NVMe™ (NVM Express™) Management Messages over MCTP Binding Specification

Supersedes: 1.0.0
Document Class: Normative
Document Status: Published
Document Language: en-US
Copyright notice

Copyright © 2015, 2018 Distributed Management Task Force, Inc. (DMTF). All rights reserved.

DMTF is a not-for-profit association of industry members dedicated to promoting enterprise and systems
management and interoperability. Members and non-members may reproduce DMTF specifications and
documents for uses consistent with this purpose, provided that correct attribution is given. As DMTF
specifications may be revised from time to time, the particular version and release date should always be
noted.

Implementation of certain elements of this standard or proposed standard may be subject to third party
patent rights, including provisional patent rights (herein "patent rights"). DMTF makes no representations
to users of the standard as to the existence of such rights, and is not responsible to recognize, disclose,
or identify any or all such third party patent right, owners or claimants, nor for any incomplete or
inaccurate identification or disclosure of such rights, owners or claimants. DMTF shall have no liability to
any party, in any manner or circumstance, under any legal theory whatsoever, for failure to recognize,
disclose, or identify any such third party patent rights, or for such party’s reliance on the standard or
incorporation thereof in its product, protocols or testing procedures. DMTF shall have no liability to any
party implementing such standard, whether such implementation is foreseeable or not, nor to any patent
owner or claimant, and shall have no liability or responsibility for costs or losses incurred if a standard is
withdrawn or modified after publication, and shall be indemnified and held harmless by any party
implementing the standard from any and all claims of infringement by a patent owner for such
implementations.

PCI-SIG, PCIe, and the PCI HOT PLUG design mark are registered trademarks or service marks of PCI-
SIG.

All other marks and brands are the property of their respective owners.

For information about patents held by third-parties which have notified the DMTF that, in their opinion,
such patent may relate to or impact implementations of DMTF standards, visit
http://www.dmtf.org/about/policies/disclosures.php.

This document’s normative language is English. Translation into other languages is permitted.
CONTENTS

42 Foreword.......................................................................................................................... 4
43 Acknowledgments ............................................................................................................ 4
44 Introduction ....................................................................................................................... 5
45 Document conventions ..................................................................................................... 5
46 Typographical conventions ............................................................................................... 5
47 ABNF usage conventions .................................................................................................. 5
48 1 Scope ............................................................................................................................ 6
49 2 Normative references .................................................................................................... 6
50 3 Terms and definitions .................................................................................................... 7
51 4 Symbols and abbreviated terms ..................................................................................... 8
52 5 Conventions ................................................................................................................... 8
53 5.1 Reserved and unassigned values .................................................................................. 8
54 5.2 Byte ordering .............................................................................................................. 8
55 6 Overview ......................................................................................................................... 8
56 7 Message Type-specific considerations ........................................................................... 9
57 7.1 Message Type number ................................................................................................. 9
58 7.2 Supported transport bindings ...................................................................................... 9
59 7.3 MCTP specification versioning and version compatibility .......................................... 9
60 7.3.1 Base specification and control protocol version compatibility ............................... 9
61 7.3.2 NVMe Management Messages over MCTP–specific version information ........... 9
62 7.3.3 Packet header version compatibility ...................................................................... 10
63 7.4 Timing specifications .................................................................................................... 10
64 7.5 Encapsulation ............................................................................................................. 10
65 7.6 Maximum message size .............................................................................................. 10
66 7.6.1 Additional semantics for MCTP fields ..................................................................... 11
67 7.7 Multiple MCTP transports ......................................................................................... 11
68 ANNEX A (informative) Notation and conventions ......................................................... 12
69 ANNEX B (informative) Change log ................................................................................. 13

71 Figures

72 Figure 1 – Generic MCTP message fields ........................................................................ 10
73
Foreword

The NVMe™ (NVM Express™) Management Messages over MCTP Binding Specification (DSP0235) was prepared by the Platform Management Components Intercommunications (PMCI Working Group) of the DMTF.

