Storage Maintenance (StorM) Working Group
Internet Draft
Intended status: Standards Track
Expires: February 2012

Frederick Knight
NetApp
M. Chadalapaka
Microsoft

August 2011

Internet Small Computer Systems Interface (iSCSI) SCSI Architecture Features Update

draft-ietf-storm-iscsi-sam-04.txt

Abstract

Internet Small Computer Systems Interface (iSCSI) is a SCSI transport protocol that maps the SCSI family of protocols onto TCP/IP. The iSCSI protocol as specified in [draft-ietf-storm-iscsi-cons-xx] (and as previously specified by the combination of RFC 3720 and RFC 5048) is based on the SAM-2 (SCSI Architecture Model – 2) version of the SCSI family of protocols. This document defines enhancements to the iSCSI protocol to support certain additional features of the SCSI protocol that were defined in SAM-3, SAM-4, and SAM-5.

This document is a companion document to [draft-ietf-storm-iscsi-cons-xx].

RFC EDITORS NOTE: The above references to [draft-ietf-storm-iscsi-cons-xx] should reference the RFC number assigned to that document, and this note should be removed.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/1id-abstracts.html.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire February, 2012.

Table of Contents

1. Introduction

The original [RFC3720] was built based on the [SAM2] model for SCSI. Several new features and capabilities have been added to the SCSI Architecture Model in the intervening years (at the time of publication of this document, SAM-5 was the current version of the SCSI Architecture Model). This document is not a complete revision of [RFC3720]. Instead, this document is intended as a companion document to [draft-ietf-storm-iscsi-cons-xx]; this document may also be used as a companion document to the combination of [RFC3720] and [RFC5048], although both of those RFCs have been obsolete by [draft-ietf-storm-iscsi-cons-xx].

RFC EDITORS NOTE: The above references to [draft-ietf-storm-iscsi-cons-xx] should reference the RFC number assigned to that document, and this note should be removed.

---

2. Definitions, Acronyms, and Document Summary

2.1 Definitions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2.2 Acronyms

ISID	Initiator Session Identifier
LU	Logical Unit
PDU	Protocol Data Unit
SAM-4	SCSI Architecture Model – 4 (see [SAM4])
SAM-5	SCSI Architecture Model – 5 (see [SAM5])
TMF	Task Management Function

2.3 New Semantics

This document specifies new iSCSI semantics. This section summarizes the contents of the document.

Section 3: The mapping of iSCSI objects to SAM-5 objects
The iSCSI node may contain both initiator and target capabilities.

Section 4: The protocol used to negotiate the use of the new capabilities described in this document.

Section 5: New Command operations
The PRI field for SCSI command priority has been added to the SCSI command PDU (see 5.1.1).

The Status Qualifier field has been added to the SCSI response PDU (see 5.2.1).
Sense data may be returned (via autosense) for any SCSI status, not just CHECK CONDITION (see 5.2.2).

Section 6: New Task Management Functions
Four new task management functions (QUERY TASK, QUERY TASK SET, I_T NEXUS RESET, and QUERY ASYNCHRONOUS EVENT have been added (see 6.2).
A new "function succeeded" response has been added (see 6.3.1).

Section 7: New Negotiation key
A new negotiation key has been added to enable the use of the new features in section 5 and section 6.

