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

Document Type: Specification

Document Status: Work in Progress - Not a DMTF Standard

Document Language: en-US

13 | Copyright notice

14 | Copyright © 2014 Distributed Management Task Force, Inc. (DMTF). All rights reserved.

39

CONTENTS

Figures

74   # Foreword

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

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

80   ## Acknowledgments

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

82   **Editor:**

83   • Tom Slaight – Intel Corporation

84   **Contributors:**

85   • Patrick Caporale – IBM

86   • John Carroll – Intel Corporation

87   • Philip Chidester – Dell

88   • Yuval Itkin – Mellanox

89   • Patrick Kutch – Intel Corporation

90   • Myron Loewen – Intel Corporation

91   • Eliel Louzoun – Intel Corporation

92   • Pat Schoeller – Hewlett-Packard Company

93   • Hemal Shah – Broadcom Corporation

94   • Bob Stevens – Dell

95

96   Introduction

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

Document conventions

Typographical conventions

101 The following typographical conventions are used in this document:

102   • Document titles are marked in *italics*.

103   • Important terms that are used for the first time are marked in *italics*.

104   • Terms include a link to the term definition in the "Terms and definitions" clause, enabling easy
105   navigation to the term definition.

106   • ABNF rules are in `monospaced font`.

ABNF usage conventions

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

110   • Literal strings are to be interpreted as case-sensitive Unicode characters, as opposed to the
111   definition in RFC5234 that interprets literal strings as case-insensitive US-ASCII characters.
112

113

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.

Unless otherwise specified, for DMTF documents this means any document version that has minor or update version numbers that are later than those for the referenced document. The major version numbers must match the major version number given for the referenced document.

Refer to the Bibliography for additional, non-normative, reference information.

DMTF DSP0223, *Generic Operations 1.0*,
http://www.dmtf.org/standards/published_documents/DSP0223_1.0.pdf

DMTF DSP0236, *Management Component Transport Protocol (MCTP) Base Specification 1.2*
http://www.dmtf.org/standards/published_documents/DSP0236_1.2.pdf

DMTF DSP0237, *Management Component Transport Protocol (MCTP) SMBus/I2C Transport Binding Specification 1.0*
http://www.dmtf.org/standards/published_documents/DSP0237_1.0.pdf

DMTF DSP0238, *Management Component Transport Protocol (MCTP) PCIe VDM Transport Binding Specification 1.0*
 http://www.dmtf.org/standards/published_documents/DSP0238_1.0.pdf

DMTF DSP0239, *Management Component Transport Protocol (MCTP) IDs and Codes 1.2*
http://www.dmtf.org/standards/published_documents/DSP0239_1.2.pdf

IETF, RFC4122, *A Universally Unique Identifier (UUID) URN Namespace*, July 2005
http://www.ietf.org/rfc/rfc4122.txt

IETF RFC5234, *ABNF: Augmented BNF for Syntax Specifications, January 2008*,
http://tools.ietf.org/html/rfc5234

152 ISO/IEC Directives, Part 2, *Rules for the structure and drafting of International Standards,*
153 http://isotc.iso.org/livelink/livelink.exe?func=ll&objId=4230456&objAction=browse&sort=subtype

3 Terms and definitions

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

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

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

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

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

169 Refer to NVMeMI (see Bibliography) for terms and definitions that are used in the NVMe Express™
170 Management Interface specification.

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

173 **3.1**
174 **Endpoint**

175 An MCTP endpoint unless otherwise specified.

176 **3.2**
177 **NVM Express™**

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

180 **3.3**
181 **NVMe™ Management Interface**

182 The NVMe Management Interface allows management entities to communicate with an NVMe non-
183 volatile memory subsystem over one or more external interfaces.

4 Symbols and abbreviated terms

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

187 **4.1**
188 **ACPI**

189 Advanced Configuration and Power Interface

190 **4.2**
191 **MCTP**
192 Management Component Transport Protocol

193 **4.3**
194 **MC**
195 Management Controller

196 **4.4**
197 **NVMe™**
198 NVM Express

199 5 Conventions

200 5.1 Reserved and unassigned values

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

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

205 5.2 Byte ordering

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

208 6 Overview

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

212 NVM Express Management Interface Commands (NVMe Management Interface Commands) are used for
213 the accessing configuration, control, and status functions in NVMe-compatible non-volatile memory
214 devices. NVMe Management Interface Commands are defined by the NVMe Management Interface
215 specification and the members of NVM Express, Inc. Refer to www.nvmexpress.org and NVMeMI in the
216 Bibliography for more information.

217 This specification only defines how NVMe™ Management Interface Commands are encapsulated in
218 MCTP Messages and transferred between MCTP Endpoints over the specified transports. These are
219 referred to in this document as NVMe Management Messages over MCTP. The definitions and semantics
220 of the NVMe Management Commands themselves are outside the scope of this specification. See the
221 Bibliography for reference to the NVMe Management Interface specification (NVMeMI).

222 The MCTP Transport Bindings that are used for NVMe Management Messages over MCTP are defined in
223 other companion specifications such as *MCTP SMBus Binding Specification* (DSP0237) and *MCTP PCIe*
224 *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 that can be retrieved 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.
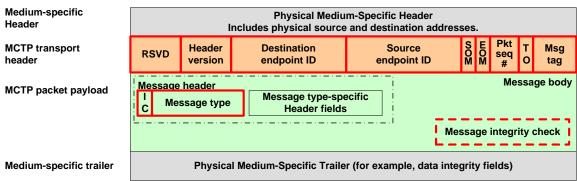
263   ### 7.4   Timing specfications

264   NVMe Management Messages over MCTP shall meet the timing specifications for MCTP Control
265   Messages for the transport binding specification for the particular transport medium.

266   ### 7.5   Encapsulation

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

269

270   **Figure 1 – Generic MCTP message fields**

271   ### 7.6   Maximum message size

272   The MCTP packet payload for NVMe Management Messages over MCTP shall be less than or equal to
273   4224 (4K+128) bytes.

274   This corresponds to a transfer of 66 MCTP packets using a baseline transmission unit of 64 bytes for the
275   MCTP packet payload.

276   The maximum message size includes the IC bit and Message Type fields plus any additional Message
277   Type–specific header fields and Message Integrity check fields, as required by NVMeMI. Refer to
278   NVMeMI for any additional restrictions on message sizes.

279   #### 7.6.1   Additional semantics for MCTP fields

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

285   ### 7.7   Multiple MCTP transports

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

290   **ANNEX A**
291   (informative)
292   **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.

- PCIe   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.

313

314 | # ANNEX B
315 | ## (informative)
316 |
317 | # Change log

Version	Date	Author	Description
1.0.0	2014-08-14		

318
319

320 # Bibliography

321   NVMeMI   NVM Express Inc., NVM Express™ Management Interface Specification 1.0, December
322   2014
323   www.nvmexpress.org

324