NVO3
Internet-Draft
Intended status: Informational
Expires: October 7, 2018

D. Worley
Ariadne

April 5, 2018

# Geneve Extensions
# draft-worley-nvo3-geneve-misc-00

Abstract

TBD

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

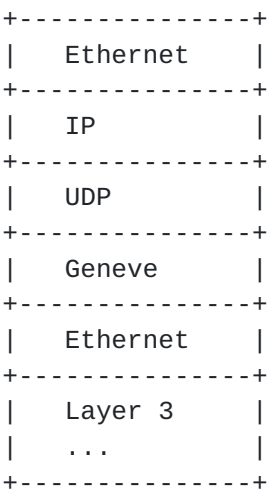
This Internet-Draft will expire on October 7, 2018.

Table of Contents

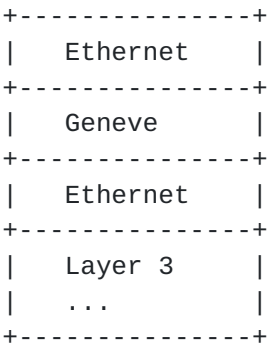


## 1. Introduction

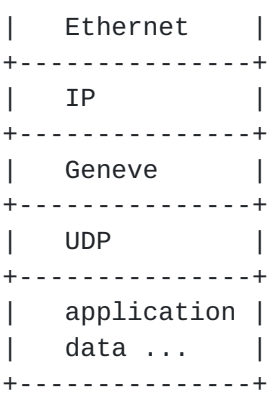
This document concerns expanding the Geneve encapsulation header to the function of a generalized encapsulation. The current Geneve proposal[I-D.ietf-nvo3-geneve] is highly extensible regarding the information that can be carried in the Geneve header, but it envisions that the encapsulated payload will be either Ethernet or a layer 3 protocol and will be carried over UDP within an IPv4 or IPv6 packet:




This document envisions that Geneve header may be used for other functions. These include tunneling at different layers of the stack, such as tunneling Ethernet over Ethernet:




or carrying additional information to be processed at various levels of the protocol stack:




remarkably little extension is needed to allow Geneve to take on this much-expanded role.

## 2. Protocol Numbers

The most important requirement for the concept of Geneve as a generalized encapsulation technique is assigning suitable protocol numbers so that Geneve can be demultiplexed at various layers of the protocol stack.[number-assignments-msg]

### 2.1. Geneve over UDP

When Geneve is used over layer 4, UDP, then there needs to be an assigned UDP destination port to specify that the UDP payload is demultiplexed as Geneve. IANA has assigned 6081 as the port number.[I-D.ietf-nvo3-geneve]

### 2.2. Geneve over IP

When Geneve is used over layer 3, IP, it needs a protocol/next header value. Protocol values are only 8 bits, but only a bit over half of the protocol values have been allocated in 30 years, so it seems that there's not a lot of pressure on the number-space, and it is reasonable to request a protocol number assignment for Geneve. Currently, the next unused IP Protocol value is 143.[protocol-numbers]

### 2.3. Geneve over Ethernet

When Geneve is used over layer 2, Ethernet, it needs an Ethertype assignment. An Ethertype assignment could be obtained from IEEE,[ieee-ethertype-reg] but it might be more expedient to obtain a SNAP Protocol Number from IANA.[RFC7042] Currently, the next unused SNAP Protocol Number is 0x0009, yielding the five-octet SNAP extension header 00-00-5E-00-09.[snap-protocol-numbers] The SNAP extension header appears in the Ethernet frame after the primary Ethernet header when the Ethertype in the Ethernet header is the OUI Extended EtherType, 0x88B7.

[IEEE_802-2014] clause 9.2.4 notes

> As discussed in 9.2.3, it is good protocol development practice to use a protocol subtype, along with a protocol version identifier in order to avoid having to allocate a new protocol identifier when a protocol is revised or enhanced. Users of the OUI Extended EtherType are, therefore, encouraged to include protocol subtype and version information in the specification of the protocol data for their protocols.

Geneve satisfies this desideratum through (1) using the first two bits of the Geneve header as a version identifier, and (2) carrying most of its data in an extensible set of options.

