To: T10 Technical Committee
From: Greg Pellegrino (Greg.Pellegrino@compaq.com)
and Rob Elliott, Compaq Computer Corporation (Robert.Elliott@compaq.com)
Date: 15 June 2001
Subject: SRP InfiniBand™ annex

Revision History

Revision 0: 5 January 2001 - first revision
Revision 1: 11 January 2001 - updates from Houston SRP meeting.
Revision 2: 18 February 2001 - updates from Orlando SRP meeting and San Francisco IBTA AWG FTF meeting. Change bars removed due to amount of changes.
Revision 3: 3 March 2001 - updates from Denver SRP and IBTA AWG joint meeting. Still no change bars. Removed most support material.
Revision 4: 28 March 2001 - updates from Dallas SRP meeting.
Revision 5: 15 June 2001 - updates from Nashua SRP meeting. Converted to FrameMaker. Change bars lost. Made target port identifier 128 bits to match SRP revision 4 and added tables for initiator and target port identifiers. Added picture of initiator in IO unit and target in processor unit. Incorporated numbers from 01-104r0.

Moved Access Controls TransportID proposal into 01-181.
Moved Extended Copy target descriptor proposal into 01-192.
Moved Alias descriptor proposal into 01-193.

Related Documents

T10/srp-r03 – SCSI over RDMA protocol revision 3 (by Ed Gardner)
T10/fcp2r06 – SCSI over Fibre Channel protocol revision 6 (by Bob Snively)

T10/00-268r8 Defining Targets/Initiators as Ports (by George Penokie)
T10/00-425r3 Long Identifiers in SPC-3, SAM-2, SBC-2 and other XOR issues (by Jim Hafner)
T10/01-104r0 SRP Protocol identifiers (by Ed Gardner)
T10/01-181r0 Access Controls TransportIDs for SBP, SRP and iSCSI (by Jim Hafner and Rob Elliott)
T10/01-192r0 SPC-3 Extended copy target descriptor for SRP (by Rob Elliott)
T10/01-193r0 SRP Alias entry designation formats (by Rob Elliott)

InfiniBand Architecture Volume 1 – General Specifications, Release 1.0
InfiniBand Architecture Volume 2 – Physical Specifications, Release 1.0
InfiniBand Architecture Volume 3 – Application of InfiniBand, Release 0.9 (draft)

Overview

This proposes topics and text for an InfiniBand annex for the SCSI over RDMA (SRP) standard.

The goal is to identify all optional InfiniBand features that must be implemented to ensure useful, interoperable SRP devices. An annex in InfiniBand Volume 1, taken from Volume 3, will describe how boot devices, a subset of SRP devices, are specifically identified and the minimum command sets that may be depended upon (the “Storage Boot Wire Protocol”).

All text outside [brackets] is part of the suggested text.

Annex A (normative)

SRP for InfiniBand™ Architecture

[footnote] InfiniBand is a trademark and service mark of the InfiniBand Trade Association.

A.1 Related documents

InfiniBand™ Architecture Volume 1 – General Specifications. Release 1.0. InfiniBand^{SM} Trade Association.

IETF RFC 2373 - IP Version 6 Addressing Architecture. R. Hinden and S. Deering. Internet Engineering Task Force.

A.2 Glossary

A.2.1 Introduction

See the InfiniBand Architecture Volume 1 glossary for full definitions of InfiniBand terms.

A.2.2 Definitions

A.2 2 1 **Channel adapter:** Device that terminates a link and executes transport-level functions. Also called a node. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 2 **Communication manager:** The software, hardware, or combination of the two that supports the communication management mechanisms and protocols used to establish and release InfiniBand connections. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 3 **Consumer:** The direct user of verbs. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 4 **Global ID (GID):** A port address used for directing packets between subnets. A GID is a valid 128-bit IPv6 address. Source and destination GIDs are carried in an optional global routing header. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 5 **Globally unique identifier (GUID):** A number that uniquely identifies a device or component. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 6 **General service interface:** An interface providing management services other than subnet management. Uses the well-known queue pair 1. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 7 **InfiniBand Port:** Location on a channel adapter to which a link connects. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 8 **IO unit:** One or more IO controllers attached to the fabric through a single channel adapter. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 9 **IO controller:** The part of an IO unit that provides IO services. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 10 **IPv6 address:** A 128-bit address constructed in accordance with IETF RFC 2373 for Internet Protocol version 6. See IETF RFC 2373.

