

*Supplement to InfiniBand™
Architecture Specification
Volume 1 Release 1.2*



*Annex A11:
RDMA IP CM Service*

September 8, 2006

Copyright © 2006 by InfiniBand™ Trade Association.
All rights reserved.

All trademarks and brands are the property of their respective owners.

This document contains information proprietary to the InfiniBand™ Trade Association. Use or disclosure without written permission by an officer of the InfiniBand™ Trade Association is prohibited.

Table 0 Revision History

Revision	Date	
0.9	09/08/2006	Draft document for general IBTA member review

LEGAL DISCLAIMER

This specification provided “AS IS” and without any warranty of any kind, including, without limitation, any express or implied warranty of non-infringement, merchantability or fitness for a particular purpose.

In no event shall IBTA or any member of IBTA be liable for any direct, indirect, special, exemplary, punitive, or consequential damages, including, without limitation, lost profits, even if advised of the possibility of such damages.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

TABLE OF CONTENTS

Annex A11:	RDMA IP CM Service	6	1
A11.1	Introduction	6	2
A11.2	Glossary	6	3
A11.3	Convention	7	4
A11.4	IP Connection Establishment	7	5
A11.4.1	Major Version (MajV) (4 bits)	8	6
A11.4.2	Minor Version (MinV) (4 bits)	8	7
A11.4.3	IP version (IPV) (4 bits)	9	8
A11.4.4	Reserved Field (Res) (4 bits)	9	9
A11.4.5	Source Port (16 bits)	9	10
A11.4.6	Source & Destination IP Addresses (128 bits)	9	11
A11.5	IP Protocol Ports Mapping into IBTA Service IDs	9	12
A11.6	Error Handling	10	13
A11.7	REJ Reasons	11	14
A11.8	Compliance Summary	12	15

LIST OF FIGURES

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39
- 40
- 41
- 42

IBTA

LIST OF TABLES

		1
		2
		3
		4
Table 0	Revision History	2
Table 1	IP Addressing CM REQ Message Private Data Format	8
Table 2	RDMA IP CM Service IBTA Service ID Format	10
Table 3	REJ ARI format	11
Table 4	REJ ARI format for RDMA IP CM Service.....	11
Table 5	RDMA IP CM Service layer ARI	12
		11
		12
		13
		14
		15
		16
		17
		18
		19
		20
		21
		22
		23
		24
		25
		26
		27
		28
		29
		30
		31
		32
		33
		34
		35
		36
		37
		38
		39
		40
		41
		42

ANNEX A11: RDMA IP CM SERVICE

A11.1 INTRODUCTION

This annex defines the RDMA IP CM Service which provides support for a socket-like connection model for RDMA-aware ULPs. The annex defines a ULP independent way for the encoding of socket's 5-tuple into IBTA CM messages. The 5-tuple consists of source and destination IP addresses and ports, and protocol number. RDMA IP CM Service allows ULPs to use the same 5-tuple that is used by sockets for IB connection set up.

The RDMA IP CM Service uses IBTA CM REQ private data. It encodes into private data the connection client (local) IP address and port, and the destination IP address. The encoding maximizes RDMA-aware ULPs usable private data.

Second, the RDMA IP CM Service defines the range of IBTA Service IDs. It provides the mapping of the destination IP port and protocol number into Service ID. This allows for a connection server to listen on the Service ID corresponding to the IP protocol port and for a connection client to specify IB connection destination Service ID corresponding to the IP protocol port for RDMA-aware ULP.

RDMA IP CM Service Provider uses Service IDs range defined by RDMA IP CM Service. The Service IDs range defined by RDMA IP CM Service indicates that a CM REQ private data is formatted according to this Annex.

There are no requirements on where the Provider of the RDMA IP CM Service belongs. IBTA CM Provider can be enhanced to support RDMA IP CM Service, the service can be provided in a layer above IBTA CM, or even incorporated into ULP CM services.

[Annex A12: Support for iSCSI Extensions for RDMA \(iSER\) on page 6](#) provides an example of the use of the RDMA IP CM Service.

A11.2 GLOSSARY

Downward Compatible

An ability of a service provider to support previous versions of a product or a protocol. Minor version downward compatible provider automatically shifts to the highest numbered previous version of a protocol minor version within a major version to ensure both sides of a communication use the same protocol version.