## 3. Geneve Protocol Type Numbers

The Geneve header contains a protocol type field which identifies the protocol of the payload of the Geneve header, i.e., the overlying protocol. The protocol type field is defined to be an Ethertype. To allow the Geneve payload to be other than layer 3 protocols (with a few layer 2 protocols specifiable through "encapsulated" Ethertypes), encodings are needed for a larger space of protocol identifers.

### 3.1. Ethernet over Geneve

When layer 2 is used over Geneve, the protocol type is 0x6558 (encapsulated Ethernet).

### 3.2. IP over Geneve

When layer 3 is used over Geneve, the protocol type is 0800 (IPv4) or 86DD (IPv6).

### 3.3. Layer 4 over Geneve

When layer 4 is used over Geneve, Geneve must be extended, because there's no defined way of representing an IP protocol/next header value directly as an Ethertype, and few or no protocols that can be represented as such a value have assigned Ethertypes.

One approach for carrying layer 4 protocols depends on the fact that all Ethertypes must have values of 0x0600 or higher ([IEEE_802-2014] clause 9.2.1). That is because Ethertypes are carried in a two-octet field in the Ethernet header, and values of that field smaller than 0x0600 are defined to specify the length of the Ethernet frame, not its Ethertype. This implies that values of the Geneve protocol type field with a first octet of 0x00 to 0x05 are not actually allocated at present and can be redefined for other uses.

This suggests extending the protocol type field so that if the first octet is 0x00, then the second octet is an IP protocol number describing the payload protocol.

#### 3.3.1. No Payload

An additional case is when there is no payload, i.e., the Geneve header contains all of the information content of the packet. In this situation, we can take use the next-header value 59, which means

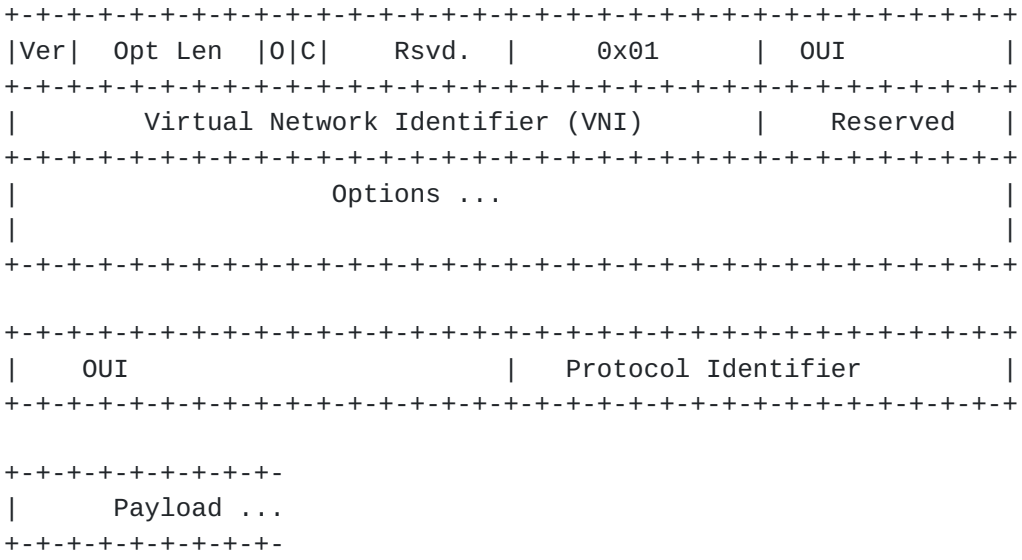
"no next header".[RFC8200] Given the encoding for IP protocol numbers, "no next header" is encoded by a Geneve protocol type of 0x003B.

### 3.3.2. Pseudoheaders for Checksums

TBD

## 3.4. SNAP Ethertypes over Geneve

But things get messier if Geneve directly encapsulates a protocol which has a SNAP protocol number, because Geneve only reserves two octets for the protocol type field. Likely the best solution[extended-msg] is allocate (1) the first octet of the protocol type field is an indicator value, 0x02, (2) the second octet of the protocol type field is the first octet of the OUI of the SNAP protocol number, and (3) as the first four octets of the Geneve payload, place the final two octets of the OUI and the two octets of the protocol identifier:




This solution allows the format of the Geneve header to remain unchanged, and the payload of the inner protocol remains aligned on a four-octet boundary. It does introduce an irregular word between the Geneve header and the payload proper, but this is upward-compatible with the current Geneve specification: a processor that does not understand the SNAP payload convention sees an Ethertype starting with 0x02 (which it does not recognize) and a payload that is four octets longer than the actual payload (which it knows it does not know how to process).