A.2 2 11 **Local ID (LID):** A port address used for directing packets within a subnet. Source and Destination LIDs are carried in every packet header. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 12 **Management datagram (MAD):** A packet used for communication to manage an InfiniBand network. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 13 **Packet:** The indivisible unit of InfiniBand Architecture data transfer and routing, consisting of one or more headers, a packet payload, and one or two CRCs. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 14 **SRP initiator port:** The SCSI initiator port as defined in SAM-2.

A.2 2 15 **SRP target port:** The SCSI target port as defined in SAM-2.

A.2 2 16 **Processor unit:** One or more consumers attached to the fabric through one or more channel adapters.

A.2 2 17 **Queue pair (QP):** An interface used for communication. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 18 **Queue pair number (QPN):** A value that identifies a queue pair within a channel adapter. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 19 **Service ID:** A value that allows a communication manager to associate an incoming connection request with the entity providing the service. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 20 **Subnet:** A set of InfiniBand ports connected via switches that have a common subnet ID and are managed by a common subnet manager. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 21 **Subnet manager:** Entity that configures and controls a subnet. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2 2 22 **Verbs:** An abstract description of the functionality of a channel adapter. An operating system may expose some or all of the verb functionality through its programming interface. See InfiniBand™ Architecture Volume 1 – General Specifications, release 1.0.

A.2.3 Acronyms

CRC: Cyclic redundancy check
GID: Global ID
GUID: Globally unique identifier
IBA: InfiniBand™ architecture
IPv6: Internet Protocol version 6
LID: Local ID
MAD: Management datagram
QP: Queue pair
QPN: Queue pair number
ROM: Read only memory
SRP: SCSI over RDMA protocol

A.3 Overview

This annex specifies requirements for mapping the SCSI over RDMA protocol (SRP) onto the InfiniBand Architecture (IBA), a transport that provides the necessary RDMA semantics.

A.4 InfiniBand Architecture overview

An IBA processor unit contains consumers and one or more channel adapters, each containing one or more ports and queue pairs (QPs) (see Figure 1).

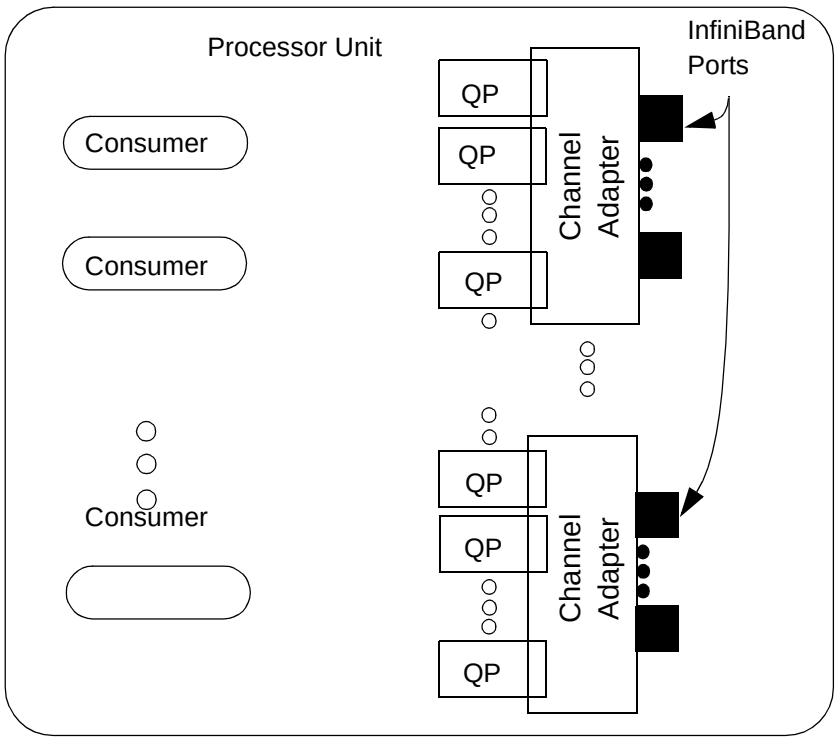

Figure 1 - Processor unit (derived from InfiniBand Architecture specification)

An IBA IO unit contains a channel adapter with one or more ports, one or more queue pairs, and one or more IO controllers (see Figure 2).

