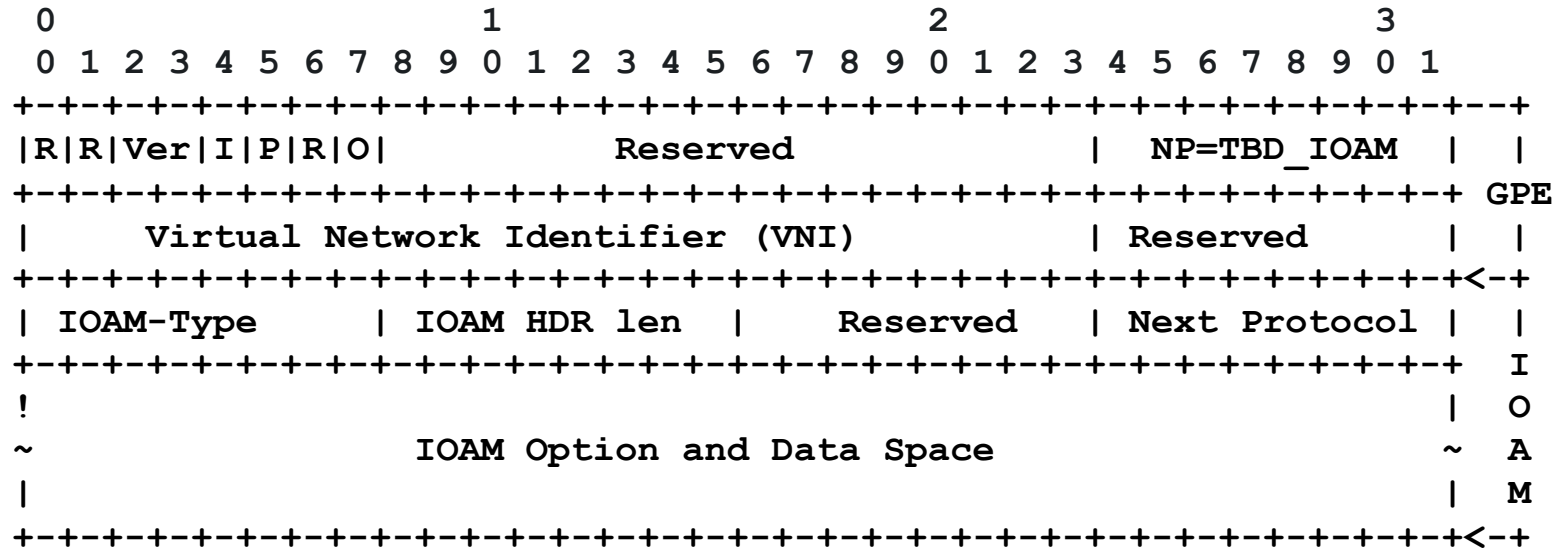
- RemainingHopCount in hops

- Max Hop Count exceeded bit, MTU exceeded bit

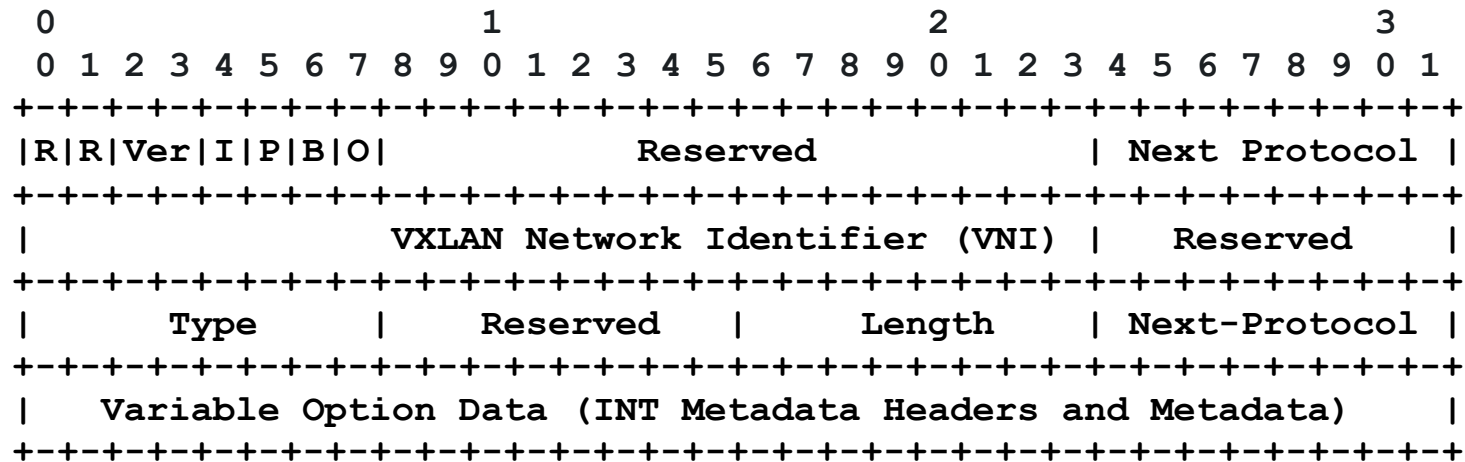
# VXLAN-GPE Encapsulation for In-situ OAM Data

