Mellanox Technologies

IPoIB registry parameters overview

Rev. 1.2

Mellanox Technologies, Inc.
350 Oakmead Parkway Suite 100
Sunnyvale, CA 94085
U.S.A.
www.mellanox.com
Tel: (408) 970-3400
Fax: (408) 970-3403
Mellanox Technologies Ltd
PO Box 586 Hermon Building
Yokneam 20692
Israel
Tel: +972-4-909-7200
Fax: +972-4-959-3245

Table of Contents

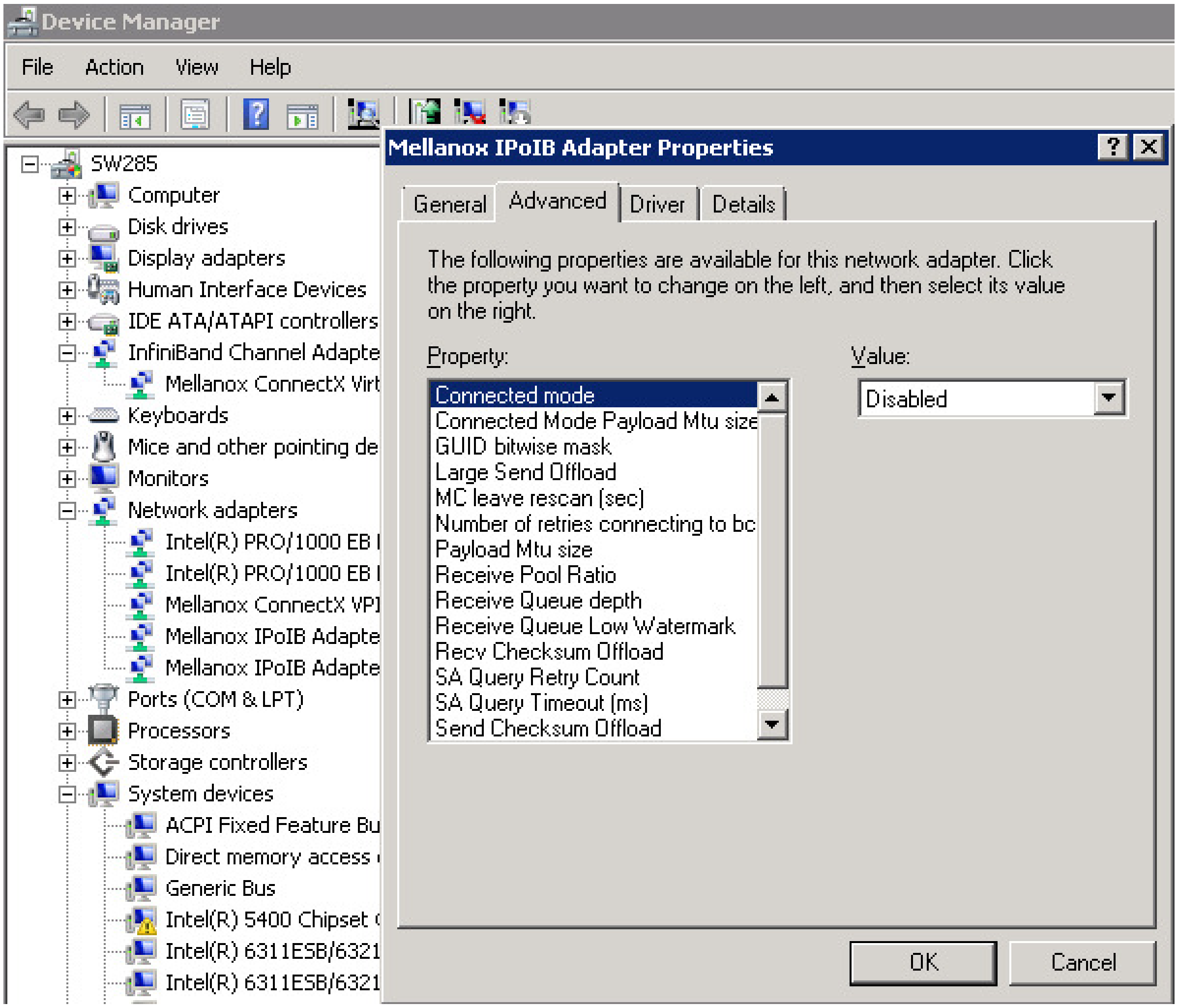
1 Introduction 4
  1.1 Controlling IPoIB adapter via GUI 4
  1.2 Controlling IPoIB adapter via registry 5
2 Features overview 5
  2.1 Connected Mode 5
  2.2 Connected Mode Payload MTU size 6
  2.3 GUID bitwise mask 6
  2.4 LSO (Large Send Offload) 6
  2.5 MC Leave Rescan 6
  2.6 Number of retries connecting to broadcast group 6
  2.7 Payload MTU 6
  2.8 Receive checksum offload 7
  2.9 Receive Pool Ratio 7
  2.10 Receive Queue Depth 7
  2.11 Receive Queue Low watermark 7
  2.12 SA query retry count 8
  2.13 SA query timeout 8
  2.14 Send checksum offloads 8
  2.15 Send Queue Depth 8

1 Introduction

Mellanox IPoIB driver' features and settings can be controlled through set of predefined parameters. These parameters can be set either via GUI of device manager (see fig.1) or directly by editing appropriate registers values (see table 1 for the table of appropriate registry values names and set of possible value)

1.1 Controlling IPoIB adapter via GUI

Run "Device Manager" (Start->run->devmgmt.msc), expand "Network Adapters", choose desired IPoIB adapter, right-click properties and then choose "Advanced" tab.