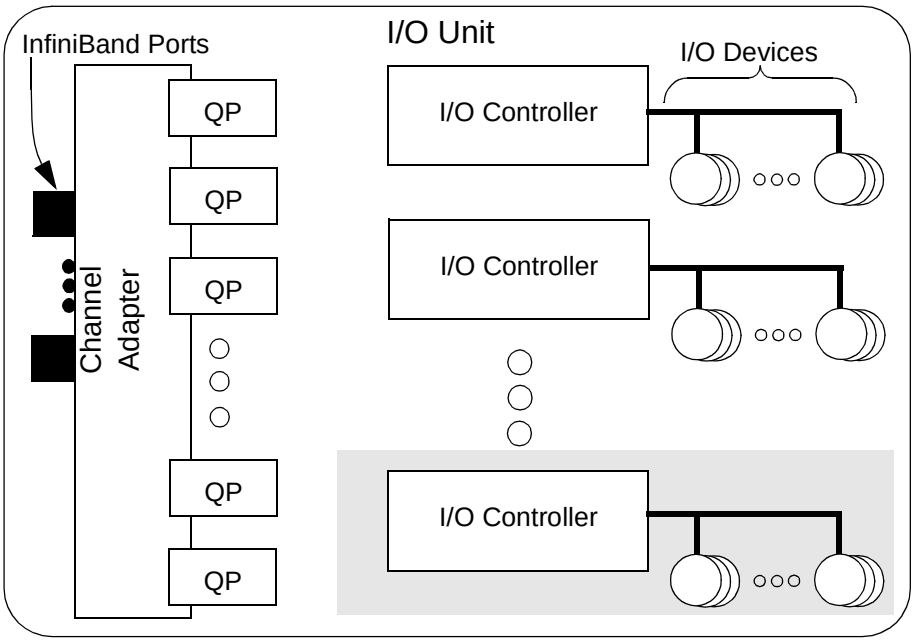

Figure 2 - IO unit (derived from InfiniBand Architecture specification)

Each InfiniBand port has a 64-bit globally unique identifier (GUID) called a port GUID. Each channel adapter has a channel adapter GUID (which is shared by all ports on the channel adapter). Each IO controller has an IO controller GUID.

Each InfiniBand port is assigned a 16-bit local ID (LID) or a range of LIDs by the subnet manager. Each InfiniBand port has one or more 128-bit global IDs (GID). Each GID is globally unique, and may be formed in part from the port GUID. A GID is an IPv6 address. The subnet manager provides GUID to GID/LID resolution.

Table 1 summarizes the IBA names (GUIDs) and addresses (IDs) relevant to SRP.

Table 1 - InfiniBand Architecture names and addresses

Name	Scope of uniqueness	Size	Description
Port GUID	worldwide	64 bits	Identifies an InfiniBand port within a channel adapter
Channel adapter GUID (Node GUID)	worldwide	64 bits	Identifies a channel adapter
IO controller GUID	worldwide	64 bits	Identifies an IO controller in an IO unit
LID	subnet	16 bits	Address assigned by the subnet manager to each InfiniBand port
GID (IPv6)	worldwide	128 bits	Address assigned by the subnet manager; (e.g. subnet prefix plus the port GUID)

A.5 SCSI architecture mapping

Figure 3 illustrates how SCSI initiator devices, initiator ports, target ports, and target devices map to InfiniBand Architecture objects. The figure also illustrates the mapping of the I-T-L nexus onto InfiniBand Architecture objects. The figure shows an initiator in a processor unit with a target in an IO unit.