draft-brockners-ippm-ioam-vxlan-gpe-01

# IOAM in VXLAN-GPE



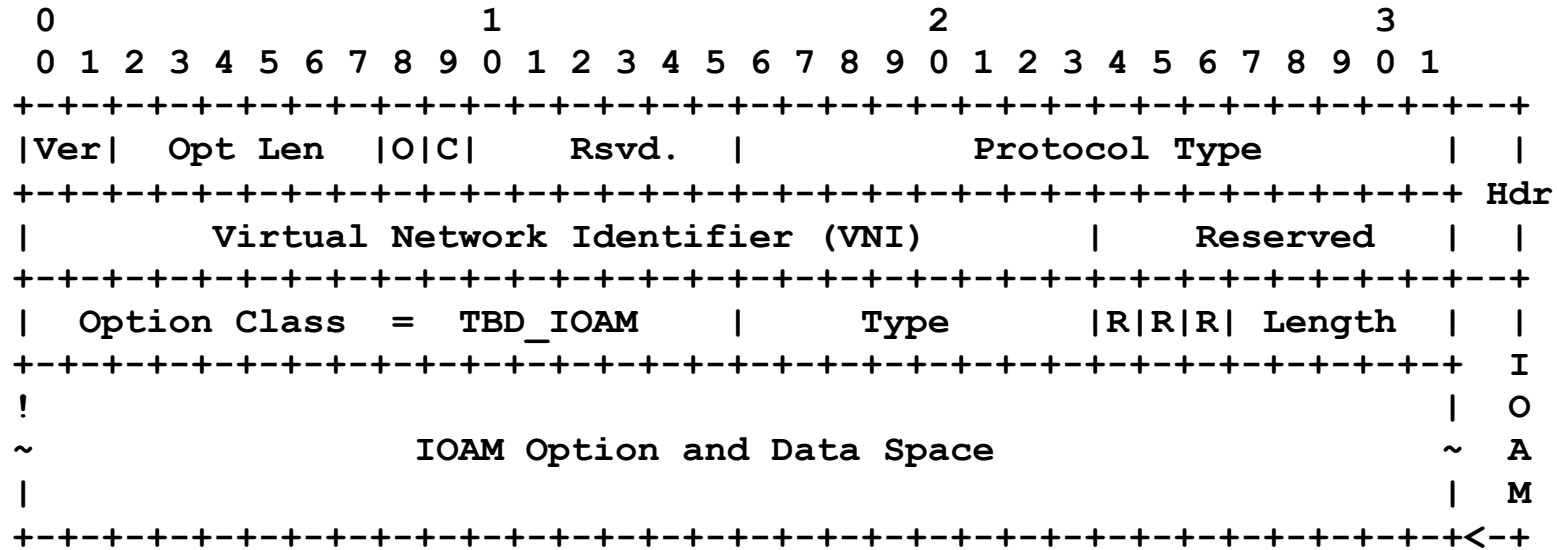
# INT over VXLAN GPE



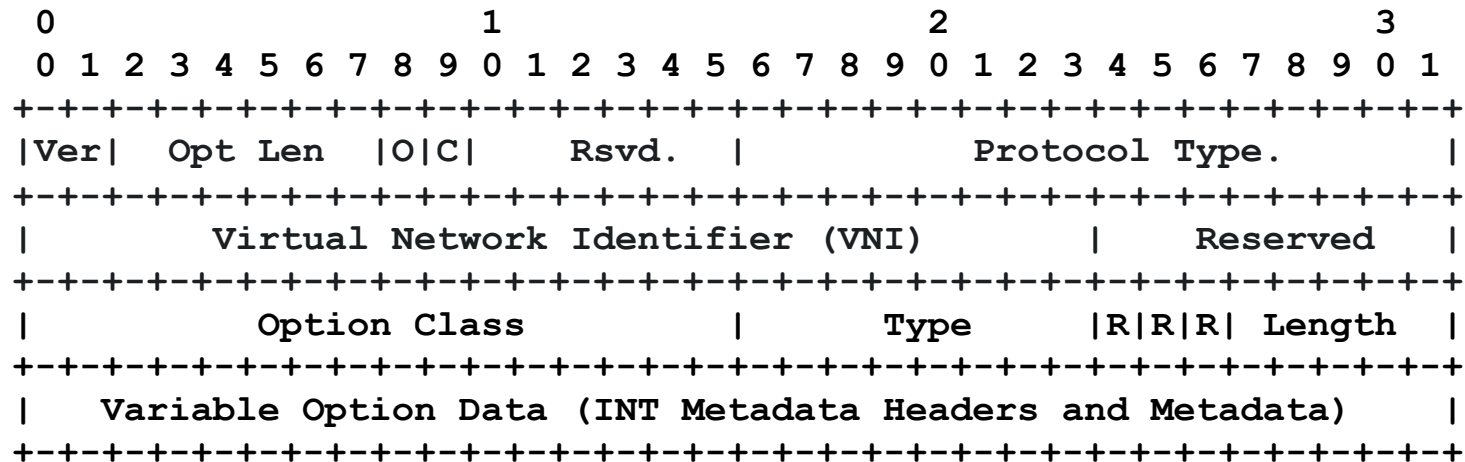
# Geneve encapsulation for In-situ OAM Data