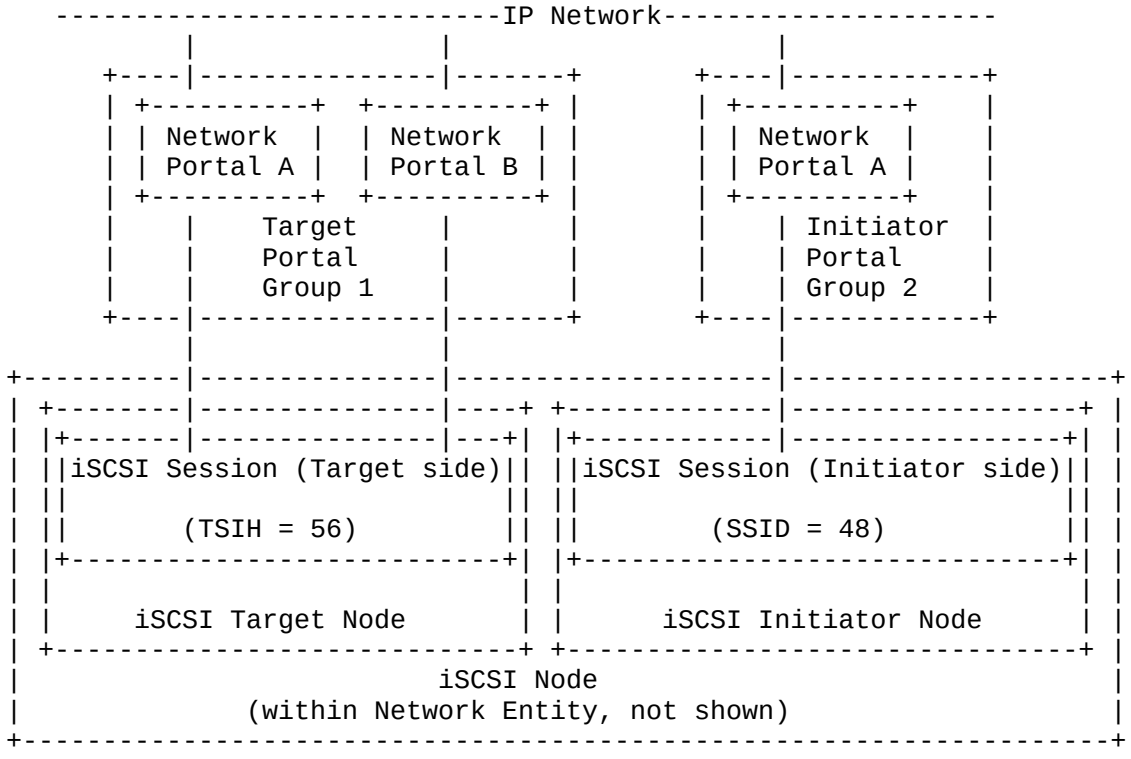
3. Terminology Mapping

The iSCSI model (defined in [RFC3720]) uses different terminology than the SCSI Architecture Model. In some cases, iSCSI uses multiple terms to describe what in the SCSI Architecture Model is described with a single term. The iSCSI terms and SAM-5 terms are not necessarily equivalent, but rather, the iSCSI terms represent examples of the objects or classes described in SAM-5 as follows:

RFCxxx Terminology	SAM-5 Terminology
Network Entity	none
iSCSI Node	SCSI Device
iSCSI Name	SCSI Device Name
iSCSI Node Name	SCSI Device Name
iSCSI Initiator Node	SCSI Initiator Device
iSCSI Initiator Name	SCSI Device Name
iSCSI Initiator Port Name iSCSI Node Name + ',i,' + ISID	SCSI Initiator Port Name
iSCSI Target Node	SCSI Target Device
iSCSI Target Name	SCSI Device Name
iSCSI Target Port Name iSCSI Node Name + ',t,' + Target Portal Group Tag	SCSI Target Port Name
iSCSI Target Portal Group	SCSI Target Port
iSCSI Initiator Node + active ISID	SCSI Initiator Port
iSCSI Initiator Name + ',i,' + ISID, iSCSI Target Name + ',t,' + Portal Group Tag	I_T Nexus
Target Portal Group Tag	Relative Port ID

RFC EDITORS NOTE: The above reference (in row 1) to [RFCxxx] should reference this RFC, and this note should be removed.

The following diagram shows an example of a combination target device and initiator device. Such a configuration may exist in a target device that implements a SCSI Copy Manager. This example shows how a session that shares Network Portals within a Portal Group may be established (see Target Portal Group 1). In addition, this example shows the Initiator using a different Portal Group than the Target Portal Group, but the Initiator Portal group sharing Network Portal A with the Target Portal Group.



4. New Feature Use

4.1 Negotiation of New Feature Use

The iSCSIProtocolLevel operational text key (see 7.1.1) containing a value of "2" or higher MUST be negotiated to enable the use of features described in this RFC.

Note that an operational value of "2" or higher for this key on an iSCSI session does not influence the SCSI level features in any way on that I_T nexus. An operational value of "2" or higher for this key permits the iSCSI-related features defined in this document to be used on all connections on this iSCSI session. SCSI level hand-shakes (e.g. commands, mode pages) eventually determine the existence or lack of various SAM-5 features available for the I_T nexus between the two SCSI end points). To summarize, negotiation of this key to "2" or higher is a necessary but not a sufficient condition of SAM-5 compliant feature usage at the SCSI protocol level.

For example, an iSCSI implementation may negotiate this new key to "2" but respond to the new task management functions (see 6.2) with a "Task management function not supported" (which indicates a SCSI error that prevents the function from being performed). In contrast, if the key is negotiated to "2", an iSCSI implementation MUST NOT reject a task management function request PDU that requests one of the new task management functions (such a reject would report an iSCSI protocol error).