DMTF is a not-for-profit association of industry members dedicated to promoting enterprise and systems management and interoperability.

Acknowledgments

The DMTF acknowledges the following individuals for their contributions to this document:

Editor:
- Tom Slaight – Intel Corporation

Contributors:
- Patrick Caporale – Lenovo
- John Carroll – Intel Corporation
- Philip Chidester – Dell
- Yuval Itkin – Mellanox Technologies
- Patrick Kutch – Intel Corporation
- Myron Loewen – Intel Corporation
- Eliel Louzoun – Intel Corporation
- Pat Schoeller – Hewlett Packard Enterprise
- Hemal Shah – Broadcom Limited
- Bob Stevens – Dell
Introduction

The NVMe™ Management Messages over MCTP Binding Specification defines a new MCTP message type used to convey NVMe™ Management Messages over MCTP to storage devices.

Document conventions

Typographical conventions

The following typographical conventions are used in this document:

- Document titles are marked in *italics*.
- Important terms that are used for the first time are marked in *italics*.
- Terms include a link to the term definition in the “Terms and definitions” clause, enabling easy navigation to the term definition.
- ABNF rules are in monospaced font.

ABNF usage conventions

Format definitions in this document are specified using ABNF (see RFC5234), with the following deviations:

- Literal strings are to be interpreted as case-sensitive Unicode characters, as opposed to the definition in RFC5234 that interprets literal strings as case-insensitive US-ASCII characters.
NVMe™ (NVM Express™) Management Messages over MCTP Binding Specification

1 Scope

The NVMe™ (NVM Express™) Management Messages over MCTP Binding Specification defines the bindings between NVMe Management Interface protocol elements and MCTP elements in order to transport NVMe Management Messages for storage devices using MCTP. The specific NVMe management message contents will be documented outside of DMTF directly by the NVMe Management Interface working group.

 Portions of this specification rely on information and definitions from other specifications, which are identified in clause 2. The following references are particularly relevant:

- DMTF DSP0236, Management Component Transport Protocol (MCTP) Base Specification, defines the MCTP transport protocol over which the NVMe over MCTP messages are to be conveyed.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated or versioned references, only the edition cited (including any corrigenda or DMTF update versions) applies. For references without a date or version, the latest published edition of the referenced document (including any corrigenda or DMTF update versions) applies.

3 Terms and definitions

In this document, some terms have a specific meaning beyond the normal English meaning. Those terms are defined in this clause.

The terms "shall" ("required"), "shall not," "should" ("recommended"), "should not" ("not recommended"), "may," "need not" ("not required"), "can" and "cannot" in this document are to be interpreted as described in ISO/IEC Directives, Part 2, Annex H. The terms in parenthesis are alternatives for the preceding term, for use in exceptional cases when the preceding term cannot be used for linguistic reasons. Note that ISO/IEC Directives, Part 2, Annex H specifies additional alternatives. Occurrences of such additional alternatives shall be interpreted in their normal English meaning.

The terms "clause," "subclause," "paragraph," and "annex" in this document are to be interpreted as described in ISO/IEC Directives, Part 2, Clause 5.

The terms "normative" and "informative" in this document are to be interpreted as described in ISO/IEC Directives, Part 2, Clause 3. In this document, clauses, subclauses, or annexes labeled "(informative)" do not contain normative content. Notes and examples are always informative elements.

Refer to DSP0236 for terms and definitions that are used across the MCTP specifications.

Refer to NVMe-MI for terms and definitions that are used in the NVMe Express™ Management Interface specification.

The terms defined in DSP0223, and DSP1001 apply to this document. The following additional terms are used in this document.

3.1 Endpoint

An MCTP endpoint unless otherwise specified.