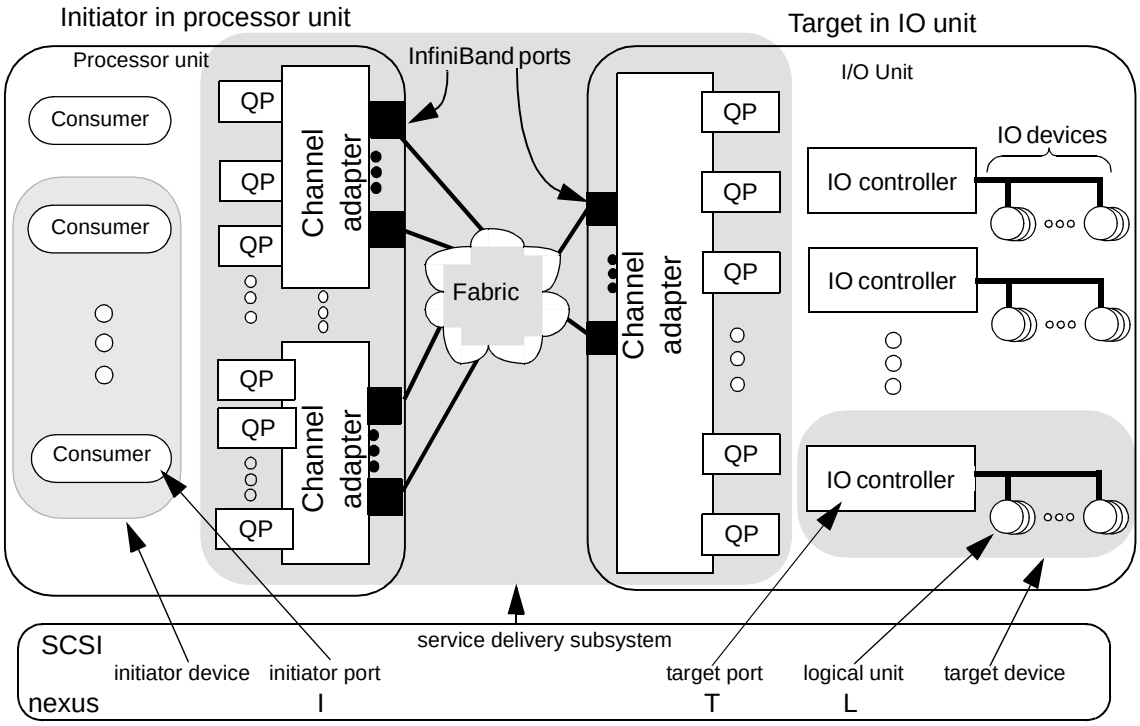

Figure 3 - Initiator in processor unit and target in IO unit

Figure 4 illustrates how SCSI initiator devices, initiator ports, target ports, and target devices map to InfiniBand Architecture objects. The figure also illustrates the mapping of the I-T-L nexus onto InfiniBand Architecture objects. The figure shows an initiator in an IO unit with a target in a processor unit.