draft-brockners-ippm-ioam-geneve-01

# IOAM in Geneve



# INT over Geneve

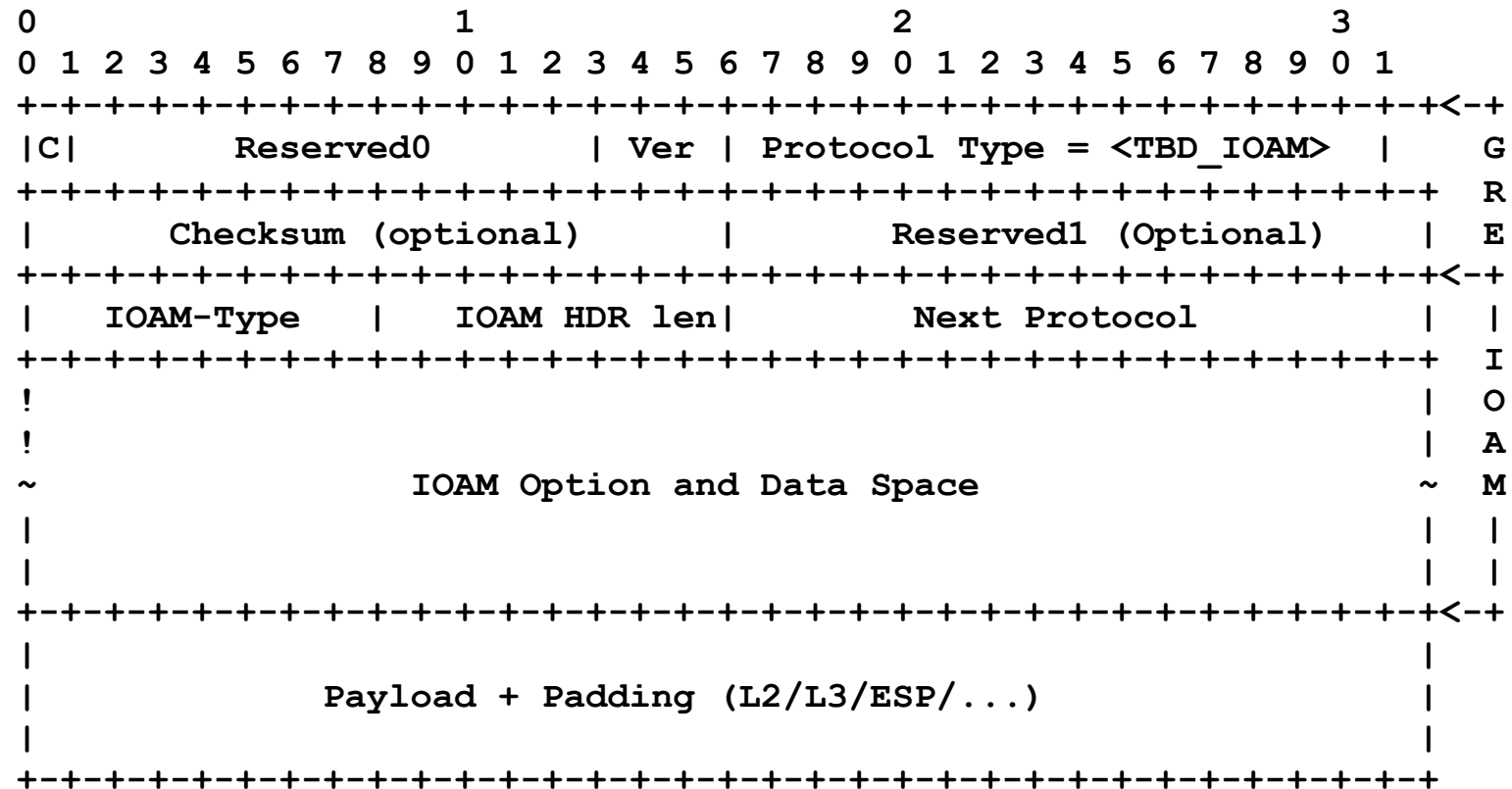




# GRE Encapsulation for In-situ OAM Data

draft-weis-ippm-ioam-gre-00

# IOAM in GRE

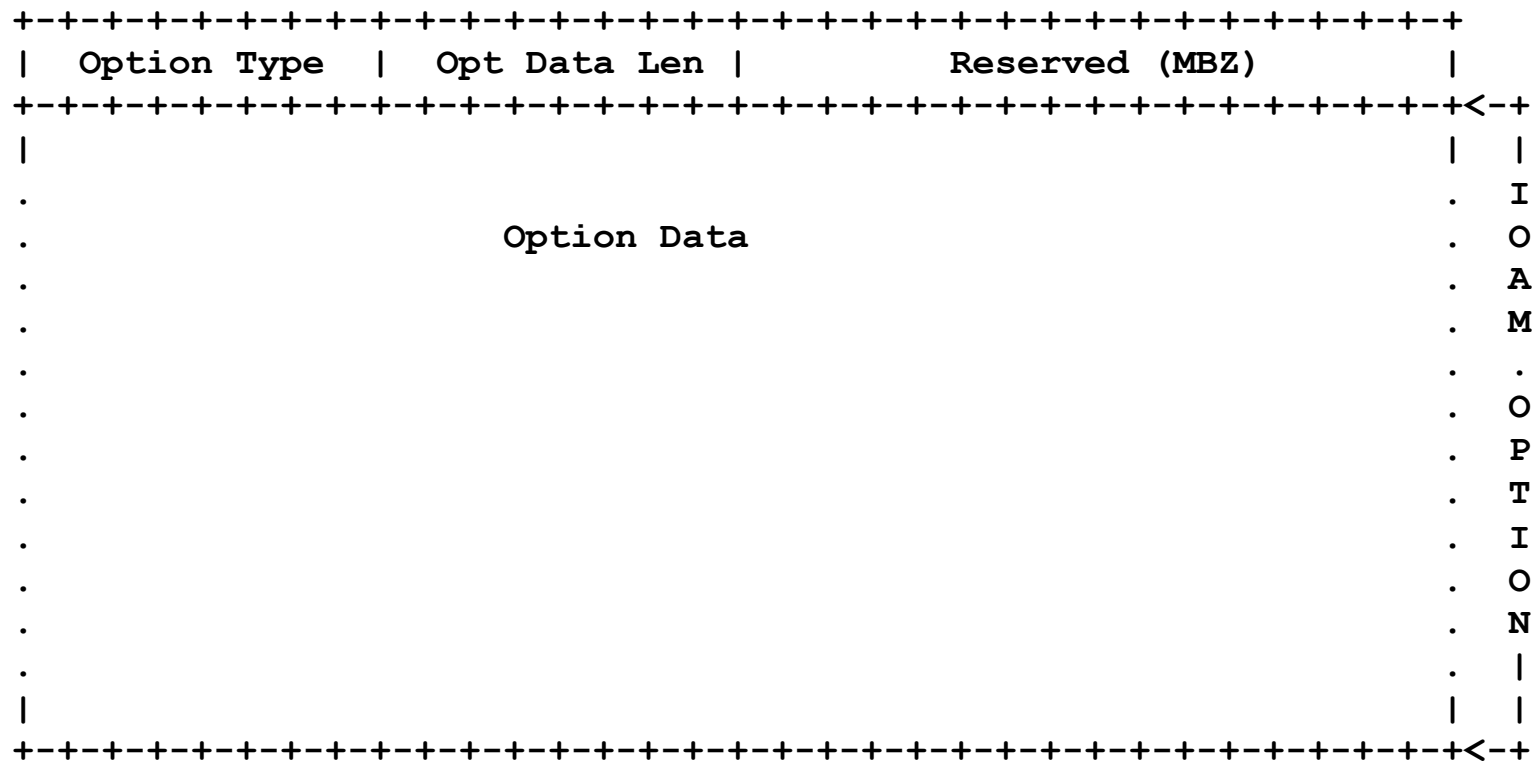


# In-situ OAM IPv6 Options

draft-ioametal-ippm-6man-ioam-ipv6-options-00

# IOAM in IPv6

- In IPv6 Hop-by-Hop Options header or IPv6 Destination Options header
- Alignment Considerations:
  - Options are aligned in a packet so that multi-octet values within the Option Data field of each option are 4-octet aligned
  - Trace-Type for Incremental Tracing Option in IPv6 MUST be selected such that the IOAM node data length is a multiple of 8-octets



# In-situ OAM raw data export with IPFIX

draft-spiegel-ippm-ioam-rawexport-00

# Using IPFIX for Raw Export of IOAM

- IPFIX/PSAMP has already defined most of the information elements needed for raw IOAM export
- Many network devices may be quite constrained in the IPFIX formats that they can support for IOAM export at high speeds
  - Small number of information elements supported
  - Alignment along 4 octet boundaries
    - Alignment constraint varies depending on the encoding used for packet snippets:
      - Fixed length :  $4n$
      - Variable length  $< 255$  :  $4n + 3$
      - Variable length  $< 65536$  :  $4n + 1$
  - Possibly a limited number of fixed templates