The following properties are available for this network adapter. Click the property you want to change on the left, and then select its value on the right.

Property:

- Connected mode
- Connected Mode Payload Mtu size
- GUID bitwise mask
- Large Send Offload
- MC leave rescan (sec)
- Number of retries connecting to bc
- Payload Mtu size
- Receive Pool Ratio
- Receive Queue depth
- Receive Queue Low Watermark
- Recv Checksum Offload
- SA Query Retry Count
- SA Query Timeout (ms)
- Send Checksum Offload

Value: Disabled

fig.1 - Controlling IPoIB adapter via GUI

1.2 Controlling IPoIB adapter via registry

One can directly change the value of an appropriate registry.

IPoIB registries are located under the following path:

HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Control\Class\{4D36E972-E325-11CE-BFC1-08002BE10318}

Under this location, one can find the subdirectories that belong to various IPoIB adapters installed and then update the desired value. The following table contains the name of each registry, its maximum, default and minimum values and the link to the description.

Important: The last value provided stands for the "GUID to MAC" conversion mask, that should be either 0 or contain exactly 6 non-zero digits using binary representation.

No.	Registry Name	Default	Min	Max
2.1	**BCJoinRetry**	50	0	1000
2.2	**CmEnabled**	FALSE	FALSE	TRUE
2.3	**CmPayloadMtu**	MAX_CM_PAYLOAD_MTU	512	65520
2.4	**Guid_mask**	0	0	252 (=0xFC)
2.5	**Iso**	0	0	1
2.6	**MCLeaveRescan**	260	1	3600
2.7	**PayloadMtu**	2044	512	4092
2.8	**RecvChksum**	ENABLED	DISABLED	BYPASS
2.9	**RecvRatio**	1	1	10
2.10	**RqDepth**	512	128	1024
2.11	**RqLowWatermark**	4	2	8
2.12	**SaRetries**	10	1	UINT_MAX
2.13	**SaTimeout**	1000	250	UINT_MAX
2.14	**SendChksum**	ENABLED	DISABLED	BYPASS
2.15	**SqDepth**	512	128	1024

Table 1 – IPoIB registries summary table

2 Features overview

2.1 Connected Mode

This value enables or disables Connected mode of IPoIB and currently unsupported for 2.1 Release (should be "disabled")

2.2 Connected Mode Payload MTU size

The maximum available size of IPoIB transfer unit. This value is relevant only when working with CM enabled and thus not supported for 2.1 release. See also section 2.7

2.3 GUID bitwise mask

The last value provided stands for the "GUID to MAC" conversion mask, that should be either 0 or contain exactly 6 non-zero digits using binary representation.

Zero (0) mask indicates its default value: 0xb' 11100111. That is, to take all except intermediate bytes of GUID to form MAC address

In a case of improper mask, the driver will use the default one.

Please, refer to http://mellanox.com/related-docs/prod_software/guid2mac_checker_user_manual.txt for more details

2.4 LSO (Large Send Offload)

Disables/Enables the LSO feature (if supported by HW). This feature has positive impact on overall performance.

Note that 4K MTU support defined as beta-version in 2.1 release, and it is not advisable to use both 4K MTU and LSO feature enabled simultaneously

2.5 MC Leave Rescan

This parameter indicates when to leave Multicast group when there were no receives for the indicated period of time (measured in seconds)

2.6 Number of retries connecting to broadcast group

The maximum number of retries to connect to BC till failure indication

2.7 Payload MTU

The maximum available size of IPoIB transfer unit. This value is relevant only when working with CM enabled and thus not supported for 2.1 release

It should be decremented by size of IPoIB header (==4B). For example, if the HCA support 4K MTU, upper threshold for payload MTU is 4092B and not 4096B.

4K MTU size improve performance also for short-sized messages, because NDIS can coalesce message of a smaller size into a big one.

Note that 4K MTU support defined as beta-version in 2.1 release, and it is not advisable to use both 4K MTU and LSO feature enabled simultaneously

If using CM mode:

MTU will be not limited by 4K threshold.

UD QP still may be used for different protocols (like ARP).

For these situations the threshold for the UD QP will take the default value

2.8 Receive checksum offload

Flags to indicate whether to offload receive checksum

Possible values:

- Disabled - No hardware checksum
- Enabled (if supported by HW) - Try to offload if the device supports it
- Bypass - Always report success (checksum bypass)

This HW offload always improve performance and is enabled by default.

2.9 Receive Pool Ratio

Initial ratio of receive pool size to receive queue depth

2.10 Receive Queue Depth

Number of receive WQEs to allocate.

2.11 Receive Queue Low watermark

Receives are indicated with NDIS_STATUS_RESOURCES when the number of "receives" posted to the RQ falls bellow this value

2.12 SA query retry count

Number of times to retry an SA query request

2.13 SA query timeout

Time, in milliseconds, to wait for a response before retransmitting an SA query request.

2.14 Send checksum offloads

Flags to indicate whether to offload send checksum

Possible values:

* Disabled - No hardware checksum
* Enabled (if supported by HW) - Try to offload if the device supports it
* Bypass - Always report success (checksum bypass)

This HW offload always improve performance and is enabled by default.

2.15 Send Queue Depth

Number of send WQEs to allocate.

Send queue depth needs to be a power of two; otherwise, it will be rounded up to the closest acceptable value