(The value 0x02 is chosen as the indicator because numerous Ethertypes with high byte 0x01 are listed in the IEEE's registration database for "Xerox (Experimental)", despite that they are not acceptable as Ethertypes.)

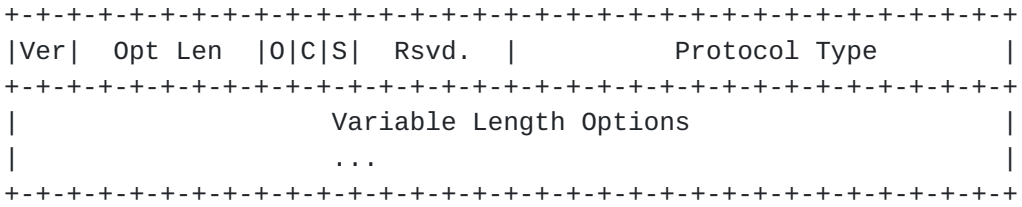
## 4. Alternate Packet Marking

"Alternate packet marking" encompasses a number of methods of inserting one or more "marking" bits in packets when they are transmitted, and when they are received, measuring the arrival times of packets with specific markings to determine flow statistics, including delay and jitter.[I-D.fmm-nvo3-pm-alt-mark][I-D.mizrahi-ippm-compact-alternate-marking]

Alternate marking is mentioned here because even if compact marking is used, one of the reserved flag bits needs to be allocated for marking.

## 5. Short Header

For some applications, there may be no need for a Virtual Network Identifier. It may be reasonable to allocate one of the header flag bits to mean "short", in that the second word of the header is suppressed. Thus, when S = 1, the format of the Geneve header is:




## 6. TCAM Support

My memory is that the question of "TCAM support" was raised a couple of IETFs ago. I take that to mean the question of whether a Geneve header can be classified as to whether it conforms to a particular "profile" or not by using a ternary-content-addressed memory, that is, by examining a fixed set of bits at fixed locations to see if they have certain fixed values. If the profile in question is the presence of a particular set of options with particular lengths, the classification is possible using TCAM.

A problem arises if we want to support profiles that allow further options beyond the specified set. It is straightforward to verify that the required option class/type values appear in the expected locations relative to the Geneve header. The problem is that there's no way to verify that the apparent options are actually within the

Geneve header -- to do that would require that the Option Length field is greater than or equal to a specified constant, and that test can't be done via TCAM.

One solution is embed in the Geneve header fixed part, and each option, a flag telling whether there is a further option in the Geneve header.[tcam-msg] Then a Geneve header can be verified to have at least N options (when N > 1) when the another-option flag is true in the header fixed part and in the first N-1 options.

(Effectively, the another-option flags represent the number of options in the Geneve header in uniary, and using TCAM it is possible to compare a number to be greater than a constant if the number is represented in unary.)

## 7. Allocation of Flag Bits

The Geneve format reserves eight bits in the first word for flag bits. Accumulating all of these proposals, there are five allocation flag bits:

O (existing) packet contains OAM information. The endpoint should direct the packet into a control queue.

C (existing) critical option is present

M (new) traffic marking for measurementSection 4

A (new) (in the fixed part) there are one or more options, (in an option) this are one or more options following this oneSection 6