# Relevant IPFIX/PSAMP Information Elements (1)

313	ipHeaderPacketSection	octetArray	default	current	<p>This Information Element carries a series of n octets from the IP header of a sampled packet, starting sectionOffset octets into the IP header. However, if no sectionOffset field corresponding to this Information Element is present, then a sectionOffset of zero applies, and the octets MUST be from the start of the IP header.</p> <p>With sufficient length, this element also reports octets from the IP payload. However, full packet capture of arbitrary packet streams is explicitly out of scope per the Security Considerations sections of <a href="#">[RFC5477]</a> and <a href="#">[RFC2804]</a>.</p> <p>The sectionExportedOctets expresses how much data was exported, while the remainder is padding.</p> <p>When the sectionExportedOctets field corresponding to this Information Element exists, this Information Element MAY have a fixed length and MAY be padded, or it MAY have a variable length.</p> <p>When the sectionExportedOctets field corresponding to this Information Element does not exist, this Information Element SHOULD have a variable length and MUST NOT be padded. In this case, the size of the exported section may be constrained due to limitations in the IPFIX protocol.</p>
410	sectionExportedOctets	unsigned16	quantity	current	<p>This Information Element specifies the observed length of the packet section (e.g., dataLinkFrameSection, ipHeaderPacketSection, ipPayloadPacketSection, mplsLabelStackSection, and mplsPayloadPacketSection) when padding is used.</p> <p>The packet section may be of a fixed size larger than the sectionExportedOctets. In this case, octets in the packet section beyond the sectionExportedOctets MUST follow the <a href="#">[RFC7011]</a> rules for padding (i.e., be composed of zero (0) valued octets).</p>

# Relevant IPFIX/PSAMP Information Elements (2)

315	dataLinkFrameSection	octetArray	default	current	<p>This Information Element carries n octets from the data link frame of a selected frame, starting sectionOffset octets into the frame.</p> <p>However, if no sectionOffset field corresponding to this Information Element is present, then a sectionOffset of zero applies, and the octets MUST be from the start of the data link frame.</p> <p>The sectionExportedOctets expresses how much data was observed, while the remainder is padding.</p> <p>When the sectionExportedOctets field corresponding to this Information Element exists, this Information Element MAY have a fixed length and MAY be padded, or MAY have a variable length.</p> <p>When the sectionExportedOctets field corresponding to this Information Element does not exist, this Information Element SHOULD have a variable length and MUST NOT be padded. In this case, the size of the exported section may be constrained due to limitations in the IPFIX protocol.</p> <p>Further Information Elements, i.e., dataLinkFrameType and dataLinkFrameSize, are needed to specify the data link type and the size of the data link frame of this Information Element. A set of these Information Elements MAY be contained in a structured data type, as expressed in <a href="#">[RFC6313]</a>. Or a set of these Information Elements MAY be contained in one Flow Record as shown in Appendix B of <a href="#">[RFC7133]</a>.</p> <p>The data link layer is defined in [ISO/IEC.7498-1:1994].</p>
408	dataLinkFrameType	unsigned16	flags	current	<p>This Information Element specifies the type of the selected data link frame. The following data link types are defined here:</p> <ul style="list-style-type: none"><li>- 0x01 IEEE802.3 ETHERNET [IEEE802.3]</li><li>- 0x02 IEEE802.11 MAC Frame format [IEEE802.11]</li></ul> <p>Further values may be assigned by IANA. Note that the assigned values are bits so that multiple observations can be OR'd together.</p> <p>The data link layer is defined in [ISO/IEC.7498-1:1994].</p>



# Relevant IPFIX/PSAMP Information Elements (3)

---

89	forwardingStatus	unsigned8	identifier	current	<p>This Information Element describes the forwarding status of the flow and any attached reasons.</p> <p>The layout of the encoding is as follows:</p> <p>MSB - 0 1 2 3 4 5 6 7 - LSB</p> <pre> +---+---+---+---+---+---+---+   Status  Reason code or flags   +---+---+---+---+---+---+---+</pre> <p>See the Forwarding Status sub-registries at <a href="http://www.iana.org/assignments/ipfix/ipfix.xhtml#forwarding-status">[http://www.iana.org/assignments/ipfix/ipfix.xhtml#forwarding-status]</a>.</p> <p>Examples:</p> <p>value : 0x40 = 64</p> <p>binary: 01000000</p> <p>decode: 01 -&gt; Forward</p> <p>          000000 -&gt; No further information</p> <p>value : 0x89 = 137</p> <p>binary: 10001001</p> <p>decode: 10 -&gt; Drop</p> <p>          001001 -&gt; Bad TTL</p>
----	------------------	-----------	------------	---------	---