Protocol Number In the Internet Protocol the protocol number specifies the protocol above IP [<http://www.iana.org>]. In the Internet Protocol version 4 (IPv4) [RFC791] there is a field, called “Protocol”, to identify the next level protocol. This is an 8 bit field. In Internet Protocol version 6 (IPv6) [RFC1883] this field is called the “Next Header” field. For, example, Transmission Control (TCP) Protocol Number is 6, User Datagram (UDP) is 17, Stream Control Transmission Protocol (SCTP) is 132.

RDMA-Aware ULP RDMA-aware ULP is an Upper Layer Protocol that explicitly uses RDMA primitives. In contrast, a ULP that is not RDMA-aware may be using RDMA implicitly by running over IPoIB or SDP.

RDMA IP CM Service IP-based Connection Management services defined in this annex.

RDMA IP CM Service Provider The provider of the RDMA IP CM Service.

RDMA IP CM Service Connection Client A connection client that uses RDMA IP CM Service and that generates connection requests.

RDMA IP CM Service Connection Server A connection server that uses RDMA IP CM Service and that responds to the connection requests.

RDMA IP CM Service Service ID Range The range of IBTA Service IDs assigned to RDMA IP CM Service. The Service IDs in the range share a prefix of 0x0000000001.

A11.3 CONVENTION

Though the IETF specification uses the same big endian byte ordering convention as defined in section [Section 1.5.1, “Byte Ordering,” on page 66](#), the rule for bit ordering is different. In IETF specification convention, for a byte, bit 0 is the most significant bit and bit 7 is the least significant bit. Similarly, for a word, bit 0 is the most significant bit and bit 31 is the least significant bit. In this annex, the IBTA convention for bit ordering and byte ordering as defined in section [Section 1.5.1, “Byte Ordering,” on page 66](#) is used.

A11.4 IP CONNECTION ESTABLISHMENT

The RDMA IP CM Service connection establishment is mapped to InfiniBand CM REQ and REJ MADs. The RDMA IP CM Service can use other CM MADs in accordance with Communication Management. See [Chapter 12: Communication Management on page 650](#).

The RDMA IP CM Service has a predefined format for the CM REQ Private Data. The formatted CM REQ private data field together with the connection Service ID convey the 5-tuple socket connection information. ULPs can define the format for the remaining CM REQ MAD (Consumer) Private Data field.

CA11-1: The CM REQ private data used by the RDMA IP CM Service during connection establishment shall conform to the format as defined in Table 1, "IP Addressing CM REQ Message Private Data Format," on page 8.

Table 1 IP Addressing CM REQ Message Private Data Format

bits	31-24		23-16		15-8	7-0
bytes						
0-3	MajV	MinV	IPV	Res	Source Port	
4-7	Source IP Address (127 – 96)					
8-11	Source IP Address (95 – 64)					
12-15	Source IP Address (63 – 32)					
16-19	Source IP Address (31 – 0)					
20-23	Destination IP Address (127 – 96)					
24-27	Destination IP Address (95 – 64)					
28-31	Destination IP Address (63 – 32)					
32-35	Destination IP Address (31 – 0)					
36-91	Consumer Private Data					

A11.4.1 MAJOR VERSION (MAJV) (4 BITS)

The current RDMA IP CM Service protocol specification requires MajV to be set to 0.

CA11-2: The RDMA IP CM Service connection server shall reject the connection request if MajV in the REQ message does not match the Major Version number it supports.

A11.4.2 MINOR VERSION (MINV) (4 BITS)

The current RDMA IP CM Service protocol specification requires MinV to be set to 0.

CA11-3: The RDMA IP CM Service protocol shall be Minor Version downward compatible.

CA11-3.1.1: RDMA IP CM Service connection server shall reject the connection request if MinV in the CM REQ message is higher than the Minor Version number it supports.

CA11-3.1.2: For a Major version number, RDMA IP CM Service Provider shall support the Minor Version numbers of an incoming CM REQ message equal to and lower than the Minor Version it supports.

A11.4.3 IP VERSION (IPV) (4 BITS)

The Internet Protocol version number of the end-point address field (fields Source IP Address and Destination IP Address). If IPV = 0x4, the IP addresses are IP version 4 format (32 bit addresses). If IPV = 0x6, then both IP addresses are IP version 6 format (128 bit addresses). All other IPV values are reserved.

CA11-4: The RDMA IP CM Service connection server shall reject the connection if IPV of the REQ message differ from the one(s) it supports.