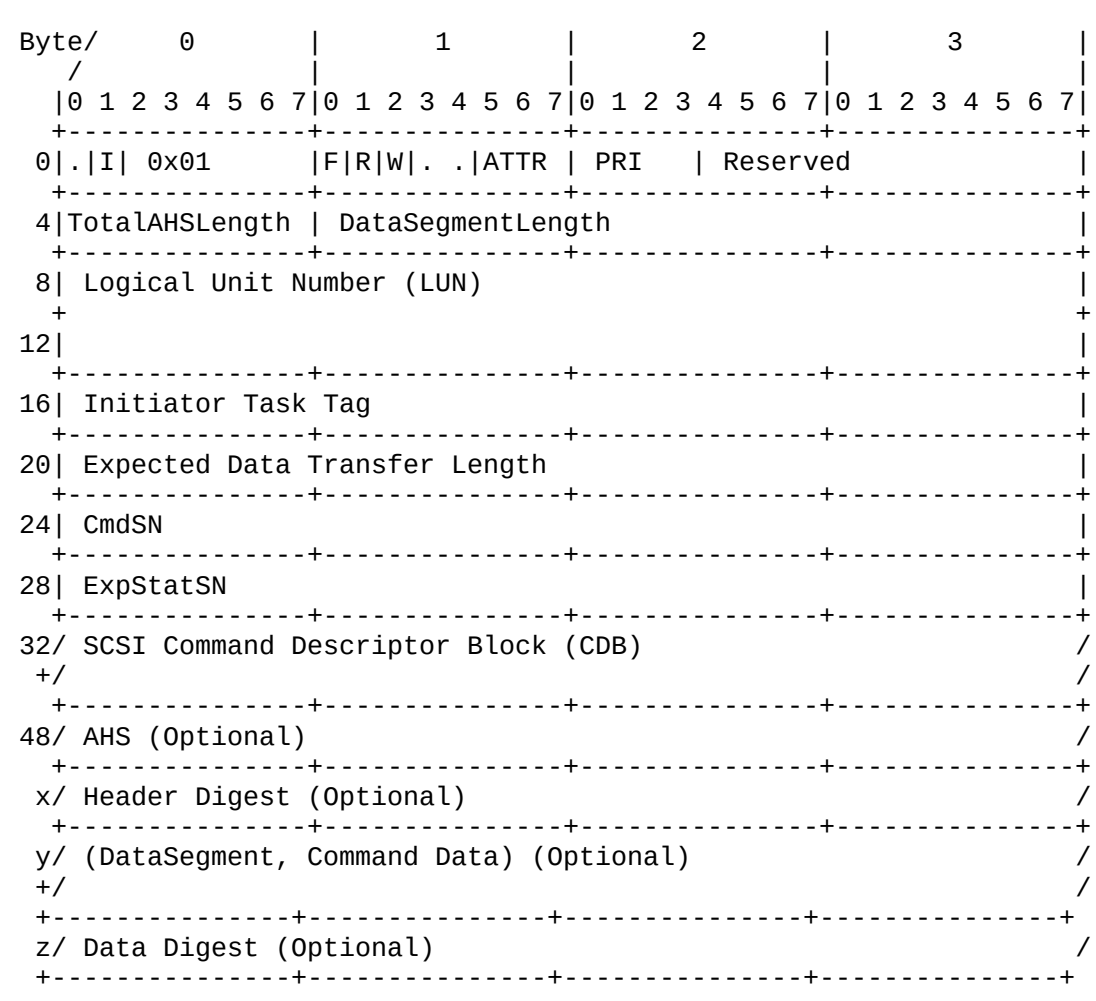
4.2 Impact on standard INQUIRY data

The negotiated value of the iSCSIProtocolLevel key is an increment from the base iSCSI version descriptor value (0960h)(see [SPC4]). If the SCSI device server returns an iSCSI version descriptor in the standard INQUIRY data, then the value returned in that iSCSI version descriptor MUST be set to the sum of the base value (0960h) plus the negotiated value of the iSCSIProtocolLevel key (for example, if the negotiated iSCSIProtocolLevel=2, then if an iSCSI version descriptor is returned in the standard INQUIRY data it is set to 0962h).