S (new) short header - the second word of the fixed part is absentSection 5

## 8. Applications

### 8.1. Packet Spraying

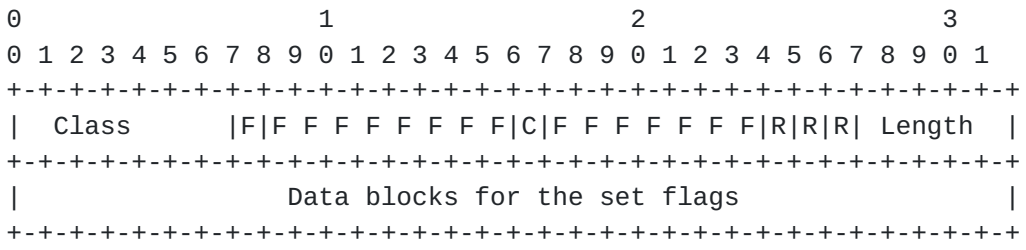
One application of these extensions is described in the draft [I-D.xiang-nvo3-geneve-packet-spray] in section 5.3, in which it is desirable to put the layer 4 header (TCP or UDP) directly after the Geneve header (which carries the "Flow Group ID" and "Sequencing Number" in an option). In these cases, the Geneve protocol type field will be 0x0006 (for TCP) or 0x0011 (for UDP).

## 8.2. OAM Headers

Within this framework, an attractive application is implementing the proposed OOAM header as a Geneve header. The current header proposal[I-D.ooamdt-rtgwg-ooam-header] exhibits some of the properties that Geneve was designed to avoid, in particular, specification of various functions by means of a limited number of flag bits in a fixed order. These limitations can be avoided by reformatting the fields of the OOAM header as a Geneve header.

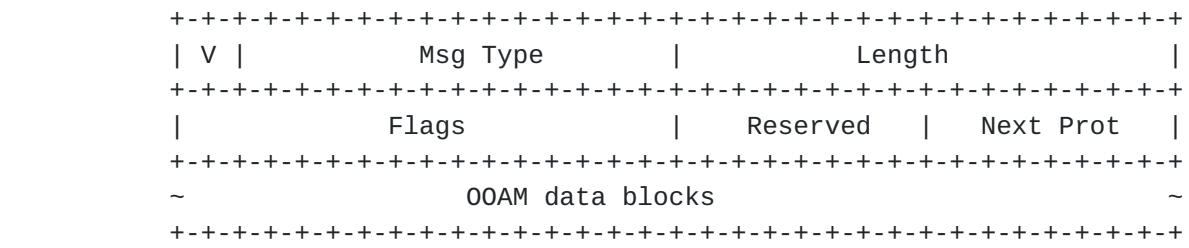
The OOAM header has three parts: a fixed part, the data blocks specified by the Flags field, and an OOAM control message.

The (potentially) sixteen flags that indicate the data blocks can be replaced by sixteen Geneve options, each of which carries as its data the data block of the corresponding flag. This has the disadvantage that if a number of flags would be specified in the current format, the Geneve header would contain the same number of words to introduce the data blocks. This can be compressed by allocating a group of 2^16 Geneve option class/type values, each of which encodes a subset of the flags, and which has as option data the sequence of data blocks for those flags:

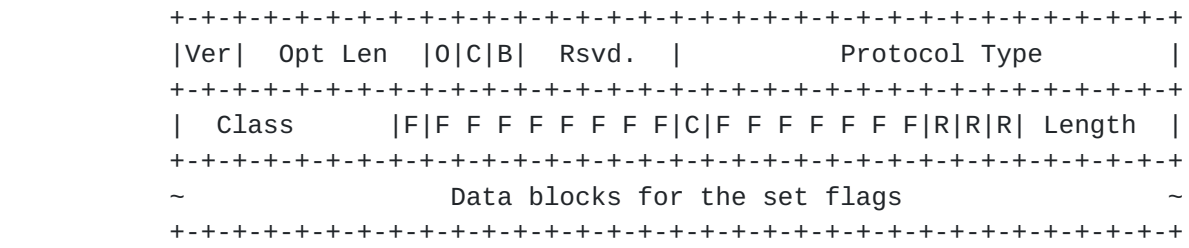



This approach does consume 2^16 of the available 2^23 option class/type values, but that is less than 1% of the available number space.

Using the S (short header) bit, the Geneve header is no longer than the current OOAM header format:

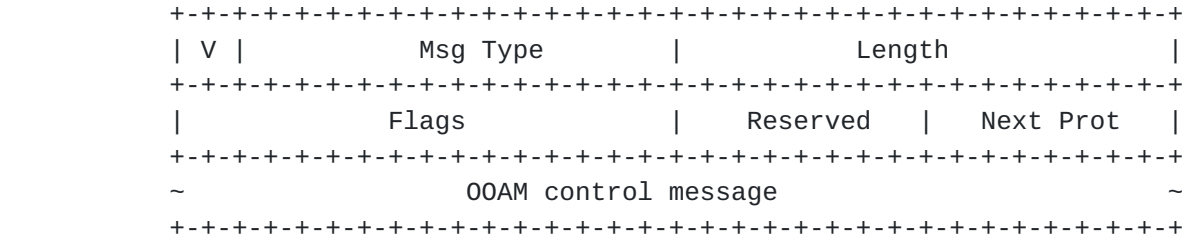



vs.