A11.4.4 RESERVED FIELD (RES) (4 BITS)

Must be 0 on transmit and ignored on receive.

A11.4.5 SOURCE PORT (16 BITS)

The Source Port is a local IP protocol port number of the client. The protocol number of the IP protocol port is encoded in the Service ID of the CM REQ message as defined in A11.5 IP Protocol Ports Mapping into IBTA Service IDs on page 9.

A11.4.6 SOURCE & DESTINATION IP ADDRESSES (128 BITS)

The Internet Protocol (IP) address for the local interface (Source IP Address) and remote interface (Destination IP Address). These can be either IPv4 or IPv6 addresses, as specified by the IPV field.

CA11-5: If Source IP Address and Destination IP Address are IPv4 addresses, as specified by the IPV field, then Source IP Address(31-0) and Destination IP Address(31-0) shall be used to transmit the source and destination IP addresses, respectively, and bits Source IP Address(127-32) and Destination IP Address(127-32) shall be set to zero.

A11.5 IP PROTOCOL PORTS MAPPING INTO IBTA SERVICE IDS

The RDMA IP CM Service defines a mapping of IP protocol port spaces into the IBTA assigned Service ID space. The Service ID of the RDMA IP CM Service must have a prefix of 0x0000000001.

The RDMA IP CM Service user shall use Service ID as defined in [Table 2 RDMA IP CM Service IBTA Service ID Format on page 10](#).

CA11-6: The Service ID of an RDMA IP CM Service shall conform to the format as defined in [Table 2 RDMA IP CM Service IBTA Service ID Format on page 10](#).

The prefix of Service IDs for RDMA-aware ULPs starts with 0x00 as all Service IDs assigned by IBTA. The rest of the prefix is zeroed out with the 5th byte of 0x01 specifying the Service ID range for the RDMA IP CM Service. The sixth byte specifies IP protocol number of the port. The last

Table 2 RDMA IP CM Service IBTA Service ID Format

Byte Location	Description	Value
0	IBTA AGN	0x00
1 to 3	Prefix of RDMA-aware Service ID range	each byte is 0x00
4	RDMA-aware ULP Service ID range	0x01
5	IP Protocol Number	see IANA
6 to 7	Destination port number	ULP port number (usually well-known)

two bytes of the Service ID are the ULP port for the IP protocol number. For example, for NFSv4-RDMA over SCTP using the default port the Service ID is 0x0000000001840801 (port 2049), and for iSCSI/iSER over TCP using the well-known port the Service ID is 0x0000000001060CBC (port 3260). The port numbers are assigned by IETF per protocol number and are administered by IANA.

A11.6 ERROR HANDLING

CA11-7: The connection server shall treat CM REQ private data as formatted as defined [Table 1 IP Addressing CM REQ Message Private Data Format on page 8](#) if a Service ID belongs to the Service ID range of the RDMA IP CM Service.

Since IBTA Communication Manager may not be aware of RDMA IP CM Service there is no requirements for IBTA Communication Manager to validate the formatting of the REQ MAD private data matches [Table 1 IP Addressing CM REQ Message Private Data Format on page 8](#) when a Service ID belongs to the Service IDs range of the RDMA IP CM Service.

CA11-8: If the connection server does not support RDMA IP CM Service, it should return a CM REJ Message indicating “Invalid Service ID” (Reason 8) to the connection request whose Service ID belongs to the Service IDs range of the RDMA IP CM Service.

The connection server can return a generic CM REJ Message indicating “Consumer Reject” (Reason 28) with the ARI indicating RDMA IP CM Service reject with the RDMA IP CM Service ARI of 0x00 to the connection request whose Service ID belongs to the Service IDs range of the RDMA IP CM Service.

If the connection client gets a CM REJ Message indicating “Consumer Reject” (Reason 28) with the ARI indicating RDMA IP CM Service reject with the RDMA IP CM Service ARI of 0x00 to the connection request this does not imply that connection server supports RDMA IP CM Service nor that there is listener on the requested Service ID.

oA11-1: If the connection server supports RDMA IP CM Service but does not accept the parameters in the formatted Private Data field as defined by [Table 1 IP Addressing CM REQ Message Private Data Format on page 8](#), it shall return a CM REJ Message indicating “Consumer Reject” (Reason 28), with the content of the REJ MAD ARI field as defined in [A11.7 REJ Reasons on page 11](#).