5. SCSI Commands

5.1 SCSI Command Additions

The format of the SCSI Command PDU is:



5.1.1 Command Priority (byte 2)

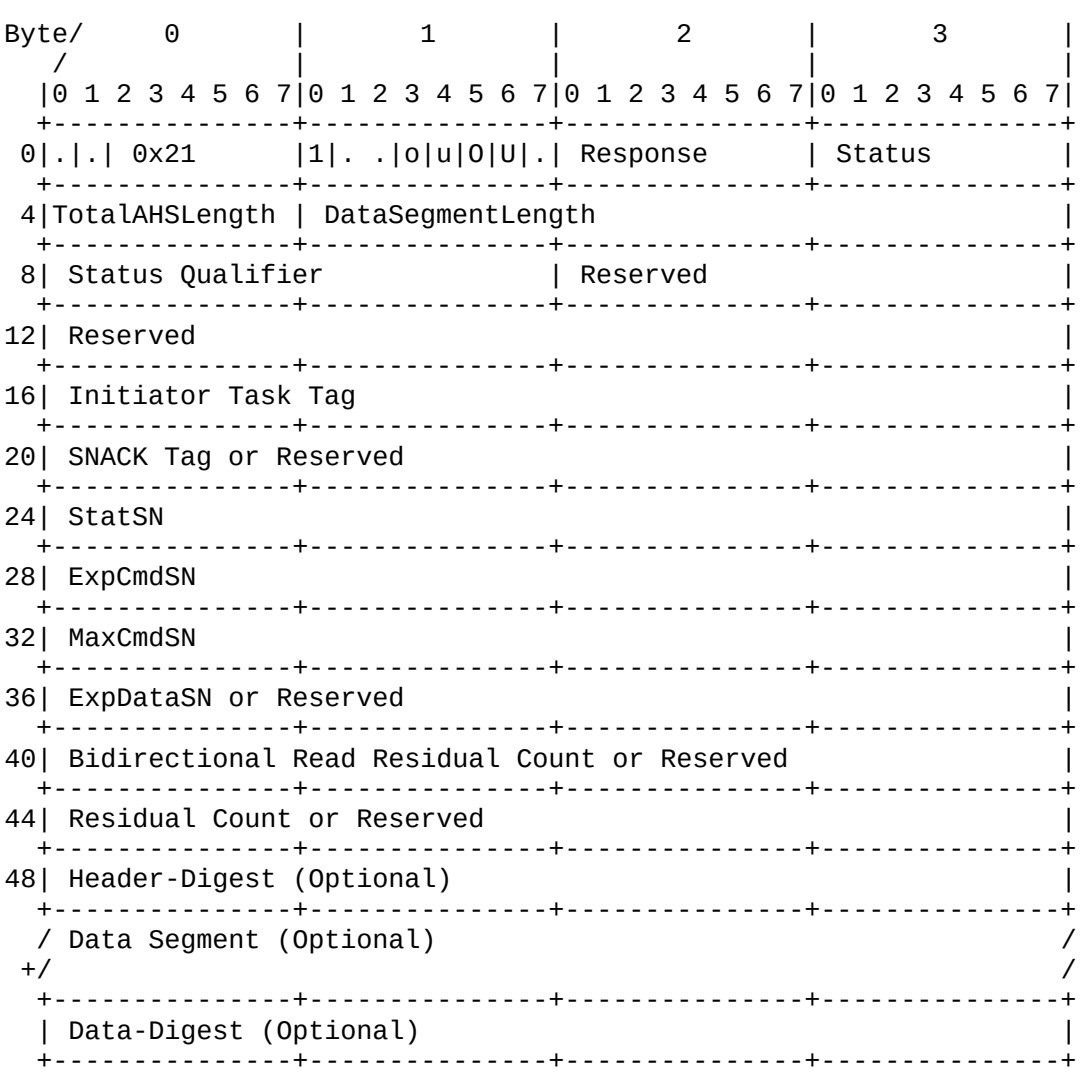
The Command Priority (PRI) specifies the relative scheduling importance of this task in relation to other SIMPLE tasks already in the task set (see [SAM4]).

Section 10, iSCSI PDU Formats of [RFC3720], requires that senders set this field to zero. A sender MUST NOT set this field to a value other than zero unless the iSCSIProtocolLevel text key defined in section 7.1.1 has been negotiated on the session with a value of "2" or higher.

This field MUST be ignored by iSCSI targets unless the iSCSIProtocolLevel text key with a value of "2" or higher as defined in section 7.1.1 was negotiated on the session.

5.2 SCSI Response Additions

The format of the SCSI Response PDU is:



5.2.1 Status Qualifier

The Status Qualifier provides additional status information (see [SAM4]).

As defined in Section 10, iSCSI PDU Formats of [RFC3720], compliant senders already set this field to zero. Compliant senders MUST NOT set this field to a value other than zero unless the iSCSIProtocolLevel text key with a value of "2" or higher as defined in section 7.1.1 was negotiated on the session.