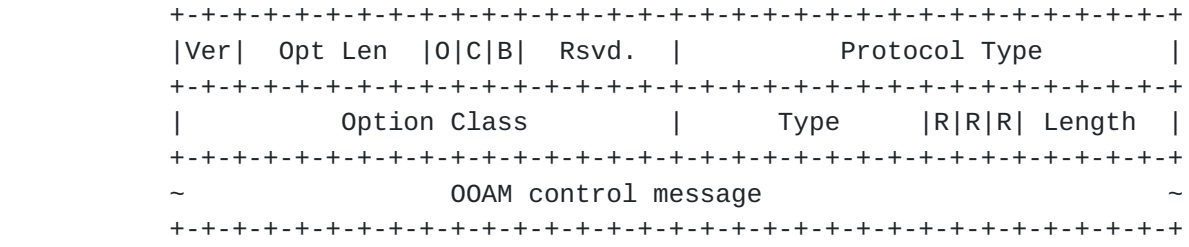



The major limitation of using a Geneve header is that it is limited to a 63 words (252 octets) of options.

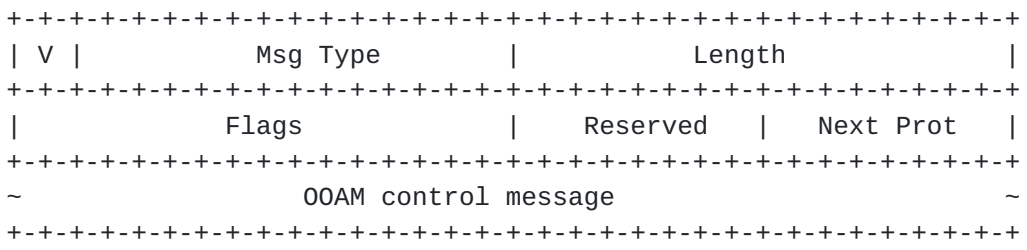
The OOAM control message could be carried within a Geneve option, although that would limit it to 31 words (124 octets). The option specifies the control message type:



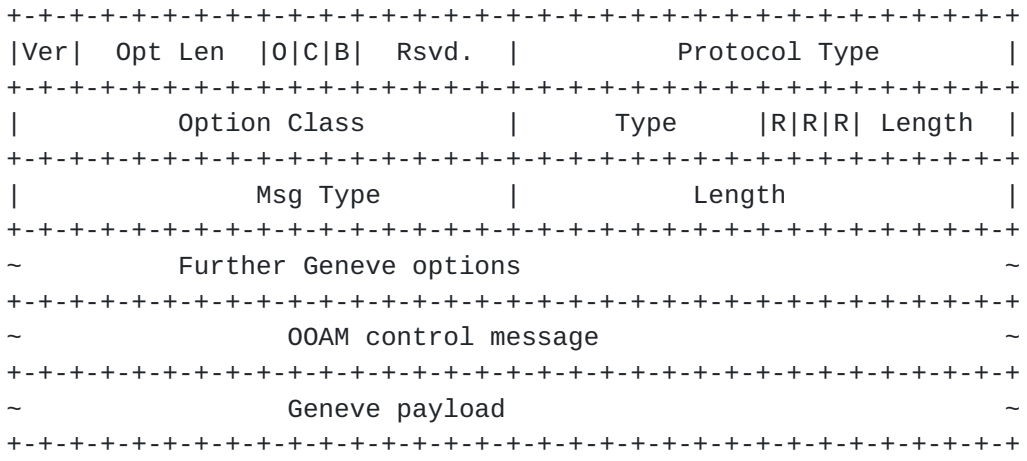

vs.