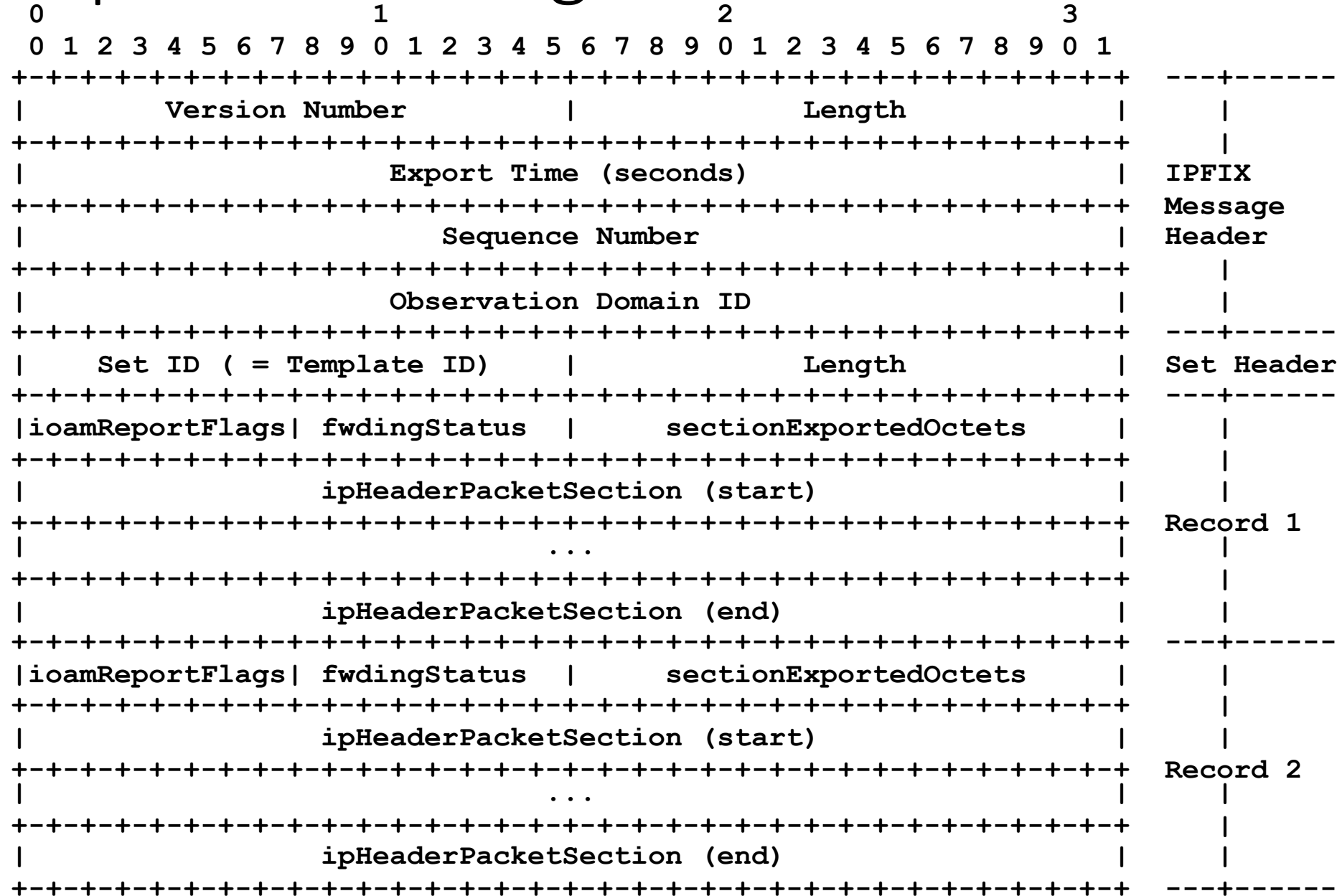
---

# Proposed new IPFIX/PSAMP Information Elements

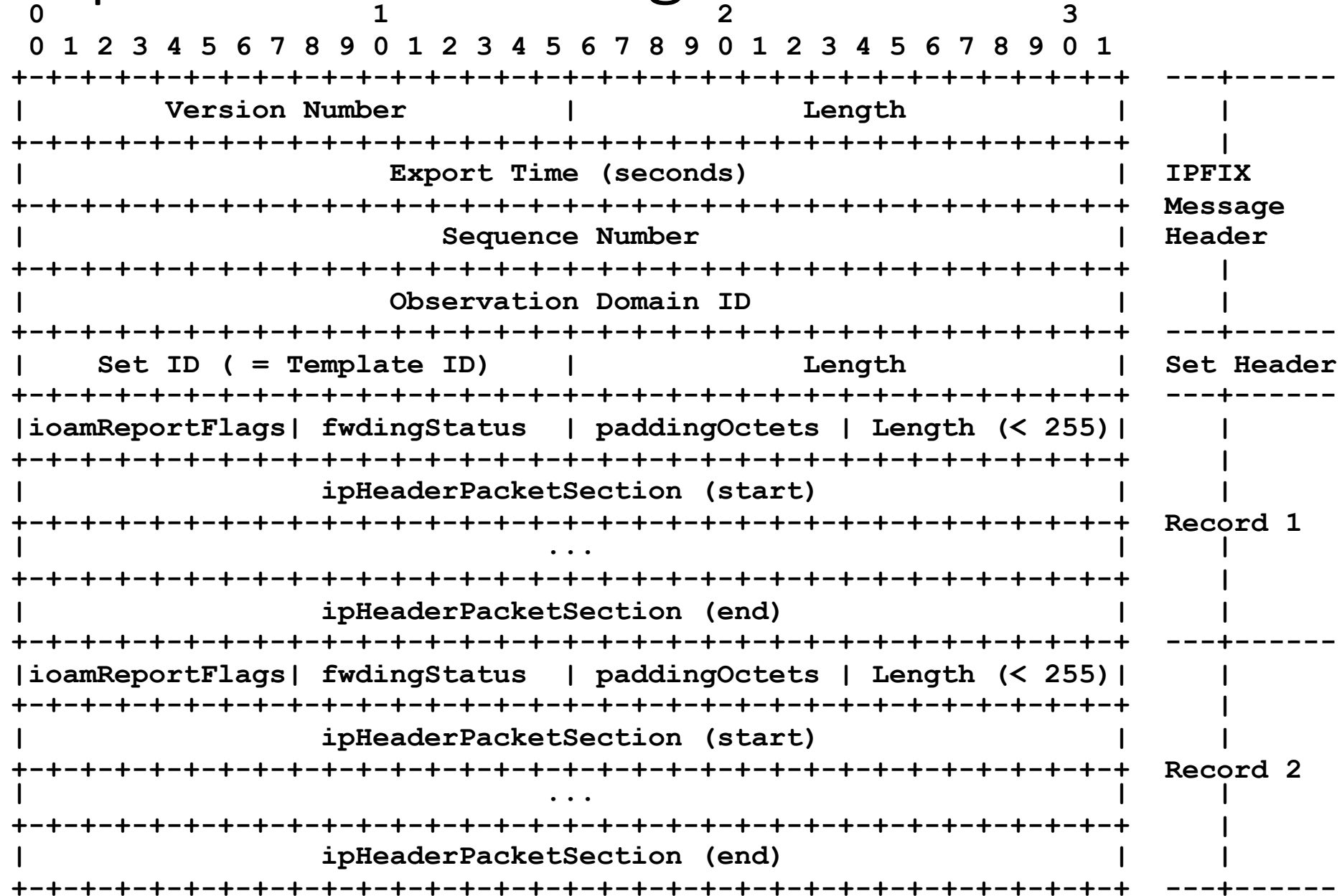
tbd	ioamReportFlags	unsigned8	flags	current	<div>This Information Element describes properties associated with an IOAM report</div> <div><div>bit</div><div>description</div><div>-----+-----</div><div>0</div><div>Dropped Association - Dropped packet of interest</div><div>1</div><div>Congested Queue Association - Indicates the presence of congestion on a monitored queue</div><div>2</div><div>Tracked Flow Association - Matched a flow of interest</div><div>3-7</div><div>Reserved</div></div>
tbd	ioamEncapsulationType	unsigned8	identifier	current	<div>0</div> <div>None: [I-D.ietf-ippm-ioam-data]</div> <div>1</div> <div>GRE: [I-D.weis-ippm-ioam-gre]</div> <div>2</div> <div>VXLAN-GPE: [I-D.brockners-ippm-ioam-vxlan-gpe]</div> <div>3</div> <div>GENEVE: [I-D.brockners-ippm-ioam-geneve]</div> <div>4</div> <div>NSH: [I-D.brockners-sfc-ioam-nsh]</div>
tbd	ioamPreallocatedTraceData	octetArray	default	current	<div>This Information Element carries n octets of IOAM Preallocated Trace data. *</div>
tbd	ioamIncrementalTraceData	octetArray	default	current	<div>This Information Element carries n octets of IOAM Incremental Trace data. *</div>
tbd	ioamE2EData	octetArray	default	current	<div>This Information Element carries n octets of IOAM E2E data. *</div>
tbd	ioamPotData	octetArray	default	current	<div>This Information Element carries n octets of IOAM POT data. *</div>

\* When ioamEncapsulationType is present and has a value other than "None", and with sufficient length, this element may also report octets from subsequent headers and payload