3.2 NVMe Express™

NVMe Express™ is an optimized register interface, command set, and feature set for PCI Express based storage. The NVMe specifications are maintained by NVMe Express, Inc.

3.3 NVMe™ Management Interface

The NVMe Management Interface allows management entities to communicate with an NVMe non-volatile memory subsystem over one or more external interfaces.
4 Symbols and abbreviated terms

The abbreviations defined in DSP0004, DSP0223, and DSP1001 apply to this document. The following additional abbreviations are used in this document.

4.1 ACPI

Advanced Configuration and Power Interface

4.2 MCTP

Management Component Transport Protocol

4.3 MC

Management Controller

4.4 NVMe™

NVM Express

4.5 NVMe-MI

NVM Express Management Interface

5 Conventions

5.1 Reserved and unassigned values

Unless otherwise specified, any reserved, unspecified, or unassigned values in enumerations or other numeric ranges are reserved for future definition by the DMTF.

Unless otherwise specified, numeric or bit fields that are designated as reserved shall be written as 0 (zero) and ignored when read.

5.2 Byte ordering

Unless otherwise specified, byte ordering of multibyte numeric fields or bit fields is “Big Endian” (that is, the lower byte offset holds the most significant byte, and higher offsets hold lesser significant bytes).

6 Overview

Non-Volatile Memory Express (NVMe) is an optimized register interface, command set, and feature set for PCI Express based storage. The NVMe Management Interface protocol may also be used for other types of non-volatile memory devices.

NVMe Express Management Interface Commands (NVMe Management Interface Commands) are used for the accessing configuration, control, and status functions in NVMe-compatible non-volatile memory devices. NVMe Management Interface Commands are defined by the NVMe Management Interface
This specification only defines how NVMe™ Management Interface Commands are encapsulated in MCTP Messages and transferred between MCTP Endpoints over the specified transports. These are referred to in this document as NVMe Management Messages over MCTP. The definitions and semantics of the NVMe Management Commands themselves are outside the scope of this specification. See the reference to the NVMe Management Interface specification (NVMe-MI).

The MCTP Transport Bindings that are used for NVMe Management Messages over MCTP are defined in other companion specifications such as MCTP SMBus Binding Specification (DSP0237) and MCTP PCIe Binding Specification (DSP0238).

7 Message Type-specific considerations

7.1 Message Type number
The Message Type number for NVMe Management Messages over MCTP messages is defined in the MCTP IDs and Codes Specification (DSP0239).

7.2 Supported transport bindings
As of this writing, use of the specified Message Type is defined for the following transport bindings:

- MCTP SMBus Binding Specification (DSP0237)
- MCTP PCIe Binding Specification (DSP0238)

7.3 MCTP specification versioning and version compatibility
Per DSP0236, the following types of versioning information can be retrieved by using the Get MCTP Version Support command:

- MCTP base specification version information
- MCTP control protocol version information
- NVMe Management Messages over MCTP-specific version information

Additionally, the MCTP packet carries the following versioning information:

- MCTP packet header version information

7.3.1 Base specification and control protocol version compatibility
Unless otherwise specified herein, NVMe Management Messages over MCTP shall meet the requirements of the base specification and control protocol that are identified by the MCTP base specification and control protocol version information, respectively, that are obtained from the endpoint using the Get MCTP Version Support command.

Endpoints that implement NVMe Management Messages over MCTP must also meet the requirements for MCTP Control Messages that are defined by the base specification.

7.3.2 NVMe Management Messages over MCTP–specific version information
The complete semantics of the differences between versions of NVMe Management Messages is left to the NVM Express Management Interface working group, and is outside the scope of this specification. However, the versioning approach should follow the major/minor/update/alpha convention as defined in the Get MCTP Version Support command in DSP0236.
7.3.3 Packet header version compatibility

The Header Version field in MCTP packets identifies the media-specific formatting used for MCTP packets. It can also indicate a level of current and backward compatibility with versions of the base specification, as specified by the header version definition in each medium-specific transport binding specification.