A11.7 REJ REASONS

In order to separate RDMA IP CM Service reject reasons from an RDMA-aware ULP that uses RDMA IP CM Service both of which use “Consumer Reject” (reason 28) the ARI field is partitioned between them as defined by [Table 3 REJ ARI format on page 11](#)

Table 3 REJ ARI format

Byte Location	Description	Value
0	Rejection Layer	0x00 - RDMA IP CM Service; 0x01 - RDMA-Aware ULP
1-71	Layer ARI	defined by the layer

CA11-9: When the RDMA CM Service connection server rejects a request, the format of the ARI field shall conform to the format as defined in [Table 4 REJ ARI format for RDMA IP CM Service on page 11](#)

Table 4 REJ ARI format for RDMA IP CM Service

Byte Location	Description	Value
0	Rejection Layer	0x00 - RDMA IP CM Service
1	The RDMA IP CM Service layer ARI	see Table 5 RDMA IP CM Service layer ARI on page 12
2	Suggested Value Length in Bytes	0x00 - no suggested value; x -suggested value present of length of x bytes
3	alignment filler	reserved
4-71	Suggested Value	undefined

CA11-10: If the connection server supports RDMA IP CM Service, then for the RDMA IP CM Service Provider connection rejection, the RDMA IP CM Service layer reject ARI shall conform to the format as defined in [Table](#)

[5 RDMA IP CM Service layer ARI on page 12](#). The RDMA IP CM Service layer ARI shall occupy the second byte (byte 1) of the CM REJ ARI field.

Table 5 RDMA IP CM Service layer ARI

ARI Code	Name	Description
0x00	Unspecified	Unspecified reason for RDMA IP CM Service rejection.
0x01	Major Version	Major version of RDMA IP CM Service REQ Message is not supported by the remote CM.
0x02	Minor Version	Minor version of RDMA IP CM Service REQ Message is not supported by the remote CM.
0x03	IPV	IP version of RDMA IP CM Service REQ Message is not supported by the remote CM.
0x04	Source IP Address	The format of the Source IP Address is inconsistent with the IP version in the RDMA IP CM Service REQ Message.
0x05	Destination IP Address	The format of the Destination IP Address is inconsistent with the IP version in the RDMA IP CM Service REQ Message.
0x06	Unknown IP Address	The Destination IP Address does not match the server IP Address.

The connection server may optionally specify the suggested value for the rejected field of RDMA IP CM Service layer ARI.

oA11-2: If the connection server supports RDMA IP CM Service, then for the RDMA IP CM Service Provider connection rejection, the rejected field suggested value in ARI shall conform to the format as defined in [Table 4 REJ ARI format for RDMA IP CM Service on page 11](#).

The connection server is allowed to reject request for security or any other reason with the *unspecified* ARI returned to the connection client. There is no suggested value for unspecified rejection reason.

CA11-11: The RDMA IP CM Service Provider shall verify the Major and Minor versions field first.

In case of a rejection due to more than one reason, the connection server can freely choose one of the errors to be reported in ARI. The specification does not mandate any priority between the various fields except for the version fields.

A11.8 COMPLIANCE SUMMARY

In order to claim compliance for the RDMA IP CMS Service, a product shall meet all requirements specified in this section.

- **CA11-1:** RDMA IP CM Service REQ private data format Page 8
- **CA11-2:** RDMA IP CM Service MajV requirement. Page 8

- CA11-3: RDMA IP CM Service MinV downward compatibility . . . Page 8 1
- CA11-3.1.1:RDMA IP CM Service MinV requirement Page 8 2
- CA11-3.1.2:RDMA IP CM Service MinV support. Page 8 3
- CA11-4: RDMA IP CM Service IPV support rules Page 9 4
- CA11-5: RDMA IP CM Service size of IP addresses. Page 9 5
- CA11-6: RDMA IP CM Service Service ID. Page 9 6
- CA11-7: RDMA IP CM Service CM REQ private data format Page 10 7
- CA11-8: CM REJ if RDMA IP CM Service is not supported Page 10 8
- oA11-1: RDMA IP CM Service REJ MAD ARI field in CM REJ. . . Page 11 9
- CA11-10:Use of RDMA IP CM Service layer ARI Page 11 10
- oA11-2: RDMA IP CM Service ARI Rejected field value. Page 12 11
- CA11-11:RDMA IP CM Service CM REQ verification order Page 12 12

13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42