This field MUST be ignored by receivers unless the iSCSIProtocolLevel text key with a value of "2" or higher as defined in section 7.1.1 was negotiated on the session.

5.2.2 Data Segment - Sense and Response Data Segment

Section 10.4.7 of [RFC3720] specifies that iSCSI targets MUST support and enable autosense. If Status is CHECK CONDITION (0x02), then the Data Segment MUST contain sense data for the failed command. While [RFC3720] does not make any statements about the state of the Data Segment when the Status is not CHECK CONDITION (0x02)(i.e., the Data Segment is not prohibited from containing sense data when the Status is not CHECK CONDITION), negotiation of the iSCSIProtocolLevel text key with a value of "2" or higher as defined in section 7.1.1 explicitly indicates that the Data Segment MAY contain sense data at any time, no matter what value is set in the Status field.

6. Task Management Functions

6.1 Existing Task Management Functions

Section 10.5 of [RFC3720] defines the semantics used to request SCSI Task Management Functions be performed. The following task management functions are defined:

1 - ABORT TASK
2 - ABORT TASK SET
3 - CLEAR ACA
4 - CLEAR TASK SET
5 - LOGICAL UNIT RESET
6 - TARGET WARM RESET
7 - TARGET COLD RESET
8 - TASK REASSIGN

6.2 Task Management Function Additions

Additional task Management function codes are listed below. For a more detailed description of SCSI task management, see [SAM5].

9 - QUERY TASK - determines if the task identified by the Referenced Task Tag field is present in the task set.

10 - QUERY TASK SET - determine if any task is present in the task set for the I_T_L Nexus on which the task management function was received.

11 - I_T NEXUS RESET - perform an I_T nexus loss function for the I_T nexus of each logical unit accessible through the I_T Nexus on which the task management function was received.

12 - QUERY ASYNCHRONOUS EVENT - determine if there is a unit attention condition or a deferred error pending for the I_T_L nexus on which the task management function was received.

These task management function requests MUST NOT be sent unless the iSCSIProtocolLevel text key with a value of "2" or higher as defined in section 7.1.1 was negotiated on the session.

Any compliant initiator that sends any of the new task management functions defined in this section MUST also support all new task management function responses (as specified in section 6.3.1).

For all of the task management functions detailed in this section, the Task Management function response MUST be returned as detailed in section 6.3 Task Management Function Response.

The iSCSI target MUST ensure that no responses for the tasks covered by a task management function are sent to the iSCSI initiator after the Task Management response except for a task covered by a TASK REASSIGN, QUERY TASK, or QUERY TASK SET.

If a QUERY TASK is issued for a task created by an immediate command then RefCmdSN MUST be that of the Task Management request itself (i.e., CmdSN and RefCmdSN are equal); otherwise RefCmdSN MUST be set to the CmdSN of the task to be queried (lower than CmdSN).

If the connection is still active (it is not undergoing an implicit or explicit logout), QUERY TASK MUST be issued on the same connection to which the task to be queried is allegiant at the time the Task Management Request is issued. If the connection is implicitly or explicitly logged out (i.e., no other request will be issued on the failing connection and no other response will be received on the failing connection), then a QUERY TASK function request may be issued on another connection. This Task Management request will then establish a new allegiance for the command being queried.

At the target a QUERY TASK function MUST NOT be executed on a Task Management request; such a request MUST result in Task Management response of "Function rejected".

For the I_T NEXUS RESET function, the target device MUST respond to the function as defined in [SAM4]. Each logical unit accessible via the receiving I_T NEXUS MUST behave as dictated by the I_T nexus loss function in [SAM4] for the I_T nexus on which the task management function was received. The target device MUST drop all connections in the session over which this function is received. Independent of the DefaultTime2Wait and DefaultTime2Retain value applicable to the session over which this function is received, the target device MUST consider each participating connection in the session to have immediately timed out, leading to FREE state. The resulting timeouts cause the session timeout event defined in [RFC3720], which in turn triggers the I_T nexus loss notification to the SCSI layer as described in [RFC3720].

6.2.1 LUN field

This field is required for functions that address a specific LU (i.e., ABORT TASK, CLEAR TASK SET, ABORT TASK SET, CLEAR ACA, LOGICAL UNIT RESET, QUERY TASK, QUERY TASK SET, and QUERY ASYNCHRONOUS EVENT) and is reserved in all others.