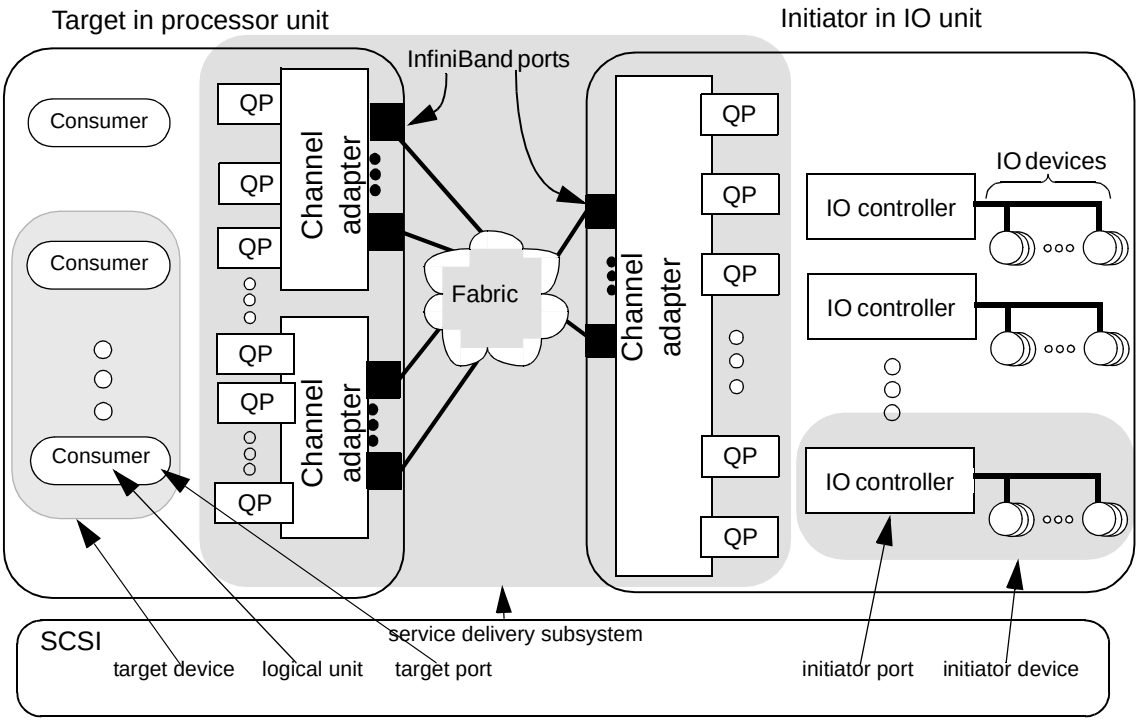

Figure 4 - Initiator in IO unit and target in processor unit

SRP initiators may be implemented in processor units or IO units.

An SRP initiator device in a processor unit is a set of consumers. An SRP initiator device in an IO unit is an IO controller.

An SRP initiator port in a processor unit is a consumer. An SRP initiator port in an IO unit is an IO controller. The initiator port identifier shall be a worldwide unique identifier, and should be constructed by concatenating a GUID such as a channel adapter GUID with an identifier extension as shown in Table 2.

Table 2 - Initiator port identifier

Byte Bit	7	6	5	4	3	2	2	0
0	GUID (e.g. channel adapter GUID)							
..								
7								
8	IDENTIFIER EXTENSION							
..								
15								

SRP targets may be implemented in processor units or IO units. In both cases, the SRP target shall include a device management agent to provide IOUnit, IOController, and ServiceEntries attributes and make available a worldwide unique IO controller GUID.

An SRP target device in a processor unit is a set of consumers. An SRP target device in an IO unit is an IO controller plus one or more IO devices.

An SRP target port in a processor unit is a consumer. An SRP target port in an IO unit is an IO controller. The target port identifier shall be a worldwide unique identifier, and shall be constructed by concatenating the IO controller GUID with an identifier extension as shown in Table 3.

Table 3 - Target port identifier

Byte Bit	7	6	5	4	3	2	2	0
0	IO CONTROLLER GUID							
..								
7								
8	IDENTIFIER EXTENSION							
..								
15								

The service delivery subsystem contains queue pairs, channel adapters, InfiniBand ports, and the InfiniBand fabric.

A.6 Communication management

A.6.1 Communication management overview

Communications managers on each InfiniBand device manage InfiniBand connections using MADs transported over the general service interface. SRP initiator and target ports shall use the active/passive (client/server) connection establishment protocol. The processor unit or IO controller containing the SRP target port shall act as the server and the processor unit or IO controller containing the SRP initiator port shall act as the client.

A.6.2 Discovering SRP target ports

To discover the service ID of an SRP target port in an IO unit, an SRP initiator port may use this sequence:
1) Retrieve the IOUnitInfo attribute from an IO unit using a DevMgtGet MAD to determine the presence and slot number of each IO controller attached to the IO unit;
2) retrieve the IOControllerProfile attributes from each IO controller, each of which includes a ServiceEntries table; and,
3) search the ServiceEntries table for entries with service names of “SRP.T10.NCITS”.

The service ID associated with each matching service name may be used in the communication management process to open InfiniBand connections to IO controllers acting as SRP target ports.

A.6.3 Establishing a connection

To establish an InfiniBand connection, the client places the service ID in a communication management Request message. The server associates the request with the appropriate SRP target port. The PrivateData field of the Request message shall include an SRP_LOGIN_REQ IU. The SRP target port may choose to refuse the connection based on the SRP_LOGIN_REQ IU content by returning a Reject message with a rea-

son code set to Consumer Reject. The PrivateData field of the Reject message shall include an SRP_LOGIN_REJECT IU.