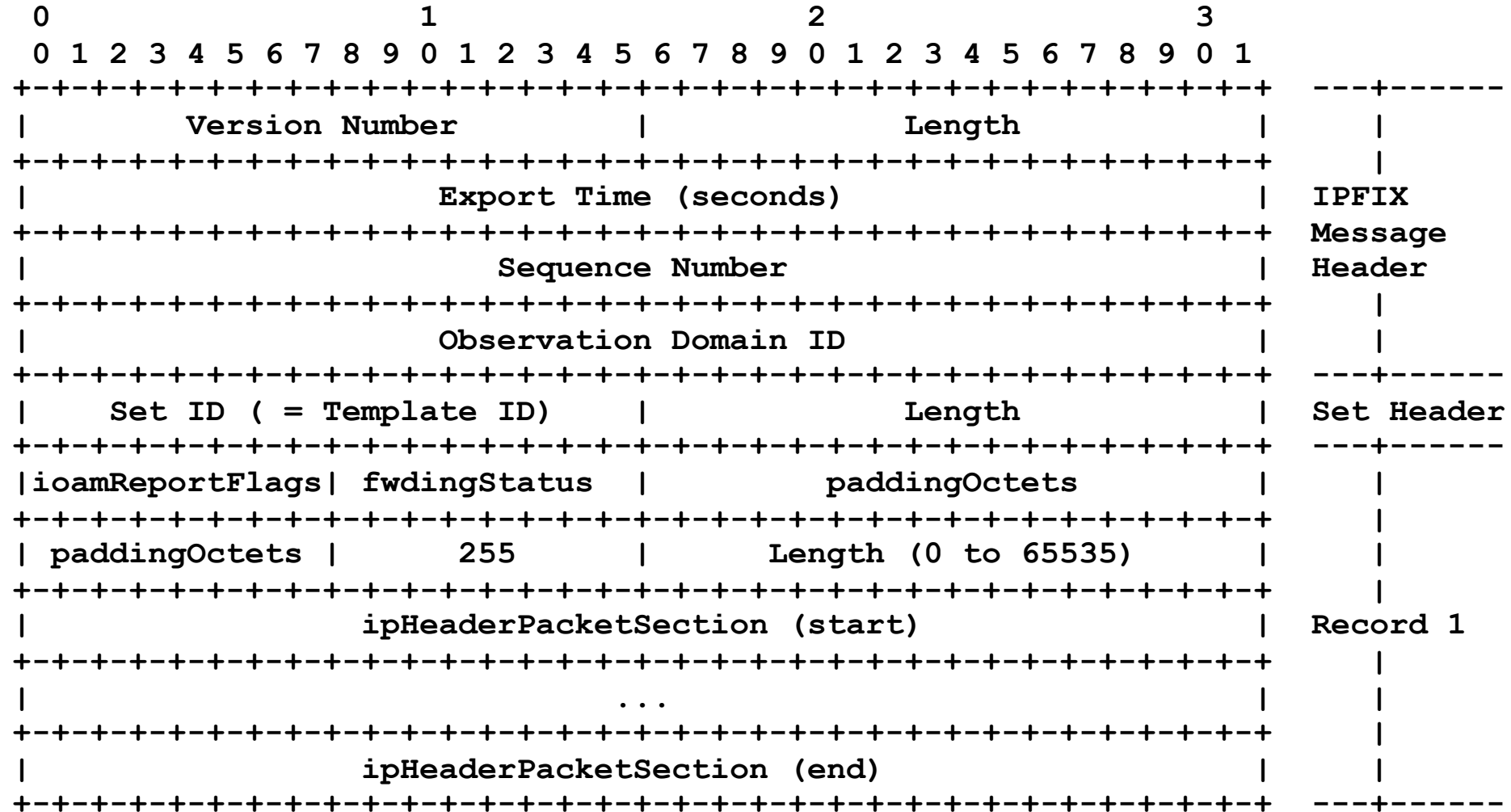
# Example: Fixed Length IP Packet



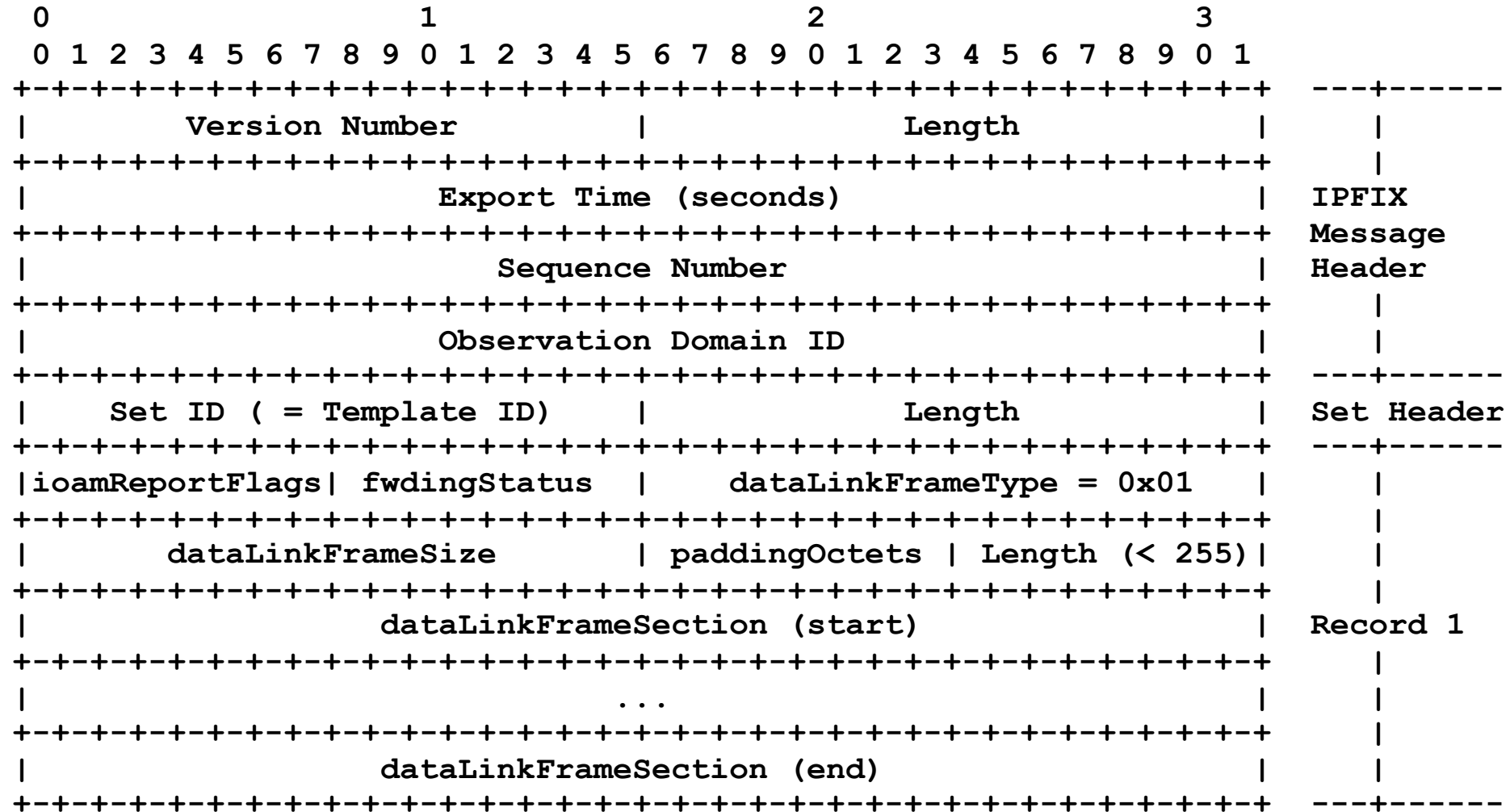
# Example: Variable Length < 255 IP Packet