An alternative approach that allows longer OOAM control messages is to place it after the Geneve header itself and before the Geneve header's payload. The option for the OOAM control message would contain only one word of data, carrying the message type and length:




vs.




In all of these forms, advantage can be taken of Geneve's more easily generalized "next protocol" fieldSection 3 and the abililty to allocate additional Geneve option class/type values for later extensibility.

However, what does not change with this approach is the fundamental work of defining the semantics of the OOAM facilities and control messages.

## 9. Revision History

[Note to RFC Editor: Please remove this entire section upon publication as an RFC.]

### 9.1. TBD

TBD

## 10. References

## 10.1. Normative References

[I-D.fmm-nvo3-pm-alt-mark]
Fioccola, G., Mirsky, G., and T. Mizrahi, "Performance Measurement (PM) with Alternate Marking in Network Virtualization Overlays (NVO3)", draft-fmm-nvo3-pm-alt-mark-01 (work in progress), March 2018.

[I-D.ietf-nvo3-geneve]
Gross, J., Ganga, I., and T. Sridhar, "Geneve: Generic Network Virtualization Encapsulation", draft-ietf-nvo3-geneve-06 (work in progress), March 2018.

[I-D.mizrahi-ippm-compact-alternate-marking]
Mizrahi, T., Arad, C., Fioccola, G., Cociglio, M., Chen, M., Zheng, L., and G. Mirsky, "Compact Alternate Marking Methods for Passive and Hybrid Performance Monitoring", draft-mizrahi-ippm-compact-alternate-marking-01 (work in progress), March 2018.

[protocol-numbers]
Internet Assigned Numbers Authority, "Protocol Numbers", October 2017, <https://www.iana.org/assignments/protocol-numbers/protocol-numbers.xhtml#protocol-numbers-1>.

[RFC7042] Eastlake 3rd, D. and J. Abley, "IANA Considerations and IETF Protocol and Documentation Usage for IEEE 802 Parameters", BCP 141, RFC 7042, DOI 10.17487/RFC7042, October 2013, <https://www.rfc-editor.org/info/rfc7042>.

[RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <https://www.rfc-editor.org/info/rfc8200>.

[snap-protocol-numbers]
Internet Assigned Numbers Authority, "SNAP Protocol Numbers", June 2017, <https://www.iana.org/assignments/ethernet-numbers/ethernet-numbers.xhtml#ethernet-numbers-6>.

## 10.2. Informative References

[extended-msg]
Worley, D., "OUI Extended Ethertypes as next-protocol", IETF NVO3 mailing list msg06235, August 2017, <https://www.ietf.org/mail-archive/web/nvo3/current/msg06235.html>.

[I-D.ooamdt-rtgwg-ooam-header]
Mirsky, G., Kumar, N., Kumar, D., Chen, M., Yizhou, L., and D. Dolson, "OAM Header for use in Overlay Networks", draft-ooamdt-rtgwg-ooam-header-04 (work in progress), March 2018.

[I-D.xiang-nvo3-geneve-packet-spray]
Xiang, H., Yu, Y., Congdon, P., and J. Wang, "Packet Spraying in Geneve Overlay Network", draft-xiang-nvo3-geneve-packet-spray-00 (work in progress), March 2018.

[ieee-ethertype-reg]
IEEE Standards Association, "IEEE-SA - Registration Authority Ethertype", January 2018, <https://standards.ieee.org/develop/regauth/ethertype/index.html>.

[IEEE_802-2014]
IEEE Computer Society, "802 - IEEE Standard for Local and Metropolitan Area Networks: Overview and Architecture", June 2014, <https://standards.ieee.org/findstds/standard/802-2014.html>.

[number-assignments-msg]
Worley, D., "Number assignments", IETF NVO3 mailing list msg06219, July 2017, <https://www.ietf.org/mail-archive/web/nvo3/current/msg06219.html>.

[tcam-msg]
Worley, D., "TCAM compatibility for Geneve", IETF NVO3 mailing list msg06142, Jun 2017, <https://www.ietf.org/mail-archive/web/nvo3/current/msg06142.html>.

## Acknowledgments

Donald Eastlake suggested the use of a SNAP protocol number.

## Author's Address

Dale R. Worley
Ariadne Internet Services
738 Main St.
Waltham, MA 02451
US

Email: worley@ariadne.com