Unless otherwise specified herein, NVMe Management Messages over MCTP shall meet the requirements that are associated with the header version value that is used with the NVMe Management Messages over MCTP, as specified by the corresponding MCTP transport binding specification. This includes meeting requirements for any transport-binding-specific MCTP Control Messages that are called out by the particular transport binding specification.

7.4 Timing specifications

NVMe Management Messages over MCTP are made up of one or more MCTP packets. Each MCTP packet shall comply with the timing, arbitration, and fairness requirements of the transport binding specifications for the media through which it passes. The MCTP endpoint may choose to negotiate longer packet payload lengths than the 64-byte baseline (when the longer packets are not blocked by bridges) provided they do not prevent other devices on the MCTP network from also meeting the transport binding specification requirements, nor prevent them from meeting the message timing specifications for their supported message types.

7.5 Encapsulation

Referring to Figure 1, the NVMe Management Messages over MCTP are carried via the MCTP packet payload of one or more MCTP packets.

7.6 Maximum message size

The MCTP message body (including IC bit, Message Type, Message type-specific header fields, message payload and message integrity check if present) for NVMe Management Messages over MCTP shall be less than or equal to 4224 (4K+128) bytes.

This corresponds to a transfer of 66 MCTP packets using a baseline transmission unit of 64 bytes for the MCTP packet payload.
The maximum message size includes the IC bit and Message Type fields plus any additional Message Type–specific header fields and Message Integrity check fields, as required by NVMe-MI. Refer to NVMe-MI for any additional restrictions on message sizes.

7.6.1 Additional semantics for MCTP fields

NVMe Management Messages over MCTP shall meet the requirements for the MCTP Message Fields per DSP0236. Additional semantics, for example whether the Tag Owner bit or Msg Tag field are to be used to identify particular message streams, or to identify request/response messages, and so on, may be specified by NVMe-MI as long as such semantics do not conflict with DSP0236 or the transport binding specifications.

7.7 Multiple MCTP transports

In order to facilitate identification of devices that are accessible via multiple transports, the endpoints in the device must support the Get Endpoint UUID MCTP command. Otherwise, this specification does not define any additional behaviors related to communicating with NVM Express™ devices that may be accessed through more than one type of MCTP transport on a given MCTP network.
ANNEX A
(informative)
Notation and conventions

A.1 Notations

Examples of notations used in this document are as follows:

- **2:N**  In field descriptions, this will typically be used to represent a range of byte offsets starting from byte two and continuing to and including byte N. The lowest offset is on the left; the highest is on the right.

- **(6)**  Parentheses around a single number can be used in message field descriptions to indicate a byte field that may be present or absent.

- **(3:6)**  Parentheses around a field consisting of a range of bytes indicates the entire range may be present or absent. The lowest offset is on the left; the highest is on the right.

- **PCle**  Underlined, blue text is typically used to indicate a reference to a document or specification called out in the "Normative references" clause or to items hyperlinked within the document.

- **rsvd**  This case-insensitive abbreviation is for "reserved."

- **[4]**  Square brackets around a number are typically used to indicate a bit offset. Bit offsets are given as zero-based values (that is, the least significant bit [LSb] offset = 0).

- **[7:5]**  This notation indicates a range of bit offsets. The most significant bit is on the left; the least significant bit is on the right.

- **1b**  The lowercase "b" following a number consisting of 0s and 1s is used to indicate the number is being given in binary format.

- **0x12A**  A leading "0x" is used to indicate a number given in hexadecimal format.
ANNEX B
(informative)

Change log

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0.0</td>
<td>2015-03-06</td>
<td></td>
</tr>
<tr>
<td>1.0.1</td>
<td>2018-08-03</td>
<td>Corrected the maximum message size text to state message payload not packet payload. Updated references to NVMe-MI.</td>
</tr>
</tbody>
</table>