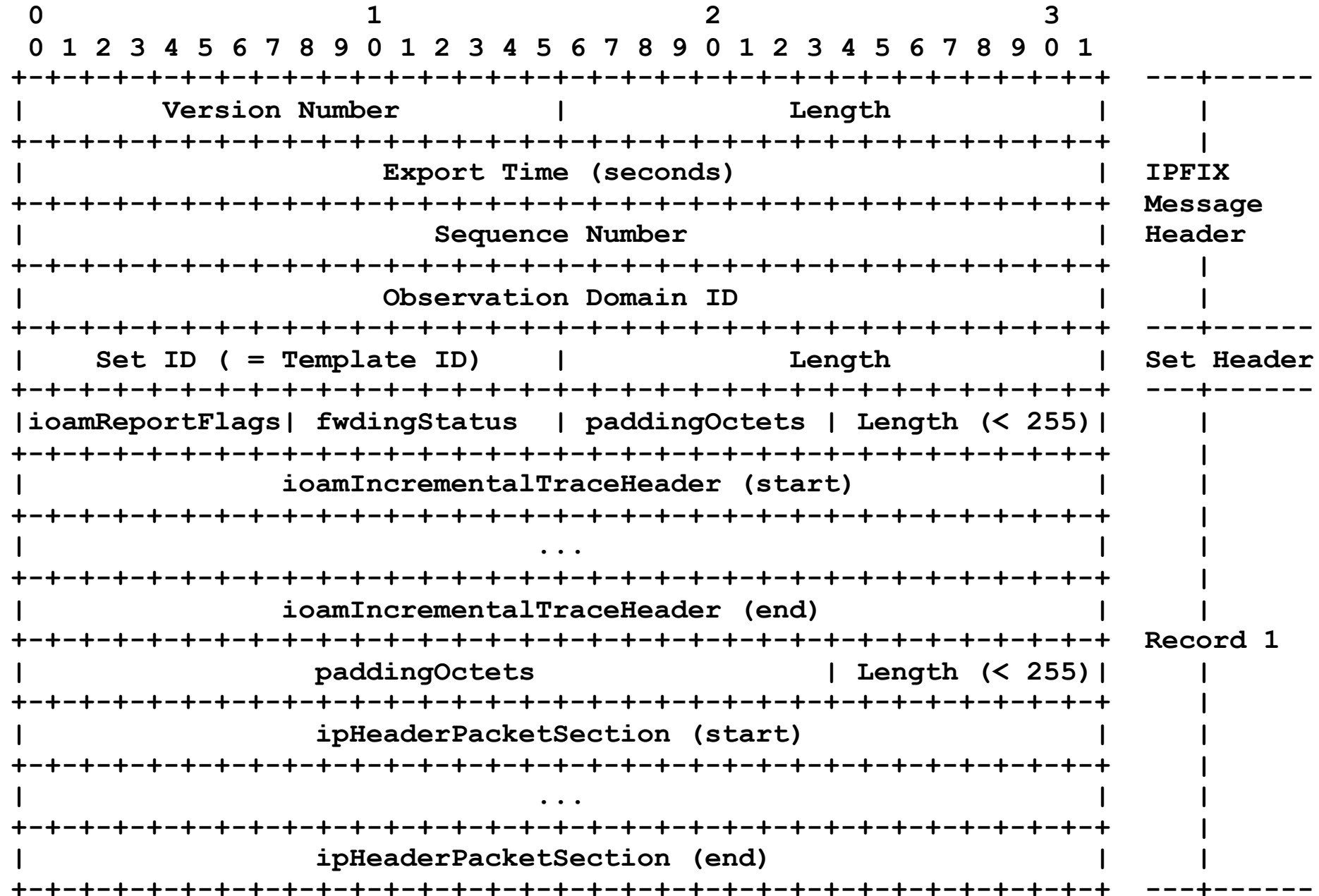
## Example: Variable Length > 255 IP Packet



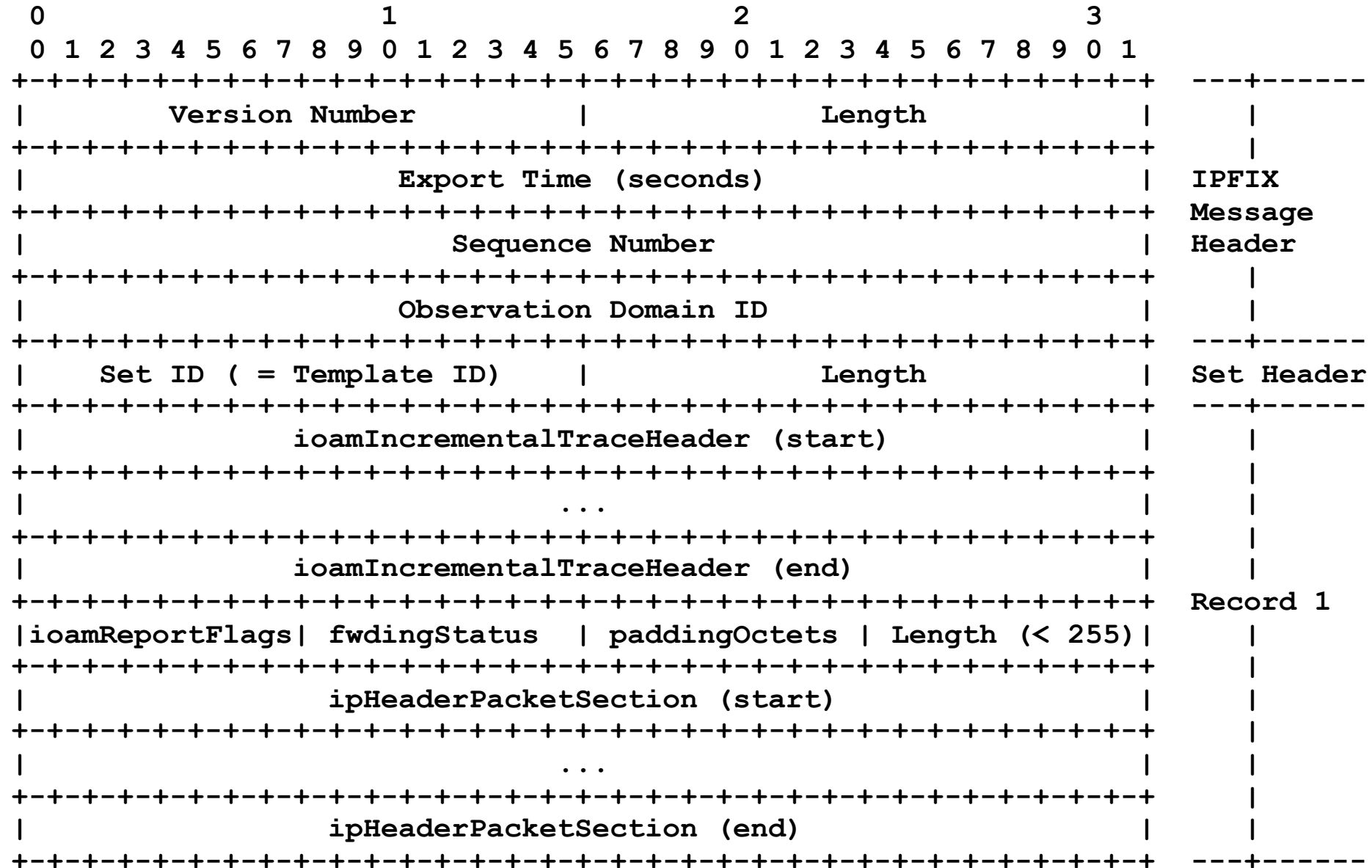
# Example: Variable Length < 255 Ethernet Packet