6.2.2 Referenced Task Tag

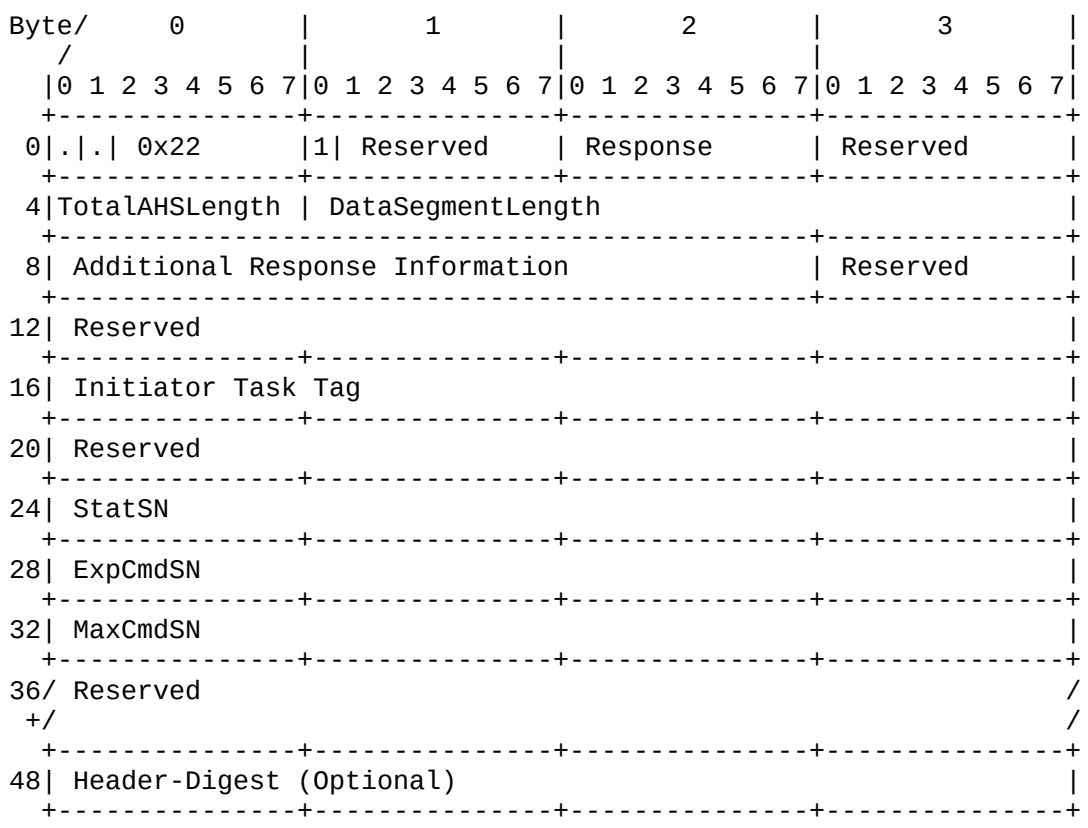
The Initiator Task Tag of the task to be aborted for the ABORT TASK function, reassigned for the TASK REASSIGN function, or queried for the QUERY TASK function. For all other functions this field MUST be set to the reserved value 0xffffffff.

6.2.3 RefCmdSN

If a QUERY TASK is issued for a task created by an immediate command then RefCmdSN MUST be that of the Task Management request itself (i.e., CmdSN and RefCmdSN are equal).

For a QUERY TASK of a task created by non-immediate command RefCmdSN MUST be set to the CmdSN of the task identified by the Referenced Task Tag field. Targets must use this field as described in section 10.6.1 of [RFC3720] when the task identified by the Referenced Task Tag field is not in the task set.

6.3 Task Management Function Responses



Section 10.6 of [RFC3720] defines the semantics used for responses to SCSI Task Management Functions. The following responses are defined in [RFC3720]:

0 - Function Complete.
1 - Task does not exist.
2 - LUN does not exist.
3 - Task still allegiant.
4 - Task allegiance reassignment not supported.
5 - Task management function not supported.
6 - Function authorization failed.
255 - Function rejected.

Responses to new task management functions (see 6.3.1) are listed below. In addition, a new task Management response is listed below. For a more detailed description of SCSI task management responses, see [SAM5].

For the functions QUERY TASK, QUERY TASK SET, I_T NEXUS RESET, and QUERY ASYNCHRONOUS EVENT, the target performs the requested Task Management function and sends a Task Management response back to the initiator.

6.3.1 Task Management Function Response Additions

The new response is listed below:

7 - Function succeeded.

In symbolic terms Response value 7 maps to the SCSI service response of FUNCTION SUCCEEDED in [SAM4].