If the server accepts the connection request and SRP login, the server returns a queue pair number (QPN) in a Response message. The PrivateData field of the Response message shall include an SRP_LOGIN_RSP IU. The SRP initiator port may choose to refuse the connection based on the SRP_LOGIN_REQ IU content by returning a Reject message with a Reason code set to Consumer Reject. The PrivateData field of the Reject message shall include an SRP_LOGIN_REJECT IU.

If the client accepts the connection reply and the SRP login response, it replies with a Ready To Use message indicating both an InfiniBand and an SRP connection are open. It may start using the connection for communication.

A.6.4 Releasing a connection

Either the SRP initiator port or SRP target port may send an SRP_LOGOUT IU with a SEND operation. The sender shall send a CM disconnect request upon receipt of an InfiniBand transport level acknowledgement to the SRP_LOGOUT IU. The sender may disconnect if its send queue has transitioned to an error state. The receiver of a LOGOUT IU shall respond with an InfiniBand transport acknowledgement and disconnect.

A.7 InfiniBand protocol requirements

SRP target ports and SRP initiator ports shall support the Reliable Connection transport service type.

SRP target ports shall implement the device management class of general management services.

SRP initiator ports and SRP target ports shall support the transport functions described in Table 4.

Table 4 - Transport operation support requirements

Transport functions	SRP initiator port	SRP target port
Send to	Mandatory	Mandatory
Send from	Mandatory	Mandatory
RDMA write to	Mandatory	Not used
RDMA write from	Not used	Mandatory
RDMA read to	Mandatory for data-out commands	Not used
RDMA read from	Not used	Mandatory for data-out commands
RDMA Write with immediate data (to or from)	Not used	Not used
ATOMIC (to or from)	Not used	Not used

IO units containing an IO controller acting as an SRP target port shall report the device management IOUnit attributes as described in Table 5.

Table 5 - IOUnit attributes for SRP target ports

Field	SRP requirements
Change ID	
Max Controllers	At least one
Option ROM	
Controller List	At least one IO controller must be present

IO controllers acting as SRP target ports shall report the device management IOControllerProfile attributes as described in Table 6.

Table 6 - IOControllerProfile attributes for SRP target ports

Field	SRP requirements
[IO controller] GUID	
Device ID	
Vendor ID	
Device Version	
Subsystem Vendor ID	
Subsystem [Device] ID	
IO Class	FF00h [per T10/01-104]
IO Subclass	609Eh [per T10/01-104]
Protocol	0108h [per T10/01-104]
Protocol Version	0001h [per T10/01-104]
Service Connections	At least one
Initiators Supported	At least one
Send Message Depth	At least one [Editor's note: probably the IOC's send queue depth]
RDMA Read Depth	[Editor's note: probably how many RRs an IOC can issue and have outstanding, i.e. CM:REQ:InitiatorDepth]
Send Message Size	[Editor's note: probably indicates how big the initiator's receive buffers need to be]
RDMA Transfer Size	At least one [Editor's note: probably the size of RDMA Writes the IOC can issue. It can always use smaller transfers]
Controller Operations Capability Mask	These bits shall be set to one: 0: ST (Send Messages to IO controllers) 1: SF (Send Messages from IO controllers) 5: WF (RDMA Write Requests from IO controllers) This bit shall be set to one by SRP target ports supporting data-out commands: 3: RF (RDMA Read Requests from IO controllers) These bits may be set to zero: 2: RT (RDMA Read Requests to IO controllers) 4: WT (RDMA Write Requests to IO controllers) 6: AT (Atomic Operations to IO controllers) 7: AF (Atomic Operations from IO controllers)
Controller Services Capability Mask	Bit 1 may be set for SRP ports with boot support: 1: SBWP Storage Boot Wire Protocol
Service Entries	At least one
ID String	

An IO controller acting as an SRP target port shall register with its Communications Manager a Service Name string of "SRP.T10.NCITS". This string is assigned an "IO SERVICE ID" type service ID by the Communications Manager.

IO controllers acting as SRP target ports shall include at least one ServiceName/ServiceID pair in the device management ServiceEntries attribute pair as described in Table 7.

Table 7 - ServiceEntries attribute pair for SRP target ports

Field	Length	SRP requirements
ServiceName_n	320	“SRP.T10.NCITS”
ServiceID_n	64	Assigned by the IO controller