# Example: Variable Length IP Packet w/ Outer Incremental Trace



# Example: Variable Length IP Packet w/ Outer Incremental Trace



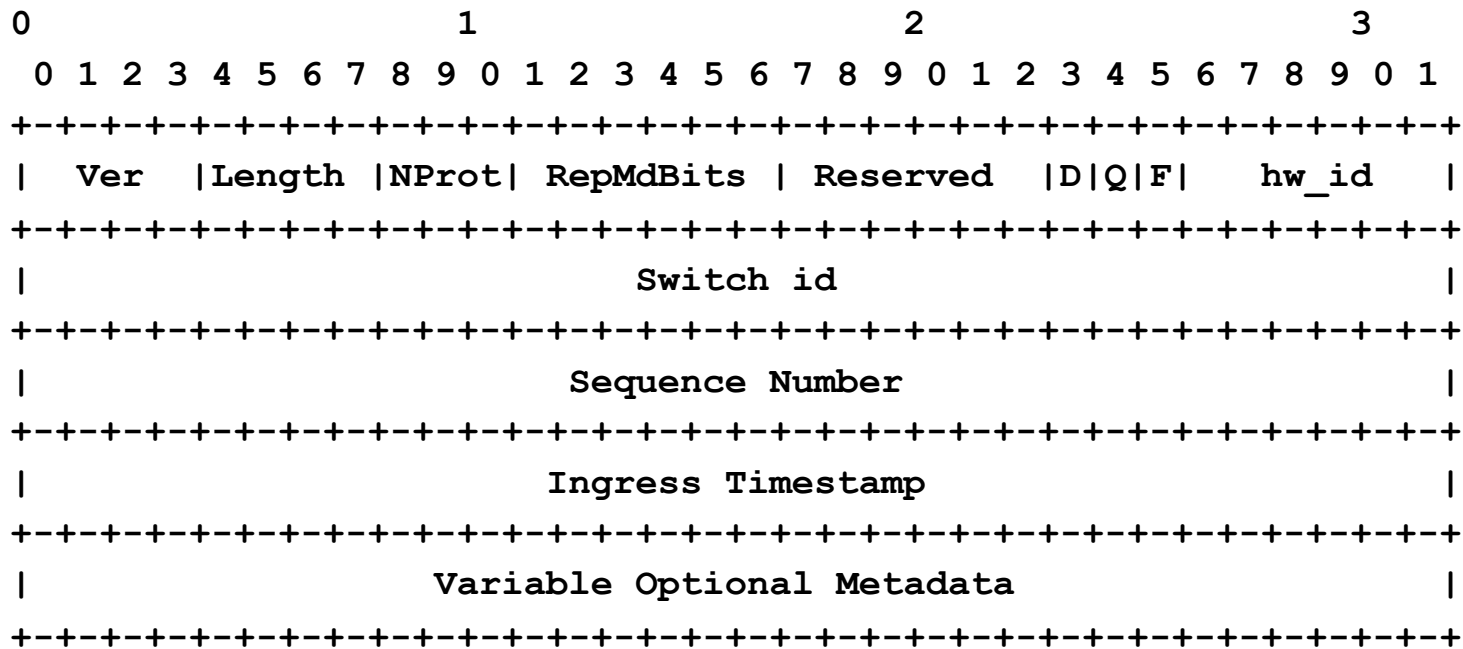


# Backup Slides

# Inefficiencies in Current IPFIX/PSAMP Definitions

- If length > 255 is required, length field takes 3 octets
- *ipHeaderPacketSection*
  - *When the sectionExportedOctets field corresponding to this Information Element does not exist, this Information Element SHOULD have a variable length and MUST NOT be padded.*
    - If fixed size encoding is used, must include *sectionExportedOctets* (16) even though there is a length in the IP header
- *dataLinkFrameSection*
  - If fixed size, must include *sectionExportedOctets* (16) just like *ipHeaderPacketSection*
  - *Further Information Elements, i.e., dataLinkFrameType and dataLinkFrameSize, are needed to specify the data link type and the size of the data link frame of this Information Element. A set of these Information Elements MAY be contained in a structured data type, as expressed in [[RFC6313](#)]. Or a set of these Information Elements MAY be contained in one Flow Record as shown in Appendix B of [[RFC7133](#)].*
    - Must include *dataLinkFrameType* (16) and *dataLinkFrameSize* (16)
      - *dataLinkFrameSize* even when a variable length encoding is used
- Padding between multiple variable length fields
- Export Time is only in units of seconds

# P4.org Telemetry Report v1.0 Header



RepBits: Report Metadata Bits indicate which optional metadata (4 octets each) is present

bit 0 (MSB): Ingress port id (16 bits) + Egress port id (16 bits)

bit 1: Hop latency

bit 2: Queue id (8 bits) + Queue occupancy (24 bits)

bit 3: Egress Timestamp (32 bits)

bit 4: Queue id (8 bits) + Drop reason (8 bits) + Padding (16 bits)

bit 5: Egress port tx utilization