The task management function response of “Function succeeded” MUST be supported by an initiator that sends any of the new task management functions (see 6.2).

For the QUERY TASK function, if the specified task is in the task set, then the logical unit returns a Response value of Function succeeded and additional response information is returned as specified in [SAM5]. If the specified task is not in the task set, then the logical unit returns a Response value of Function complete.

For the QUERY TASK SET function, if there is any command present in the task set from the specified I_T_L nexus, then the logical unit returns a Response value of Function succeeded. If there are no commands present in the task set from the specified I_T_L nexus, then the logical unit returns a Response value of Function complete.

For the I_T NEXUS RESET function, after completion of the events described in section 6.2 for this function, the logical unit returns a Response value of Function complete. However, because the target drops all connections, the Service Response (defined by [SAM4]) for this SCSI task management function may not be reliably delivered to the issuing initiator port.

For the QUERY ASYNCHRONOUS EVENT, if there is a unit attention condition or deferred error pending for the specified I_T_L nexus, then the logical unit returns a Response value of Function succeeded and additional response information is returned as specified in [SAM5]. If there is no unit attention or deferred error pending for the specified I_T_L nexus then the logical unit returns a Response value of Function complete.

6.4 Task Management Requests Affecting Multiple Tasks

Section 4.1 of [RFC5048] defines the notion of "affected tasks" in multi-task abort scenarios. This section adds to the list include in that section by defining the tasks affected by the I_T NEXUS RESET function.

I_T NEXUS RESET: All outstanding tasks received on the I_T nexus on which the function request was received for all logical units accessible to the I_T nexus.

Section 4.1.2 of [RFC5048] and section 4.1.3 of [RFC5048] identify semantics for task management functions that involve multi-task abort operations. If an iSCSI implementation supports the I_T NEXUS RESET function, it MUST also support the protocol behavior as defined in those sections and follow the sequence of actions as described in those sections when processing the I_T NEXUS RESET function.

7. Login/Text Operational Text Keys

7.1 New Operational Text Keys

7.1.1 iSCSIProtocolLevel

Use: LO, IO
Irrelevant when: SessionType = Discovery
Senders: Initiator and Target
Scope: SW

iSCSIProtocolLevel=<numerical-value-from-0-to-31>

Default is 1.
Result function is Minimum.

This key is used to negotiate the use of iSCSI features that require different levels of protocol support (e.g., PDU formats, end node semantics) for proper operation.

Negotiation of the iSCSIProtocolLevel key to a value corresponding to an RFC indicates that both negotiating parties are compliant to the RFC in question, and agree to support the corresponding PDU formats and semantics on that iSCSI session.

An operational value of iSCSI ProtocolLevel = "x" on an iSCSI session requires that the iSCSI protocol semantics on that iSCSI session be a logical superset of the capabilities in all RFCs that have claimed values of an iSCSIProtocolLevel less than "x".

An iSCSIProtocolLevel key negotiated to “0” indicates that the implementation does not claim a specific iSCSI protocol level.

An iSCSIProtocolLevel key negotiated to "2" or higher is required to enable use of features defined in this RFC.

If the negotiation answer is ignored by the acceptor, or the answer from the remote iSCSI end point is key=NotUnderstood, then the features defined in this RFC, and the features defined in any RFC requiring a key value greater than "2" MUST NOT be used.

8. Security Considerations

At the time of writing this document does not introduce any new security considerations other than those described in [RFC3720]. Consequently, all the iSCSI-related security text in [RFC3723] is also directly applicable to this document.

9. IANA Considerations

This document modifies or creates a number of iSCSI-related registries. The following iSCSI-related registries are modified:

1. iSCSI Task Management Functions Codes

   Name of the existing registry: "iSCSI TMF Codes"

   Additional entries:

   9, QUERY TASK, [RFCxxx]

   10, QUERY TASK SET, [RFCxxx]

   11, I_T NEXUS RESET, [RFCxxx]

   12, QUERY ASYNCHRONOUS EVENT, [RFCxxx]

   RFC EDITORS NOTE: The above reference to [RFCxxx] should reference this RFC, and this note should be removed.
   ---

2. iSCSI Login/Text Keys

   Name of the existing registry: "iSCSI Text Keys"

   Fields to record in the registry: Assigned value and its associated RFC reference:

   iSCSIProtocolLevel, [RFCxxx]

   RFC EDITORS NOTE: The above references to [RFCxxx] should reference this RFC, and this note should be removed.

This document creates the following iSCSI-related registries for IANA to manage.

3. iSCSI Protocol Level

   Name of new registry: “iSCSI Protocol Level”

Namespace details: Numerical values from 0 to 31

Information that must be provided to assign a new value: An IESG-approved standards track specification defining the semantics and interoperability requirements of the proposed new value and the fields to be recorded in the registry.

Assignment policy:

The assignments of these values must be coordinated with the ANSI/INCITS T10 committee; therefore review by an expert that maintains an association with that committee is required prior to IESG approval of the associated specification. After creation of the registry, values are to be assigned sequentially (for example, any value greater than 4 will not be assigned until after the value 4 has been assigned).

1 and 3-31: range reserved by IANA for assignment in this registry.

Fields to record in the registry: Assigned value, and its associated RFC reference.

0, [RFCxxx]

2, [RFCxxx]

RFC EDITORS NOTE: The above references to [RFCxxx] should reference this RFC, and this note should be removed.

Allocation Policy:

Expert review ([IANA]) and Standards Action ([IANA])

4. iSCSI Task Management Response Codes

Name of new registry: "iSCSI TMF Response Codes"

Namespace details: Numerical values that can fit in 8 bits.

Information that must be provided to assign a new value: An IESG-approved specification defining the semantics and interoperability requirements of the proposed new value and the fields to be recorded in the registry.

Assignment policy:

If the requested value is not already assigned, it may be assigned to the requester.

8-254: Range reserved by iANA for assignment in this registry.

Fields to record in the registry: Assigned value, Operation Name, and its associated RFC reference.

0x0, Function complete, [RFC3720]

0x1, Task does not exist, [RFC3720]

0x2, LUN does not exist, [RFC3720]

0x3, Task still allegiant, [RFC3720]

0x4, Task allegiance reassignment not supported, [RFC3720]

0x5, Task management function not supported, [RFC3720]

0x6, Function authorization failed, [RFC3720]

0x7, Function succeeded, [RFCxxx]

255, Function rejected, [RFC3720]

RFC EDITORS NOTE: The above reference to [RFCxxx] should reference this RFC, and this note should be removed.

Allocation Policy:

Standards Action ([IANA])

10. References

10.1 Normative References

[RFC2119] Bradner, S. "Key Words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3720] Satran, J., Meth, K., Sapuntzakis, C., Chadalapaka, M., and E. Zeidner, "Internet Small Computer Systems Interface (iSCSI)", RFC 3720, April 2004.

[RFC3723] Aboba, B., Tseng, J., Walker, J., Rangan, V., and Travostino, F., "Securing Block Storage Protocols over IP", RFC 3723, April 2004.

[RFC5048] Chadalapaka, M., "Internet Small Computer System Interface (iSCSI) Corrections and Clarifications", RFC 5048, October 2007.

[draft-ietf-storm-iscsi-cons-xx] Chadalapaka, M., Satran, J., Kalman, M., “iSCSI Protocol (consolidated)”, RFC xxx, Date 2011.

RFC EDITORS NOTE: The above references to [draft-ietf-storm-iscsi-cons-xx] and [RFCxxx] should reference the RFC number assigned to that draft, and this note should be removed.

[IANA] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

[SAM2] T10/1157D, SCSI Architecture Model - 2 (SAM-2).

[SAM4] ISO/IEC 14776-414, SCSI Architecture Model - 4 (SAM-4).

[SAM5] T10/2104D rev r04, SCSI Architecture Model - 5 (SAM-5), Committee Draft.

[SPC4] T10/1731D rev r23, SCSI Primary Commands – 4 (SPC-4), Committee Draft.

10.2 Additional Reference Sources

For more information on the SCSI Architecture Model, contact the T10 group at http://www.t10.org.

11. Acknowledgements

The Storage Maintenance (STORM) Working Group in the Transport Area of the IETF has been responsible for defining these additions to the iSCSI protocol (apart from other relevant IP Storage protocols). The editor acknowledges the contributions of the entire working group.

The following individuals directly contributed to identifying [RFCxxx] issues and/or suggesting resolutions to the issues clarified in this document: David Black, Rob Elliott. This document benefited from all of these contributions.

RFC EDITORS NOTE: The above reference to [RFCxxx] should reference this RFC, and this note should be removed.

Author's Addresses

Frederick Knight
7301 Kit Creek Road
P.O. Box 13917
Research Triangle Park, NC 27709, USA
Phone: +1-919-476-5362
Email: knight@netapp.com

Mallikarjun Chadalapaka
Microsoft
One Microsoft Way
Redmond, WA 98052 USA
Email: cbm@chadalapaka.com